UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Dealing with Distraction : : Learning, Attention, and the Neural Encoding of Natural Auditory Scenes

Permalink

https://escholarship.org/uc/item/28x8b41x

Author

Caporello, Emily Laura

Publication Date

2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Dealing with Distraction: Learning, Attention, and the Neural Encoding of Natural Auditory Scenes

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Neurosciences

by

Emily Laura Caporello

Committee in Charge:

Professor Timothy Gentner, Chair
Professor Steven Hillyard
Professor Jill Leutgeb
Professor John Reynolds
Professor Tatyana Sharpee

2013

Copyright

Emily Laura Caporello, 2013

All Rights Reserved.

The Dissertation of Emily Laura Caporello is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2013

TABLE OF CONTENTS

Signature Page..... iii

Table of Contents..... iv

List of Figures..... v

List of Tables..... vi

Acknowledgements..... vii

Vita..... viii

Abstract..... ix

Introduction..... 1
I.1 Auditory Plasticity..... 2
I.2 Auditory Attention..... 8
I.3 Conclusion..... 24
I.4 Acknowledgements 26
I.5 Bibliography..... 27

Chapter 1..... 45
1.1 Introduction..... 45
1.2 Results…..... 47
1.3 Methods..... 58
1.4 Discussion..... 67
1.5 Acknowledgements 73
1.6 Bibliography..... 83

Chapter 2..... 89
2.1 Introduction..... 89
2.2 Results…..... 90
2.3 Methods..... 97
2.4 Discussion..... 111
2.5 Acknowledgements 117
2.6 Bibliography..... 124

Conclusion..... 132

LIST OF FIGURES

Figure 1.1: Avian auditory system connectivity, continued 74

Figure 1.2: Stimuli, continued..... 75

Figure 1.3: Neural responses to stimulus class, continued..... 76

Figure 1.4: Behavioral training and neural responses, continued..... 77

Figure 1.5: Effect of noise varies with magnitude of song, continued..... 78

Figure 1.6: Effect of reward category on neural response, continued..... 79

Figure 1.7: Effect of stimulus frequency on neural responses, continued..... 80

Figure 2.1: Auditory attention behavior in European starlings, continued..... 118

Figure 2.2: Chronic recording implantation, continued..... 119

Figure 2.3: Attention-driven modulation, continued..... 120

Figure 2.4: Attention response bias, continued..... 121

Figure 2.5: Competitive interaction model by cell type, continued..... 122

Figure 2.6: Electrophysiological properties, continued..... 123

LIST OF TABLES

Table 1.1: Summary of recorded units, continued..... 81

Table 1.2: Mean (± SEM) spike rates, continued..... 82

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Timothy Gentner for his support as chair of my committee. I would also like to thank Professor Steven Hillyard, Professor Jill Leutgeb, Professor John Reynolds, and Professor Tatyana Sharpee for their support as members of my committee.

The Introduction, in part, has been submitted for publication of the material as it may appear in Hearing Research, 2013, under authorship of Emily L. Caporello and Timothy Q. Gentner. Emily Laura Caporello was the primary investigator and author of this paper.

Chapters 1 and 2, in part, are currently being prepared for publication of the material as it may appear under authorship of Emily L. Caporello and Timothy Q. Gentner. Emily Laura Caporello was the primary investigator and author of this material.

VITA

2005 Bachelor of Arts in Cellular and Molecular Neuroscience

Johns Hopkins University, Baltimore, MD

2013 Doctor of Philosophy in Neurosciences

University of California at San Diego, La Jolla, CA

ABSTRACT OF THE DISSERTATION

Dealing with Distraction: Learning, Attention, and the Neural Encoding of Natural Auditory Scenes

by

Emily Laura Caporello

Doctor of Philosophy in Neurosciences

University of California, San Diego, 2013

Professor Timothy Gentner, Chair

My thesis work addresses the cortical processing of behaviorally relevant and irrelevant auditory signals using ethologically relevant stimuli. This issue is investigated over two time courses: after long-term behavioral training, and in real

time during the allocation of selective attention. Both studies examine the responses of single- and multi-units recorded extracellularly in the European Starling. I find that behavioral meaning affects neural processing in several regions, and provide evidence that neural modulation induced by behavioral valence involves not only an enhancement in the response to target stimuli, but also an active suppression of ignored stimuli.

Introduction

Our environment is filled with a rich array of sounds, often more than we can listen to at once. However, at a given time, any of these sounds could potentially provide information about our surroundings, and it is essential for us to be able to listen to any of them as we choose. Over time, certain types of stimuli, such as human speech, may prove to be consistently relevant to our behavior, and it may be additionally advantageous to detect these types of stimuli. Therefore, in the brain, we might expect to see mechanisms that allow for both the flexible allocation of attention to any sound in our environment at a given time, and also additional mechanisms that aid in processing signals that are reliably informative.

My thesis work explores the encoding of behaviorally relevant and irrelevant stimuli on these two timescales. In order to understand how the brain works in its natural state, it is important to study behaviors that are as ethologically relevant as possible while maintaining experimental control. By using a well-established model of vocal communication and a one of the strongest animal models of human language, the European starling (Knudsen and Gentner, 2010; Abe and Watanabe, 2011; Kiggins et al., 2012). Overall, these results suggest that experience leads to changes in neural processing that enhance the neural responses to behaviorally relevant natural signals. I was able to use invasive single unit recording techniques to study the processing of rich natural vocalizations of this species which, like human speech, are both

behaviorally relevant to the animal and specifically represented in the brain. In my first project, I examined how the processing of behaviorally relevant natural sounds is affected by the presence of irrelevant noise. In my second project, I explored how the short-term allocation of attention to a complex auditory scene affects auditory processing during behavior.

To put this work into relevant context, the following paragraphs provide background for long term plasticity in the auditory system and neural correlates of auditory attention.

I.1 Auditory Plasticity

The evidence for long term neural changes resulting from experience with relevant stimuli is well documented in the auditory domain. Numerous studies have shown that repeatedly pairing a stimulus with relevance, either through behavior or direct neurotransmitter release leads to long term changes in the neural representation of the paired stimulus.

Indeed, the relevance of certain types of stimuli is reflected in the presence of brain regions that are devoted to processing these stimulus categories. For instance, in humans, there are regions dedicated to processing human voices (Belin et al., 2000; Chang et al., 2010), and other environmental sounds (Doehrmann et al., 2008). Similar selectivity for conspecific vocalizations can be found in non-human primates (Wollberg and Newman, 1972; Winter and Funkenstein, 1973; Wang et al., 1995;

Binder, 2000; Tian et al., 2001; Petkov et al., 2008; Rauschecker and Scott, 2009), and songbirds (Leppelsack and Vogt, 1976; Bonke et al., 1979; Scheich et al., 1979; Katz and Gurney, 1981; Margoliash, 1986, 1983; Doupe and Konishi, 1991; Hausberger and Cousillas, 1996; Stripling et al., 2001; Gentner and Margoliash, 2003; Grace et al., 2003; Hsu et al., 2004; Atiani et al., 2009). Additionally, in the high-level visual system of primates, there are regions dedicated to the processing of faces (Gross et al., 1972; Desimone et al., 1984). This type of selectivity suggests that with experience, the repeated behavioral relevance of certain stimuli leads to an increase in the neural real estate devoted to their sensory processing.

Learning-Induced Adult Plasticity in the Auditory Cortex

The mechanisms underlying the selectivity for these types of natural stimuli have not been precisely worked out, but research with simpler stimuli indicates that this type of selectivity is likely a result of many interactions where these stimuli are relevant to behavior. For instance, after rodents learn to pair certain sounds with reward, changes in the primary auditory cortex can be observed that reflect an enhancement in the neural representation of the relevant stimulus (Recanzone et al., 1993; Bao and Chang, 2004; Polley et al., 2004, 2006; Rutkowski and Weinberger, 2005; Blake et al., 2006, see Ohl and Scheich, 2005 for a review).

Understanding how learning interacts with existing general biases for conspecific signals is crucial to understanding the coding of complex natural signals. Beyond the neural changes within primary regions, there is direct evidence that the

stimulus categories represented in higher level regions are also dependent upon learning and experience. For instance in the visual domain, training primates to categorize shapes based on specific feature dimensions results in neural selectivity across groups along relevant features (Freedman et al., 2001, 2003; Sigala and Logothetis, 2002; Freedman and Assad, 2006).

Experience dependent plasticity also plays a large role in shaping neural responses to natural stimuli. For instance, in maternal mice, auditory responses to pup calls are qualitatively different than those evoked in virgin mice (Galindo-Leon et al., 2009). The majority of work on natural stimulus selectivity and learning, however, with has been done in the songbird system. Songbirds are an excellent model system to study the auditory processing of natural signals (Theunissen and Shaevitz, 2006; Knudsen and Gentner, 2010), because like humans, songbirds use complex, learned vocalizations to modulate conspecific behaviors in a variety of natural social contexts (Catchpole and Slater, 1995; Kroodsma and Miller, 1996), often in the midst of environmental noise or the chorus of other birds. A mentioned earlier, several auditory forebrain regions in songbirds are selective for the general acoustic characteristics of conspecific songs. While the auditory cortex of songbirds is not arranged in the 6-layered structure typical of mammalian cortex, it shares a cellular level homology with mammalian auditory cortex (Wang et al., 2010).

In songbirds, neural selectivity for conspecific song in the auditory system is dependent upon early experience (Phan et al., 2006; Raksin et al., 2012; Hauber et al., 2013; van der Kant et al., 2013) and neural discriminability of vocal sounds develops

simultaneously with behavioral discrimination performance (Miller-Sims and Bottjer, 2012). Within regions selective for conspecific vocalizations, neural responses are modified as a result of learning (Thompson and Gentner, 2010) and recent experience (Terleph et al., 2008), with stronger representations for learned stimuli observed in the caudal mesopallium (CM; Gentner and Margoliash, 2003; Woolley and Doupe, 2008). Furthermore, neural responses within CM contain more information about relevant songs than novel ones (Jeanne et al., 2011), and populations of neurons show improved encoding for learned songs (Jeanne et al., 2013). The neural mechanisms underlying complex stimulus selectivity are likely also contributing to the learned representations for speech in high-level human cortical regions (Liebenthal et al., 2010).

Mechanisms of Learning Induced Plasticity

One mechanism thought to induce these long term changes in the auditory system is through neurotransmitters associated with attention and reward, such as acetylcholine and dopamine (Sarter et al., 2005). Pairing tones with dopamine release via ventral tegmental area stimulation (Bao et al., 2001) or acetylcholine release via nucleus basalis stimulation (Metherate and Weinberger, 1989; Edeline and Weinberger, 1993; Bakin and Weinberger, 1996; Froemke et al., 2007) leads to the types of overrepresentations of paired stimuli seen in learning. Not only are the frequencies of the paired tones improved, but also the temporal components (Kilgard and Merzenich, 1998a; Bao and Chang, 2004). Details on mechanisms of Ach

activation that may lead to plasticity in the auditory system have been reviewed (Rasmusson, 2000; Suga and Ma, 2003)

Overall these plasticity studies put forward a mechanism where attending to certain relevant stimuli repeatedly leads to an enhancement of the response to those stimuli (for reviews see Jääskeläinen et al., 2007; Pienkowski and Eggermont, 2011; Scheich et al., 2011).

Benefits of Learning-Dependent Plasticity

Presumably, the changes that occur as a result of experience are intended to improve the discrimination and detectability of relevant signals. Indeed, neural changes within the auditory cortex are associated with improvements in the detection of relevant auditory stimuli (Reed et al., 2011). While there is less data on this direct causal link in the case of selectivity for natural signals such as speech, the fact that throughout normal childhood development humans become expert in recognizing and categorizing the set of phonemes that carry information in one's native language (Kuhl et al., 2006) indicates that improvements in speech comprehension co-occur with developments in language areas.

Understanding Stimulus Selectivity: The Role of Irrelevant Stimuli

The acoustic environments in which such relevant sounds, such as speech, occur are often contaminated by background noise and other potentially informative auditory signals. When conversing with a friend on a crowded bus, for example, the

speech of one's friend must be heard amidst the noise of the road, engine, and other conversations. Despite this, very little is understood about the influence of behaviorally irrelevant sounds on the neural processing of those relevant to behavior, especially in the context of natural stimuli.

Irrespective of learning, there are several mechanisms by which background noise could degrade signal detection. Simply adding noise to natural signals leads to increases in spurious spikes and deletions of informative spikes in the avian auditory cortex (Narayan et al., 2007), making neural discriminability of natural sounds more difficult. Additionally, tones lying outside the receptive field of a neuron can modulate the response to an excitatory tone when played simultaneously. Two-tone suppression occurs peripherally in the cochlea (Sellick and Russell, 1979; Cheatham and Dallos, 1989; Ruggero, 1992) and auditory nerve (Sachs and Kiang, 1968; Cheatham and Dallos, 1989; Ruggero, 1992), and in more complicated forms in primary auditory cortex (Sutter et al., 1999). Thus, even sounds which do not overlap in frequency space with an auditory target might reduce its neural representation.

What is not clear is how experience detecting signals in noise affects the neural processing of targets and distracters. Recent evidence has shown that just as we have seen with excitatory stimuli, learning can also lead to a change in the suppressive abilities of irrelevant stimuli. During and immediately following trials where animals detect tones amidst noisy distracters, changes in A1 receptive fields show increased suppression at uninformative frequencies, co-occurring with enhanced excitability at informative frequencies (Fritz et al., 2005a, 2007b). These results indicate that

learning not only results in the enhancement of neural responses to relevant stimuli, but also in reductions in the responses to irrelevant stimuli. This evidence suggests that learning to detect natural signals in noise should lead to similar changes in neural responses not only to the target stimuli, but in the suppressive nature of the noise itself.

I.2 Auditory Attention

Our past experience may crucial for choosing which sounds in our environment, such as speech, are informative, but once that target is chosen, there are further processes that must occur in order to hear the desired target in a noisy environment. In the example case of conversing with a friend on a bus, one must "tune out" the road, engine and other ambient noise, and "focus on" the stream of speech being produced by one's friend rather than that from the talkative stranger nearby. The challenge of attending selectively to a stream of speech amid one or more competing streams of speech is referred to as the cocktail party problem (CPP; Cherry, 1953). Humans and non-humans can solve species-typical versions of the CPP, but how these solutions are implemented in the activity patterns of single neurons, neural populations, and well-defined neural circuits is not known. Gaining such knowledge will require the use of animal models that both permit direct control over attention driven behaviors and invasive experimental techniques unavailable in humans.

Broad-scale Neural Correlates of Auditory Attention

Given that selective attention can modulate a wide range of auditory responses at multiple processing stages, it is not surprising to find that neural correlates to attention, measured with EEG and MEG, appear at multiple stages throughout the sensory processing stream. Attention-dependent EEG modulation can be observed as soon as 20 msec after stimulus onset (Woldorff et al., 1987; Woldorff and Hillyard, 1991), an effect thought to originate from primary auditory cortex (Woldorff et al., 1993). Magnetoencephalography (MEG), functional imaging (fMRI), and positron emission tomography (PET) data further support early cortical gain in the contralateral temporal lobe of attended tones and speech (O'Leary et al., 1996; Grady et al., 1997; Ding and Simon, 2012), both within primary (Fujiwara et al., 1998; Jäncke et al., 1999) and secondary auditory regions (Petkov et al., 2004; Woods et al., 2009; Okamoto et al., 2011). Early dichotic listening studies, where subjects attend to tone streams in one ear while ignoring input to the other ear, indicated that selective attention can modify auditory event-related potentials (ERPs). Specifically, the N1 negative component (80-110 msec after sound onset), P2 component (150-200 msec after onset) and P3 component (250-450 msec after onset) are larger when auditory stimuli presented to the contralateral ear are attended versus ignored (Hillyard et al. 1973, Picton and Hillyard 1974). The components differ in their sources and their sensitivities to different aspects of attention tasks. N1 and P2 components can be modified by the amplitude, location, and stimulus parameters of acoustic stimuli (Hillyard et al., 1973). The N1 component is thought to originate at least in part within

the primary auditory cortex (Knight et al. 1981, Giard et al. 1988), while the P2 component is thought to originate in secondary cortices anterior to A1 (Rif et al., 1991). Attention modifies the magnitude of the N1 and P2 components elicited by all stimuli (targets and distracters) in the attended channel (Hillyard et al., 1973; Picton and Hillyard, 1974; Hansen and Hillyard, 1978). The P3 component, however, is modified only by behaviorally relevant stimuli, is largely invariant to changes in stimulus parameters (Hillyard et al., 1973), occurs in multiple sensory modalities (Klinke et al., 1968; Barlow, 1969), and is elicited even during behaviorally relevant stimulus omissions (Picton and Hillyard, 1974). The P3 component is thought to originate in parietal regions (Picton and Hillyard, 1974; Downar et al., 2000), and to reflect decision making, arousal, or motor response (reviewed in Pritchard, 1981).

Selective auditory attention can also modify neural activity in subcortical sensory regions such as the inferior colliculus (Rinne et al., 2008) and possibly the dorsal cochlear nucleus (Kaltenbach, 2006), presumably through the massive descending corticofugal projections from auditory cortex (Winer, 2005). There is some debate over whether selective attention affects can be observed in the medial olivocohlear bundle (recently reviewed in Tanedo and Bhagat 2008), but recent studies confirm the presence of attention-dependent modulations in the magnitude of evoked otoacoustic emissions (EOAEs); (Maison et al., 2001; de Boer and Thornton, 2007; Smith et al., 2012). Thus, it is reasonable to conclude that almost any area of the central auditory system can be modulated by attention. One potential advantage to modifying responses at the level of the brainstem, or even the cochlea, is that speech

harmonics or other first order features useful for scene segregation may be differentially enhanced (Anderson and Kraus, 2011). Collectively, the evidence is consistent with early, stimulus-sensitive modulation associated in modality-specific regions, and later, stimulus-insensitive modulation associated in frontal and parietal regions.

Task demands also affect attention-related activity. Both location and pitch can be used to solve the cocktail party problem (Spieth et al., 1954; Bregman, 1990; Darwin and Hukin, 2000a, 2000b; Woods et al., 2001), and attending to either the spectral features of speech or its location activate distinct brain regions. Attention to location results in increased activity in regions associated with spatial processing, such as posterior temporal cortex and posterior parietal regions, while attention to spectral features of speech leads to increased activity in auditory- and speech-specific regions such as anterior and superior temporal cortex (Ahveninen et al., 2006; Hill and Miller, 2010). The distinct activity patterns elicited by spatial and non-spatial attention to speech also correspond to similar “what” vs. “where” pathway activity seen in auditory discrimination tasks with non-linguistic stimuli (Alain et al., 2001; Maeder et al., 2001; Woods and Alain, 2001; Degerman et al., 2006; Paltoglou et al., 2011). Different brain regions are affected even when altering the discrimination between non-spatial auditory features, such as duration vs. pitch (Brechmann and Scheich, 2005) high vs. low frequency (Paltoglou et al., 2009), or pitch vs. loudness (Rinne et al., 2007). A similar effect seen in bimodal selective attention studies, where subjects are presented with both visual and auditory stimuli and perform discriminations in one

modality or another (Petkov et al. 2004; Johnson and Zatorre 2005; Johnson and Zatorre 2006; Woods et al. 2009; Degerman et al. 2007; Salmi et al. 2007). In each case, the regions showing increased activation above baseline correlate to the features that are behaviorally relevant for the task, or, in tasks where attention is not being explicitly measured, the features that are likely being attended to. Similar effects have been observed when measuring the strength of gamma oscillations over temporal lobe during auditory (versus visual) attention (Sokolov et al., 2004), rather than a gain in overall neural activity. Increased gamma activity has been associated with attention in other modalities, such as vision (Tiitinen et al., 1993; Tallon-Baudry et al., 1997).

In line with these results, it may not be surprising that attending to speech stimuli, as opposed to non-linguistic stimuli, activates regions such as the superior temporal sulcus and gyrus that are known to be involved in language processing (Hari et al., 1989; Belin et al., 2000; Binder et al., 2004; Alho et al., 2006; Pasley et al., 2012; Rämä et al., 2012), often more prominently within the left hemisphere (Jäncke, Mirzazade, and Shah 1999; Hugdahl et al. 2003; Coch, Sanders, and Neville 2005; Jäncke et al. 2001; Hill and Miller 2010; Alho et al. 2003, but see Ding and Simon 2012). An early electrocorticography (ECog) study in humans similarly found a gain in gamma band power within the left hemisphere secondary auditory cortex during a phoneme discrimination task (Crone et al., 2001), and a more recent EEG study found increased gamma band activity in the left temporal lobe during the correct detection of degraded words (Hannemann et al., 2007). This well-documented left-hemisphere specialization for speech processing is thought to underlie the frequently observed bias

in subjects to attend to speech presented to the right ear ("right ear advantage", see Hugdahl et al. 2009 for a review, Kompus et al. 2012), since auditory information is processed in the hemisphere contralateral to peripheral stimulation.

Within specific brain regions, attending to speech improves neural discriminability. A recent electrocorticography study measured activity in the posterior superior temporal lobe during selective attention to one of two simultaneously presented speakers (Mesgarani and Chang, 2012). In this region, local field potentials (LFP) in the high gamma range (75-150Hz) were not only modified by attention, but the evoked response was more likely to resemble the response to the attended speaker alone. Compellingly, this effect was absent during error trials, indicating that poorly discriminable neural responses in cortex are associated with an inability to detect a lingual target in a noisy environment. The neural mechanisms that underlie these changes in representational fidelity, like many observed effects of attention, remain largely unknown.

The foregoing evidence indicates that different brain regions may be differentially modulated by auditory attention depending on the nature of specific stimuli and task demands, e.g., detecting a difference in space or in pitch. Other studies also note, however, consistent attention-dependent modulation of the *same* regions regardless of task parameters including superior parietal cortex, dorsolateral frontal cortex, the pre-central sulcus, and middle and superior temporal cortex (Paus et al., 1997; Zatorre et al., 1999; Degerman et al., 2006; Wu et al., 2007; Salmi et al., 2009; Hill and Miller, 2010; Westerhausen et al., 2010; Falkenberg et al., 2011). A

recent study of spatial attention using high-density electroencephalography found that alpha band activity (8-12Hz) originating in posterior parietal regions could be used to determine the identity of an attended speech source among two distracters (Kerlin et al., 2010). Similar activation of superior parietal cortex has been reported for fMRI data during attentional switching in visual tasks (Sauseng et al., 2005; Shomstein and Yantis, 2006), suggesting a for a role of parietal cortex in spatial attention across modalities (Giesbrecht et al. 2003; Shomstein and Yantis 2006; Sauseng et al. 2005; Rihs, Michel, and Thut 2009). Such regions may be part of a general "attention network", either obligatory for selective attention or reflective of modulation at processing stages or task demands shared by multiple tasks.

In addition to different demands between tasks, difficulty with the same task can also affect attentional modulation of auditory responses. In particular, frontal cortex activation, which is often observed independent of task parameters, appears to depend on task difficulty or attentional load (Altmann et al., 2008; Bidet-Caulet et al., 2010). Studies that have observed a lack of prefrontal cortex modulation during attention to continuous speech (Alho et al., 2003, 2006) have used the idea of attentional load to argue that attending to continuous speech is less dependent on prefrontal cortices than other tasks because humans are very experienced in attending to continuous speech.

One additional aspect of natural auditory signals that is important to consider in the context of attention is their temporally dynamic organization. Unlike many artificial auditory stimuli such as tones and noise, or visual objects that are constant

over time, the features of an auditory object (pitch, amplitude, spectral contrast) fluctuate across time. Indeed, it is these fluctuations that carry the information in natural signals. We have already noted that expectations can influence auditory stream segregation, but it is worth noting that in the context of many vocal communication signals, including speech, the time scales for the expectations are dynamic. That is, the signal occurring at any given time may set up expectations about future signals at a range of possible times. While stimulus predictability has been shown to effect single unit processing of natural signals in songbird auditory regions (Gill 2008), in general very little is known about the influence of expectation on attention mechanisms. It is reasonable to predict, however, that attention may bias temporal expectation so that particular portions of natural auditory signals become more informative than others. Recent work supports these ideas in showing that adults (Sanders et al., 2002) and children as young as 3 (Astheimer and Sanders, 2012) attend most carefully to word onsets, which are highly informative about speech content. Moreover, attention probes presented during speech elicit larger N1 ERPs during word onsets (Astheimer and Sanders, 2009), than during the least predictable portions of speech (Astheimer and Sanders, 2011). Neural (ECog) data are consistent with a role for temporal dynamics in attention, showing that low frequency oscillations entrained to attended auditory stimuli are dependent upon the predictability of the attended stream (Besle et al., 2011).

Overall, the evidence supports a flexible model of attention where the loci of attentional modulation are highly affected by the acoustic parameters, temporal

dynamics, and behavioral relevance of the stimuli, as well as the ultimate task-related goals of the animal. Because of the dynamic nature of natural communication signals such as speech, the factors that are most useful for separating target objects from distracters could oscillate throughout the listening period. Furthermore, the predictability of speech fluctuates in time, allowing for periods where attentional modulation may be more or less necessary for successful signal detection. In general, the complex and variable nature of speech makes it unlikely that attentional modulation is consistent over the stimulus duration in either its magnitude or in the brain regions it affects. Parsing out the complexity of attentional modulation to natural signals requires a closer look at the circuitry modulated by attention, and for that we look now to studies in animal models.

Animal Models of Attention

Recent years have seen an increasing interest in investigating auditory streaming and attention in animal models. Besides allowing for more invasive experimental techniques that can address the neural mechanisms underlying attention, many animal species regularly encounter difficult cocktail-party-like environments in which selection of a relevant target stream would likely require attention (see Bee, 2008 for a review). For instance, king penguins reside in large noisy colonies of several thousand vocal birds, yet young penguins can identify the distant call of their parent at volume levels 6 decibels below background noise (Aubin and Jouventin, 1998). The difficulty of this recognition task is made clear by the many strategies

employed by the penguins to solve it quickly, such as modifying acoustic parameters of the call to separate it from background noise, and limiting calls to geographically specific “rendezvous areas” (Aubin, 2004). Frogs, like humans, can use the frequency separation between simultaneous communication calls as a perceptual cue for stream segregation (Nityananda and Bee, 2011). In somewhat more disturbing example, Vampire bats, who will often feed on the same prey animal several nights in a row, can detect faint breathing sounds and use them to identify individuals (Gröger and Wiegrebe, 2006). Controlled laboratory experiments on European starlings, a songbird species, indicate that subjects can learn to recognize an individual song even when it is always simultaneously presented with another song, and that subjects can recognize learned songs masked by a loud songbird chorus (Hulse et al., 1997). Both ferrets and European starlings have been trained to perform stream segregation tasks similar to those used to study the process in humans (MacDougall-Shackleton et al., 1998; Ma et al., 2010). Even singing insects, such as crickets, face a version of the cocktail party problem in their natural environments. Crickets rely on several adaptations to identify conspecifics in noisy environments, including neural gain control mechanisms that reduce the influence of background noise from relevant call signals (Schmidt and Römer, 2011).

The Advantages to Studying Animals

One benefit of developing non-human models of auditory attention is the application of techniques that allow researchers to investigate the neural mechanisms

underlying attention, such as extracellular and intracellular electrophysiology, calcium and voltage imaging, viral transfection, and transgenetics. These techniques could help us understand more about the attentional effects observed in human auditory attention studies, such as the role of corticofugal pathways underlying subcortical attention-dependent changes in neural activity (Ryan et al., 1984; Suga and Ma, 2003) the existence of an anatomical hierarchy of auditory attention modulation as reported in the visual domain (Yin et al., 2008), and the role of neurotransmitters such as acetylcholine (Metherate and Weinberger, 1989; Bakin and Weinberger, 1996; Kilgard and Merzenich, 1998a, 1998b; Metherate et al., 2012). Single neuron data can also allow us to investigate the applicability of models proposed to solve the cocktail party problem and selective attention, such as those involving sparse coding (Asari et al., 2006), normalization (Reynolds and Heeger, 2009), receptive field modulation (Mesgarani et al., 2010), and others (Sagi et al., 2001; Jääskeläinen et al., 2011).

