Title
Crossmodal influences in mouse auditory cortex during passive stimulation and an audiovisual behavior

Author
Morrill, Ryan James

Publication Date
2020

Peer reviewed|Thesis/dissertation

Crossmodal influences in mouse auditory cortex during passive stimulation and an audiovisual behavior

by
Ryan James Morrill

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Michael P. Stryker
6F464CB16FDD412...

Michael P. Stryker
Chair

DocuSigned by:
Andrea Hasenstaub
2C04B1E8...

Andrea Hasenstaub

DocuSigned by:
Massimo Scanziani
DocuSigned by:

Massimo Scanziani

DocuSigned by:
Christoph Schreiner
99D647545A30488...

Christoph Schreiner, MD, PhD

Committee Members

To my parents, Ken and Susan

Acknowledgments

I have many people to thank for supporting me through graduate school. My adviser, Andrea Hasenstaub, has done more than anyone to shape my thinking about science. She has an incredibly sharp and careful mind and has made my science more rigorous and thorough. During my time in her lab, she gave me significant independence in developing and carrying out projects. More often than not scientific projects fail, and many of mine have too; learning this lesson early in my career has made me a wiser and humbler scientist. I owe her thanks for all her years of support. My thesis committee, Michael Stryker, Massimo Scanziani and Christoph Schreiner, have been crucial advocates during my time at UCSF. They have graciously given their time to provide feedback on scientific projects and have also offered welcome career advice and kind of words of encouragement. Prior to starting at UCSF, I had several influential mentors including Asif Ghazanfar, Cory Miller, Xintian Hu and Bo Hong. I started my scientific career in Asif's lab at Princeton University. His support helped me start my scientific journey on the right foot, and his infectious curiosity for big, exciting questions remains an inspiration.

The past and current members of the Hasenstaub Lab have provided both scientific support as well as camaraderie during stressful times when experiments were not working. I want to thank Biz Phillips, Nerissa Hoglen, Phil Larimer, James Bigelow, Tim Olsen, Jefferson DeKloe, Mauricio de La Parra, George Luu, Hanna Butler-Struben and Stephanie Bazarini. In particular, I would like to thank my co-authors of the work in Chapter 3, James Bigelow, and Jefferson DeKloe. James helped me develop many of the analyses presented in Chapter 3, providing both code and a wealth of knowledge and wisdom, always delivered with acerbic wit. Jefferson was invaluable in getting our high-throughput mouse behavior project off the ground, spending countless days training mice to perform difficult

behavioral tasks. This was demanding and thankless work and without it the project in Chapter 3 would have never come to fruition.

I would also like to thank many friends I have made at UCSF, who have offered support and scientific advice from their own unique perspectives: Stefan Lemke, Perry Spratt, Lucas Tian, Kenny Kay, Hannah Gemrich, Kwadwo Opoku-Nsiah, Jiggy Athilingam, Jermyn See, and Eszter Kish. Regular pho dinners with Stefan and Perry were a welcome salve from the stresses of the day, and the assorted adventures with Lucas, Kenny, Kwadwo and others helped take my mind off of science.

Chelsea Dill, my partner, has been a remarkable pillar of support in my life over the past four years. Nothing is easy during a pandemic, and during the process of preparing this dissertation, she has been a wellspring of love, thoughtful advice, and encouragement. From reminders to get outside and enjoy the fresh air to her homecooked meals to long conversations about the future, Chelsea has been invaluable in helping me produce this work.

Finally, my family has always provided me with the encouragement and incredible support to follow my passions. My younger brother, Drew, is also a PhD-bound scientist, studying physics at the University of Colorado. His thoughtful demeanor, intelligence and thirst for adventure have always been an inspiration, and I strive to emulate his wholehearted embrace of life in its full range. My father, Ken, taught me to think analytically about the world, encouraging me from a young age to tinker with the world around me and seek explanations at the mechanistic level. My mother, Susan, taught me to seek a life marked by care and compassion, and to consider how actions impact the world in ways large and small. They never discouraged me from taking risks or embarking on adventures with uncertain outcomes, and this helped me embrace the scientific endeavor. For these reasons, I have dedicated this dissertation to my parents.