What Animal Models of Auditory Processing Have Told Us so Far

The application of extracellular electrophysiology in auditory attention research originated decades ago in the classic work of David Hubel and colleagues (Hubel et al., 1959). In a largely anecdotal study, the authors report data from neurons in secondary auditory cortices of awake unrestrained cats that responded to complex sounds only when the cat appeared to be attending to it (often inferred from gaze orientation). The observed results were intriguing, but at the time little was known about the processing of complex sounds in high-level cortical regions or controlling

attention in non-human species. As the authors write at the end of their paper, "attention is an elusive variable that no one has as yet been able to quantify." Perhaps for these reasons, follow-up work on auditory processing in awake, behaving animals came only years later and focused primarily on simpler stimuli (such as clicks, tones, and white noise) driving neural responses in regions such as A1 that were better understood functionally and anatomically. These studies consistently reported an increase in the single-neuron response to a behaviorally-relevant stimulus during task engagement compared to passive or ignored conditions (Miller et al. 1972; Hocherman et al. 1976; Pfingst, Connor, and Miller 1977; Benson and Hienz 1978; for a comprehensive review see Fritz et al. 2007). Decades of research on learning-dependent plasticity in the auditory cortex, such as in the studies mentioned above, has provided us with ample evidence that the response properties of auditory neurons are capable of changing to better respond to relevant stimuli.

Evidence for Active Enhancement of Targets

The enhanced response to an attended stimulus over an unattended stimulus could be accomplished in a variety of ways. A theory developed early on as a result of the first EEG dichotic listening studies involved a specific auditory attention network that would represent a "trace" of the attended target, against which other auditory inputs would be compared to (Näätänen and Michie, 1979; Näätänen, 1992). This was in contrast with a gain model, also proposed early on in auditory attention research (Hillyard et al., 1973; Woldorff et al., 1993) and similar to those proposed in visual

attention (Desimone and Duncan, 1995; Hillyard et al., 1998), in which the firing rates of cells that respond to the attended stimulus increase, while the firing rates of cells that respond to the ignored stimulus decrease. Another more recently proposed version is the tuning model, where attention also modulates the tuning properties of individual cells (Jääskeläinen et al., 2007) in addition to general gain-model modulation. This model shares similarity with a recently proposed model of visual attention (Reynolds and Heeger, 2009). In support of the tuning model, Fritz and colleagues developed experiments that demonstrated rapid changes in individual neuron response properties during a difficult tone detection task (Fritz et al., 2003). Shortly after the onset of behavior, the receptive fields of ferret A1 neurons changed to enhance task-relevant frequencies. These changes developed rapidly and could be reshaped by introducing a new behaviorally relevant frequency. In follow-up experiments using chord targets with several relevant frequencies, the authors were able to show that the cells' receptive fields could change to enhance responses to several discrete, behaviorally relevant frequencies. Stimulus reconstruction data from neural activity in these ferrets indicate that attention affects the coding properties of A1 neurons (Mesgarani et al., 2008), and recent human studies using EEG and fMRI further support attention-dependent tuning modulation (Kauramäki et al., 2007; Okamoto et al., 2007; Ahveninen et al., 2011).

Evidence for Active Suppression of Distracters

Interestingly, Fritz and colleagues not only saw enhancement of relevant frequencies, but also suppression of irrelevant frequencies (Fritz et al., 2005b, 2007b). This suppression of the single-unit response to irrelevant sounds has been confirmed in other experiments (Otazu et al., 2009), and is consistent with imaging data in humans (Hopf et al., 2006; Paltoglou et al., 2009).

Active suppression has been observed in classic human attention EEG studies as well. P200 magnitude, for instance, is larger in response to ignored tones that when the same tones are presented in the absence of auditory attention (Alho et al., 1987; Rif et al., 1991). the amount of attention-induced suppression in the receptive field is dependent upon attentional load, with suppression occurring most strongly in low signal-to-noise environments (Atiani et al., 2009; Neelon et al., 2011). Most recently, a clever visual paradigm separating the influence of enhancement and suppression to targets and distracters, respectively, demonstrated not only that suppression increased during attention trials, but that the overall magnitude of suppression correlated with behavioral performance (Bridwell and Srinivasan, 2012).

These results indicate that selective attention involves both an increase in gain in the sensory response to the target and an increase in distracter suppression. The mechanisms that underlie these processes in auditory attention, particularly to natural stimuli, are largely unknown.

Towards a neurobiological model for attention to natural signals

One overriding theme to draw from the wealth of human-targeted studies reviewed above is that selective auditory attention cannot be operationalized in a single way. Attention modulates responses (both neural and behavioral) at multiple levels in processing pathways and in multiple ways depending on the interaction of behaviorally relevant stimuli, task demands, experience, and expectations. Further development of animal models for attention therefore necessitate the incorporation of natural, behaviorally relevant stimuli into behavioral designs the control and isolate specific aspects of attention so that underlying neural circuits can be isolated and their modulation by attention determined.

To date, the majority of non-human auditory attention and task engagement experiments have employed simple or synthetic stimuli due to their easy interpretability and analytical benefits. As detailed above, however, behavioral relevance plays an increasingly clear role in the neural encoding of complex signals. Additionally, recent studies in these regions have found that neural responses to song features are tolerant to natural variation (Meliza and Margoliash, 2012), amplitude (Billimoria et al., 2008), and spectro-temporal modulation (Moore et al., 2013), which may relate to mechanisms of stream segregation. In neurons that are affected by masking, the specific changes observed in the response to natural signals might indicate targets of auditory attention.

Although the acoustic complexity of natural signals presents a range of technical challenges, they also bestow several advantages. Natural communication

signals, in particular, allow researchers the opportunity to investigate more complex and ethologically relevant auditory behaviors that have adapted in the context of natural acoustic scenes such as individual recognition or mate choice. In principle, these natural functions of communication signals can be exploited for the study of attention. To date, the predominant means of manipulating attention has been to measure neural responses to the same stimuli presented during passive listening and while the animal is actively using them in some kind of auditory task (Fritz et al., 2003; Otazu et al., 2009). While attention almost certainly varies across these two conditions, other behavioral-state variables, such as motivation, arousal, reward expectation, and non-attention driven task demands vary as well. Indeed, rapid shifts in receptive field gain commonly attributed to attention (Fritz et al., 2003), can be recapitulated by varying task demands (David et al., 2012).

Dissociating the effects of attention from those of reward, working memory, arousal, and motor control is difficult. Researchers in the visual system have spent years carefully developing tasks that can control for and measure the allocation of attention (Yantis, 1998). The hallmark of selective attention is that in a highly stimulating environment, the amount of sensory input is simply too much to process simultaneously (James, 1890). Targets are chosen from the scene and attended to, while other parts of the scene are ignored. Task difficulty, captured in measures such as percent accuracy, is often used as evidence that the task requires attention to solve, and attention-related effects have been shown to scale with difficulty (Cate et al., 2009; Falkenberg et al., 2011). Another, perhaps more foolproof design includes

"catch trials" which test the subject's knowledge of the ignored parts of the scene. This is the animal behavior equivalent to asking human subjects if they noticed words or other aspects of the ignored speech in a dichotic listening task, noted by Hillyard and colleagues in their 1973 study as a crucial part of attention-related experimental design (Hillyard et al., 1973). Logic follows that if subjects cannot accurately report on the distracter stream, but can report on the target stream, they are selectively attending to the target. Recently presented experiments using well-controlled auditory tasks indicate that there is growing interest in animal models of auditory attention, particularly in the role of the prefrontal cortex (Rodgers and DeWeese, 2012; Yin et al., 2012), and attention to natural communication signals (Caporello and Gentner, 2012). Furthermore, the application of natural communication signals in auditory attention tasks may provide insights into the mechanisms of attention that play a role in stream segregation and solving the cocktail party problem.

I.3 Conclusion

Throughout the fields of plasticity and attention, we find that despite numerous advancements in uncovering the neural mechanisms underlying the processing of relevant and irrelevant stimuli, there are still many questions left to answer. In the case of auditory attention, there is much work to be done investigating the neural circuits that give rise to the strong, consistent broad-scale modulations seen in human subjects. In the case of plasticity, where much more work has been done on the neural

circuitry and cellular mechanisms related to learning-induced stimulus selectivity, questions still remain regarding where along the auditory hierarchy learning can first modulate stimulus selectivity. And in both fields, little is known about the role irrelevant stimuli play in shaping neural responses.

In the first half of my thesis, I show an emergence of selectivity for natural stimuli in the ascending songbird auditory cortex, and also a novel discovery of response suppression induced by a non-preferred stimulus (noise), which affects cortical regions in a hierarchical manner. Additionally, I show that after associating noise with irrelevance through learning, the suppressive effects of noise are enhanced early on in the cortical hierarchy.

In the second half of my thesis, I show that the neural responses in high levels of auditory cortex are modulated by selective attention to natural vocalizations. In particular, I show that within a subclass of these neurons, responses are related to the active suppression of the ignored stimulus. Furthermore, the magnitude of this suppression is dependent upon the performance of the animal on a trial by trial basis Overall, my work explores not only the way behaviorally relevant stimuli are processed, but also how behaviorally irrelevant stimuli are processed and how their presence in a complex, natural auditory scene affects target stimulus coding. These results move us closer to understanding the circuitry that underlies the essential ability of humans and other animals to detect relevant stimuli in noisy, naturalistic environments.

I.4 Acknowledgements

The Introduction, in part, has been submitted for publication of the material as it may appear in Hearing Research, 2013, under the authorship of Emily L. Caporello and Timothy Q. Gentner. Emily Laura Caporello was the primary investigator and author of this paper.

I.5 Bibliography

Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. Nature neuroscience 14:1067–1074.

Ahveninen J, Hämäläinen M, Jääskeläinen IP, Ahlfors SP, Huang S, Lin F-H, Raij T, Sams M, Vasios CE, Belliveau JW (2011) Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. Proceedings of the National Academy of Sciences of the United States of America 108:4182–4187.

Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin F-H, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated "what" and "where" pathways in human auditory cortex. Proceedings of the National Academy of Sciences of the United States of America 103:14608–14613.

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) "What" and "where" in the human auditory system. Proceedings of the National Academy of Sciences of the United States of America 98:12301–12306.

Alho K, Töttölä K, Reinikainen K, Sams M, Näätänen R (1987) Brain mechanism of selective listening reflected by event-related potentials. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 68:458–470.

Alho K, Vorobyev V a, Medvedev S V, Pakhomov S V, Roudas MS, Tervaniemi M, Van Zuijen T, Näätänen R (2003) Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. Cognitive Brain Research 17:201–211.

Alho K, Vorobyev V a, Medvedev S V, Pakhomov S V, Starchenko MG, Tervaniemi M, Näätänen R (2006) Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. Brain research 1075:142–150.

Altmann CF, Henning M, Döring MK, Kaiser J (2008) Effects of feature-selective attention on auditory pattern and location processing. NeuroImage 41:69–79.

Anderson S, Kraus N (2011) Sensory-Cognitive interaction in the neural encoding of speech in noise: a review. J Am Acad Audiol 21:575–585.

Asari H, Pearlmutter B a, Zador AM (2006) Sparse representations for the cocktail party problem. Journal of neuroscience 26:7477–7490.

Astheimer LB, Sanders LD (2009) Listeners modulate temporally selective attention during natural speech processing. Biological psychology 80:23–34.

Astheimer LB, Sanders LD (2011) Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia 49:3512–3516.

Astheimer LB, Sanders LD (2012) Temporally selective attention supports speech processing in 3- to 5-year-old children. Developmental cognitive neuroscience 2:120–128.

Atiani S, Elhilali M, David S V, Fritz JB, Shamma SA (Univeristy of M (2009) Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron 61:467–480.

Aubin T (2004) Penguins and their noisy world. Anais da Academia Brasileira de Ciências 76:279–283.

Aubin T, Jouventin P (1998) Cocktail-party effect in king penguin colonies. Proceedings of the Royal Society B: Biological Sciences 265:1665–1673.

Bakin JS, Weinberger NM (1996) Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. Proceedings of the National Academy of Sciences of the United States of America 93:11219–11224.

Bao S, Chan V, Merzenich M (2001) Cortical remodelling induced by activity of ventral tegmental dopamine neurons. Nature 412:79-83.

Bao S, Chang EF (2004) Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. Nature neuroscience 7:974–981.

Barlow JS (1969) Some observations on the electrophysiology of timing in the nervous system. Electroencephalography and clinical neurophysiology 27:545.

Bee M a (2008) Finding a mate at a cocktail party: Spatial release from masking improves acoustic mate recognition in grey treefrogs. Animal behaviour 75:1781–1791.

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

Benson DA, Hienz RD (1978) Single-Unit activity in the auditory cortex of monkeys selectivy attending left vs. right ear stimuli. Brain Research 159:307–320.

Besle J, Schevon C a, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. Journal of neuroscience 31:3176–3185.

Bidet-Caulet A, Mikyska C, Knight RT (2010) Load effects in auditory selective attention: evidence for distinct facilitation and inhibition mechanisms. NeuroImage 50:277–284.

Billimoria CP, Kraus BJ, Narayan R, Maddox RK, Sen K (2008) Invariance and sensitivity to intensity in neural discrimination of natural sounds. Journal of neuroscience 28:6304–6308.

Binder J (2000) The new neuroanatomy of speech perception. Brain : a journal of neurology 123 Pt 12:2371–2372.

Binder JR, Liebenthal E, Possing ET, Medler D a, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nature neuroscience 7:295–301.

Blake DT, Heiser MA, Caywood M, Merzenich MM (2006) Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. Neuron 52:371–381.

Bonke D, Scheich H, Langner G (1979) Responsiveness of Units in the Auditory Neostriatum of the Guinea Fowl (Numida meleagris) to Species-Specific Calls and Synthetic Stimuli. Journal of Comparative Physiology A 255:243–255.

Brechmann A, Scheich H (2005) Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cerebral cortex 15:578–587.

Bregman AS (1990) Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press.

Bridwell DA, Srinivasan R (2012) Distinct attention networks for feature enhancement and suppression in vision. Psychological science 23:1151–1158.

Caporello EL, Gentner TQ (2012) Selective attention to natural auditory streams in single cortical neurons of the European starling. In: Society for Neuroscience Meeting, pp 368.29. New Orleans, LA.

Catchpole CK, Slater P (1995) Bird song: Biological themes and variations.

Cate AD, Herron TJ, Yund EW, Stecker GC, Rinne T, Kang X, Petkov CI, Disbrow EA, Woods DL (2009) Auditory attention activates peripheral visual cortex. PloS one 4:e4645.

Chang EF, Rieger JW, Johnson KO, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech representation in human superior temporal gyrus. Nature neuroscience 13:1428–1432.

Cheatham MA, Dallos P (1989) Two-tone suppression in inner hair cell responses. Hearing research 40:187–196.

Cherry EC (1953) Some Experiments on the Recognition of Speech, with One and Two Ears. Journal of the Acoustical Society of America 25:975–979.

Coch D, Sanders LD, Neville HJ (2005) An event-related potential study of selective auditory attention in children and adults. Journal of cognitive neuroscience 17:605–622.

Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. Clinical Neurophysiology 112:565–582.

Darwin CJ, Hukin RW (2000a) Effectiveness of spatial cues, prosody, and talker characteristics. Journal of the Acoustical Society of America 107:970–977.

Darwin CJ, Hukin RW (2000b) Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. The Journal of the Acoustical Society of America 108:335–342.

David S V, Fritz JB, Shamma SA (Univeristy of M (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proceedings of the National Academy of Sciences of the United States of America 109:2144–2149.

De Boer J, Thornton ARD (2007) Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. Hearing research 233:117–123.

Degerman A, Rinne T, Pekkola J, Autti T, Jääskeläinen IP, Sams M, Alho K (2007) Human brain activity associated with audiovisual perception and attention. NeuroImage 34:1683–1691.

Degerman A, Rinne T, Salmi J, Salonen O, Alho K (2006) Selective attention to sound location or pitch studied with fMRI. Brain research 1077:123–134.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons int he macaque. Journal of Neuroscience 4:2051–2062.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annual review of neuroscience 18:193–222.

Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. Journal of neurophysiology 107:78–89.

Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF (2008) Probing category selectivity for environmental sounds in the human auditory brain. Neuropsychologia 46:2776–2786.

Doupe AJ, Konishi M (1991) Song-selective auditory circuits in the vocal control system of the zebra finch. Proceedings of the National Academy of Sciences of the United States of America 88:11339–11343.

Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. Nature neuroscience 3:277–283.

Edeline J-M, Weinberger NM (1993) Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. Behavioral neuroscience 107:82–103.

Falkenberg LE, Specht K, Westerhausen R (2011) Attention and cognitive control networks assessed in a dichotic listening fMRI study. Brain and cognition 76:276–285.

Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. Nature 443:85–88.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science (New York, NY) 291:312–316.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. The Journal of neuroscience : the official journal of the Society for Neuroscience 23:5235–5246.

Fritz JB, Elhilali M, David S V, Shamma SA (Univeristy of M (2007a) Auditory attention--focusing the searchlight on sound. Current opinion in neurobiology 17:437–455.

Fritz JB, Elhilali M, Shamma SA (Univeristy of M (2005a) Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hearing research 206:159–176.

Fritz JB, Elhilali M, Shamma SA (Univeristy of M (2005b) Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. Journal of neuroscience 25:7623–7635.

Fritz JB, Elhilali M, Shamma SA (Univeristy of M (2007b) Adaptive changes in cortical receptive fields induced by attention to complex sounds. Journal of neurophysiology 98:2337–2346.

Fritz JB, Shamma SA (Univeristy of M, Elhilali M, Klein DJ (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature neuroscience 6:1216–1223.

Froemke RC, Merzenich MM, Schreiner CE (2007) A synaptic memory trace for cortical receptive field plasticity. Nature 450:425–429.

Fujiwara N, Nagamine T, Imai M, Tanaka T, Shibasaki H (1998) Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. Cognitive Brain Research 7:99–109.

Galindo-Leon EE, Lin FG, Liu RC (2009) Inhibitory plasticity in a lateral band improves cortical detection of natural vocalizations. Neuron 62:705–716.

Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. Nature 424:669–674.

Giard M-H, Perrin F, Pernier J, Peronnet F (1988) Several attention-related wave forms in auditory areas: a topographic study. Electroencephalography and clinical neurophysiology 69:371–384.

Giesbrecht B, Woldorff MG, Song AW, Mangun GR (2003) Neural mechanisms of top-down control during spatial and feature attention. NeuroImage 19:496–512.

Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. Journal of neurophysiology 89:472–487.

Grady CL, Meter JW Van, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP (1997) modulation of activity in primary and secondary auditory cortex. NeuroReport 8:2511–2516.

Gröger U, Wiegrebe L (2006) Classification of human breathing sounds by the common vampire bat, Desmodus rotundus. BMC biology 4:18.

Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. Journal of neurophysiology 35:96–111.

Hannemann R, Obleser J, Eulitz C (2007) Top-down knowledge supports the retrieval of lexical information from degraded speech. Brain research 1153:134–143.

Hansen JC, Hillyard SA (1978) Endogenous brain potentials associated with selective auditory attention. Electroencephalography and clinical Neurophysiology:277–290.

Hari R, Hfimfilfiinen M, Kaukoranta E, Mfikelfi J, Tiihonen J (1989) Selective listening modifies activity of the human cortex. Experimental Brain Research 74:463–470.

Hauber ME, Woolley SMN, Cassey P, Theunissen FE (2013) Experience dependence of neural responses to different classes of male songs in the primary auditory forebrain of female songbirds. Behavioural brain research 243:184–190.

Hausberger M, Cousillas H (1996) Categorization in birdsong : from behavioural to neuronal responses. Behavioral Processes 35:83–91.

Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cerebral cortex 20:583–590.

Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the human brain. Science 182:177–180.

Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philosophical transactions of the Royal Society of London Series B, Biological sciences 353:1257–1270.

Hocherman S, Benson DA, Goldstein MH, Heffner HE, Hienz RD (1976) Evoked unit activity in auditory cortex of monkeys performing a selective attention task. Brain Research 117:51–68.

Hopf J-M, Boehler CN, Luck SJ, Tsotsos JK, Heinze H-J, Schoenfeld MA (2006) Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. Proceedings of the National Academy of Sciences of the United States of America 103:1053–1058.

Hsu A, Woolley SMN, Fremouw TE, Theunissen FE (2004) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 24:9201–9211.

Hubel DH, Henson CO, Rupert A, Galambos R, Series N, May N (1959) “Attention” Units in the Auditory Cortex. Science 129:1279–1280.

Hugdahl K, Thomsen T, Ersland L, Morten Rimol L, Niemi J (2003) The effects of attention on speech perception: An fMRI study. Brain and Language 85:37–48.

Hugdahl K, Westerhausen R, Alho K, Medvedev S V, Laine M, Hämäläinen H (2009) Attention and cognitive control: unfolding the dichotic listening story. Scandinavian journal of psychology 50:11–22.

Hulse SH, MacDougall-Shackleton S a, Wisniewski a B (1997) Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (Sturnus vulgaris). Journal of Comparative Psychology 111:3–13.

Jääskeläinen IP, Ahveninen J, Andermann ML, Belliveau JW, Raij T, Sams M (2011) Short-term plasticity as a neural mechanism supporting memory and attentional functions. Brain research 1422:66–81.

Jääskeläinen IP, Ahveninen J, Belliveau JW, Raij T, Sams M (2007) Short-term plasticity in auditory cognition. Trends in neurosciences 30:653–661.

James W (1890) The principles of psychology. H. Holt and Company.

Jäncke L, Buchanan TW, Lutz K, Shah NJ (2001) Focused and nonfocused attention in verbal and emotional dichotic listening: an FMRI study. Brain and language 78:349–363.

Jäncke L, Mirzazade S, Shah NJ (1999) Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. Neuroscience letters 266:125–128.

Jeanne JM, Sharpee TO, Gentner TQ (2013) Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns. Neuron 78:352–363.

Jeanne JM, Thompson J V, Sharpee TO, Gentner TQ (2011) Emergence of learned categorical representations within an auditory forebrain circuit. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:2595–2606.

Johnson JA, Zatorre RJ (2005) Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. Cerebral cortex 15:1609–1620.

Johnson JA, Zatorre RJ (2006) Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. NeuroImage 31:1673–1681.

Kaltenbach J a (2006) The dorsal cochlear nucleus as a participant in the auditory, attentional and emotional components of tinnitus. Hearing research 216-217:224–234.

Katz LC, Gurney ME (1981) Auditory responses in the zebra finch's motor system for song. Brain research 221:192–197.

Kauramäki J, Jääskeläinen IP, Sams M (2007) Selective attention increases both gain and feature selectivity of the human auditory cortex. PloS one 2:e909.

Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a “cocktail party”. Journal of neuroscience 30:620–628.

Kiggins JT, Comins JA, Gentner TQ (2012) Targets for a comparative neurobiology of language. Frontiers in evolutionary neuroscience 4:6.

Kilgard MP, Merzenich MM (1998a) Plasticity of temporal information processing in the primary auditory cortex. Nature neuroscience 1:727–731.

Kilgard MP, Merzenich MM (1998b) Cortical Map Reorganization Enabled by Nucleus Basalis Activity. Science 279:1714–1718.

Klinke R, Fruhstorfer H, Finkenzeller P (1968) Evoked responses as a function of external and stored information. Electroencephalography and clinical neurophysiology 25:119–122.

Knight RT, Hillyard SA, Woods DL, Neville HJ (1981) The effects of frontal cortex lesions on event-related potentials during auditory selective attention. Electroencephalography and Clinical Neurophysiology 52:571–582.

Knudsen DP, Gentner TQ (2010) Mechanisms of song perception in oscine birds. Brain and language 115:59–68.

Kompus K, Specht K, Ersland L, Juvodden HT, Van Wageningen H, Hugdahl K, Westerhausen R (2012) A forced-attention dichotic listening fMRI study on 113 subjects. Brain and language 121:240–247.

Kroodsma D, Miller E (1996) Ecology and evolution of acoustic communication in birds.

Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, Iverson P (2006) Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science 9:F13–F21.

Leppelsack H-J, Vogt M (1976) of Comparative Responses of Auditory Neurons in the Forebrain of a Songbird to Stimulation with Species-specific Sounds *. Journal of Comparative Physiology 107:263–274.

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

Ma L, Micheyl C, Yin P, Oxenham AJ, Shamma SA (Univeristy of M (2010) Behavioral measures of auditory streaming in ferrets (Mustela putorius). Journal of Comparative Psychology 124:317–330.

MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W (1998) Auditory scene analysis by European starlings (Sturnus vulgaris): perceptual segregation of tone sequences. The Journal of the Acoustical Society of America 103:3581–3587.

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

Maison S, Micheyl C, Collet L (2001) Influence of focused auditory attention on cochlear activity in humans. Psychophysiology 38:35–40.

Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. The Journal of neuroscience : the official journal of the Society for Neuroscience 3:1039–1057.

Margoliash D (1986) Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. The Journal of neuroscience : the official journal of the Society for Neuroscience 6:1643–1661.

Meliza CD, Margoliash D (2012) Emergence of selectivity and tolerance in the avian auditory cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:15158–15168.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:1–5.

Mesgarani N, David S V, Fritz JB, Shamma SA (Univeristy of M (2008) Phoneme representation and classification in primary auditory cortex. The Journal of the Acoustical Society of America 123:899–909.

Mesgarani N, Fritz JB, Shamma SA (Univeristy of M (2010) A computational model of rapid task-related plasticity of auditory cortical receptive fields. Journal of computational neuroscience 28:19–27.

Metherate R, Intskirveli I, Kawai HD (2012) Nicotinic filtering of sensory processing in auditory cortex. Frontiers in behavioral neuroscience 6:44.

Metherate R, Weinberger NM (1989) Acetylcholine produces stimulus-specific receptive field alterations in cat auditory cortex. Brain research 480:372–377.

Miller JM, Sutton D, Pfingst BE, Ryan AF, Beaton RD, Gourevitch G (1972) Single cell activity in the auditory cortex of rhesus monkeys: behavioral dependency. Science 177:449–451.

Miller-Sims VC, Bottjer SW (2012) Development of auditory-vocal perceptual skills in songbirds. PloS one 7:e52365.

Moore RC, Lee T, Theunissen FE (2013) Noise-invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise Kording KP, ed. PLoS Computational Biology 9:e1002942.

Näätänen R (1992) Attention and brain function. Hillsdale, NJ: Lawrence Erlbaum Associates.

Näätänen R, Michie PT (1979) Early selective-attention effects on the evoked potential: a critical review and reinterpretation. Biological Psychology 8:81–136.

Narayan R, Best V, Ozmeral EJ, McClaine E, Dent M, Shinn-Cunningham BG, Sen K (2007) Cortical interference effects in the cocktail party problem. Nature neuroscience 10:1601–1607.

Neelon MF, Williams J, Garell PC (2011) Elastic Attention: Enhanced, then Sharpened Response to Auditory Input as Attentional Load Increases. Frontiers in human neuroscience 5:41.

Nityananda V, Bee M a (2011) Finding your mate at a cocktail party: frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. PloS one 6:e21191.

O'Leary DS, Andreason NC, Hurtig RR, Hichwa RD, Watkins GL, Ponto LL, Rogers M, Kirchner PT (1996) A positron emission tomography study of binaurally and dichotically presented stimuli: effects of level of language and directed attention. Brain and language 53:20–39.

Ohl FW, Scheich H (2005) Learning-induced plasticity in animal and human auditory cortex. Current opinion in neurobiology 15:470–477.

Okamoto H, Stracke H, Bermudez P, Pantev C (2011) Sound processing hierarchy within human auditory cortex. Journal of cognitive neuroscience 23:1855–1863.

Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C (2007) Attention improves population-level frequency tuning in human auditory cortex. The Journal of Neuroscience 27:10383–10390.

Otazu GH, Tai L-H, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. Nature neuroscience 12:646–654.

Paltoglou AE, Sumner CJ, Hall D a (2009) Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. Hearing research 257:106–118.

Paltoglou AE, Sumner CJ, Hall D a (2011) Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. The European journal of neuroscience 33:1733–1741.

Pasley BN, David S V, Mesgarani N, Flinker A, Shamma SA (Univeristy of M, Crone NE, Knight RT, Chang EF (2012) Reconstructing speech from human auditory cortex. PLoS biology 10:e1001251.

Paus T, Zatorre RJ, Hofle N, Caramanos Z, Gotman J, Petrides M, Evans AC (1997) Time-Related Changes in Neural Systems Underlying Attention and Arousal During the Performance of an Auditory Vigilance Task. Journal of Cognitive Neuroscience 9:392–408.

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. Nature neuroscience 7:658–663.

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

Pfingst BE, Connor TAO, Miller JM (1977) Brain Response Plasticity of Neurons in Auditory Cortex of the Rhesus Monkey. Experimental Brain Research 29:393–404.

Phan ML, Pytte CL, Vicario DS (2006) Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. Proceedings of the National Academy of Sciences of the United States of America 103:1088–1093.

Picton TW, Hillyard SA (1974) Human auditory evoked potentials. II: effects of attention. Electroencephalography and clinical Neurophysiology 36:191–199.

Pienkowski M, Eggermont JJ (2011) Cortical tonotopic map plasticity and behavior. Neuroscience and biobehavioral reviews 35:2117–2128.

Polley DB, Heiser M a, Blake DT, Schreiner CE, Merzenich MM (2004) Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. Proceedings of the National Academy of Sciences of the United States of America 101:16351–16356.

Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. Journal of neuroscience 26:4970–4982.

Pritchard WS (1981) Psychophysiology of P300. Psychological bulletin 89:506–540.

Raksin JN, Glaze CM, Smith S, Schmidt MF (2012) Linear and nonlinear auditory response properties of interneurons in a high-order avian vocal motor nucleus during wakefulness. Journal of neurophysiology 107:2185–2201.

Rämä P, Relander-Syrjänen K, Carlson S, Salonen O, Kujala T (2012) Attention and semantic processing during speech: an fMRI study. Brain and language 122:114–119.

Rasmusson DD (2000) The role of acetylcholine in cortical synaptic plasticity. Behavioural brain research 115:205–218.

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

Recanzone GH, Schreiner CE, Merzenich MM (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. The Journal of Neuroscience 13:87–103.

Reed AC, Riley J, Carraway RS, Carrasco A, Perez CA, Jakkamsetti V, Kilgard MP (2011) Cortical map plasticity improves learning but is not necessary for improved performance. Neuron 70:121–131.

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185.

Rif J, Hari R, Matti SH, Sams M (1991) Auditory attention affects two different areas in the human supratemporal cortex. Electroencephalography and clinical Neurophysiology 79:464–472.

Rihs T a, Michel CM, Thut G (2009) A bias for posterior alpha-band power suppression versus enhancement during shifting versus maintenance of spatial attention. NeuroImage 44:190–199.

Rinne T, Balk MH, Koistinen S, Autti T, Alho K, Sams M (2008) Auditory selective attention modulates activation of human inferior colliculus. Journal of neurophysiology 100:3323–3327.

Rinne T, Kirjavainen S, Salonen O, Degerman A, Kang X, Woods DL, Alho K (2007) Distributed cortical networks for focused auditory attention and distraction. Neuroscience letters 416:247–251.

Rodgers C, DeWeese MR (2012) Neuronal correlates of selective auditory attention in rat prefrontal and sensory cortices. In: Society for Neuroscience Meeting, pp 599.20. New Orleans, LA.

Ruggero MA (1992) Responses to sound of the basilar the mammalian of cochlea. Current opinion in neurobiology 2:449–456.

Rutkowski RG, Weinberger NM (2005) Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. Proceedings of the National Academy of Sciences of the United States of America 102:13664–13669.

Ryan AF, Miller JM, Pfingst BE, Martin GK (1984) Effects of reaction time performance on single-unit activity in the central auditory pathway of the rhesus macaque. Journal of Neuroscience 4:298–308.

Sachs MB, Kiang NY (1968) Two-tone inhibition in auditory-nerve fibers. The Journal of the Acoustical Society of America 43:1120–1128.

Sagi B, Nemat-Nasser SC, Kerr R, Hayek R, Downing C, Hecht-Nielsen R (2001) A biologically motivated solution to the cocktail party problem. Neural computation 13:1575–1602.

Salmi J, Rinne T, Degerman A, Salonen O, Alho K (2007) Orienting and maintenance of spatial attention in audition and vision: multimodal and modality-specific brain activations. Brain structure & function 212:181–194.

Salmi J, Rinne T, Koistinen S, Salonen O, Alho K (2009) Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. Brain research 1286:155–164.

Sanders LD, Newport EL, Neville HJ (2002) Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nature neuroscience 5:700–703.

Sarter M, Hasselmo ME, Bruno JP, Givens B (2005) Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. Brain Research Reviews 48:98-111

Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR, Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human EEG alpha activity. The European journal of neuroscience 22:2917–2926.

Scheich H, Brechmann A, Brosch M, Budinger E, Ohl FW, Selezneva E, Stark H, Tischmeyer W, Wetzel W (2011) Behavioral semantics of learning and crossmodal processing in auditory cortex: the semantic processor concept. Hearing research 271:3–15.

Scheich H, Langner G, Bonke D (1979) Responsiveness of units in the auditory neostriatum of the guinea fowl (Numida meleagris) to species-specific calls and synthetic stimuli. Journal of Comparative Physiology ? A 132:257–276.

Schmidt AKD, Römer H (2011) Solutions to the cocktail party problem in insects: selective filters, spatial release from masking and gain control in tropical crickets. PloS one 6:e28593.

Sellick PM, Russell IJ (1979) Two-tone suppression in cochlear hair cells. Hearing research 1.

Shomstein S, Yantis S (2006) Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. Journal of neuroscience 26:435–439.

Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415:318–320.

Smith DW, Aouad RK, Keil A (2012) Cognitive task demands modulate the sensitivity of the human cochlea. Frontiers in psychology 3:30.

Sokolov A, Pavlova M, Lutzenberger W, Birbaumer N (2004) Reciprocal modulation of neuromagnetic induced gamma activity by attention in the human visual and auditory cortex. NeuroImage 22:521–529.

Spieth W, Curtis JF, Webster JC (1954) Responding to one of two simultaneous messages. Journal of the Acoustical Society of America 26:391–396.

Stripling R, Kruse AA, Clayton DF (2001) Development of Song Responses in the Zebra Finch Caudomedial Neostriatum : Role of Genomic and Electrophysiological Activities ABSTRACT : Journal of neurobiology 48:163–180.

Suga N, Ma X (2003) Multiparametric corticofugal modulation and plasticity in the auditory system. Nature reviews Neuroscience 4:783–794.

Sutter ML, Schreiner CE, McLean M, O'Connor KN, Loftus WC (1999) Organization of Inhibitory Frequency Receptive Fields in Cat Primary Auditory Cortex. Journal of neurophysiology 82:2358–2371.

Tallon-Baudry C, Bertrand O, Delpuech C, Permier J (1997) Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans. The Journal of Neuroscience 17:722–734.

Tanedo JR, Bhagat SP (2008) Attentional Control of Cochlear Micromechanics: The Role of the Medial Olivocochlear Efferent System. Perspectives on Hearing and Hearing Disorders: Research and Diagnostics 12:17–21.

Terleph TA, Lu K, Vicario DS (2008) Response properties of the auditory telencephalon in songbirds change with recent experience and season. Tanimoto H, ed. PloS one 3:e2854.

Theunissen FE, Shaevitz SS (2006) Auditory processing of vocal sounds in birds. Current opinion in neurobiology 16:400–407.

Thompson J V, Gentner TQ (2010) Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. Journal of neurophysiology 103:1785–1797.

Tian B, Reser DH, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science (New York, NY) 292:290–293.

Tiitinen H, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, Näätänen R (1993) Selective attention enhances the auditory 40-Hz transient response in humans. Nature 364:59–60.

Van der Kant A, Derégnaucourt S, Gahr M, Van der Linden A, Poirier C (2013) Representation of early sensory experience in the adult auditory midbrain: implications for vocal learning. PloS one 8:e61764.

Wang X, Merzenich MM, Beitel R, Schreiner CE, Beitel R, Schreiner E (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset : temporal and spectral characteristics Representation of a Species-Specific Vocalization in the Primary Auditory Cortex of the Common Marmoset : Tempo. Journal of neurophysiology 74:2685–2706.

Wang Y, Brzozowska-Prechtl A, Karten HJ (2010) Laminar and columnar auditory cortex in avian brain. Proceedings of the National Academy of Sciences of the United States of America 107:12676–12681.

Westerhausen R, Moosmann M, Alho K, Belsby S-O, Hämäläinen H, Medvedev S V, Specht K, Hugdahl K (2010) Identification of attention and cognitive control networks in a parametric auditory fMRI study. Neuropsychologia 48:2075–2081.

Winer JA (2005) Decoding the auditory corticofugal systems. Hearing research 207:1–9.

Winter P, Funkenstein HH (1973) The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. (Saimiri sciureus). Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale 18:489–504.

Woldorff MG, Gallen CC, Hampson S a, Hillyard S a, Pantev C, Sobel D, Bloom FE (1993) Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proceedings of the National Academy of Sciences of the United States of America 90:8722–8726.

Woldorff MG, Hansen JC, Hillyard SA (1987) Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential. Electroencephalography and Clinical Neurophysiology Supplement 40:146–154.

Woldorff MG, Hillyard SA (1991) Modulation of early auditory processing during selective listening to rapidly presented tones. Electroencephalography and clinical Neurophysiology 79:170–191.

Wollberg Z, Newman JD (1972) Auditory Cortex of Squirrel Monkey: Response Patterns of Single Cells to Species-Specific Vocalizations. Science 175:212–214.

Woods DL, Alain C (2001) Conjoining three auditory features: an event-related brain potential study. Journal of cognitive neuroscience 13:492–509.

Woods DL, Alain C, Diaz R, Rhodes D, Ogawa KH (2001) Location and frequency cues in auditory selective attention. Journal of Experimental Psychology 27:65–74.

Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I, Kang X (2009) Functional maps of human auditory cortex: effects of acoustic features and attention. PloS one 4:e5183.

Woolley SC, Doupe AJ (2008) Social context-induced song variation affects female behavior and gene expression. PLoS biology 6:e62.

Wu C-T, Weissman DH, Roberts KC, Woldorff MG (2007) The neural circuitry underlying the executive control of auditory spatial attention. Brain Research 1134:187–198.

Yantis S (1998) Control of Visual Attention. In: Attention (Pashler HE, ed), pp 223–292. Psychology Press.

Yin P, Fritz JB, Shamma SA (Univeristy of M (2012) Responses of ferret frontal cortex neurons in a long-term auditory memory task. In: Society for Neuroscience Meeting, pp 460.29. New Orleans, LA.

Yin P, Mishkin M, Sutter ML, Fritz JB (2008) Early stages of melody processing: stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R. Journal of neurophysiology 100:3009–3029.

Zatorre RJ, Mondor TA, Evans AC (1999) Auditory Attention to Space and Frequency Activates Similar Cerebral Systems. NeuroImage 10:544–554.

Chapter 1

1.1 Introduction

To guide behavior, vertebrate nervous systems must encode the changing set of natural signals that carry behaviorally relevant information. Because these signals are often complex, their representation is thought to require the integration of several simpler features. The canonical model for feature integration in cortex is hierarchical convergence, in which the weighted combination of responses from neurons tuned to simpler features explains the response of neurons tuned to more complex features. These classical "receptive fields" (Hubel and Wiesel, 1959, 1962) model a wide variety of responses in visual cortices, but their generality is not well understood. Moreover, observations in both the visual and auditory systems (Imig and Adria´n, 1977; Fuzessery and Feng, 1982; Morrone et al., 1982; Carandini and Heeger, 1994; Nelken et al., 1994; Sutter et al., 1999a; Fitzpatrick, 2000; Freeman et al., 2002; Bonin et al., 2005; Zoccolan et al., 2005; Schneider and Woolley, 2011) show that excitatory stimuli (often very weak excitatory stimuli) can modify a neuron's response to other stimuli in a non-linear fashion. Such responses cannot be predicted by receptive field models, and so are termed "non-classical". One mechanism thought to underlie some non-classical receptive fields is a second form of feature integration called divisive normalization (Heeger, 1991, 1992; Carandini et al., 1997), where instead of adding together weighted responses, one response is normalized (divided by) another.

Behavior also shapes cortical responses. Indeed it is difficult to understand the representation of natural signals without also accounting for natural behavior, because the set of signals relevant to an animal at any given time depends, at least in part, on experience. The auditory cortex in particular shows robust experience-dependent plasticity (Galambos, 1956; Recanzone et al., 1993) and this plasticity is tied directly to receptive field changes (Bakin and Weinberger, 1990; Fritz et al., 2003). Studies showing such changes have typically used sine tones (Galambos, 1956; Bakin and Weinberger, 1990; Recanzone et al., 1993; Fritz et al., 2003; Rutkowski and Weinberger, 2005; Polley et al., 2006) or other artificial stimuli (Bao and Chang, 2004; Fritz et al., 2005, 2007), however, and so it is difficult to understand them in the context of either classical (hierarchical) or non-classical (suppressive) models for natural feature integration.

Here, we investigate how learning shapes responses to natural and artificial signals across the songbird auditory cortex (Wang et al., 2010), and how these responses fit into classical or non-classical response models. Songbirds and primates share a similar functional hierarchy of cortical regions in which the acoustic characteristics of conspecific vocal communication signals selectively drive responses in regions post-synaptic to the primary thalamo-recipient zones (Winter and Funkenstein, 1973; Leppelsack and Vogt, 1976; Margoliash, 1983; Doupe and Konishi, 1991; Belin et al., 2000; Scott, 2000; Petkov et al., 2008). In songbirds, auditory experience significantly alters the representations of conspecific songs in multiple regions throughout the auditory cortex (Gentner and Margoliash, 2003;

Thompson and Gentner, 2010; Jeanne et al., 2011). The encoding principles that support the representation of these complex acoustic signals are not understood. Our results here support a hierarchical organization for auditory cortex that includes classic receptive fields tuned to the complex acoustic structure of natural songs, and demonstrate a novel, non-classical, form of response suppression in the secondary regions. In at least one sub-region of the songbird cortex, Field L1, this response suppression is learning-dependent and stimulus selective.

1.2 Results

We recorded auditory evoked responses from a total of 274 sites (150 single-units, 124 multi-units; Table 1.1; Methods) in four different forebrain regions (L2a, L1, L3 and CM; Fig. 1.1) of wild-caught, urethane anesthetized starlings (n=9). At each site, we presented the same set of 26 stimuli that included all Song+Silence, Noise+Silence, and Song+Noise stimuli (Fig. 1.2; Methods). On average, both song and noise were effective auditory stimuli in all regions, evoking mean firing rates significantly above the spontaneous firing rates measured when no stimulus was presented (Table 1.2).

The design of our stimulus set and its combination with operant recognition training allows us to examine how different regions respond to song, noise, and their combination, and how such responses within the different regions are affected by recognition learning. To do this, we divide the analysis into two parts. First, we

characterize the changes in general response tuning across the auditory forebrain in the absence of any controlled learning, using the song and noise stimuli and their combination. In the second part, we examine how responses to song, noise, and their combination are affected by explicit training in recognition learning task.

Song Selectivity Across Regions

To characterize how tuning for general acoustic properties of conspecific song differs across auditory forebrain regions independent of controlled learning, we examined the stimulus-evoked responses recorded at 66 auditory sites (Table 1.1) in forebrain regions L2a, L1, L3 and CM of untrained starlings (n=5). Figure 1.3 (a-d) shows examples of spiking activity evoked by Song+Silence, Noise+Silence, and Song+Noise stimuli at single unit sites in Fields L2a, L1, L3 and CM, and highlights the variability in overall spike rates and responses to different stimuli between regions. To compare responses to different stimuli between each region, we parsed songs into constituent motifs and converted the raw spike rates for each motif to z-scores centered on the mean response to all motifs and noise at each site (Methods), and then averaged over all sites in a given region.

Figure 1.3 (e-h) shows population data for each region. In Fields L2a and L1, the mean normalized spike rates evoked by the Song+Silence and Noise+Silence stimuli (Fig. 1.3e,f, blue and red bars) were not significantly different from one another. In L3 and CM, however, Song+Silence stimuli evoked significantly higher spike rates than the Noise+Silence stimuli (L3: $p<0.0001$, CM: $p = 0.0010$, Tukey-

Kramer post-hoc on rmANOVA; Fig. 1.3g,h, blue and red bands). Because the Noise+Silence and Song+Silence stimuli are matched in power and cover the same range of frequencies (Methods), simple differences in acoustic power or frequency tuning of the recorded neurons cannot explain these observed biases in the different regions. Instead these response differences likely reflect tuning for the spectrotemporal features of song that emerges along the processing pathway. The increasing response bias for song relative to noise between the thalamo-recipient Field L2a and its efferent targets, particularly L3 and CM, is consistent with data from other songbird species (Sen et al., 2001; Grace et al., 2003; Amin et al., 2007).

To examine how these two general classes of signals, song and noise, with very different acoustic structures and behavioral significance are integrated, we also presented the Song+Noise stimuli. Recall that each Song+Noise stimulus was the combination of one Song+Silence stimulus and one Noise+Silence stimulus (Methods), resulting in alternating, non-overlapping frequency bands of song and noise. A Song+Noise stimulus has twice the total acoustic power of a Song+Silence or Noise+Silence stimulus. A neuron with a classical receptive field, where responses are the weighted sum of simpler inputs, would produce a response to Song+Noise stimuli greater than to Song+Silence and Noise+Silence stimuli alone. Consistent with this model, the spike rates elicited by L2a neurons in response to Song+Noise stimuli are on average significantly higher than those evoked by both the Song+Silence and Noise+Silence stimuli ($p<0.0001$ both cases, Tukey-Kramer post-hoc after rmANOVA; Fig. 1.3a,e). Responses in L1 show a similar pattern, although the spike

rates evoked by the Song+Noise stimuli were not significantly greater than those to song or noise alone ($p=0.1250$, rmANOVA; Fig. 1.3b,f).

In contrast, L3 and CM showed qualitatively different responses to the combined noise and song. In these two regions the Song+Noise stimuli evoked significantly lower spike rates than the Song+Silence stimuli (L3, $p<0.0001$ and CM $p=0.0010$, Tukey-Kramer post-hoc on rmANOVA; Fig. 1.3c,d,g,h). This result is surprising given that the Noise+Silence stimuli by themselves evoke moderately significant excitatory responses in these regions (Table 1.2), on average driving firing rates above baseline levels. Thus, linear summation cannot account for the observed suppression. Furthermore, the suppression in L3 and CM is specific to the noise stimuli. In principle, it might be adding any stimulus to the silent frequency bands in the Song+Silence stimuli would lead to a reduced firing rate. To test this possibility, we recorded neural responses elicited by unmodified song segments (Full Song, i.e., without any frequency bands removed or replaced; see Methods) in a subset of neurons (L1 n=16, L3 n=17, CM n=46). There were no significant differences in any regions between the evoked responses to Full Song and Song+Silence stimuli (all regions $p>0.05$, student t-test; Data not shown). Thus, neurons in L3 and CM appear to possess nonlinear response properties that allow some excitatory stimuli to modulate the response to other excitatory stimuli. This type of complex response profile suggests non-classical normalization akin to that observed in the visual system (Morrone et al., 1982; Heeger, 1992; Carandini and Heeger, 1994; Carandini et al.,

1997; Zoccolan et al., 2005) and elsewhere in the auditory system using narrowband tone stimuli (Nelken et al., 1994; Sutter et al., 1999a).

Learning Alters General Song Selectivity in L1

To ascertain how learning might affect the observed pattern of song selectivity or response suppression, we trained a second group of starlings (N=4) using an operant conditioning procedure to classify four Song+Noise stimuli (Methods). In all four training stimuli presented to a given bird, the same noise always occurred in the same five frequency bands (Fig. 1.4a,b). To respond correctly, therefore, the birds had to rely on the five frequency bands containing song. As a result, behavioral relevance for each bird was tied to the specific frequency bands that carried song information and to the spectral features therein that distinguished each song exemplar. The frequency band that contained song information and the explicit song exemplars used for training were counter balanced across birds (Fig. 1.4b). Subjects learned this task easily, showing accurate classification with d' > 1 consistently after a mean of 31.8 ± 3.4 100-trial blocks. After performance reached asymptote (mean d' over final 10 blocks = 3.10 ± 1.40), we anesthetized each bird with urethane and recorded stimulus-evoked extracellular action potentials from sites (n=208; Table 1.1) in Fields L1, L2a, L3 and CM (Fig. 1.2). At each recording site, we presented the same set of 26 modified song and noise stimuli as described for the untrained birds. The only difference between the trained and untrained subject groups was the familiarity of the

specific spectrotemporal features in each song and each animal's experience in extracting those features from specific, i.e. task-relevant, frequency bands.

In regions L2a, L3, and CM, we observed nearly identical response preferences for the Song+Silence, Song+Noise and Noise+Silence stimuli from sites in trained and untrained birds (Fig. 1.4c,e,f). At sites in Field L1, however, learning significantly altered the responses evoked by the different stimulus classes ($p=0.0116$, mixed effects rmANOVA, interaction between training and stimulus type; Fig. 1.4d). While there were no significant differences between mean spike rates evoked by Song+Silence, Song+Noise, and Noise+Silence stimuli at L1 sites in untrained birds, in trained birds both the Song+Silence and Song+Noise stimuli evoked significantly higher spike rates than the Noise+Silence stimuli ($p<0.0001$, Tukey-Kramer post-hoc on rmANOVA, Fig. 1.4d). This "learned" pattern of responses closely resembles the pattern observed in L3 and CM in untrained and trained birds (Fig. 1.3g,h; Fig. 1.4e,f). As in the case of L3 and CM neurons, this pattern cannot be explained by noise suppressing responses below baseline (Fig. 1.4g). Thus, learning to recognize song in the presence of noise produces a bias in L1 neurons for song over noise, and induces non-classical normalization of the song response by noise reminiscent of that observed in regions L3 and CM of untrained subjects.

Suppression and Facilitation Vary with Motif Response Strength

Previous studies show that the degree of response normalization varies with stimulus drive, with normalization having, in absolute terms, a smaller effect on

stimuli that drive a neuron poorly, and a greater effect on stimuli that drive the neuron more strongly (at least until a certain response threshold; Morrone et al., 1982; Freeman et al., 2002; Bonin et al., 2005).To address whether this is true for our data set, for each site we divided the rank-ordered firing rates to the Song +Silence versions all motifs into 10 bins (see Methods), then calculated the mean response to the same groups of motifs when noise was present (Song+Noise versions). These comparisons are shown in Figure 1.5. In all regions, the effect of noise on firing rate varied significantly and systematically as a function of motif response strength (Figure 1.5a-d, $p<0.0001$in all cases, rmANOVA).

In L2a, combining noise with Song+Silence motifs facilitates responses (above the response to noise alone) for most motif groups (Fig. 1.5a). The added noise never suppresses L2a responses, and the response facilitation is strongest for those Song+Silence motifs that evoke the weakest responses. In contrast, and consistent with our previous observations (Figs. 1.3,1.4), combining noise and song had much different effects in L1, L3, and CM (Fig. 1.5b-d). For Song+Silence motifs that elicit firing rates equal to or less than the Noise+Silence response, combining noise with song results in a response equal to noise alone in these regions (suggesting masking rather than facilitation). More importantly, for Song+Silence motifs that elicit responses greater than Noise+Silence, the combination of these two stimuli is suppressive (i.e. rates are reduced relative to Song+Silence), and the strength of this suppression is strongest for the Song+Silence motifs that evoke the strongest responses.

Given the mechanisms of noise addition observed in these post-thalamorecipient regions – masking when song responses are small, and suppression when song responses are large - we might predict that a neuron which responds strongly to Noise+Silence (at a spike rate greater than most Song+Silence motifs) would, due to the masking effect of noise on low-response motifs as described above, respond to many Song+Noise motifs at a rate approximately equal to noise alone. Therefore, when responses are collapsed over all motifs, the response profile would show approximately equal response drive for Noise+Silence and Song+Silence stimulus on average, and a slightly larger response to Song+Noise stimuli, as was observed in L1 data from untrained birds (Fig 1.3f). However, if the Noise+Silence response is sufficiently small compared to the Song+Silence response, the average response to each stimulus class would mostly reflect the suppressive effect of combining noise with a song stimulus that elicits a stronger response. In this case Song+Noise stimuli would appear to elicit responses, on average, less than Song+Silence stimuli and greater than Noise+Silence stimuli (as observed in L1 data from trained birds, Fig. 1.4d, and all L3 and CM data, Figs 1.3g,h, 1.4e,f). Therefore, the changes in the response profile observed in trained and untrained L1 data can be described by an increased distance between the response to Song+Silence motifs and the response to Noise+Silence.

Indeed, in Field L1 there is a significant effect of experience in the overall magnitude of noise suppression ($p = 0.0307$, rmANOVA; Fig. 1.5e). In theory, this learning-dependent change may reflect an increase in the strength of the neural

response evoked by Song+Silence motifs, a decrease in the strength of the neural response evoked by noise, an increase in the magnitude of noise suppression for a given Song+Silence response, or a combination of these three possibilities. Distinguishing between these possible mechanisms of altering stimulus selectivity requires a comparison of absolute spike rates elicited by L1 neurons in trained and untrained subjects. However, since many variables likely contributed to differences in absolute spike rates across these populations, differentiating between these mechanisms will require additional experimentation.

Recognition Learning and Noise Suppression Interact in L1

Previous work indicates that training can induce response biases for behaviorally relevant songs and motifs in several higher order regions in the starling auditory forebrain (Gentner and Margoliash, 2003; Thompson and Gentner, 2010; Jeanne et al., 2011). To examine how learning might bias neuronal responses to specific stimuli in the present study, we grouped evoked responses according to whether the stimulus was rewarded, unrewarded, or not presented (i.e. unfamiliar) during training (Fig. 1.6). To focus initially on any differential responses to behaviorally relevant motifs (Fig. 1.6a-d), we collapsed responses for each pair of Song+Noise and Song+Silence stimuli originally constructed from same motif (i.e., differing only by the presence or absence of noise; see Methods). Sites in Fields L3 and L2a showed no significant differences in the spike rates evoked by rewarded, unrewarded, and unfamiliar motif stimuli. Consistent with previous results (Gentner

and Margoliash, 2003; Jeanne et al., 2011), CM sites showed a significant modulation in mean spike rates between rewarded, unrewarded, and unfamiliar motifs ($p < 0.0001$ rmANOVA; Fig. 1.6d), with rewarded motifs eliciting the strongest responses. Responses also differed significantly between rewarded, unrewarded, and unfamiliar motifs in Field L1 ($p=0.0042$ rmANOVA; Fig. 1.6b), but in the direction opposite that for CM, with rewarded motifs eliciting the lowest firing rates.

Having established that learning can modify responses in CM and L1, we next asked what components of the stimuli might drive these learning effects. Because the birds learned to recognize song in the presence of non-overlapping noise, it is possible that the learning-driven response biases observed in CM and L1 are specific to only the Song+Noise stimuli. Alternatively, because only the song (not the noise) provides information about the correct operant response during training, the effects of learning may be indifferent to the presence or absence of noise in the Song+Noise and Song+Silence stimuli. This was the case in CM, where the average increase in spike rates observed for Song+Silence over Song+Noise stimuli (Fig. 1.4f) was similar regardless of whether the song components were rewarded, unrewarded, or unfamiliar (Fig. 1.6e). In L1, however, the significantly lower spike rates evoked by rewarded motifs on average (Fig. 1.6b) was observed only in the presence of noise (Fig. 1.6f), while responses to unrewarded and unfamiliar motifs were indifferent to the presence of noise ($p=0.0078$ rmANOVA, interaction between stimulus class and training class). Thus, the learning-dependent response properties of L1 reflect the behaviorally

relevant acoustics of the rewarded songs and the way in which those informative song components combine with non-overlapping noise.