The text of Chapter 2 is, in full, a reprint of the material as it appears in: Morrill, Ryan J., and Andrea R. Hasenstaub. 2018. “Visual Information Present in Infragranular Layers of Mouse Auditory Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 38 (11): 2854–62.

Crossmodal influences in mouse auditory cortex during passive stimulation and an audiovisual behavior

Ryan James Morrill

Abstract

To enable flexible behavior, the brain utilizes signals from across the sensory modalities, merging some streams while filtering out others in a context-dependent manner. These processes occur at many levels in the sensory hierarchy, but the cerebral cortex appears to play a prominent role in enabling dynamic use of crossmodal sensory information. This dissertation explores the interactions of auditory and visual sensory processing in the auditory cortex (ACtx) of the awake mouse using two experimental approaches. First, the influence of visual stimuli on neural firing in ACtx is investigated using multisite probes to sample activity across cortical layers. Visual stimuli elicit spiking responses in both primary and secondary ACtx. Through fluorescent dye electrode track tracing and optogenetic identification using layer-specific markers, these responses are revealed to be largely restricted to infragranular layers and particularly prominent in layer 6. Presentation of drifting visual gratings show that responses are not orientation-tuned, unlike visual cortex responses. The deepest cortical layers thus appear to be an important locus for crossmodal integration in ACtx. Second, to test the influence of modality-specific attention on ACtx stimulus processing, a novel audiovisual (AV) go/no-go rule-switching task for mice is presented. Translaminar ACtx extracellular recordings from mice performing the task show that attentional state modulates responses to AV stimuli. On average, single-unit firing rates (FRs) in deep and middle cortex are reduced during auditory attention in response to task-relevant stimuli, although a smaller population of units increases FRs. Pre-stimulus activity also decreases when behavior is guided by the

auditory rule and appears to account for much of the change in stimulus-evoked activity. This general reduction in activity does not impair decoding with a PSTH-based pattern classifier, but instead increases mutual information encoding efficiency in the deep, putatively-excitatory neurons. Analysis of spectrotemporal receptive field (STRF) nonlinearities calculated from stimuli delivered between behavioral trials suggests that attending to sound increases the selectivity of neurons for STRF-defined sound features. These results suggest that modality-specific attention can act on ACtx in through rapid, context-dependent shifts in activity level as well as information processing.

Table of Contents

List of Figures

Chapter 1. Introduction

1.1 The interplay of the senses

To interpret and act effectively on the constant barrage of sensory signals it receives, the brain must merge sensory information from across the modalities in a flexible manner. In some cases, this entails combining signals to strengthen a percept, while in others it requires filtering out irrelevant signals. In speech perception, for example, the visual signal of lip movements aids in resolving ambiguous auditory speech cues (Sumby and Pollack 1954; Calvert et al. 1997; Schwartz, Berthommier, and Savariaux 2004), but auditory and visual speech signals that are obviously incongruent fail to elicit neural signatures of integration (Fairhall and Macaluso 2009). That sensory signals from different modalities are merged to improve perception is well-documented in the psychophysical literature; task accuracy, detection thresholds and reaction times are all improved when participants performing difficult tasks are provided with cues from multiple modalities (Nozawa, Reuter-Lorenz, and Hughes 1994; Grant and Seitz 2000; Lovelace, Stein, and Wallace 2003; Diederich and Colonius 2004). Physiological and anatomical work suggests that the brain begins to merge the senses at surprisingly early stages of cortical sensory processing (Foxe and Schroeder 2005; Ghazanfar and Schroeder 2006; Kayser and Logothetis 2007). Bringing the senses together allows the brain to utilize the statistical regularities of signals that frequently co-occur in space and time to make better predictions about the environment (Chandrasekaran et al. 2009).