Task-Relevant Frequency Bands are Biased in L1

One hallmark of experience dependent plasticity in mammalian primary auditory cortex is the over-representation of neurons tuned to sine tone-frequencies made relevant by behavioral training. In this study, information (song) required to solve the discrimination task was restricted to specific frequencies (Fig. 1.4b). This design enabled us to ask whether auditory neurons in trained birds were biased for behaviorally relevant frequencies (i.e., frequency bands that contained song during training) while still using stimuli with natural features to control behavior. We label the frequency bands that contained song during training "task-relevant", and the frequency bands that contained noise during training "task-irrelevant". Recall that the set of frequency bands made relevant by the song recognition learning was counterbalanced across birds (Fig. 1.4b, Methods). Thus, half of all the Song+Silence stimuli contained song in the task-relevant frequencies, while the other half contained song in the task-irrelevant frequencies (Fig. 1.2a). Similarly, the two Noise+Silence stimuli have noise in either the task-relevant or -irrelevant frequency bands for each subject (Fig. 1.2a). At sites in L1, stimuli with acoustic power in only the task-relevant bands evoked significantly higher firing rates than stimuli with equal acoustic power in only the task-irrelevant bands ($p=0.0146$, student t-test; Fig. 1.7b), regardless of whether the frequency bands contained song or noise (Fig. 1.7e). No other region

showed systematic response differences between the task-relevant and -irrelevant frequency bands (Fig. 1.7a,c,d). Because the actual frequency bands containing task-relevant information (song) were counterbalanced across subjects, it is unlikely that these results can be explained by a bias for specific frequencies within Field L1. However, as a control, we compared responses evoked by stimuli with power in only odd-numbered frequency bands to those with power in only the even-numbered bands, and found no significant differences in the L1 ($p=0.2588$, student t-test; data not shown).

1.3 Methods

Subjects. Nine adult (5 female, 4 male) European starlings, Sturnus vulgaris, wild-caught in southern California, served as subjects. Prior to any training or testing, starlings were housed in large mixed-sex aviaries with food and water available ad libitum. Daily light-dark schedules always matched natural photoperiods. All procedures were conducted in accordance with the University of California, San Diego IACUC approval.

Stimuli. We derived stimuli from six samples of starling song (duration: 9.1s – 10.7s), each drawn without overlap from libraries of the longer continuous song bouts of four male starlings. The portions of the frequency spectrum in each stimulus that contained song were directly controlled. We first restricted the overall frequency range of each of song stimulus by computing its spectrogram and setting the real portion of

each coefficient outside the desired overall frequency range (800Hz - 9000Hz) to zero. We then divided the remaining frequencies into ten bands (equally-spaced logarithmically) and set all the coefficients in either the odd or even bands of each song to zero, and inverted the spectrogram to recover the filtered time-domain waveform. The foregoing filtering process produced stimuli with alternating frequency bands of birdsong and silence, which we refer to as Song+Silence stimuli (Fig. 2). From each of the six original song stimuli, we created two Song+Silence stimuli (12 total), one with song in the odd bands (and silence in the even bands) and one with song in the even bands (and silence in the odd bands). All 12 Song+Silence stimuli were scaled to equal total power (mean RMS = 70.0 dB). We applied the same filtering process to a 10-s bandpass (800 - 9000Hz) white noise signal to produce a pair of noise signals with alternating frequency bands of noise and silence, which we refer to as Noise+Silence stimuli (Fig. 1.2). The average power in each noise band was matched to the mean RMS for the corresponding bands averaged over all six original song samples (mean RMS per band = 59.3 dB, range = [49.4 65.4]. Finally, we produced a third set of stimuli, referred to as Song+Noise stimuli (Fig. 1.2), by combining the twelve Song+Silence stimuli with the complementary Noise+Silence stimuli so that the non-overlapping frequency bands alternated between song and noise. All stimuli used in this study were novel to the subjects at the start of the experiment. All signal processing was done using custom routines in Matlab.

Behavioral Song Recognition Training. Four subjects (2 male, 2 female) were trained to recognize subsets of the Song+Noise stimuli using operant conditioning

techniques. For training, subjects were isolated in sound attenuation chambers (Acoustic Systems, Austin TX) equipped with an operant panel (Fig. 1.4a). Subjects learned to peck at fixed locations on the panel (response port) to received food through a series of shaping procedures (Gentner and Hulse, 2000; Gentner and Fenn, 2006). We then trained the subjects to classify four Song+Noise stimuli divided into two classes (two stimuli per class) using a "Go/No-go" operant procedure (Gentner and Margoliash, 2003; Gentner and Fenn, 2006). Subjects initiated a training trial by pecking at the response port triggering the presentation of one of four Song+Noise stimuli, which then played to completion. After the stimulus presentation ended, the subject could respond either by pecking into the response port within 2s ("go" response) or by withholding a peck ("no-go" response). We rewarded "go" responses to two of the four training stimuli (rewarded class) with 2-s access to food. A go response to either of the remaining two stimuli (unrewarded class) initiated variable duration time out (range: 10s – 40s), during which the house light was extinguished and food remained inaccessible. We did not reinforce correct no-go responses. Responses during the stimulus presentation had no operant consequence. The stimulus exemplar presented for each trial was selected at random with replacement. Starlings of both sexes readily learn to classify of conspecific songs with these procedures (Gentner and Margoliash, 2003; Gentner and Fenn, 2006). In the present study, because the white noise in each Song+Noise stimulus was identical, the only information available to the subject for correct classification was the specific acoustic information contained in the five bands of conspecific song. We refer to those bands

containing informative acoustic signals (i.e. song) as task-relevant bands and those containing no information (i.e. noise) as task-irrelevant. The frequency bands deemed task-relevant and the song samples assigned to each class were counterbalanced across subjects. Access to the food hopper was controlled entirely by the subject's response on each trial. Water was available ad libitum. Subjects completed a mean of 6.07 ± 1.32 100-trial blocks per day. At the time of recording, subjects had completed an average of 102.5 ± 10.9 100-trial blocks (range: 76.8-129.9).

Electrophysiology. One day prior to recording, we anesthetized each bird (4 trained, 5 untrained) with isoflurane (1.5-3.0%) and attached a metal pin to the skull with dental acrylic. We verified full auditory and behavioral recovery in the operant apparatus prior to recording. On the recording day, the animal was anesthetized (20% urethane, 7ml/kg IM, in 3 injections over 90 minutes), wrapped in a cloth jacket, then head-fixed to a stereotaxic apparatus inside a sound-attenuating recording chamber (Acoustic Systems) similar to that used for operant training. Urethane has been reported to induce a "moderate" reduction in the total number of Field L sites excited by both tones and song stimuli, though temporal response profiles remain intact (Capsius and Leppelsack, 1996). The effects of urethane are more pronounced in CM, but again appeared as a "linear" decrease in the response magnitude rather than a change in selectivity (Capsius and Leppelsack, 1996). Thus, while urethane may affect overall auditory responsiveness in these regions, there is no evidence that its presence alters stimulus selectivity in any region tested.

A small craniotomy and durotomy approximately 2400 um rostral and 1000 um lateral to the bifurcation of the saggital sinus permitted electrode access to the target brain region. We coated a 16-channel tetrode array (model A2x2-tet-3mm-150-150, site area 121 or 312 μm^2; NeuroNexus Technologies) with fluorescent diI crystals (Sigma-Aldrich) before lowering the electrode into the brain. Continuous voltage extracellular waveforms were amplified by a 16-channel AC amplifier (AM systems Inc.), and digitized (Spike 2, CED) for later offline analysis. To locate auditory sites [criterion: mean firing rate 2 standard deviations above baseline or trial-by-trial spike train correlation 2 standard deviations above baseline], we pseudo-randomly presented Song+Noise, Song+Silence, and Noise+Silence stimuli, and segments of unfamiliar conspecific song as the electrode was slowly advanced. When a putative auditory neuron was isolated, we played all Song+Noise, Song+Silence, and Noise+Silence stimuli to the anesthetized animal. At each recording site we presented, randomly without replacement, several repetitions (mean 7.67 ± 0.13) of the 26 song and noise stimuli used in the present analysis. In a subset of neural recordings, we also presented the six original samples of starling song with all frequency information intact. Software (Spike2, CED) controlled stimulus playback and recorded time aligned extracellular voltage traces at all electrode sites. Offline, we extracted putative action potentials from the waveform using threshold-triggered template matching, and clustered putative action potential waveforms with a combination of PCA and changes in waveform shape over time (Spike2, CED). We confirmed isolation quality using the ISI distribution of each putative single or multi-unit. Waveform clusters that could be

well-separated from other clusters and had less than 1% refractory-period violations were considered single-units; clusters with greater than 1% refractory period violations or poor separation in PCA-space were considered multi-units. The soma of Field L neurons can be small (5-10 um; Saini and Leppelsack, 1981; Fortune and Margoliash, 1992) and connected by gap junctions (Saini and Leppelsack, 1977), making it extremely difficult to isolate single units (e.g. Schmidt and Konishi, 1998; Nieder and Klump, 1999a, 1999b; Sen et al., 2001; Grace et al., 2003; Cousillas et al., 2005). Because obtaining a reasonable sample of responding in L2a required inclusion of multi-unit sites, we also include single-units and multi-units from all the other auditory regions sampled (Table 1.1). Altering these single/multi-unit criteria, or restricting the analyses to only single units does not qualitatively change the reported results.

Histology. After the recording session, subjects were given a lethal dose of sodium pentobarbital and perfused trans-cardially with 10% neutral-buffered formalin. We extracted the brain and transferred it to 30% sucrose in PBS for cryoprotection. We then sectioned the brain into 40-50 um parasaggital slices (Fig. 1.1b-d) and mounted them on gelatin-coated slides. We photographed the diI-labeled electrode tracks , then stained the tissue with nissl and/or gallyas myelin stain to reveal the cytoarchitectonic borders of Field L subregions and CM. Due to restrictions in the lengths of our electrodes, we were unable to sample from the ventral-most portion of Field L. After coverslipping, the nissl and gallyas-stained tissue was photographed, and the fluorescent and bright-field images were co-registered using digital image

software (Photoshop CS4). We verified the absolute location of each electrode track and terminus in high-resolution (40x) images (Fig. 1.1c). We then identified the boundaries of each brain region crossing the electrode track and their depths along the penetration (Fig. 1.1d). We used the depth noted during each recording to identify the brain region at each recording site. The assignment of each site to a specific sub-region was done blind to knowledge of any response characteristics for the site. Occasional distortion of the hippocampus during electrode lowering caused discrepancies between recorded depth and observed depth in slices. In these cases, nearby slices showing intact hippocampus were used to approximate hippocampal depth at the site of penetration (n = 52 penetrations, mean discrepancy between histology depth and recorded depth was 22 ± 130 um). We recorded too few neurons from Field L2b to include in this analysis. Recording sites in CM ranged from 1000 to 1250μm from midline, which we consider to be primarily in the lateral CM (CLM), but note that there are no distinct cytoarchitectonic borders available to distinguish between CLM from its medial counterpart (CMM).

Behavioral Analysis. We quantified behavioral performance using d-prime [d' = Z(HR) - Z(FA)] where HR (hit rate) is the proportion of go responses to rewarded stimuli, FA (false alarm rate) is the proportion of go responses to unrewarded stimuli, and Z() denotes the standard z-score. We calculated d' for each block of 100 trials. Chance behavior is reflected in a d' value of 0, with values greater than zero representing above-chance performance.

Firing Rate Analysis. Several studies indicate that the motif is an appropriate timescale for vocal recognition learning (Gentner and Hulse, 2000). We manually identified motif boundaries within the stimuli that contained song (Song+Silence and Song+Noise stimuli) using acoustic features such as short periods of silence. We calculated the firing rate of each cell for each motif by summing the number of action potentials during a motif over all trials and dividing by the motif duration and number of trials. The firing rates varied greatly over the population of units recorded, due to both the inclusion of single- and multi-units and the variability in baseline and evoked firing rates by region (see Table 1.2). To preserve the variability in the neural responses across stimuli, while removing the large variability in the average firing rates between sites, we converted the spike rates of each cell to z-scores,

$$Eq\ 1.\ \ z = \frac{x-\mu}{\sigma},$$

where the difference between the firing rate elicited by any given stimulus (x) and the mean firing rate elicited by all stimuli (□) is divided by the standard deviation in the firing rates elicited by all stimuli (□) at the site. Because our interest here is in describing how firing rates to different song stimuli are modulated in various ways, we do not include the non-stimulus driven firing rates in our Z-score calculations. We note, however, that including these baseline rates in Z-scores calculations does not alter the pattern of reported results.

To be consistent throughout our different analyses, we use the motif-duration firing rate as the basic response unit in each population analysis shown, even when the analysis does not require a breakdown to motif level. In analyses where we compare larger stimulus categories composed of many motifs (i.e., 'Rewarded' vs. 'Unrewarded' stimuli), we define a cell's response to a given stimulus category as the mean response elicited by the motifs in that stimulus group. The results do not change significantly when firing rates calculated over longer timescales, such as full songs, are used instead. All responses to Noise+Silence stimuli were calculated over the full duration of the stimulus, since there are no motif boundaries in these stimuli. We used a repeated-measures ANOVA test and the Tukey-Kramer post-hoc test ($a=0.05$) to compare differences between groups.

Rank-ordered Motif Grouping. To determine how the interaction between noise and song varied with the strength of the song-driven response, we ordered the Song+Silence motifs by the magnitude of their evoked firing rate, then split the motifs into ten groups, each with as close to the same number of motifs as possible. We then compared the mean firing rates evoked by each group of Song+Silence motifs with the mean firing rates evoked by the same motifs in the Song+Noise context, to get a within-cell measure of how added noise affects motif responses as a function of motif drive. Note that because the motifs comprising each group varied between neurons (i.e., the motifs that drove one neuron best were be different from those that drove another neuron best), the assortment of the motifs in each group is not consistent over the population. To normalize responses for population-level analysis, we defined the

mean response to the Song+Silence motif group that elicited the largest response as the cell's maximum rate, and scaled the firing rates to other motif groups by this maximum rate. Of course, the maximum rate defined in this way is relative to the set of motifs tested, and does not preclude the possibility that another stimulus might evoke an even stronger response.

1.4 Discussion

This study describes a qualitative transformation in the encoding of behaviorally relevant signals across the auditory forebrain. In the thalamorecipient forebrain region L2a, the combination of behaviorally relevant acoustic information (birdsong) and irrelevant acoustic information (noise) is facilitative. In secondary regions, L1, L3 and CM, the combination of these signals is suppressive. In L1, the strength of this suppression is modulated by learning to target specific songs and specific spectral regions within those songs that carry behaviorally relevant information.

Hierarchy of Song Selectivity and Modulation of Song Responses by Noise

For all songbirds, correct recognition of conspecific songs plays a vital role in social behaviors (Catchpole and Slater, 1995; Kroodsma and Miller, 1996). We observed a categorical shift in the strength of the neuronal responses elicited by noise and song at different levels in the forebrain. In the primary thalamo-recipient zone

(Field L2a) neurons are driven strongly by both noise and song, while neurons in secondary regions respond most strongly to songs and poorly, if at all, to noise. This shift from a non-selective to song-selective auditory response is consistent with previous data from songbirds (Sen et al., 2001; Grace et al., 2003; Amin et al., 2007), and mirrors increases in selectivity for vocal signals observed in secondary cortical regions of other species (Belin et al., 2000; Petkov et al., 2008). As others have noted (Sen et al., 2001), the change in song-selectivity between L2 and higher regions likely reflects a change in the predominant attributes of receptive fields (RFs) in each region. Many neurons in L2a are well characterized by temporally and spectrally punctate receptive fields, responding maximally to acoustic energy at a single frequency (Theunissen et al., 2000; Sen et al., 2001; Woolley et al., 2009; Kim and Doupe, 2011). The increased response to Song+Noise relative to Song+Silence stimuli (or Noise+Silence) in Field L2a can be explained by the increase in overall stimulus power for the Song+Noise stimuli. Thus, the L2a response pattern is consistent with classically motivated receptive field models (Hubel et al., 1959; Movshon et al., 1978) in which the response of a given neuron is taken to reflect the weighted convergence of inputs tuned to simpler stimulus features.

The receptive fields of neurons in the efferent targets of L2a display increasingly complex filter properties, preferring stimuli that are extended in time or spectral bandwidth (or both) (Sen et al., 2001; Kim and Doupe, 2011), and become increasingly difficult to model with linear techniques (Theunissen et al., 2000; Gentner and Margoliash, 2003). To improve the performance of receptive field models in these

regions, researchers have advocated a variety of modifications including the use of different stimuli (Klein et al., 2000; Theunissen et al., 2000; Depireux et al., 2001), stimulus representations (Gill et al., 2008), or fitting techniques (Sharpee et al., 2004; Gill et al., 2006), to achieve marginal gains in the predictability of responses. Our results demonstrate that spectral integration in L1, L3, and CM is inherently non-linear: Noise alone is mildly excitatory in these regions, but suppressive when combined with song, and the magnitude of this suppression is nonlinearly related to strength of the song response. Both these findings challenge the basic assumptions of feature integration that underlie classical receptive fields. Because our song and noise did not overlap in frequency, the suppression can be appreciated as a form of frequency convergence that mirrors the cross-frequency suppression for pure tones observed in A1 (Imig and Adria´n, 1977; Nelken et al., 1994; Schulze and Langner, 1999; Sutter et al., 1999b), but for complex signals. More generally, these non-linear response characteristics are reminiscent of "suppressive surround" phenomena described in the visual cortex (Morrone et al., 1982; Heeger, 1992; Carandini et al., 1997; Freeman et al., 2002) suggesting common mechanisms for stimulus processing across sensory modalities.

Learning-Dependent Response Mechanisms in L1

In the present experiment, the increase in song-selectivity from L2a to higher regions was observed in both trained and untrained birds, suggesting that the hierarchical encoding of species-typical signals develops either innately or as the

product of normal experience. Because all subjects were at least one year old when captured, we cannot address these possible origins. We can say, however, that learning during adulthood affects how this functional hierarchy encodes subsets of stimuli. In untrained birds, Field L1 neurons showed no difference in the strength of responses elicited by song or noise. In birds that had learned to recognize several different combined Song+Noise stimuli, L1 neurons showed a strong preference on average for song over noise. Because recording sites from the trained and untrained birds overlapped extensively (Methods; Fig. 1.1d), anatomical sampling variations are not likely to explain the increased song selectivity in L1 of trained birds. Instead, we attribute these differences to song recognition learning. Consistent with this change in the bias for song over noise, L1 also shows a significant learning-dependent change in magnitude of the non-linear suppression of song responses by noise (Figs. 1.4d, 1.5e).

Our learning results are not restricted to comparisons between birds. One strong benefit of our stimulus design is that we could observe potential learning-driven changes within the same neuron (or site) by presenting a range of familiar and unfamiliar stimuli. Previously we've shown that training starlings on simple song recognition tasks produces changes in the strength of the spiking responses evoked by familiar and unfamiliar songs and motifs throughout CM (Gentner and Margoliash, 2003; Jeanne et al., 2011) and NCM (Thompson and Gentner, 2010). The data from CM presented here are consisted with these previous reports, and we can now add L1 to the growing list of forebrain auditory regions that are similarly modified by

learning. Based on its position in the ascending sensory hierarchy, L1 may the earliest region to show these learning-dependent response characteristics.

One common explanation for the effects of learning on neural responses is that the inputs to a given neuron are reweighted to bias those that carry task relevant information, thereby shifting the receptive field so that behaviorally relevant signals to evoke stronger responses (Fritz et al., 2003, 2005, 2007). Consistent with this, a variety of behaviorally-induced frequency biases have been observed in mammals trained with tones (Bakin and Weinberger, 1990; Recanzone et al., 1993; Polley et al., 2006). In the present results, we observed a bias in Field L1 for the specific frequency bands that contained task-relevant information (song acoustics) during training (Fig. 1.7b,e). This bias was present regardless of the signal in the task-relevant bands, suggesting a marked bias in a subset of frequency channels consistent with the mammalian frequency biases.

Our results demonstrate, however, that learning alters more than the weighting of frequency channels in the context of classical receptive fields. In trained subjects, the rewarded motifs evoked significantly lower spike rates in L1 than unrewarded and unfamiliar motifs. A similar effect of associative learning is found in NCM (Thompson and Gentner, 2010; Thompson et al., 2013). What is most remarkable about this learning effect is that these L1 response differences are not tied exclusively to the spectrotemporal features of the rewarded songs, which are necessary for successful recognition, but rather to how those specific features combine with irrelevant noise in independent frequency channels. Because neither the unrewarded

nor the unfamiliar motifs were significantly suppressed by the same noise (Fig. 1.6f), we conclude that the non-classical suppressive responses identified in L1 are modifiable by learning to target specific acoustic features. This flexibility suggests that the processing power of these non-classical mechanisms has been greatly underestimated in other systems, particularly for processing complex natural signals where at any given time various features may be task-irrelevant.

Conclusion

In all post-thalamorecipient regions (L1, L3, and CM), regardless of training condition, there appears to be a common mechanism for combining song with noise in independent frequency channels. Song+Silence motifs that evoke a weak response from a neuron in one of these regions (i.e. a firing rate below that evoked by noise alone) are essentially masked when combined with noise, as the response to the combined signal is indistinguishable from that to Noise+Silence. In contrast, combining the same noise with a Song+Silence motif that evokes a response greater than Noise+Silence leads to a suppressed response intermediate to that evoked by noise or song alone. The strength of this noise suppression is proportional to the strength of the initial response to the motif (Fig. 5b-d). This response suppression is consistent with divisive normalization in visual cortex (Heeger, 1991, 1992) where the output of a neuron's receptive field is normalized (divided) by the activity pooled from other neurons.

Divisive normalization is a powerful, generalized, gain control mechanism (Kouh and Poggio, 2008; Carandini and Heeger, 2012) that can prevent responses in noisy or distracting environments from growing so strong that the neuron's dynamic range is saturated. The same mechanism can account for contrast normalization (Carandini et al., 1997), including in that observed in primary auditory cortex (Rabinowitz et al., 2011), and has been used to model the enhanced stimulus responses that accompany selective visual attention (Reynolds and Heeger, 2009). Indeed, its wide utility has led to speculation that divisive normalization is a canonical computation in cortex (Kouh and Poggio, 2008; Carandini and Heeger, 2012), and our observations here are consistent with that interpretation. Understanding the role of divisive normalization in the detection, attention to, and learning of natural auditory signals requires much more extensive examination, but the basic outline of a functional cortical network that uses a combination of hierarchical selectivity and gain control mechanisms to produce robust responses to behaviorally relevant natural signals is apparent.

1.5 Acknowledgements

Chapter 1, in part, is currently being prepared for publication of the material as it may appear under the authorship of Emily L. Caporello and Timothy Q Gentner. Emily Laura Caporello was the primary investigator and author of this material.

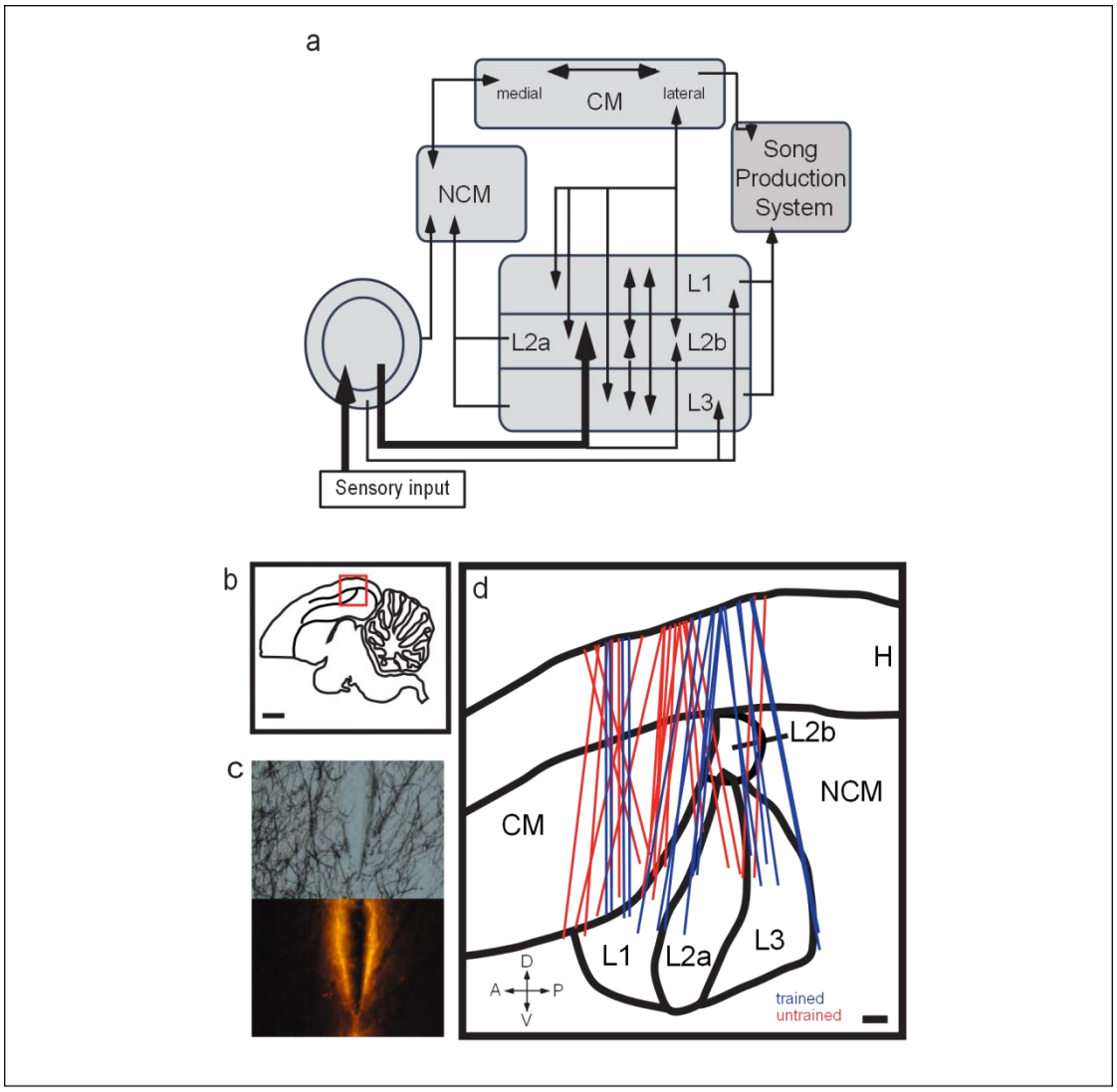


Figure 1.1. Avian auditory system connectivity and histological data. (a) Schematic diagram of the connectivity between regions in the avian auditory system . (b) Schematic of whole parasaggital section indicating area of interest (red square, Fig 1d). Scale bar 2mm. (c) Example image of electrode terminus taken at 40x magnification under brightfield (top, gallyas myelin stain) and fluorescence (bottom, diI; see Methods). (d) Schematic indicating approximate position of electrodes in recordings from naïve (red) and trained (blue) birds. Tracks are taken from slices approximately between 800µm and 1300µm lateral from the midline. Saggital section in inset is located at approximately 1000µm lateral to midline. Scale bar 200µm. H, Hippocampus; NCM, Caudomedial Nidopallium; CM, Caudolateral Mesopalium; D, dorsal; A, anterior; P, Posterior; V, ventral.

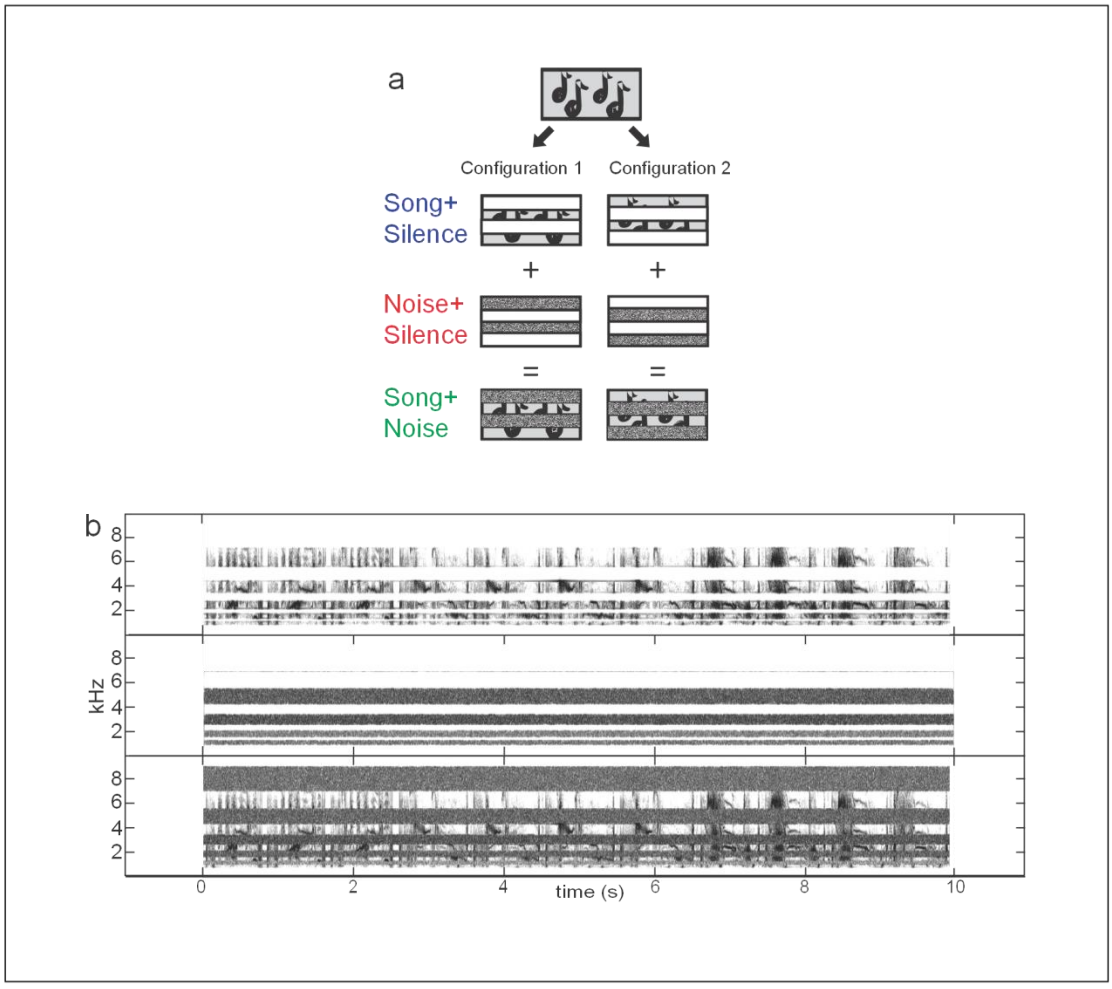


Figure 1.2. Stimuli. (a) Schematic of stimulus construction. Full songs (n = 6) or white noise were filtered in the frequency domain to produce stimuli with alternating frequency bands containing song and silence (Song+Silence stimuli), noise and silence (Noise+Silence stimuli), or song and noise (Song+Noise stimuli; see Methods). Within each of these categories, stimuli could have one of two possible configurations depending on which frequency bands were modified. (b) Spectrograms of three example stimuli: one Song+Silence stimulus (top), one Noise+Silence stimulus (middle), and a Song+Noise stimulus (bottom). The level of acoustic power at a given time and frequency is indicated by the darkness of the spectrogram.

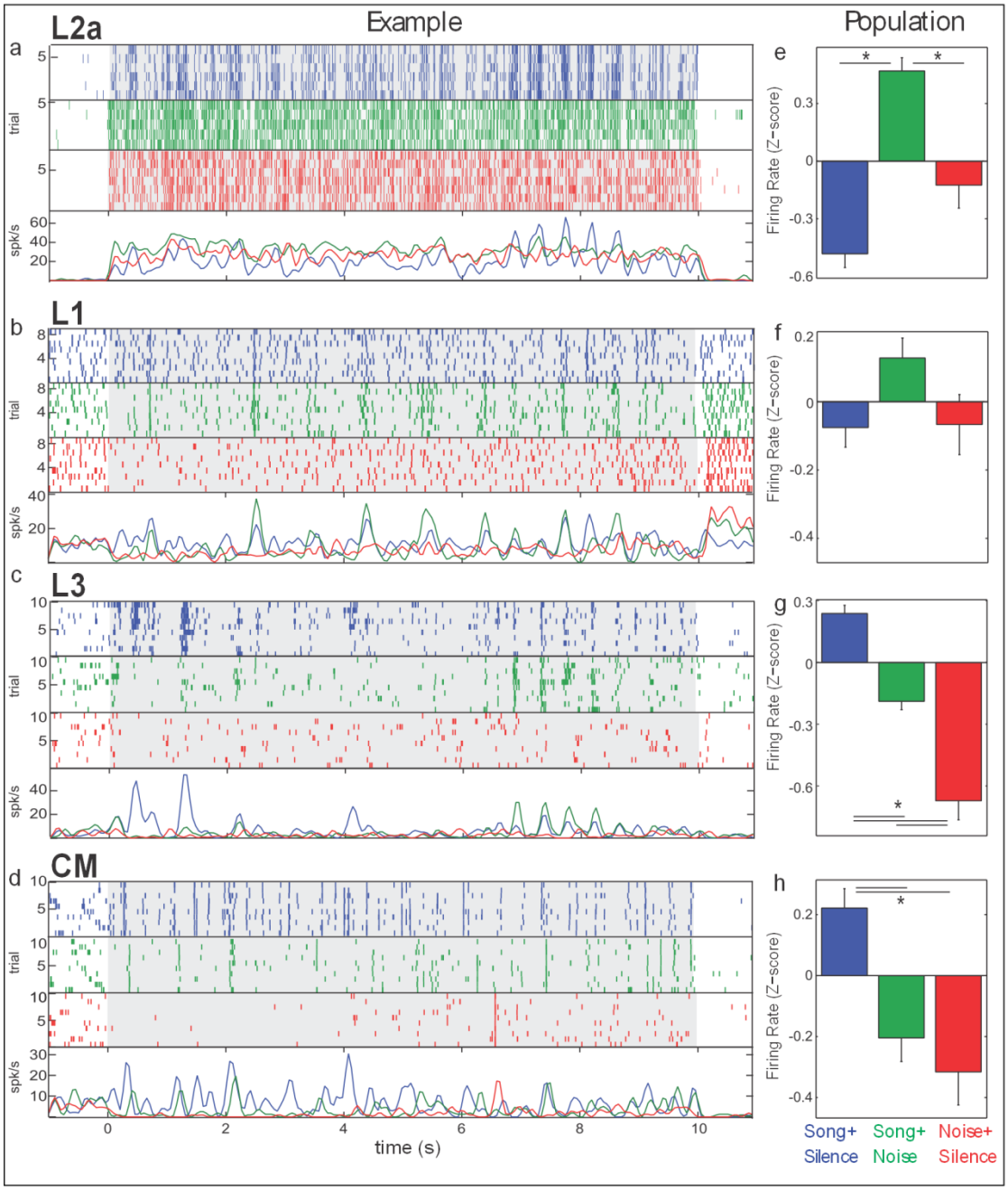


Figure 1.3. Neural responses to stimulus class by brain region. (a-d) Electrical activity recorded from single units in region L2a (a), L1 (b), L3(c), and CM(d) shown in response to the Song+Silence stimulus (blue), the Song+Noise stimulus (green), and the Noise+Silence stimulus (red). Offset response observed in L1 example cell not necessarily typical for units recorded in this region. (e-h) Mean (±SEM) normalized response rates over population of units recorded in L2a (e), L1 (f), L3(g), and CM(h) evoked by Song+Silence (So+Si, blue bar) Song+Noise (So+N, green bar) and Noise+Silence (N+Si, red bar) stimuli. There is a significant main effect of stimulus class within L2a ($p<0.0001$), L3 ($p<0.0001$), and CM ($p = 0.0010$) populations (rmANOVA). Horizontal lines below asterisks indicate significant differences between stimulus classes (L2a: So+Si < So+N, So+N > N+Si; L3: all comparisons significant; CM: So+Si > So+N, So+Si > N+Si; Tukey-Kramer post-hoc a=0.05).

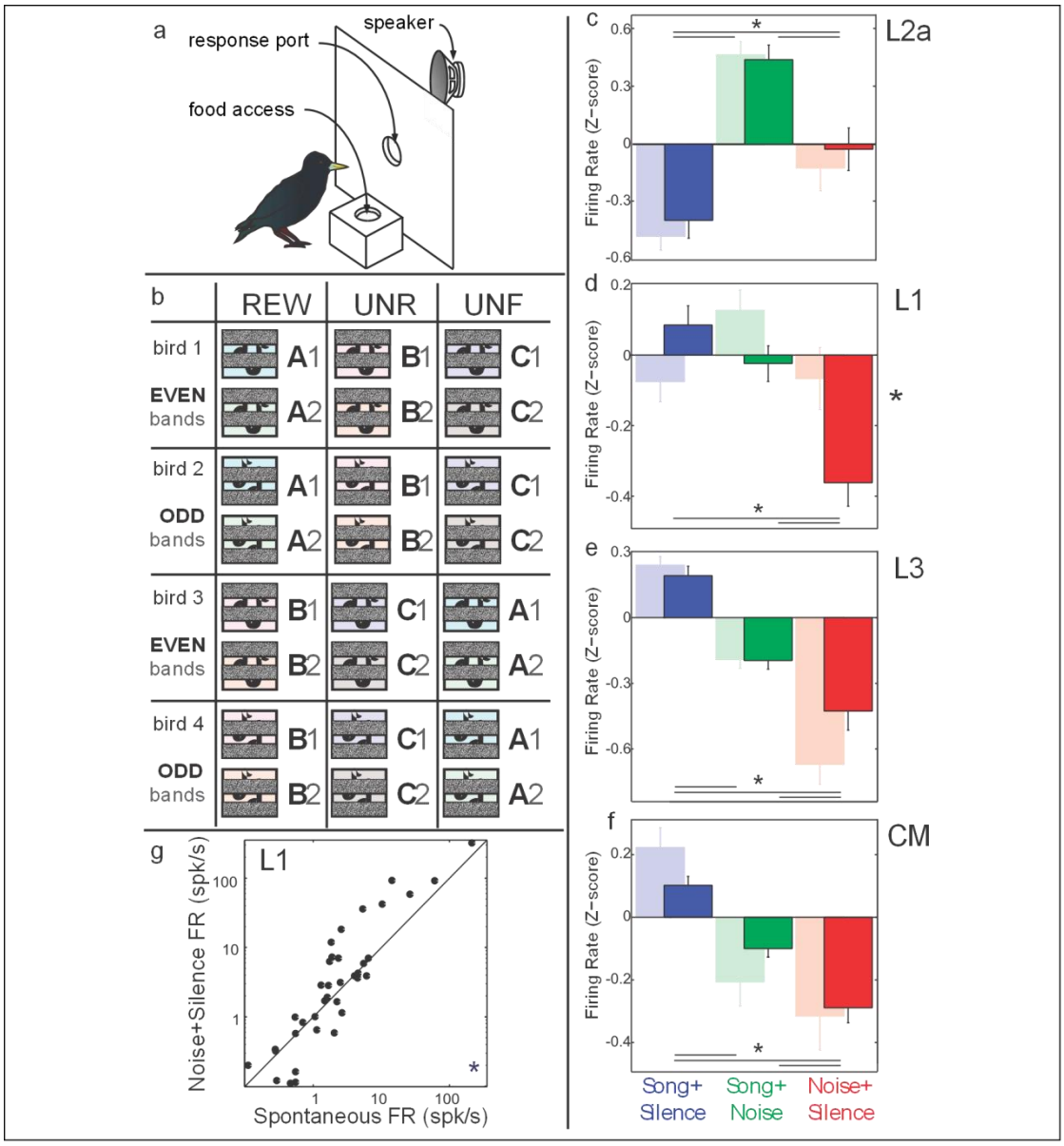


Figure 1.4. Behavioral training and neural responses from trained subjects. (a) Schematic of operant panel used for behavioral training. (b) Training schedules for each subject (n=4) trained in this study. A single Song+Noise banding configuration was chosen for training. Banding configuration and song class familiarity were alternated for different subjects. (c-f) Mean (±SEM) normalized response rates over populations of units recorded in L2a(c), L1(d), L3(e), and CM(f) from trained birds, evoked by all Song+Silence (blue bar), all Song+Noise (green bar), and all Noise+Silence (red bar) stimuli. All regions showed a significant main effect of stimulus class (all regions $p<0.0001$, rmANOVA). Horizontal lines below asterisks indicate significant differences between stimulus classes (L2a: all comparisons significant; L1: So+Si > N+Si, So+N > N+Si; L3: all comparisons significant; CM: all comparisons significant; Tukey-Kramer post-hoc $a=0.05$). Asterisk to the right of the bar plot in (d) indicates a significant interaction between stimulus class and training condition in L1 data recorded from trained and untrained subjects ($p = 0.0116$, rmANOVA). No other regions had a significant main effect of or interaction with t raining condition ($p > 0.05$, rmANOVA). (g) Spontaneous activity vs. evoked firing rates in response to Noise+Silence stimuli for L1 neurons from trained subjects ($p=0.0198$, student t-test).

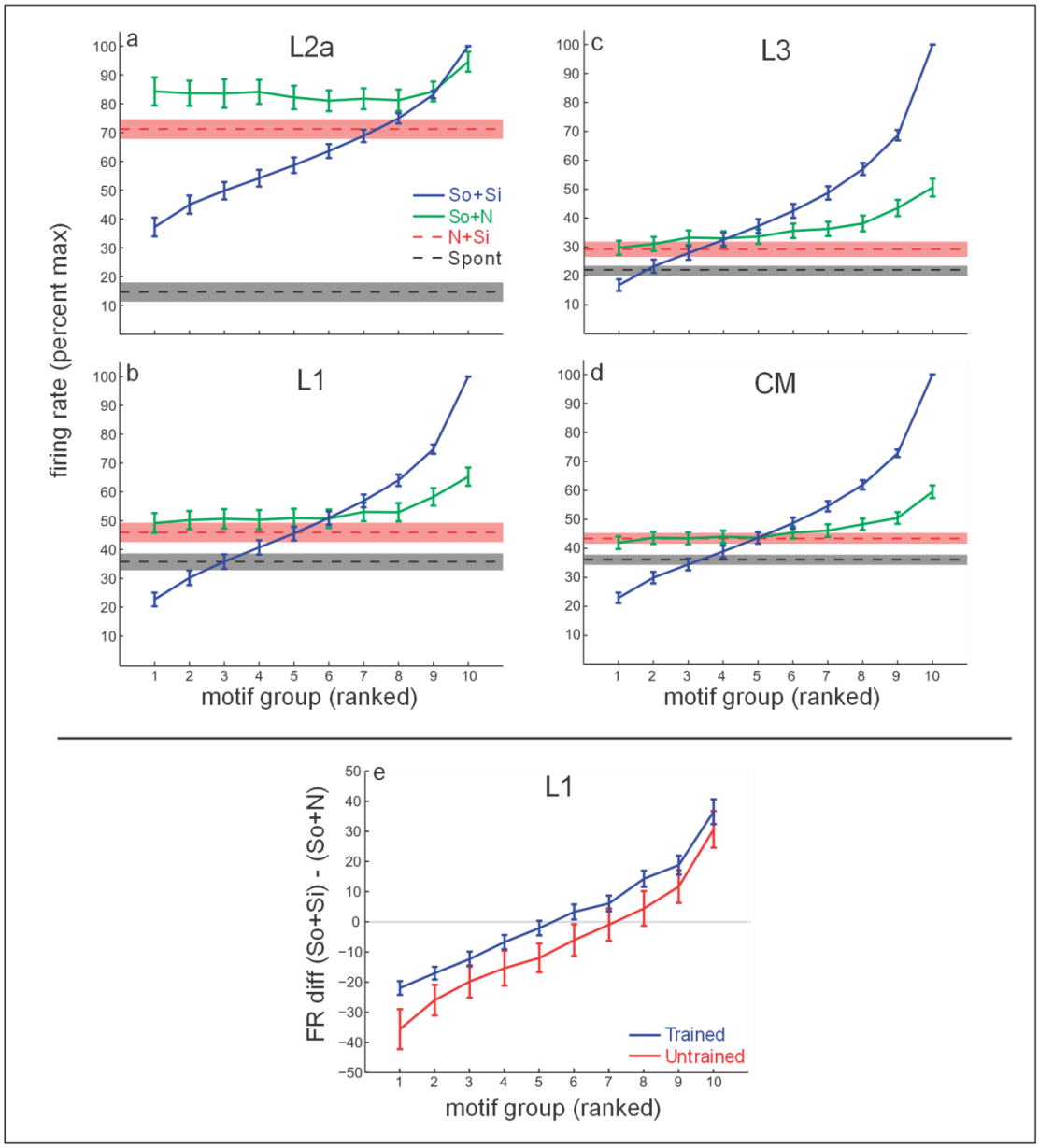


Figure 1.5. Effect of noise varies with magnitude of song motif response. (a-d) Mean (±SEM) responses to rank-ordered motif groups with and without noise (So+Si, blue; So+N, green), averaged across all cells from trained and untrained subjects in L2a(a), L1(b), L3(c), and CM(d). Mean (±SEM) response to N+Si shown in red, mean (±SEM) baseline rate shown in black. Firing rates are normalized within a cell to the maximum response elicited by a motif group. The difference between the So+N and So+Si responses to each motif group significantly varies as a function of motif group in each region ($p < 0.0001$, rmANOVA, all regions). (e) Mean (±SEM) difference between rank-ordered motif groups presented with and without noise ((So+Si)-(So+N)) in L1. Data from trained subjects are shown in blue, untrained subjects shown in red. Experience significantly reduces the amount of noise suppression across motif groups ($p = 0.0307$, ANOVA).

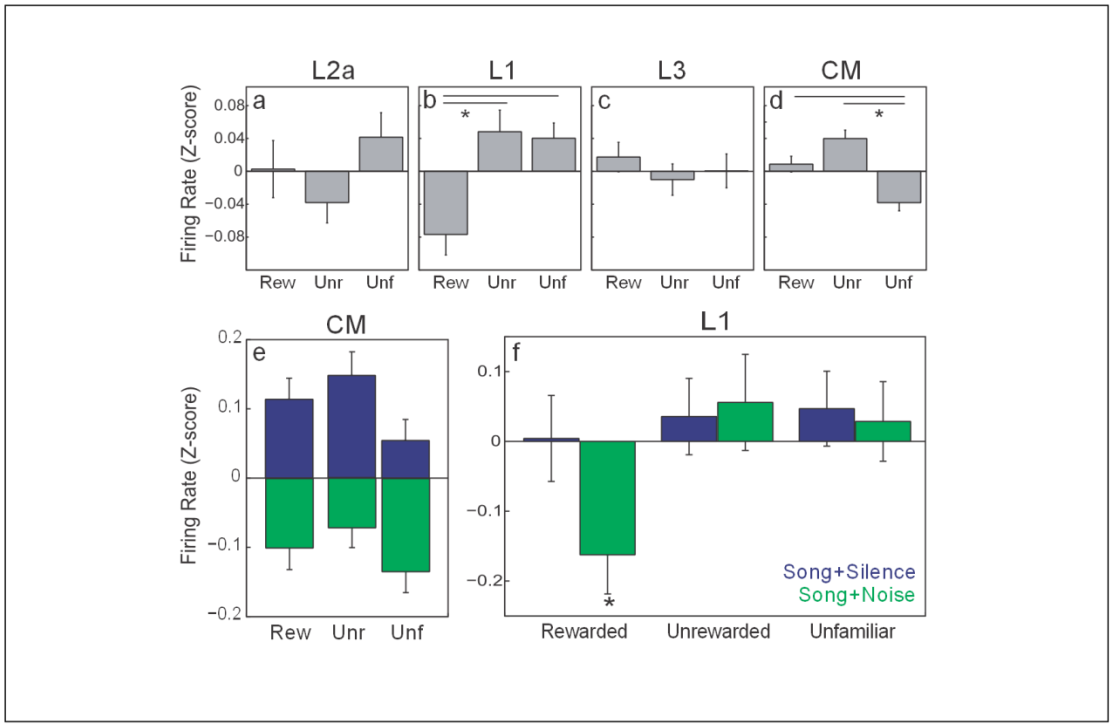


Figure 1.6. Effect of reward category on neural response from trained subjects. (a-d) Mean (±SEM) normalized response rates over populations of units recorded in L2a (a), L1 (b), L3 (c), and CM (d) from trained birds, evoked by all stimuli containing Rewarded songs (4 So+Si stimuli, 4 So+N stimuli), all stimuli containing Unrewarded songs (4 So+Si stimuli, 4 So+N stimuli), and all stimuli containing Unfamiliar songs (4 So+Si stimuli, 4 So+N stimuli). There is a significant main effect of reward category within regions L1 (b, $p=0.0042$) and CM ($p<0.0001$, rmANOVA). Horizontal lines above asterisk indicate significant comparisons between reward category means from L1 (b) and CM (d) data, (L1: Rew > Unr, Rew > Unf; CM: Rew > Unf, Unr > Unf; Tukey-Kramer post-hoc $a=0.05$). (e-f) Mean (±SEM) normalized response rates of evoked by each reward category (Rewarded, Unrewarded, Unfamiliar), split by general stimulus class (Song+Silence, blue bars, Song+Noise, green bars). In CM (e), there is no significant interaction between reward category and general stimulus class. However, in L1, the interaction is significant ($p=0.0078$, rmANOVA).

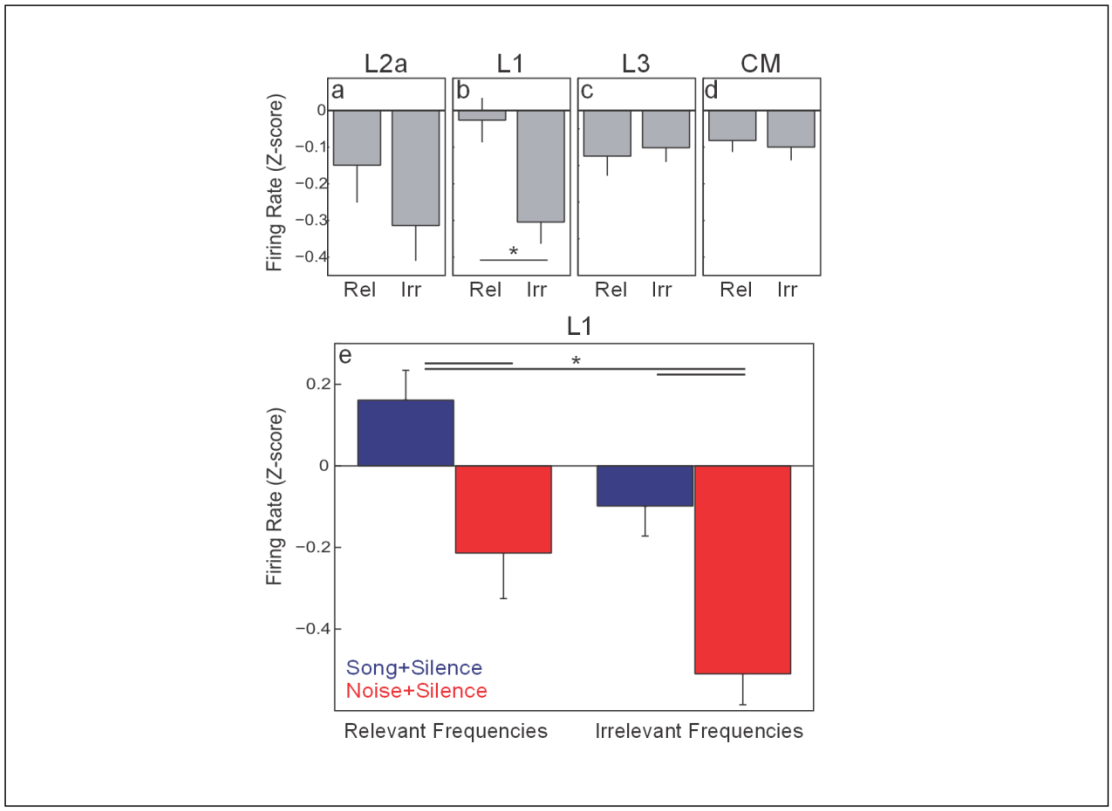


Figure 1.7. Effect of stimulus frequency on neural responses from trained subjects. (a) Mean (±SEM) normalized response rates over populations of units recorded in L2a(a), L1(b), L3(c), and CM(d) from trained birds, evoked by stimuli containing power in only Relevant (Rel) or Irrelevant (Irr) frequency bands (collapsed over all Song+Silence and Noise+Silence stimuli). Relevant frequency bands are defined as the frequency bands which contained song during training, while Irrelevant frequency bands are those that contained noise during training. Horizontal line below asterisk indicates a significant difference between L1 frequency categories ($p = 0.0146$, rmANOVA). (e) Mean (±SEM) normalized response rates of L1 sites evoked by each frequency category, split by general stimulus class (Song+Silence, blue bars, Noise+Silence, red bars). Horizontal lines below asterisk indicate significant Post-hoc comparisons (So+Si Rel > N+Si Rel, So+Si Rel > N+Si Irr, So+Si Irr > N+Si Irr; Tukey-Kramer a=0.05).

Table 1.1. Summary of recorded units in auditory regions of trained and untrained birds.

Brain Region	Trained			Naive			Grand Total
	Single Units	Multi-Units	Total	Single Units	Multi-Units	Total	
Field L2a	4	15	19	0	11	11	30
Field L1	18	19	37	13	5	18	55
Field L2	22	16	38	18	5	23	67
CM	78	36	114	3	11	14	128
Total	122	86	208	34	32	66	274

Table 1.2. Mean (± SEM) spike rates (spikes/sec) evoked by the twelve Song+Silence and two Noise+Silence stimuli at all sites in four forebrain regions. P-values reflect comparison (paired t-test) between responses to a given stimulus and spontaneous rates measured when no stimulus was presented.

Brain Region	Baseline Firing Rate	Firing Rate by Stimulus Class			
		Song+Silence		Noise+Silence	
Field L2a	26.08 ± 12.21	68.11 ± 19.62	p = 0.0017	80.09 ± 24.29	p = 0.0031
Field L1	5.89 ± 1.33	11.26 ± 2.40	p = 0.0010	11.41 ± 2.73	p = 0.0041
Field L3	7.22 ± 1.63	16.48 ± 4.19	p = 0.0011	11.51 ± 2.76	p = 0.0029
CM	6.31 ± 0.98	11.81 ± 2.52	p = 0.0019	11.02 ± 2.82	p = 0.0247

1.6 Bibliography

Amin N, Doupe A, Theunissen FE (2007) Development of selectivity for natural sounds in the songbird auditory forebrain. Journal of neurophysiology 97:3517–3531.

Bakin JS, Weinberger NM (1990) Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. Brain research 536:271–286.

Bao S, Chang EF (2004) Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. Nature neuroscience 7:974–981.

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

Bonin V, Mante V, Carandini M (2005) The suppressive field of neurons in lateral geniculate nucleus. The Journal of neuroscience : the official journal of the Society for Neuroscience 25:10844–10856.

Capsius B, Leppelsack H-J (1996) Influence of urethane anesthesia on neural processing in the auditory cortex analogue of a songbird. Hearing research 96:59–70.