The merging of cues from across the senses is not necessarily an automatic process. Instead, multimodal integration is modulated in the service of behavioral goals. A fundamental task of the brain is to decide when and how to merge these signals. Evidence

suggests that the brain does this near-optimally, weighting the relative contribution of cues from different sensory modalities to a decision by their reliability (Ernst and Banks 2002). More generally, learning and attention are both known to determine when and how multimodal signals are integrated (Driver and Spence 1998; Mozolic et al. 2008; Talsma et al. 2010), and crossmodal responses in brain regions traditionally considered unisensory can be driven by learning and behavioral demands (Brosch, Selezneva, and Scheich 2005). In some cases, instead of integrating across sensory streams, the brain must do the opposite: filter out task-irrelevant stimuli, a process likely undergirded by attentional mechanisms but with direct implications for multisensory processing.

In this dissertation, the interplay of the sensory modalities is addressed using two complementary experimental approaches. In Chapter 2, the organization of visual responses in the mouse auditory cortex is investigated. Using multisite probes to span the cortical depth, we test whether a passively-viewed visual stimulus evokes neural responses in the auditory cortex and find that there is a strong depth bias for spiking responses in the infragranular layers. In Chapter 3, we train mice on an audiovisual rule-switching task to determine whether attention to sound or vision affects stimulus processing in the auditory cortex. We find that crossmodal attention modulates activity levels before and during stimulus presentation, and that the efficiency of stimulus encoding improves while the mouse attends to sound. Subsequent sections of the current chapter provide context for this work with an overview of relevant scientific literature.

1.2 Neural bases of multisensory integration: historical context

Everyday experiences are seldom perceived through a single sensory modality. Take an act as simple as opening a door, for example. Feeling the shape of the doorknob in your hand, seeing its metallic luster, and hearing the creak of the door's hinges: this experience

is simultaneously haptic, visual, and auditory. These sensory signals reach the brain through distinct anatomical pathways, but quickly come together to generate perceptual experience. Despite the multisensory nature of our reality, the sensory systems have typically been studied one at a time (Ghazanfar and Schroeder 2006). How and where the senses merge in the brain is an ongoing and evolving area of neuroscientific research, with much progress made in the last several decades.

Early studies on sensory cortical receptive field properties noted instances in which stimuli of a non-preferred modality generated suprathreshold spiking responses in the cat visual cortex (Horn 1965; Murata, Cramer, and Bach-y-Rita 1965; Spinelli, Starr, and Barrett 1968; Bental, Dafny, and Feldman 1968). This work provided clues that even primary sensory cortical function may not be strictly unisensory. It was the pioneering work of Meredith, Stein and others recording from superior colliculus (SC) neurons in the cat that began to formalize principles of multisensory processing (Meredith and Stein 1983; Stein et al. 1989; Meredith and Stein 1986b; Meredith, Nemitz, and Stein 1987; Meredith and Stein 1986a; Stanford and Stein 2007). The SC, located in the midbrain, controls orienting behaviors based on sensory inputs, and its intermediate and deep layers contain neurons that respond to visual, auditory and somatosensory inputs. Recordings from these layers identified multisensory units with responses to stimuli from more than one modality. For a given SC multisensory unit, spatial receptive fields for each of these modalities are typically in register, such that a unit with visual responses to one region of space will respond to sounds or touch coming from the same approximate location. When stimuli from multiple sensory modalities are presented together, their responses are sometimes supperaditive, eliciting more spiking activity than the sum of the responses to each modality presented individually (Stanford and Stein 2007; Stein and Stanford 2008). Superadditivity is typically

observed when the response elicited by a single modality would be particularly weak; for example, a dim flash of light and a faint sound when paired in spatial register may elicit a superadditive response. On the other hand, subadditivity is observed with the coincident presentation of two particularly strong sensory stimuli. Together, these phenomena relate to what is known as the principle of inverse effectiveness for multisensory integration, stating that as the strength of individual stimuli increase, the strength of the multisensory effect decreases. This work also demonstrated that the magnitude of multisensory SC processing is dependent on the temporal coincidence of multiple signals, and when signals are presented further apart in time, integrative effects are diminished (Meredith, Nemitz, and Stein 1987).