Carandini M, Heeger DJ (1994) Summation and division by neurons in primate visual cortex. Science (New York, NY) 264:1333–1336.

Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nature reviews Neuroscience 13:51–62.

Carandini M, Heeger DJ, Movshon JA (1997) Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex. J Neurosci 17:8621–8644.

Catchpole CK, Slater P (1995) Bird song: Biological themes and variations.

Cousillas H, Leppelsack H-J, Leppelsack E, Richard J-P, Mathelier M, Hausberger M (2005) Functional organization of the forebrain auditory centres of the European starling: a study based on natural sounds. Hearing research 207:10–21.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-Temporal Response Field Characterization With Dynamic Ripples in Ferret Primary Auditory Cortex. J Neurophysiol 85:1220–1234.

Doupe AJ, Konishi M (1991) Song-selective auditory circuits in the vocal control system of the zebra finch. Proceedings of the National Academy of Sciences of the United States of America 88:11339–11343.

Fitzpatrick D (2000) Seeing beyond the receptive field in primary visual cortex. Current opinion in neurobiology 10:438–443.

Fortune ES, Margoliash D (1992) Cytoarchitectonic Organization and Morphology of Cells of the Field L Complex in Male Zebra Fincehs (Taenopygia guttata). Journal of Comparative Neurology 325:388–404.

Freeman TCB, Durand S, Kiper DC, Carandini M (2002) Suppression without inhibition in visual cortex. Neuron 35:759–771.

Fritz JB, Elhilali M, Shamma SA (Univeristy of M (2005) Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hearing research 206:159–176.

Fritz JB, Elhilali M, Shamma SA (Univeristy of M (2007) Adaptive changes in cortical receptive fields induced by attention to complex sounds. Journal of neurophysiology 98:2337–2346.

Fritz JB, Shamma SA (Univeristy of M, Elhilali M, Klein DJ (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nature neuroscience 6:1216–1223.

Fuzessery ZM, Feng AS (1982) Frequency selectivity in the anuran auditory midbrain: Single unit responses to single and multiple tone stimulation. Journal of Comparative Physiology ? A 146:471–484.

Galambos R (1956) Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. J Neurophysiol 19:424–437.

Gentner TQ, Fenn KM (2006) Recursive syntactic pattern learning by songbirds. Nature:1–5.

Gentner TQ, Hulse SH (2000) Perceptual classification based on the component structure of song in European starlings. The Journal of the Acoustical Society of America 107:3369–3381.

Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. Nature 424:669–674.

Gill P, Zhang J, Woolley SMN, Fremouw T, Theunissen FE (2006) Sound representation methods for spectro-temporal receptive field estimation. Journal of computational neuroscience 21:5–20.

Gill PR, Woolley SMN, Fremouw TE, Theunissen FE (2008) What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. Journal of neurophysiology 99:2809–2820.

Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. Journal of neurophysiology 89:472–487.

Heeger DJ (1991) Nonlinear model of the Neural Responses in Cat Visual Cortex. In: Computational Models of Visual Processing, pp 119–133. Cambridge, MA: MIT Press.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Visual Neuroscience 9:181–197.

Hubel DH, Henson CO, Rupert A, Galambos R, Series N, May N (1959) "Attention" Units in the Auditory Cortex. Science 129:1279–1280.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. Journal of Physiology 148:574–591.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology 160:106–154.

Imig TJ, Adria´n HO (1977) Binaural columns in the primary field (A1) of cat auditory cortex. Brain Research 138:241–257.

Jeanne JM, Thompson J V, Sharpee TO, Gentner TQ (2011) Emergence of learned categorical representations within an auditory forebrain circuit. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:2595–2606.

Kim G, Doupe A (2011) Organized representation of spectrotemporal features in songbird auditory forebrain. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:16977–16990.

Klein DJ, Depireux DA, Simon JZ, Shamma SA (Univeristy of M (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. Journal of computational neuroscience 9:85–111.

Kouh M, Poggio T (2008) A canonical neural circuit for cortical nonlinear operations. Neural computation 20:1427–1451.

Kroodsma D, Miller E (1996) Ecology and evolution of acoustic communication in birds.

Leppelsack H-J, Vogt M (1976) of Comparative Responses of Auditory Neurons in the Forebrain of a Songbird to Stimulation with Species-specific Sounds *. Journal of Comparative Physiology 107:263–274.

Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. The Journal of neuroscience : the official journal of the Society for Neuroscience 3:1039–1057.

Morrone MC, Burr DC, Maffei L (1982) Functional Implications of Cross-Orientation Inhibition of Cortical Visual Cells. I. Neurophysiological Evidence. Proceedings of the Royal Society B: Biological Sciences 216:335–354.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283:79–99.

Nelken I, Prut Y, Vaadia E, Abeles M (1994) Population responses to multifrequency cortex : One- and two-parameter sounds in the cat auditory families of sounds. Hearing Research 72:206–222.

Nieder a, Klump GM (1999a) Adjustable frequency selectivity of auditory forebrain neurons recorded in a freely moving songbird via radiotelemetry. Hearing research 127:41–54.

Nieder A, Klump GM (1999b) Time course of simultaneous masking in the starling's auditory forebrain. Experimental Brain Research 124:311–320.

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

Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. Journal of neuroscience 26:4970–4982.

Rabinowitz NC, Willmore BDB, Schnupp JWH, King AJ (2011) Contrast gain control in auditory cortex. Neuron 70:1178–1191.

Recanzone GH, Schreiner CE, Merzenich MM (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. The Journal of Neuroscience 13:87–103.

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185.

Rutkowski RG, Weinberger NM (2005) Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. Proceedings of the National Academy of Sciences of the United States of America 102:13664–13669.

Saini KD, Leppelsack HA (1977) Cell and Tissue Neuronal Arrangement in the Auditory Field L of the Neostriatum of the Starling. Cellular Tissue Research 316:309–316.

Saini KD, Leppelsack H-J (1981) Cell types of the auditory caudomedial neostriatum of the starling (Sturnus vulgaris). The Journal of comparative neurology 198:209–229.

Schmidt MF, Konishi M (1998) Gating of auditory responses in the vocal control system of awake songbirds. Nature neuroscience 1:513–518.

Schneider DM, Woolley SMN (2011) Extra-classical tuning predicts stimulus-dependent receptive fields in auditory neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:11867–11878.

Schulze H, Langner G (1999) Auditory cortical responses to amplitude modulations with spectra above frequency receptive fields: evidence for wide spectral integration. Journal of comparative physiology A, Sensory, neural, and behavioral physiology 185:493–508.

Scott SK (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. Journal of neurophysiology 86:1445–1458.

Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. Neural computation 16:223–250.

Sutter ML, Schreiner CE, McLean M, O'connor KN, Loftus WC (1999a) Organization of Inhibitory Frequency Receptive Fields in Cat Primary Auditory Cortex. J Neurophysiol 82:2358–2371.

Sutter ML, Schreiner CE, McLean M, O'Connor KN, Loftus WC (1999b) Organization of Inhibitory Frequency Receptive Fields in Cat Primary Auditory Cortex. Journal of neurophysiology 82:2358–2371.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. The Journal of Neuroscience 20:2315–2331.

Thompson J V, Gentner TQ (2010) Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. Journal of neurophysiology 103:1785–1797.

Thompson J V, Jeanne JM, Gentner TQ (2013) Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. Journal of neurophysiology 109:721–733.

Wang Y, Brzozowska-Prechtl A, Karten HJ (2010) Laminar and columnar auditory cortex in avian brain. Proceedings of the National Academy of Sciences of the United States of America 107:12676–12681.

Winter P, Funkenstein HH (1973) The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. (Saimiri sciureus). Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale 18:489–504.

Woolley SMN, Gill PR, Fremouw TE, Theunissen FE (2009) Functional groups in the avian auditory system. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:2780–2793.

Zoccolan D, Cox DD, DiCarlo JJ (2005) Multiple object response normalization in monkey inferotemporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 25:8150–8164.

Chapter 2

2.1 Introduction

Natural environments are filled with a rich array of sensory information. Since the capacity of our sensory systems is limited, we cannot simultaneously perceive every part of a rich, complex scene, and instead we select certain relevant stimuli to focus on at the expense of perceiving everything else. This process, known as selective attention, has been studied by researchers for decades. From that large body of research, a basic property of the neural processes underlying selective attention can be distilled – responses to attended features are enhanced in sensory regions, while responses to irrelevant features are suppressed. However, the effect of attention on auditory neurons at the single-cell level remains largely unknown, particularly in the most ethologically relevant setting: attending to conspecific vocalizations.

Here, we demonstrate a novel naturalistic auditory attention task that requires subjects, European starlings, to detect a target birdsong in the presence of an irrelevant birdsong. We directly measured the effects of auditory attention by recording from neurons in the high-level auditory area CM while subjects were engaged in the attention task.

We show that birds can perform these naturalistic auditory attention tasks, that high-level auditory neurons are modulated by attention, and that the nature of the modulation is different depending on the cell type recorded. Specifically, we show that the response of putative fast-spiking interneurons to a pair of songs are better modeled

as a linear combination of the individual responses to the songs than wide spiking cells, and that the fast-spiking cells respond more like the ignored stimulus than the attended stimulus during correct attention trials. Overall this data supports a model of auditory attention where the ignored stimulus is suppressed by interneurons in high-level auditory regions.

2.2 Results

Birds can Perform Auditory Attention Tasks

We successfully trained 8 birds on a selective auditory attention task. The task requires the bird to attend to one of two simultaneously presented 2 second birdsong clips. The clips are presented from the same speaker, but are separated in frequency space, with one song restricted to a low frequency band, and the other song restricted to a high frequency band (see Methods, Figure 2.1a,b). To direct attention to a single song target within the presented pair, we trained the birds to perform a two alternative-choice delayed match to sample task. The bird initially hears one song clip (the sample), then, after a 500ms pause, a second song clip (the target) is played. If the target is identical to the sample, the trial is considered a "match" trial, and the subject must respond to the left peck port in order to receive a food reward. Conversely, if the target is different from the sample, the trial is considered a "non-match" trial and the subject must respond to the right peck port in order to receive a food reward (training

phase 1, Figure 2.1b,c). The target and sample are always restricted to the same frequency range, either high or low. Thus, by listening to the sample, the animal is cued to the relevant frequency band necessary to solve the task.

Initially, the sample and target stimuli were presented alone. Once the animal performed well above chance on this basic phase, however, we introduced a distracter song clip played during the target portion of the trial, exactly overlapping the presentation of the target stimulus. The distracter song clip was always chosen from the opposite frequency range from the sample and target. Initially, the distracter was presented at a very low volume (training phase 2+, Figure 2.1d). Once the subject was able to perform above chance with a low volume distracter, the volume of the distracter was increased, until the mean volume of the distracter was equal to that of the target and sample (end phase, Figure 2.1e). This task was very difficult for the birds, with 5 out of 13 birds unable to perform the task at the full distracter volume level. For the 8 birds that were able to perform the full task, mean performance at the end of training, once plateaued, was at d prime = 0.890 (last 3000 trials; range: 0.361-1.410; Figure 2.1f). In a subset of fully trained birds (n=3), we probed the animals with several distracter volumes (Figure 2.1g). Performance significantly varied with distracter strength ($r = -0.48$, $p = 0.018$).

High-level Auditory Neurons are Modulated by Attention to Natural Streams

After behavioral training, we implanted 16 channel extracellular recording electrodes into the caudal mesopallium (CM, Figure 2.2a,b; see Methods) of subjects

and recorded action potentials in this region while they performed the attention task. The task was identical to that in the training sessions, with a few modifications to adjust for electrophysiology demands. During a recording session, a limited number of stimuli were chosen for presentation (see Methods) to maximize the number of repeats for a given stimulus (Figure 2.2c-e). High frequency and low frequency targets were presented in blocks of 34 trials. For neurophysiological control purposes, a fraction of trials (n=10) were presented without any distracters. For behavioral control purposes, a fraction of catch trials (n=4) were presented, which ensured the animal was attending to the appropriate frequency band. We only included units recorded during sessions with above average behavioral performance for analysis (see Methods), yielding a total of 198 song pairs recorded from 117 units.

Our first hypothesis was that directing attention to either component of a stimulus pair would modulate responses in CM, a high-level auditory region selective for conspecific song (Gentner and Margoliash, 2003; Grace et al., 2003; Hsu et al., 2004; Atiani et al., 2009) whose neural coding is known to be affected in the long term by the behavioral meaning of the stimulus (Gentner and Margoliash, 2003; Woolley and Doupe, 2008, Jeanne et al., 2011, 2013). To test assess the amount of modulation between “attend high” and “attend low” conditions, we calculated the correlation coefficient between the mean response to the stimulus pair (target+distracter) when the bird attended to the high song component and the mean response when the bird attended to the low component (see Methods). Since these two conditions, “attend high” and “attend low”, had different stimuli preceding the presentation of the song

pair (see Figure 2.2d,e), it is possible that a portion of the modulation observed between conditions could be explained by the influence of the preceding stimulus (Figure 2.3a,b). To control for this, it was crucial for us to relate any modulation seen in the attention condition to the amount of modulation observed when the trials were presented passively. We calculated the correlation coefficient between the passive presentations of the stimulus pair when it was preceded by the high frequency sample (identical stimuli to "attend hi" trials) and when it was preceded by the low frequency sample (identical stimuli to "attend lo" trials). For a given stimulus pair, if the modulation due to attention was greater than that due to the difference in stimulus properties, the correlation coefficient calculated from attention trials should be lower than that calculated from passive trials. Indeed, over the population of recorded neurons, that is what we observed (Figure 2.3c, $p = 0.0017$).

To assess whether an individual cell demonstrated a significant attention effect, we calculated the mean squared error (MSE) between the mean response to "attend hi" and "attend lo" trials. Again, we related this modulation to the passive condition by subtracting the MSE from the passive trials from the MSE from attention trials (see Methods). If the MSE difference was significantly different between passive and attention conditions, the stimulus pair was considered to be significantly modulated by attention (156 out of 188 song pairs, 83%, Figure 3d). The effect size of the attention modulation was determined using the wilcoxon sign rank statistic (see Methods).

Encoding of Attended and Ignored Stimuli by Cell Type

Previous studies have shown that attending to one of a pair of components, such as speech or visual objects (Reynolds et al., 1999; Mesgarani and Chang, 2012), can produce a neurophysiological response more similar to the target than the distracter component. Since attention modulation is only apparent at specific times during the response (see Figure 2.3a,b), we determined time bins where the firing rates that were significantly different when attending selectively to a song within a pair (see Methods). Within the time bins significantly modified by attention, we assessed the similarity between a unit's response to the song pair presented together and to each song component presented alone. To do this, we calculated the MSE between the response to a stimulus presented alone and attending to that song within a pair. We compared this to the MSE between the response to the song alone and ignoring the song in a pair (see Methods, Figure 2.4a). This comparison, which we call the Response Bias, yields a positive value if there is less error in the representation of the target song than the distracter, indicating the response to the song pair was more representative of the song being attended to (see Methods). Conversely, a negative value indicates a response that more closely resembles the song being ignored. Within individual song pair responses, MSE differences significantly different from zero were considered to reflect responses significantly more like the target (positive) or distracter (negative). Interestingly, over the entire population, we got a wide variety of responses (62 significantly negative, 34 significantly positive, paired nonparametric) with a mean result that was slightly but significantly negative (-0.0031; Figure 2.4b),

indicating the average response in the population to a stimulus pair was more similar to the distracter than the target ($p = 0.0009$, Mann–Whitney U test).

To investigate whether different cell types could account for this variability, we assessed the spikewidths of our population of all recorded units (including those recorded during poor behavior sessions and non-auditory units). Previous studies have shown that due to differences in action potential kinetics, parvalbumin-positive GABAergic interneurons have narrow spikewidths (McCormick and Prince, 1985; Connors and Gutnick, 1990; Kawaguchi and Kubota, 1993; Martina and Jonas, 1997; Martina et al., 1998; Nowak et al., 2003; Erisir et al., 2013). This population of narrow-spiking units can be identified in the extracellular recording of action potentials (Henze et al., 2000; Gold et al., 2006), and they are modulated by attention differently than broad-spiking neurons (Mitchell et al., 2007). Parvalbumin positive cells have been identified in the avian forebrain (Wang et al., 2010). Within our data, the distribution of single unit spikewidths was significantly bimodal ($p = 0.0047$, Hartigan's dip test; Figure 3c), with a cluster boundary between narrow and wide spiking cells at 230 μs (see Methods), consistent with previous data (Mitchell et al., 2007). Furthermore, the baseline and driven firing rates of narrow spiking cells are significantly greater than wide spiking cells (mean baseline FR wide= 6.6011 spk/sec, narrow = 13.6696 spk/sec, $p<0.0001$, Mann–Whitney U test; mean driven FR wide = 9.8491 spk/sec, narrow = 21.3245 $p<0.0001$, Mann–Whitney U test), consistent with previous analyses of putative interneurons (Mitchell et al., 2007).

When we separated out the MSE difference by spikewidths, we found that the negative bias seen in our data set is explained by the narrow spiking cells, which were strongly and significantly negative (mean = -0.0085, $p<0.0001$, Mann–Whitney U test; Figure 4d). Furthermore, 27 out of 34 narrow unit song pairs with significant MSE differences were negative. Wide unit song pairs were not significantly different from zeros (mean = 0.0004, $p=0.3814$, Mann–Whitney U test; Figure 2.4d), with 27 significant MSE diff units positive and 35 significantly negative. The narrow population was not only significantly negative, but also significantly different from the wide cell population ($p=0.0017$, Mann–Whitney U test; Figure 4d). This narrow cell effect went away when we analyzed incorrect responses (mean = -0.0005, $p = 0.7199$, Mann–Whitney U test; Figure 2.4e); while wide cells remained indifferent from zero (mean = -0.0007, $p = 0.7199$, Mann–Whitney U test; wide vs narrow $p =0.9337$, Mann–Whitney U test; Figure 2.4e). Additionally, fewer song pairs elicited significant MSE differences (44 wide, 26 narrow). These data show that narrow spiking putative interneurons respond significantly more similarly to the distracter during attention (see example narrow spiking cell in Figure 2.4a), while wide cells show no significant response to the target or distracter as a population.

Competitive Interaction Model by Cell Type

One potential reason that the mean wide cell population shows no strong similarity to either the distracter or target is that the response to the pair cannot be well explained as a weighted combination of the target and distracter in the wide cells,

while it can be explained this way in the narrow cells. To test this, we created a linear model that describes the passive response to the stimulus pair as a weighted average of the passive response to the two song components presented alone. We then assessed the accuracy of the fit by calculating the correlation coefficient between the fit and actual response to the stimulus pair. We found that the correlation between the model fit and the actual response was significantly higher in the narrow cells ($p < 0.0001$, Mann–Whitney U test, median narrow $r=0.1739$, median wide $r = 0.4817$; Figure 2.5).

Electrophysiological Properties Associated with Attention Effect Size

We found that the effect size of attentional modulation (Figure 3d) in the narrow spiking units was positively correlated with two factors – the temporal sparseness of the response attention effect size correlates with temporal sparseness (Figure 6a; $r=0.341$, $p = 0.0034$), and the degree of suppression of the baseline firing rate in the attention condition (vs. passive condition, Figure 2.6b; $r = 0.274$, $p = 0.0187$). The temporal sparseness was not significantly different in attention and passive conditions. No factors were significantly correlated with the attention effect size in wide cells (temporal sparseness: $r=0.038$, $p=0.70$; baseline firing rate difference: $r = 0.1$, $p = 0.30$; data not shown)

2.3 Methods

Subjects. Eight adult European starlings, Sturnus vulgaris, wild-caught in southern California, served as subjects. Prior to any training or testing, starlings were housed in large mixed-sex aviaries with food and water available ad libitum. Daily light-dark schedules always matched natural photoperiods. All procedures were conducted in accordance with the University of California, San Diego IACUC approval.

Stimuli. We derived a basis set of 37 clips of conspecific birdsong (duration = 2.0 seconds), each drawn without overlap, from libraries of longer continuous song bouts of (n=7) male starlings. Singers were recorded in locations greater than 1000 miles from the locations of subjects, and each song was unfamiliar to the subjects prior to training. Each clip was band-pass filtered to 800 Hz – 9000Hz (check), which roughly matches the center of the hearing range of the starling species (Dooling et al., 1986). Clips were then transformed into each of two frequency ranges (800 Hz –3634 Hz and 4439 Hz – 9000 Hz), yielding two pitch-shifted versions of the original stimulus, one high frequency and one low frequency. Note that the two versions of a given stimulus are identical in their spectrotemporal profile, but completely separated in frequency space (supplemental figure). The amplitude of each clip was normalized to equal mean power (mean RMS = AU). This process yielded a total of 37 high frequency and 37 identical low frequency 2-second song clips.

Behavioral training. Subjects (n = 8) were trained on a 2AC match to sample task using operant conditioning techniques. Subjects were isolated in sound attenuation chambers (Acoustic Systems, Austin TX) equipped with an operant panel

(see Figure 2.1b). Subjects learned to peck at fixed locations on the panel (response port) to received food through a series of shaping procedures (Gentner and Hulse, 2000; Gentner and Fenn, 2006). Once trained to use the operant panel, subjects began training on the match to sample task. Subjects initiated a match to sample trial by pecking at the center response port, triggering the presentation of a single high frequency or low frequency song clip ("sample" stimulus). After a 500ms pause, a second song clip in the same frequency domain as the sample was presented ("target" stimulus). If the target stimulus was the same as the sample stimulus, a peck into the left response port was considered "correct". If the target stimulus was different than the sample stimulus, a subject peck into the right response port was considered "correct". Correct trials were rewarded on a variable reinforcement schedule with 3 second access to food, and additionally reinforced with a flickering LED within the port occurring after a correct peck with 100% probability. Responses to the opposite peck port were considered "incorrect" trials and were reinforced with a time-out period during which the house light was extinguished an trials were unavailable. We did not reinforce a failure to respond to either the left or right peck port during the response window or pecks to left or right peck ports prior to stimulus end. The stimulus exemplar presented for each trial was selected (at random with replacement). In the present study, because the sample and target stimuli are presented in the same frequency range (high or low), the only information available to the subject for correct classification is within the frequency range of the sample (see Figure 2.1c). Once performance on this basic task was achieved (performance was greater than 95%

confidence intervals around chance), we gradually introduced an additional stimulus played simultaneously with the target (distracter). The distracter was always in the opposite frequency range from the sample and target (i.e., if the sample and target were "high frequency" stimuli, the distracter was a "low frequency" stimulus; see Figure 1d). The distracter was presented from the same speaker as the sample and target. Initially, the distracter was introduced at a significantly reduced volume (45 dB SPL), and gradually brought up to a mean volume equal to the target and sample (65 dB SPL or 60 dB SPL). Within a block (range: 1 to 30 trials) the sample song remained the song. Non-match songs and distracters were chosen randomly trial to trial. To acclimate the birds to the chronic recording schedule, random periods occurred where the subject was unable to interact with the operant apparatus. These passive periods were indicated to the subject by illuminating LEDS within each of the three ports, without flicker, for the duration of the period. Access to the food hopper was controlled entirely by the subject's response on each trial. Water was available ad libitum. Subjects were considered to learn the task when (criteria). At the end of their training, subjects had completed an average of 152 1000-trial blocks (range: 39-273). In addition to the 8 subjects who completed training, 5 subjects were unable to achieve criterion at full volume level.

Chronic Chamber Acclimation. Following behavioral training, subjects (n=5) were relocated to a modified sound attenuation chamber which allowed for simultaneous behavioral performance and recording of extracellular action potentials. The chamber contained a functionally identical operant panel to the training chambers.

Additional modifications include a 32 channel motorized commutator (Tucker Davis Technologies) affixed to the top of the sound-proof chamber and a microphone (Earthworks Audio) for audio monitoring. For further information on the chronic recording apparatus see Knudsen and Gentner, 2013. Operant control and data acquisition were managed by custom Spike-2 (Cambridge Electronic Design) scripts in combination with a Power 1401 input-output device. Subjects were given time to acclimate to the new environment until performance on the task was at criterion.

Microdrive Implantation. Once acclimated to the chronic recording chamber, subjects (n=5) were implanted with a custom electrode microdrive assembly (Scripps Institute of Oceanography Machine Shop) consisting of a permanently implanted metal base plate, a screw microdrive with adjustable rostral-caudal positioning, and plastic outer housing. For an in depth description of the electrode microdrive assembly see Knudsen and Gentner, 2013. Each bird was anesthetized with isoflurane, and an incision was made in the scalp. The scalp was resected and outer layers of the skull were removed as necessary to identify the bifurcation of the Y sinus. This histological marker was used to identify the proper stereotaxic coordinates for the caudal mesopallium of the left hemisphere (2300-2600 rostral of the Y sinus bifurcation, 1000-1250 lateral to the midline). A small craniotomy was made over the target region and a small area of dura was resected to aid electrode penetration. The base plate was then permanently affixed to the skull with dental cement, and the microdrive was positioned above the target region. A 1x16 channel single shank electrode array (NeuroNexus Technologies) attached to the microdrive was slowly lowered into the

brain to a depth of approximately 800 µm. A thin layer of silicone gel (3-4680, Dow Corning, Midland, MI; Jackson & Muthuswamy, 2008) was applied to the area surrounding the electrode penetration. Two secondary electrodes, one ground and one reference, were constructed by attaching a small length of 0.003" diameter PtIr wire (2-4mm in length) to a longer section of fine, PVC-insulated 310 wire (Pacific Wire and Cable Corp., Santa Ana, CA). The 310 wire ends of the secondary electrodes were then soldered to the ground and reference leads, respectively, on the electrode array, and the PtIr ends were permanently implanted into the brain via two separate craniotomies approximately 3mm rostral to the implant site. Finally, a 16 channel headstage (20x gain; HST/16V-G20, Plexon) affixed to the outer housing was connected to the implanted electrode array. In subsequent electrode penetrations within a subject, the outer housing was removed, the electrode array was retracted and the microdrive was moved slightly in the rostral-caudal axis, always remaining within the reported range. Prior to each penetration, the electrode array was coated with fluorescent dye (diI) to aid in track identification during histology. Upon recovery from anesthesia, the bird was given food and water ad libitum for several hours.

Chronic Recording Initiation. Recordings were initiated by tethering the bird via a cable connected to the headstage on one end and the commutator on the other. The output of the commutator went to a 16 channel microelectrode amplifier (model 3600, A-M Systems, Inc., Sequim, WA) which bandpassed (low cutoff: 300Hz, high cutoff: 5kHz) and amplified the signal to a user-specified gain (2000x-10000x). The

signal was then sent to a CED Power1401 analog-to-digital converter which digitized each channel at [19-25 kHz].

Stimulus Selection. Recordings were initiated by playing, randomly with replacement (time spacing between), all 72 “high frequency” and “low frequency” song clips while electrical activity from all recorded channels were monitored. If no action potentials sensitive to these stimuli were apparent in the recorded channels, the electrode was slowly lowered until auditory activity was observed. Due to the nature of electrophysiology research, we restricted the stimulus set used for the behavioral trials within a recording session to allow for greater repetitions of stimulus presentations. The entire stimulus library was presented passively to the animal until a subset of stimuli were chosen so that the subset consisted of both “high frequency” and “low frequency” song clips, and at least one stimulus elicited a response from the auditory unit(s) isolated at the time. This subset was then presented passively and pseudorandomly until a total 5-10 presentations of each stimulus had been completed. The responses of each isolated unit to this subset of stimuli were then analyzed. From the subset of stimuli presented, two “high frequency” stimuli and two “low frequency” stimuli were chosen as primary stimuli for the recording day. Primary stimuli were chosen to maximize, in order, the number of units eliciting a response to at least one stimulus, and the variability of the response within a unit to each of the stimuli.

Chronic Recording Behavior. The four primary stimuli were used to build 2 block types of n=34 trials each (Figure 2.2c-e). Block type 1 consisted of a high-frequency sample, Block type 2 consisted of a low-frequency sample. To create a

block, one stimulus was chosen as the sample, the remaining stimulus in that frequency range was chosen as the non-match target, and one stimulus from the opposite frequency range was chosen as the primary distracter. The breakdown of the 34 trials was as follows: 10 match trials with primary distracter, 10 non-match trials with primary distracter, 5 match trials with no distracter, 5 non-match trials with no distracter, and 4 catch trials. Catch trials are a special case where the match or non-match stimulus is presented in the opposite frequency range of the sample. These trials allow us to test whether the bird is solving the task as intended – by discriminating within the frequency range of the sample – or by an alternative strategy of listening for the spectrotemporal features of the sample, regardless of the frequency range.

Trials within a block were presented randomly with a minimum inter-trial interval of 2.0 seconds. If the bird responded correctly to a trial, it was removed from the block. If the bird did not respond to a trial, it was either immediately repeated or added to end of block. If the bird responded incorrectly, the trial was added to the end of the block. This structure ensured that upon completion of a block, each of the 34 trials from the block had been correctly completed once. In addition to the active trials, each trial from the block was also presented randomly without replacement to the passive animal, either before or after the behavioral session. Passive sessions were indicated to the animal, as in the training condition, by the illumination of LEDs within the peck ports. Video monitoring confirmed the subject did not attempt to perform trials during this time.

Upon completion of a block, another block of trials was initiated. Typically, the next block would consist of a sample from the opposite frequency range of the prior sample stimulus. For instance, if the bird had previously performed a block of trials with a "high frequency" sample, the next block would consist of trials with a "low frequency" sample. Occasionally, especially if performance on the prior block was poor, the sample stimulus frequency range of the following block would remain the same. Consistent above-chance behavioral performance while implanted was obtained from n=2 birds.

Histology. Neural tissue was preserved in 10% formalin and then cryoprotected in a solution of 30% sucrose in phosphate buffer solution. Tissue was then frozen, sectioned on a freezing microtome (40um thickness), and mounted onto glass slides. Fluorescence images were taken of each section to visualize electrode penetrations. Alternating sections were then processed with either a nissl stain to visualize cell bodies, or gallyas stain to visualize myelin tracks. Images were taken of the stained slides and co-registered with fluorescent images to localize electrode tracks within the established cytoarchitectonic boundaries of the auditory regions in the avian forebrain. Only units recorded within CM were included in this study.

Spike sorting. Offline, we extracted threshold-triggered putative action potentials from the raw electrode recording, and separated them from noise and other units with a combination of PCA analysis and changes in waveform shape over time (Spike2, CED). We confirmed isolation quality using the ISI distribution of each putative single or multi-unit. Well-separated waveform clusters with less than 1%

refractory-period violations were considered single-units; Well-separated waveform clusters with greater than 1% refractory period violations were considered multi-units. We recorded a total of 500 song pairs from 332 units (288 single units, 44 multi-units).

Data Inclusion. Only neurons that responded to at least one primary stimulus used in the associated behavioral session were included for study. A significant auditory response was considered a mean firing rate over the entire stimulus greater than 1 standard deviation above baseline, or a trial by trial correlation in firing pattern greater than 1 standard deviation above baseline. This second criterion was added because several cells were seen to respond sparsely throughout the stimulus, but do so in a highly time locked manner. Additionally, for all analyses relevant to behavior, only neurons that were recorded during good behavioral sessions were included. Good behavioral sessions are defined as sessions where the mean performance to all trials is greater than the 95% confidence intervals around chance performance, and catch trial performance is above chance (d' = 0). Since there were relatively few catch trials per recording session, it was impractical to calculate confidence bounds on catch trial performance on a per session basis. This left us with a total of 198 song pairs recorded from 117 units (112 single units, 5 multi-units) in n=2 subjects.

Spike Width. We calculated the mean trough-to-peak action potential (spike) width for all single units recorded, regardless of behavioral performance (n=288). Using a clustering algorithm (calculated with Euclidean inner-squared distance, Matlab statistics toolbox,) we segregated the narrow cell population. The boundary between narrow spiking and wide spiking cells was determined to be 230 μs. Within

the eligible population of 112 single units, 46 were narrow-spiking and 66 were wide-spiking.

Stimulus Pair Analysis. Within trials that included the components of a specific pair of high and low stimuli, comparisons were made between several trial types ("contexts") presented to the animal. Our main interest was in the neural response during the target portion of trials that were performed correctly by the bird ("correct attention" condition), specifically between the stimulus pair preceded by a high frequency cue (type A_hi) and stimulus pair preceded by a low frequency cue (type A_lo). For control purposes, we also assessed responses to these contexts when the bird performed the trial incorrectly ("incorrect attention" condition; AX_hi, AX_lo) and when the bird was presented the trial passively ("passive" condition; P_hi, P_lo). Additional contexts of interest included trials where the components of the stimulus pair were presented without distracters, such as high stimulus alone preceded by high frequency cue (type AS_lo) and low stimulus alone preceded by low frequency cue (type AS_lo). Only the target portion of each trial was analyzed (sample and break periods are not included in analysis). See figure.

PSTH calculation. Within a context, the response to each single trial was binned at 1ms and convolved with an exponential filter (tau=5ms) prior to calculating the mean peri-stimulus time histogram. We included the portion of the neural response beginning at target onset and ending 200ms after target offset.

PSTH correlation. We calculated the correlation coefficient between the mean PSTH in the P_hi context and the mean PSTH in the P_lo context (Pcorr), and

compared this to the correlation coefficient between the mean PSTH in the A_hi context and the mean PSTH in the A_lo context (Acorr).

Attention Modulation Effect. Within a stimulus pair, we determined the magnitude of the attention effect of a given neuron by comparing modulation in the passive and correct attention conditions. We first calculated the mean PSTH in the P_hi and P_lo contexts. We then calculated the mean squared error (MSE) within each bin across these two contexts. This gave us the time-dependent baseline modulation incurred by the different preceding sample stimuli in the two contexts. We repeated this calculation for the attention condition, taking the MSE between the A_hi and A_lo PSTHs. To determine whether the attention-dependent modulation was different from the passive condition, we performed a paired nonparametric test on the "passive" and "correct attention" MSE at each time bin. Effect size for each unit was determined using the wilcoxon sign rank statistic ($Z/\sqrt{nsamples}$)

Significantly Modulated Bins. To determine the time bins where significant modulation occurred during the attention condition, we calculated the discriminability between a unit's response to a song pair during "attend high" and "attend low" conditions within each non-overlapping 10ms bin from stimulus onset until 500 ms after stimulus end. Bins where the d prime between "attend high" and "attend low" trial by trial firing rates were significantly different than a shuffled distribution (n=1000 shuffles, alpha = 0.05) were considered significantly modulated.

Solo Stimulus Similarity. We calculated the mean PSTH to A_hi, A_lo, AS_hi, and AS_lo. We first calculated the error between attending to the target and presenting the target alone (MSE_targ).

$$\textit{Eq 1}.\ MSE_{targ} = (PSTH_{A_hi} - PSTH_{AS_hi})^2 + (PSTH_{A_lo} - PSTH_{AS_lo})^2$$

We then calculated the reciprocal error between attending to the target and the distracter alone (MSE_dist), which was similarly defined:

$$\textit{Eq 2}.\ MSE_{dist} = (PSTH_{A_hi} - PSTH_{AS_lo})^2 + (PSTH_{A_lo} - PSTH_{AS_hi})^2$$

We took the difference between these the two error vectors to get a measure of similarity between target and distracter stimulus over time.

$$\textit{Eq 3}.\ MSE_{diff} = MSE_{dist} - MSE_{targ}$$

Positive values represent more error representing the distracter than the target, indicating the response is more similar to the target alone than the distracter alone. Negative values represent a response more similar to the distracter alone than the target alone. To assess whether or not a unit responded significantly more like the target or distracter during the response, we tested whether MSE_{diff} was significantly different from zero using a nonparametric test.

To compare across different attention conditions, the MSE_{diff} was calculated for incorrect attention and passive conditions by substituting incorrect and passive trials for the correct attention trials ($PSTH_{A_hi}$ and $PSTH_{A_lo}$). To compare across different stimulus pairs and units, we then normalized each MSE_diff vector by the maximum absolute MSE_{diff} value across all conditions. The mean MSE_{diff} was then taken for each condition, yielding a mean normalized MSE_{diff} value for each attention condition.

Competitive Interaction Model. To describe a unit's response to the stimulus pair as a combination of the responses to each of the song components alone, we modeled the response to the song pair as a linear combination of each song alone:

$$\textit{Eq 4}.\ PSTH_{P_pair} = \mathbf{a}*PSTH_{PS_hi} + \mathbf{b}*\ PSTH_{PS_lo} + \mathbf{c}$$

where $PSTH_{P_pair}$ is the mean response to the passively presented stimulus pair ($PSTH_{PS_hi}$ and $PSTH_{PS_lo}$), $PSTH_{PS_hi}$ is the response to the passively presented high frequency stimulus, and $PSTH_{PS_lo}$ is the response to the passively presented low frequency stimulus. We solved for coefficients **a**, **b**, and intercept constant **c** to yield the optimal fit. We tested the accuracy of the fit by calculating the correlation coefficient between the model fit ($PSTH_{fit_pair}$), and the actual response ($PSTH_{P_pair}$) (Figure 2.5).

Temporal Sparseness. We used previously established method of calculating temporal sparseness in the songbird auditory system (Coleman and Mooney, 2004).

This method simply calculates the proportion of time bins during stimulus presentation that contain firing rates above the baseline rate, and subtracts this fraction from one.

$$Eq\ 5.\ TS = 1 - f_{resp}$$

Where f_{resp} is the fraction of bins that respond above baseline firing. This calculation results in a range of values from 0 to 1 with increasing values representing increasingly sparse responses.

2.4 Discussion

While European starlings and other avian species have been shown to perform cocktail party-like tasks in the controlled laboratory (MacDougall-Shackleton et al., 1998; Narayan et al., 2007), no studies have previously shown these birds to be capable of performing selective auditory attention tasks. However, given the acoustic demands of their natural environment, which can be very noisy, their successful performance on the task used in this study should not be regarded as terribly surprising, particularly when ethologically relevant stimuli are employed.

The neurophysiological results of our study indicate that auditory attention modulates the activity of single units in CM, a secondary cortical auditory region selective for conspecific stimuli. This result supports recent studies of attention to

speech in humans (Crone et al., 2001; Mesgarani and Chang, 2012), which demonstrate attention dependent modulations of LFP activity in brain regions sensitive to human speech.

Interestingly, we found differential attentional modulation in two groups of cells identified by their action potential widths. Fast-spiking inhibitory interneurons and broad-spiking neurons have been previously separated in extracellular recordings from a number of different brain regions (Fox and Ranck, 1981; Buzsáki and Eidelberg, 1982; Wilson et al., 1994; Csicsvari et al., 1999; Rao et al., 1999; Frank et al., 2001; Constantinidis and Goldman-Rakic, 2002; González-Burgos et al., 2005; Hasenstaub et al., 2005), including cortical sensory regions (Mountcastle et al., 1969; Simons, 1978; McCormick and Prince, 1985; Gur et al., 1999; Bruno and Simons, 2002; Swadlow, 2003; Yazaki-Sugiyama et al., 2009; Yokota et al., 2011) and avian sensory-motor regions (Rasmusson, 2000; Meitzen et al., 2009; Goldberg and Fee, 2010; Shea et al., 2010; Raksin et al., 2012). In addition to the intrinsic differences in response properties noted in many of these studies, narrow and wide spiking neurons are also differentially affected by experience (Woloszyn and Sheinberg, 2012), task demands (Hussar and Pasternak, 2009, 2012; Quirk et al., 2009; Kaufman et al., 2010) and attention (Mitchell et al., 2007; for a recent review see (Merchant et al., 2012). For instance, in the auditory domain, narrow-spiking cortical neurons recorded from awake-behaving primates are significantly more sensitive to stimulus categories than broad-spiking cells (Tsunada et al., 2012). Additionally, parvalbumin positive neurons in the cortex are directly modulated by sources of acetylcholine such as the nucleus

basalis (Koós and Tepper, 2002; Froemke et al., 2007), and have been shown to modulate broad-spiking neuron activity consistent with gain modulation (Atallah et al., 2012), which may implicate a specific role for this cell class in attention-dependent neural activity.

While this is the first study to differentiate between cell types using extracellular spikewidths in the avian auditory cortex, parvalbumin has been shown to stain richly throughout the avian auditory cortex, particularly within CM and the field L complex (Wang et al., 2010). Parvalbumin staining elucidates the layered organization of the avian auditory cortex, which closely mirrors that of mammalian A1 (Brückner et al., 1994; Cruikshank et al., 2001). Within CM, two layers are clearly denoted by parvalbumin staining: a thin CM-externus band, which stains heavily for parvalbumin, and a thicker CM-internus, which stains more lightly. Parvalbumin positive cells in these regions have very short processes (Braun et al., 1985), consistent with a role as local interneurons. The two CM layers are not visible in our nissl or gallyas stained tissue, so we are unable to differentiate the location of our neurons between these layers, however we suspect due to the positioning of our electrodes that the majority of cells we recorded were within CM-internus.

In this study, narrow-spiking cells were shown to respond significantly more like the ignored stimulus in a song pair than the target. Interestingly, this effect completely disappears during incorrect attention trials. If this cell class can indeed be identified as GABA-ergic interneurons, this result suggests an attention-dependent suppression of the ignored stimulus, and it suggests that incorrect performance is

correlated with a lack of ability to suppress the distracter. This result supports a larger body of work (Alho et al., 1987; Rif et al., 1991; Atiani et al., 2009; Neelon et al., 2011; Bridwell and Srinivasan, 2012) which suggests that a key component of auditory attention is the active suppression of acoustic distracters.

This is not the first study to observe an increase in neural responses to ignored stimuli. The P200 EEG response, associated with auditory attention (Picton and Hillyard, 1974), has been shown to increase in magnitude in response to irrelevant tones, compared to attended tones or when attention was directed away from auditory stimuli (Alho et al., 1987; Rif et al., 1991). This increase is theorized to represent an active suppression of auditory distracters during demanding auditory attention tasks. Furthermore, this active suppression has been shown more directly in the visual system, where distracters presented near the target location have been shown to be actively suppressed during attention (Hopf et al., 2006), and trial by trial distracter suppression magnitude is associated with correct performance (Bridwell and Srinivasan, 2012). This type of active suppression, in combination with amplification of the target, would lead to the overall gain control mechanism of attention observed repeatedly in sensory attention (Hillyard et al., 1998; Bosman and Womelsdorf, 2009; Reynolds and Heeger, 2009).

Within the narrow cell class, we found two neurophysiological properties that positively correlate with the observed attention effect size: the temporal sparseness within the neural response to the stimulus pair, and the degree of suppression in baseline firing between passive and active states. A change in baseline firing between

behavioral states has been previously reported in cortical auditory areas (Otazu et al., 2009; Knudsen and Gentner, 2013), and cells that experience a greater baseline change between states may be generally more affected by task demand. The mechanisms underlying the correlation between with sparseness and attention effect size is less clear, but the results indicate that more selective narrow spiking cells are more affected by attention.

While a similar proportion of broad-spiking cells showed significant attentional modulation (62 of 72 song pairs from narrow cells and 87 of 108 song pairs from wide cells), we were unable to correlate this modulation with any neurophysiological properties. This could be due to the fact that the "broad" cell class actually incorporates several cell types, morphological or functional, that simply cannot be differentiated by extracellular properties such as spike width. Classical anatomical work identifies at least four morphologically separable cell types in the CM region (Saini and Leppelsack, 1981), and more groups may be functionally separable.

Additionally, within the broad-spiking class, the response to the stimulus pair cannot be well modeled as a combination of the responses to the song components alone. Unlike the narrow-spiking cell responses, which were fit significantly better by this model, it appears broad-spiking cells respond quite differently to the pair of stimuli together than either component alone. This could be due to many factors, such as the nonlinear response properties which are notoriously difficult to measure in this region (Theunissen et al., 2000; Gill et al., 2008; Meliza and Margoliash, 2012).

While attention-dependent modulation in wide cells does not appear to reflect an overall amplification of the target or distracter stimulus, attention still significantly modified the responses of 81% of broad-spiking cells. The question remains, then, what does this modulation relate to? We suspect that the key lies in the specific timepoints used by the subjects to solve the task – since the subject is given 2 seconds to detect the target in the stimulus pair, it is possible and even likely that the subject is listening for specific features at specific time points, thus, the attentional state of the animal is likely to be fluctuating throughout the target/distracter presentation. Recent human work, for instance, indicates that attention to speech fluctuates throughout the course of target stream detection, with greater modulation occurring at information-rich periods such as word onsets (Sanders et al., 2002; Astheimer and Sanders, 2009, 2011, 2012; Besle et al., 2011). In primary auditory regions, masking songs with noise leads to the addition of spurious spikes and a loss of informative portions spikes (Narayan et al., 2007). Perhaps the time points where these changes occur would be likely targets of auditory attention in noisy environments.

Cursory attempts to relate wide cell modulation with fluctuations in signal to noise ratio of the target within our dataset have proven unsuccessful, and we expect to follow up with experiments that carefully control the allocation of attention in time. We hope that this work, in combination with future work further exploring attention-dependent modulation in various cell classes, will lead to a greater mechanistic understanding of attention to natural auditory signals.

2.5 Acknowledgements

Chapter 2, in part, is currently being prepared for publication of the material as it may appear under authorship of Emily L. Caporello and Timothy Q. Gentner. Emily Laura Caporello was the primary investigator and author of this material.

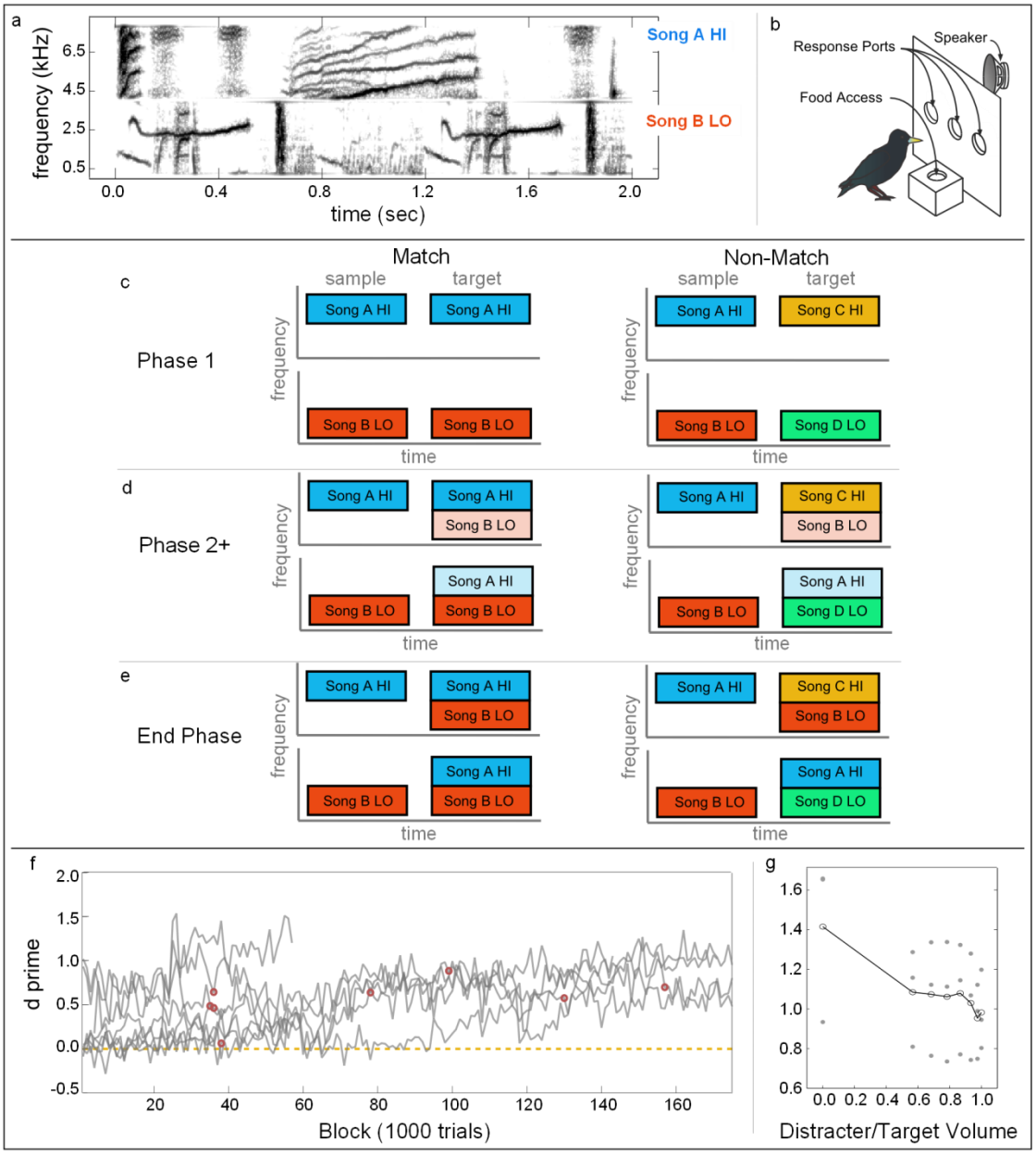


Figure 2.1. Auditory Attention Behavior in European Starlings. (a) Sample spectrogram of two song clips used for behavioral training, one restricted to a high frequency band, and one restricted to a low frequency band. (b) Schematic of operant training panel.(c) schematic of behavioral training phase one. Four trial types are shown, high frequency sample followed by matching target (upper left), high frequency target followed by non-match target (upper right), low frequency sample followed by matching target (lower left), and low frequency sample followed by non-match target (lower right). (d) schematic of behavioral training phases two and higher. During the target period, a distracter song clip was played at a low volume. The distracter volume was gradually increased until the mean volume was equal to that of the target (end phase, e). (f) Acquisition of behavior for n=8 successfully trained subjects. Chance performance at d prime = 0 indicated with yellow dotted line. Red circles indicate block at which end phase training began for each subject. (g) Performance of n=3 birds on various distracter/target volume ratios. Mean performance shown with black solid line.

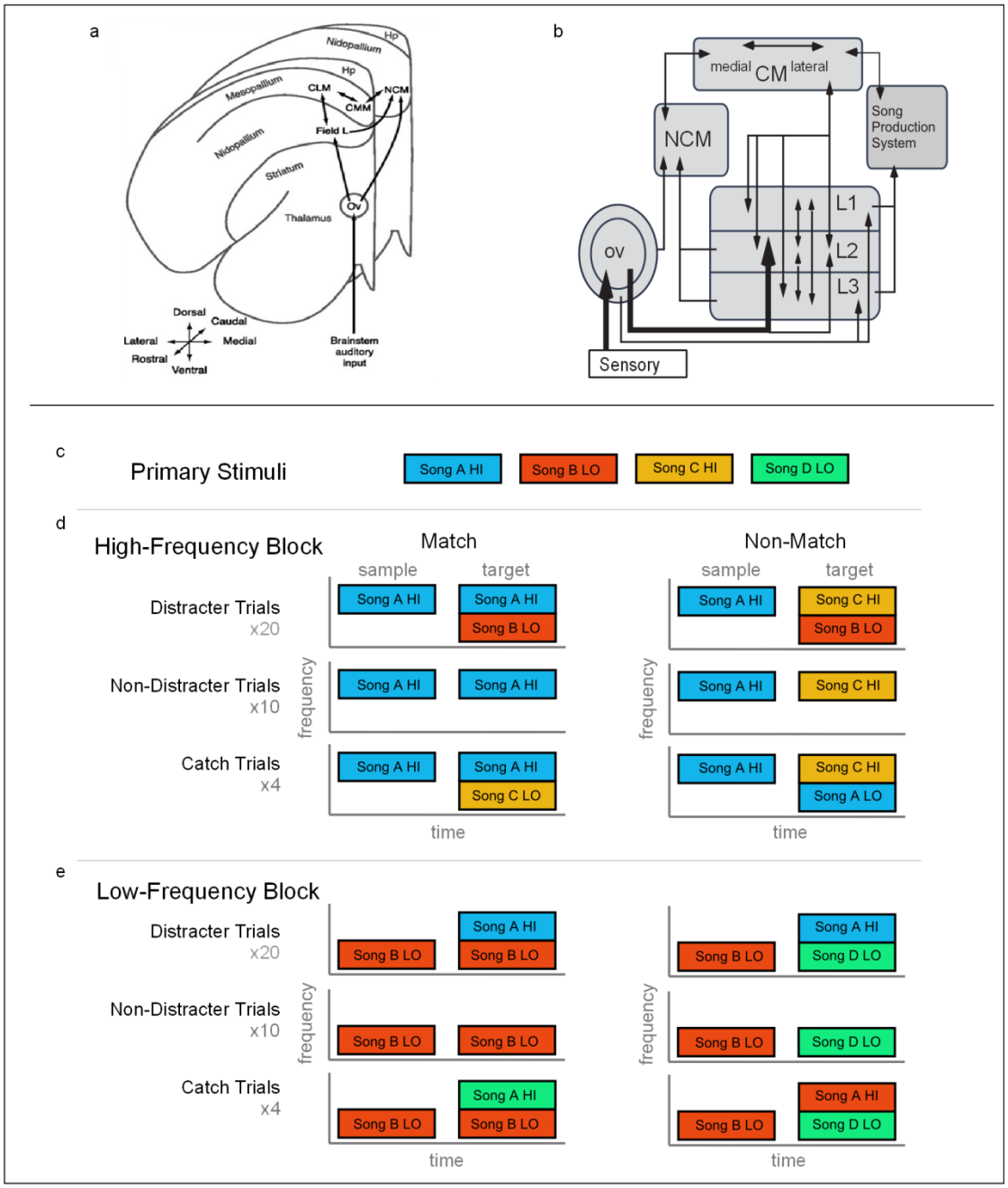


Figure 2.2. Chronic recording implantation and behavioral paradigm. Avian auditory pathway (a) and connectivity schematic (b) show sensory input from the thalamic nucleus ovoidalus (Ov) to field L and its subparts (L1, L2, and L3), the earliest cortical auditory regions. From there, auditory signals are sent to higher level regions NCM, CM (medial and lateral), and song production nuclei. Extracellular recording electrodes were implanted into CM. (c-e) Chronic recording schematic. Four primary stimuli, two high frequency and two low frequency, are used to create behavioral blocks (c). Example high frequency (d) and low frequency (e) blocks containing full volume distracter trials, non-distracter trials, and catch trials

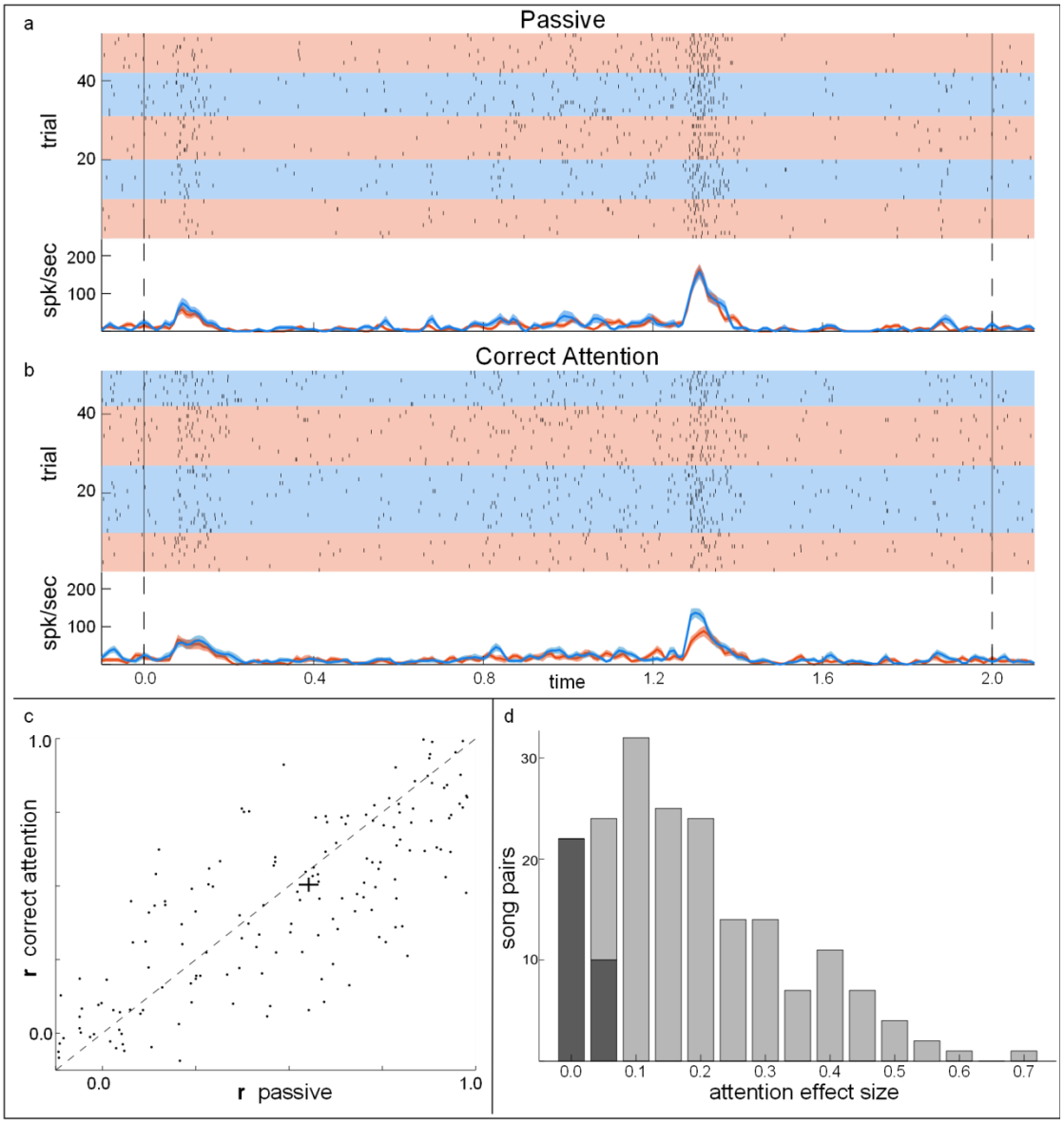


Figure 2.3. Attention-Driven Modulation of Auditory Neuron Responses. Sample neural response to a pair of stimuli during passive (a) and correct attention (b) trials. Pink indicates trials preceded by low frequency samples, blue indicates trials preceded by high frequency samples. Rasters are shown in chronological order of recording. PSTHs indicate mean (solid line) and standard error (shading). (c) Population (n= 188 song pairs) data showing the correlation coefficient between responses to passive trials preceded by low vs high frequency samples versus the correlation coefficient between correct attention trials preceded by low vs high frequency samples. Crossbars indicate population mean and standard error. (f) Histogram of attention effect sizes. Song pairs significantly modulated by attention are shown in light grey.