Since this seminal work in the SC, it has been suggested that the multisensory integrative functions of these neurons are dependent on signals from the cortex (Jiang et al. 2001; Stein et al. 2002); without corticotectal signals originating in association areas, multisensory enhancement of both orientation behavior and neuronal SC response is abolished. This finding approximately coincided with an apparently renewed interest in the cortical bases of multisensory integration. While the merging of senses in cortical association areas has long been appreciated (e.g., Pandya and Seltzer 1982; Goldman-Rakic 1988), the work most relevant to this dissertation concerns the crossmodal and multisensory effects observed in cortical areas traditionally considered unisensory. Over the past two decades, research in auditory (Ghazanfar et al. 2005; Kayser et al. 2005; Bizley et al. 2007; Lakatos et al. 2007), visual (Watkins et al. 2006; Wang et al. 2008) and somatensory (Y.-D. Zhou and Fuster 2004; Lemus et al. 2010) cortices from a variety of species began to show that a small but meaningful fraction of activity is modulated at low latencies by stimuli from another modality. In the early aughts, this led to calls for revision

of the view that sensory cortical processing begins with strictly unisensory representation in the primary regions (Shimojo and Shams 2001; Foxe and Schroeder 2005; Ghazanfar and Schroeder 2006). Attempts to relate cortical multisensory integration to the principles derived from studies on the SC have generally suggested, perhaps unsurprisingly, that cortical multisensory convergence plays by a more nuanced set of rules (Allman et al. 2008), but the requirements of temporal and spatial congruence for optimal multisensory integration have been observed (Macaluso, Frith, and Driver 2000).

Determining the functional role for crossmodal signals at such early stages in the cortical processing hierarchy is an area of active research. Several possibilities will be reviewed here to provide additional context for our finding that visual stimuli evoke spiking responses in mouse primary and secondary auditory cortex (Chapter 2). One intriguing proposition is that crossmodal convergence in the cortex may serve to provide complementary signals to each modality based on their informational strengths and weaknesses, a concept known as the "modality appropriateness hypothesis" (Welch and Warren 1980). Visual and somatosensory domains, for example, may provide more reliable spatial information when compared to audition. On the other hand, the auditory domain may provide inherently more reliable information about the timing of events. Information from complementary sensory specializations may be routed crossmodally to early sensory cortex to create a faster and more accurate representation of the environment. Research in the ferret has shown that simultaneous presentation of an auditory stimulus with a visual stimulus increases the amount of spatial information encoded about that event in the auditory cortex (Bizley and King 2008, 2009). On the other hand, psychophysical work in humans suggests that the temporal features of sound stimuli can critically influence visual perception (Recanzone 2003; Watanabe and Shimojo 2001; Shams, Kamitani, and Shimojo

2000), and that such perceptual effects are mediated by cross-talk between auditory and visual cortices (Mishra et al. 2007). Another possible role for early crossmodal cortical integration is a "push-pull" mechanism of allocating resources to the sensory system that most requires them, based on moment-to-moment changes in stimulus salience. These mechanisms have been observed in the study of attention in multimodal tasks (Laurienti et al. 2002; Kawashima and O'Sullivan 1995) as well as within-modality attention (Pinsk, Doniger, and Kastner 2004). In contrast with the putatively top-down mode of resource allocation based on behavioral demands, crossmodal signals to sensory cortex may provide an early bottom-up mechanism to suppress activity in one sensory domain when a strong stimulus occurs in others. A third possibility is that crossmodal sensory responses in early cortical stations provide a latent route for information that can be used during multisensory learning. Crossmodal signals can appear and disappear based on learned behavioral context (Brosch, Selezneva, and Scheich 2005). Under this scenario, anatomical pathways exist between the sensory systems (Henschke et al. 2015) primarily for synaptic refinement, to be strengthened when integrating across modalities becomes behaviorally advantageous. Of course, these hypotheses are not mutually exclusive, and all three may be at play simultaneously depending on the systems in question. The reliance of some forms of subcortical multisensory integration on the integrative mechanisms of the cortex (Stein et al. 2002; Jiang et al. 2001), as well as the apparent ubiquity of crossmodal influence in even early cortical sensory systems (Ghazanfar and Schroeder 2006), supports a basic hypothesis of this dissertation: a primary role of the cerebral cortex is to bring signals together from within and across the modalities in the service of behavioral goals.