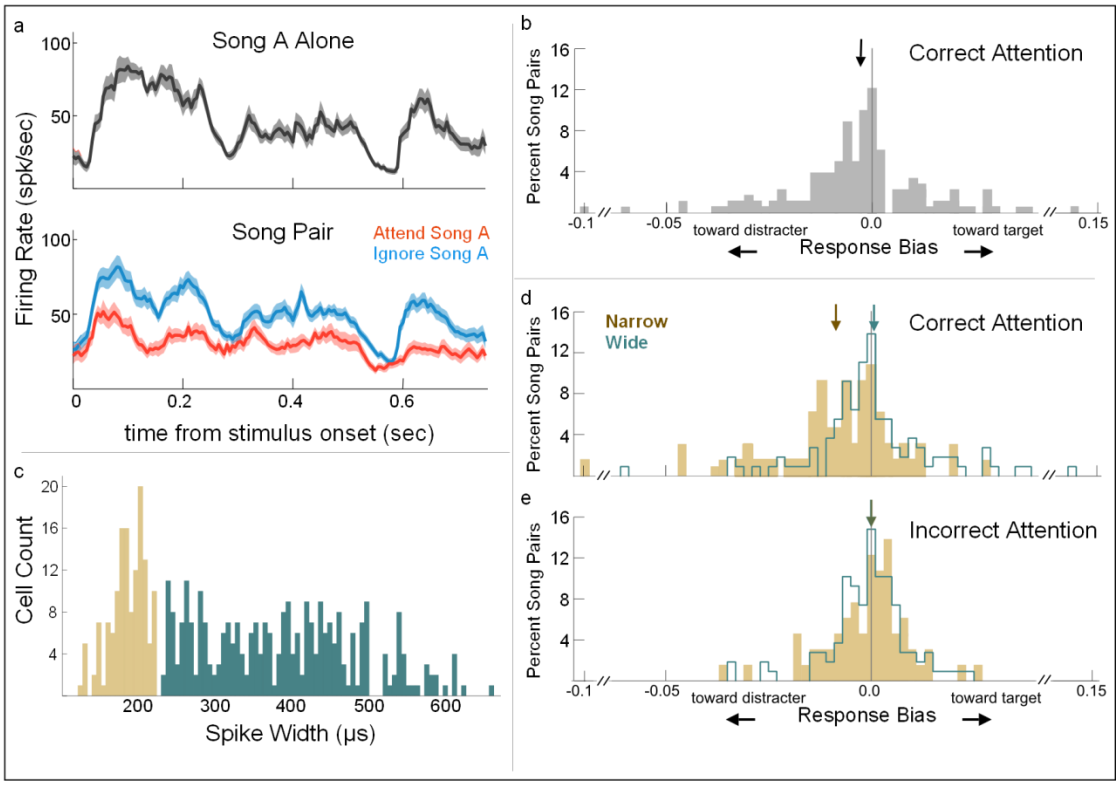


Figure 2.4. Attention Response Bias. (a) Sample cell response to song A alone (black) and song A presented within a stimulus pair when attended (red) or ignored (blue). Mean response shown in solid line, shading indicates standard error over all trials. (b) population response bias during correct attention trials. Negative values indicate attention responses more similar to distracter, positive values indicate attention responses more similar to the target. Mean of distribution indicated with black arrow. (c) spike width distribution. Yellow group indicates the narrow-spiking cell class, green indicates the wide-spiking cell class. Correct attention (d) and incorrect attention (e) Response Bias split by cell class. Narrow spiking cells shown in solid yellow bars, wide spiking cells shown in green outline. Means of distributions shown with arrows.

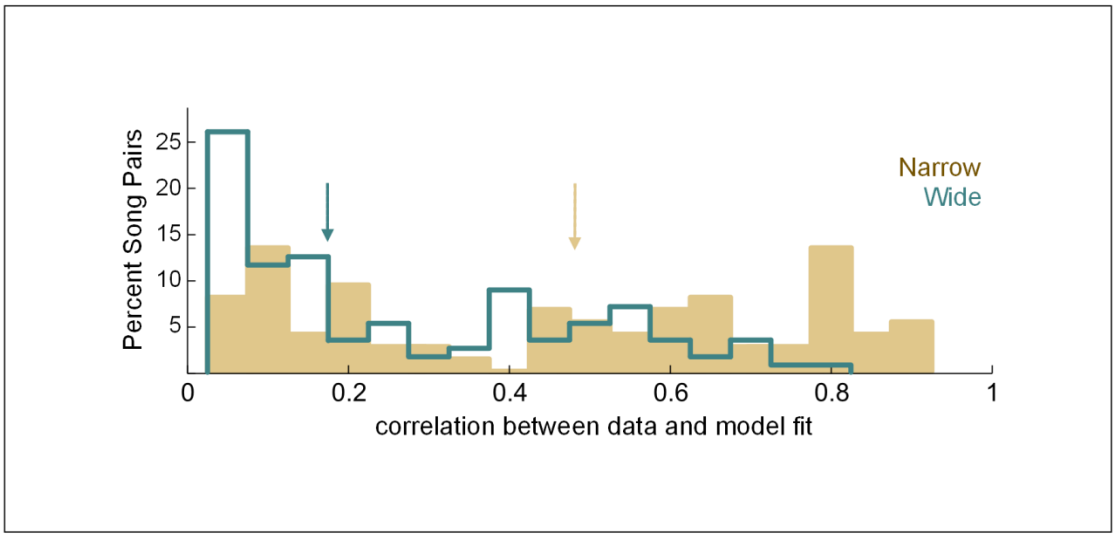


Figure 2.5. Competitive Interaction Model by Cell Type. Distribution of correlation coefficients between passive response to the stimulus pair, and the model prediction based on responses to each stimulus alone. Narrow cell population shown in solid yellow bars, wide cell population shown in green outline. Median values indicated with arrows.

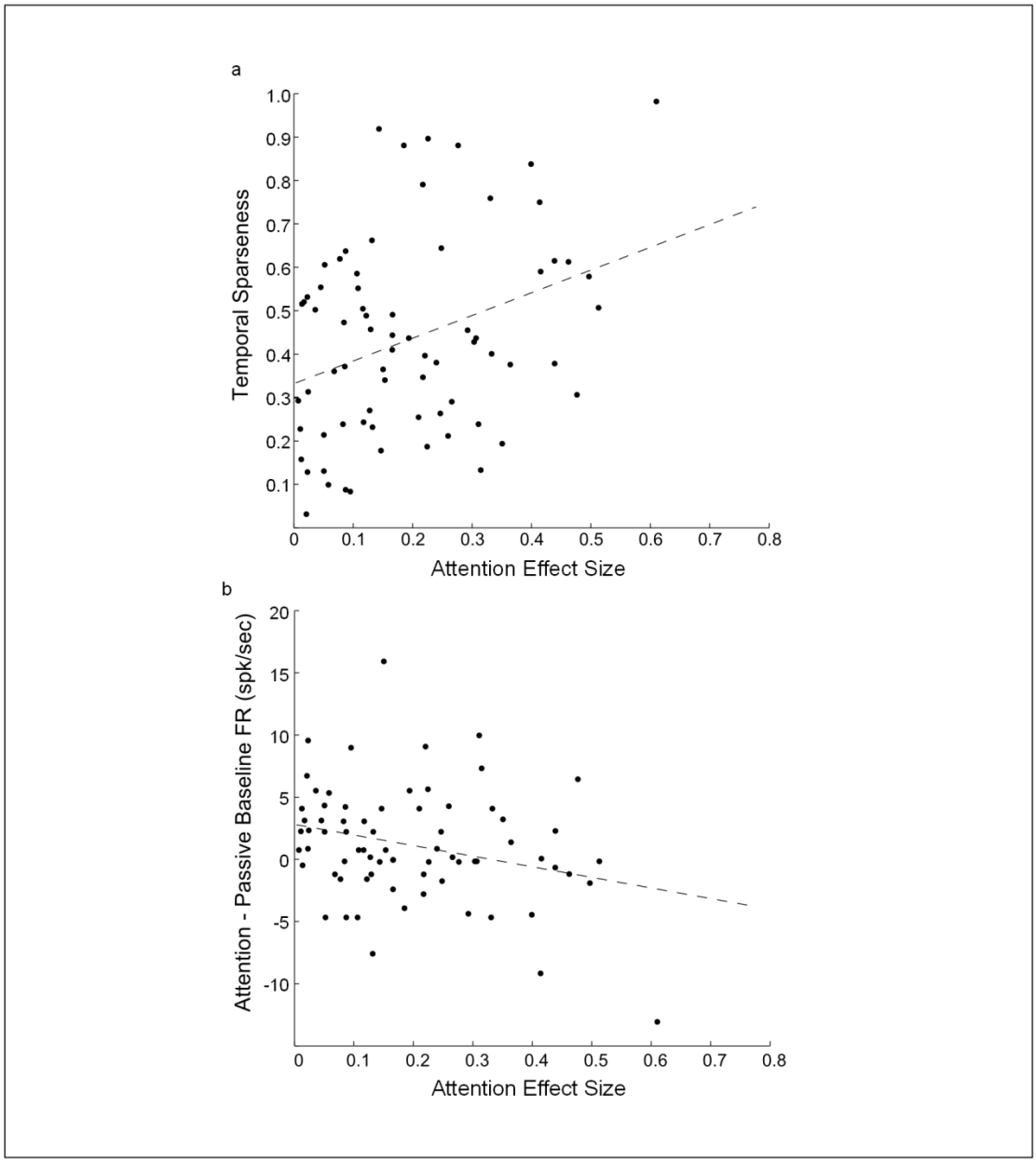


Figure 2.6. Electrophysiological Properties Associated with Attention Effect Size. Attention effect size vs. temporal sparseness (a) and attention-related baseline firing rate suppression (b) for narrow cell population. Each point represents one stimulus pair. Linear fit indicated with dashed line.

2.6 Bibliography

Alho K, Töttölä K, Reinikainen K, Sams M, Näätänen R (1987) Brain mechanism of selective listening reflected by event-related potentials. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 68:458–470.

Astheimer LB, Sanders LD (2009) Listeners modulate temporally selective attention during natural speech processing. Biological psychology 80:23–34.

Astheimer LB, Sanders LD (2011) Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia 49:3512–3516.

Astheimer LB, Sanders LD (2012) Temporally selective attention supports speech processing in 3- to 5-year-old children. Developmental cognitive neuroscience 2:120–128.

Atallah B V, Bruns W, Carandini M, Scanziani M (2012) Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. Neuron 73:159–170.

Atiani S, Elhilali M, David S V, Fritz JB, Shamma SA (Univeristy of M (2009) Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. Neuron 61:467–480.

Besle J, Schevon C a, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. Journal of neuroscience 31:3176–3185.

Bosman C, Womelsdorf T (2009) Neuronal Signatures of Selective Attention – Synchronization and Gain Modulation as Mechanisms for Selective Sensory Information Processing. In: From Attention to Goal Directed Behavioral, pp 3–28.

Braun K, Scheich H, Schachner M, Heizmann CW (1985) Distribution of parvalbumin , cytochrome oxidase activity and 14C-2-deoxyglucose uptake in the brain of the zebra finch I . Auditory and vocal motor systems. Cell and Tissue Research:101–115.

Bridwell DA, Srinivasan R (2012) Distinct attention networks for feature enhancement and suppression in vision. Psychological science 23:1151–1158.

Brückner G, Seeger G, Brauer K, Härtig W, Kacza J, Bigl V (1994) Cortical areas are revealed by distribution patterns of proteoglycan components and parvalbumin in the Mongolian gerbil and rat. Brain Research 658:67–86.

Bruno RM, Simons DJ (2002) Feedforward mechanisms of excitatory and inhibitory cortical receptive fields. The Journal of neuroscience : the official journal of the Society for Neuroscience 22:10966–10975.

Buzsáki G, Eidelberg E (1982) Direct afferent excitation and long-term potentiation of hippocampal. Journal of neurophysiology 48:597–607.

Coleman MJ, Mooney R (2004) Synaptic transformations underlying highly selective auditory representations of learned birdsong. The Journal of neuroscience : the official journal of the Society for Neuroscience 24:7251–7265.

Connors BW, Gutnick MJ (1990) Intrinsic firing patterns of diverse neocortical neurons. TIMS 13:99–104.

Constantinidis C, Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. Journal of neurophysiology 88:3487–3497.

Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. Clinical Neurophysiology 112:565–582.

Cruikshank SJ, Killackey HP, Metherate R (2001) Parvalbumin and calbindin are differentially distributed within primary and secondary subregions of the mouse auditory forebrain. Neuroscience 105:553–569.

Csicsvari J, Hirase H, Czurkó A, Mamiya A, Buzsáki G (1999) Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving Rat. The Journal of neuroscience : the official journal of the Society for Neuroscience 19:274–287.

Dooling RJ, Okanoya K, Downing J, Hulse SH (1986) Hearing in the starling (Sturnus vulgaris) Absolute thresholds and critical ratios. Bulletin of the Psychonomic Society 24:462–464.

Erisir A, Lau D, Rudy B, Leonard CS (2013) Function of Specific K + Channels in Sustained High-Frequency Firing of Fast-Spiking Neocortical Interneurons. Journal of neurophysiology 82:2476–2489.

Fox SE, Ranck JBJ (1981) Electrophysiological characteristics of hippocampal complex-spike cells and theta cells. Experimental Brain Research 41:399–410.

Frank LM, Brown EN, Wilson MA (2001) A Comparison of the Firing Properties of Putative Excitatory and Inhibitory Neurons From CA1 and the Entorhinal Cortex. Journal of neurophysiology 86:2029–2040.

Froemke RC, Merzenich MM, Schreiner CE (2007) A synaptic memory trace for cortical receptive field plasticity. Nature 450:425–429.

Gentner TQ, Fenn KM (2006) Recursive syntactic pattern learning by songbirds. Nature:1–5.

Gentner TQ, Hulse SH (2000) Perceptual classification based on the component structure of song in European starlings. The Journal of the Acoustical Society of America 107:3369–3381.

Gentner TQ, Margoliash D (2003) Neuronal populations and single cells representing learned auditory objects. Nature 424:669–674.

Gill PR, Woolley SMN, Fremouw TE, Theunissen FE (2008) What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. Journal of neurophysiology 99:2809–2820.

Gold C, Henze D a, Koch C, Buzsáki G (2006) On the origin of the extracellular action potential waveform: A modeling study. Journal of neurophysiology 95:3113–3128.

Goldberg JH, Fee MS (2010) Singing-related neural activity distinguishes four classes of putative striatal neurons in the songbird basal ganglia. Journal of neurophysiology 103:2002–2014.

González-Burgos G, Krimer LS, Povysheva N V, Barrionuevo G, Lewis D a (2005) Functional properties of fast spiking interneurons and their synaptic connections with pyramidal cells in primate dorsolateral prefrontal cortex. Journal of neurophysiology 93:942–953.

Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. Journal of neurophysiology 89:472–487.

Gur M, Beylin A, Snodderly DM (1999) Physiological Properties of Macaque V1 Neurons are Correlated With Extracellular Spike Amplitude , Duration , and Polarity. Journal of neurophysiology 82:1451–1464.

Hasenstaub A, Shu Y, Haider B, Kraushaar U, Duque A, McCormick D a (2005) Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. Neuron 47:423–435.

Henze DA, Borhegyi Z, Csicsvari J, Mamiya A, Harris KD, Buzsáki G (2000) Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo. Journal of neurophysiology 84:390–400.

Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philosophical transactions of the Royal Society of London Series B, Biological sciences 353:1257–1270.

Hopf J-M, Boehler CN, Luck SJ, Tsotsos JK, Heinze H-J, Schoenfeld MA (2006) Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. Proceedings of the National Academy of Sciences of the United States of America 103:1053–1058.

Hsu A, Woolley SMN, Fremouw TE, Theunissen FE (2004) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 24:9201–9211.

Hussar CR, Pasternak T (2009) Flexibility of sensory representations in prefrontal cortex depends on cell type. Neuron 64:730–743.

Hussar CR, Pasternak T (2012) Memory-guided sensory comparisons in the prefrontal cortex: contribution of putative pyramidal cells and interneurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:2747–2761.

Jackson N, Muthuswamy J (2008) Artificial dural sealant that allows multiple penetrations of implantable brain probes. Journal of neuroscience methods 171:147–152.

Jeanne JM, Sharpee TO, Gentner TQ (2013) Associative Learning Enhances Population Coding by Inverting Interneuronal Correlation Patterns. Neuron 78:352–363.

Jeanne JM, Thompson J V, Sharpee TO, Gentner TQ (2011) Emergence of learned categorical representations within an auditory forebrain circuit. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:2595–2606.

Kaufman MT, Churchland MM, Santhanam G, Yu BM, Afshar A, Ryu SI, Shenoy K V (2010) Roles of monkey premotor neuron classes in movement preparation and execution. Journal of neurophysiology 104:799–810.

Kawaguchi Y, Kubota Y (1993) Correlation of physiological subgroupings of nonpyramidal cells with parvalbumin- and calbindinD28k-immunoreactive neurons in layer V of rat frontal cortex Correlation of Physiological Subgroupings of Nonpyramidal Cells With Parvalbumin- and CalbindinD28k. Journal of neurophysiology 70:387–396.

Knudsen DP, Gentner TQ (2013) Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. Journal of neurophysiology 109:1690–.

Koós T, Tepper JM (2002) Dual cholinergic control of fast-spiking interneurons in the neostriatum. Journal of neuroscience 22:529–535.

MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W (1998) Auditory scene analysis by European starlings (Sturnus vulgaris): perceptual segregation of tone sequences. The Journal of the Acoustical Society of America 103:3581–3587.

Martina M, Jonas P (1997) Functional differences in Na+ channel gating between fast-spiking interneurones and principal neurones of rat hippocampus. The Journal of physiology 505 (Pt 3:593–603.

Martina M, Schultz JH, Ehmke H, Monyer H, Jonas P (1998) Functional and molecular differences between voltage-gated K+ channels of fast-spiking interneurons and pyramidal neurons of rat hippocampus. The Journal of neuroscience : the official journal of the Society for Neuroscience 18:8111–8125.

McCormick DA, Prince DA (1985) Two types of muscarinic response to acetylcholine in mammalian cortical neurons. Proceedings of the National Academy of Sciences of the United States of America 82:6344–6348.

Meitzen J, Weaver AL, Brenowitz E a, Perkel DJ (2009) Plastic and stable electrophysiological properties of adult avian forebrain song-control neurons across changing breeding conditions. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:6558–6567.

Meliza CD, Margoliash D (2012) Emergence of selectivity and tolerance in the avian auditory cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:15158–15168.

Merchant H, De Lafuente V, Peña-Ortega F, Larriva-Sahd J (2012) Functional impact of interneuronal inhibition in the cerebral cortex of behaving animals. Progress in neurobiology 99:163–178.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:1–5.

Mitchell JF, Sundberg K a, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron 55:131–141.

Mountcastle VB, Talbot WH, Sakata H, Hyvarinen J (1969) Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys: Neuronal periodicity and frequency discrimination. Journal of neurophysiology 32:452–484.

Narayan R, Best V, Ozmeral EJ, McClaine E, Dent M, Shinn-Cunningham BG, Sen K (2007) Cortical interference effects in the cocktail party problem. Nature neuroscience 10:1601–1607.

Neelon MF, Williams J, Garell PC (2011) Elastic Attention: Enhanced, then Sharpened Response to Auditory Input as Attentional Load Increases. Frontiers in human neuroscience 5:41.

Nowak LG, Azouz R, Sanchez-Vives M V, Gray CM, McCormick D a (2003) Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. Journal of neurophysiology 89:1541–1566.

Otazu GH, Tai L-H, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. Nature neuroscience 12:646–654.

Picton TW, Hillyard SA (1974) Human auditory evoked potentials. II: effects of attention. Electroencephalography and clinical Neurophysiology 36:191–199.

Quirk MC, Sosulski DL, Feierstein CE, Uchida N, Mainen ZF (2009) A defined network of fast-spiking interneurons in orbitofrontal cortex: responses to behavioral contingencies and ketamine administration. Frontiers in systems neuroscience 3:13.

Raksin JN, Glaze CM, Smith S, Schmidt MF (2012) Linear and nonlinear auditory response properties of interneurons in a high-order avian vocal motor nucleus during wakefulness. Journal of neurophysiology 107:2185–2201.

Rao SG, Williams G V, Goldman-rakic PS (1999) Isodirectional Tuning of Adjacent Interneurons and Pyramidal Cells During Working Memory: Evidence for Microcolumnar Organization in PFC. Journal of neurophysiology 81:1903–1916.

Rasmusson DD (2000) The role of acetylcholine in cortical synaptic plasticity. Behavioural brain research 115:205–218.

Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. Journal of neuroscience 19:1736–1753.

Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185.

Rif J, Hari R, Matti SH, Sams M (1991) Auditory attention affects two different areas in the human supratemporal cortex. Electroencephalography and clinical Neurophysiology 79:464–472.

Saini KD, Leppelsack H-J (1981) Cell types of the auditory caudomedial neostriatum of the starling (Sturnus vulgaris). The Journal of comparative neurology 198:209–229.

Sanders LD, Newport EL, Neville HJ (2002) Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nature neuroscience 5:700–703.

Shea SD, Koch H, Baleckaitis D, Ramirez J-M, Margoliash D (2010) Neuron-specific cholinergic modulation of a forebrain song control nucleus. Journal of neurophysiology 103:733–745.

Simons DJ (1978) Response properties of vibrissa units in rat SI somatosensory neocortex. Journal of neurophysiology 41:798–820.

Swadlow H a (2003) Fast-spike interneurons and feedforward inhibition in awake sensory neocortex. Cerebral cortex (New York, NY : 1991) 13:25–32.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. The Journal of Neuroscience 20:2315–2331.

Tsunada J, Lee JH, Cohen YE (2012) Differential representation of auditory categories between cell classes in primate auditory cortex. The Journal of physiology 590:3129–3139.

Wang Y, Brzozowska-Prechtl A, Karten HJ (2010) Laminar and columnar auditory cortex in avian brain. Proceedings of the National Academy of Sciences of the United States of America 107:12676–12681.

Wilson FAW, O'Scalaidhe SP, Goldman-Rakic PS (1994) Functional synergism between putative y-aminobutrate-containing neurons and pyramidal neurons in prefrontal cortex. Proceedings of the National Academy of Sciences 91:4009–4013.

Woloszyn L, Sheinberg DL (2012) Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. Neuron 74:193–205.

Woolley SC, Doupe AJ (2008) Social context-induced song variation affects female behavior and gene expression. PLoS biology 6:e62.

Yazaki-Sugiyama Y, Kang S, Câteau H, Fukai T, Hensch TK (2009) Bidirectional plasticity in fast-spiking GABA circuits by visual experience. Nature 462:218–221.

Yokota T, Eguchi K, Hiraba K (2011) Functional properties of putative pyramidal neurons and inhibitory interneurons in the rat gustatory cortex. Cerebral cortex (New York, NY : 1991) 21:597–606.

Conclusion

This work has investigated the role of behaviorally informative and uninformative stimuli after weeks of learning and during brief periods of attentional focus. In particular, this work has focused on how irrelevant stimuli are processed in the cortex, how learning and attention affects this processing, and how they influence the processing of informative stimuli.

By measuring song-evoked and noise-evoked spiking responses in four regions of the songbird auditory cortex, we found that an irrelevant stimulus, white noise, has very different effects on the processing of natural conspecific song throughout the cortical auditory hierarchy, even when the white noise is separated in frequency space from song stimuli. Early on in cortical processing, the noise stimulates auditory neurons and has an excitatory effect when combined with birdsong. In high-level brain regions, however, the noise fails to elicit strong neural responses on its own while drastically suppressing responses to song stimuli, in a manner consistent with inhibitory portions non-classical receptive fields. We demonstrate that the song selectivity is learning-dependent in Field L1, an efferent target of the thalamo-recipient Field L2a, and that this bias results from the selective suppression of task-relevant acoustic signals by irrelevant ones. We conclude that learning modifies non-classical suppressive responses to integrate task-relevant and irrelevant information, yielding selective representations of meaningful natural signals.

When examining how behavioral meaning affects responses in real time through a selective attention task, we found again that the relevance of a stimulus affects cortical processing. In particular, we found that the responses to irrelevant stimuli were enhanced within a putative interneuron cell class during trials where attention was allocated correctly to the target. This result implies that the ignored stimulus is being actively suppressed, presumably to facilitate processing of the attended target. In future work, we hope to clarify the mechanisms that underlie selective attention to auditory signals.

Overall, we find that behavioral meaning drastically effects not only the cortical processing of informative stimuli, but also stimuli which are irrelevant for behavior. This paints an overall picture of the brain's role not at as one dedicated to replicating the external world, but instead devoted to manipulating the external reality into one that amplifies behaviorally relevant stimuli and mutes irrelevant stimuli. My work demonstrates that this process requires not only a selective amplification of target stimuli, but an active and effortful suppression of distracters.