Much of the work cited above was performed in monkeys (Ghazanfar et al. 2005; Kayser et al. 2005; Kayser, Petkov, and Logothetis 2008), humans (Calvert et al. 1997; Mishra

et al. 2007) and ferrets (Allman et al. 2008; Bizley et al. 2007; Bizley and King 2009). This work has been foundational, but at present the ability for cell type-specific manipulations and recordings in any of these organisms is highly limited. On the other hand, the mouse has long been a genetic model organism, and a suite of genetic and molecular tools for circuit dissection developed over the last two decades has spurred its enormous rise in popularity in the field of systems neuroscience (Callaway 2005). Of particular relevance here, the mouse has become a common model organism for the study of topics such as auditory (Linden et al. 2003) and visual (Niell and Stryker 2008) cortical sensory processing. Nevertheless, research on the integration of these modalities in the mouse remains limited (Banks et al. 2011; Iurilli et al. 2012; Olcese, Iurilli, and Medini 2013; Song et al. 2017). A first step toward establishing the mouse as a model for the multisensory integration is to determine the anatomical and functional interactions between its sensory systems (Banks et al. 2011), setting a baseline for how we might perform future experiments aimed at circuit dissection. It is partly with this goal in mind that we conducted the experiments described in Chapter 2.

1.3 Cortical layers and crossmodal communication

The cerebral cortex is traditionally divided up into six layers on the basis of cellular morphology, molecular markers and afferent and efferent neuronal connections. While these layers have long been known to neuroanatomists, understanding of how each layer may contribute to cortical computational remains limited (Adesnik and Naka 2018). The laminar organization provides an intriguing framework for cortical multisensory integration. Existing models differentiate between cortical inputs which are ‘feedforward’ versus ‘feedback’ (Felleman and Van Essen 1991; Douglas and Martin 2004; Rockland and Pandya 1979). In the classical model, feedforward inputs terminate in layer 4 (L4), and

provide sensory drive from thalamic regions or, in the case of higher-order cortex, inputs from earlier cortical stations. Note that recent work suggests that infragranular cells (L5, L6) also receive direct, low-latency thalamocortical inputs (Constantinople and Bruno 2013), which carry feedforward drive in early sensory stations. Feedback inputs are those that provide modulation, carrying extramodal signals from other sensory modalities, the motor system, frontal cortex, limbic regions and so on. These are thought to arrive primarily in the supragranular (L1, L2/3) and infragranular (L5, L6) layers.

Crossmodal inputs to auditory cortex appear to be largely feedback in nature, and their laminar organization reflects this. Visual input to monkey auditory association cortex can be seen in the current source density (CSD) signal in the supragranular and infragranular layers, exhibiting a bilaminar feedback pattern (Schroeder and Foxe 2002). In the same study, auditory CSD modulation appears earliest in L4, consistent with a dissociation of a feedforward auditory signal and a modulatory feedback visual signal. A recent study using 7 Tesla functional magnetic resonance imaging in humans to resolve cortical depths at the submillimeter scale suggests that visual modulation of the auditory cortex primarily occurs in the deepest cortical layers (Gau et al. 2020). These physiological findings are supported by anatomical tracing work in the mouse. Visual cortical inputs to the mouse auditory cortex originate in V2 (both lateral, V2L and medial, V2M) and terminate primarily in L1 and L6 (Banks et al. 2011), consistent with a feedback-type input. These results, along with those presented in Chapter 2, suggest that the infragranular layers may be a site of convergence for crossmodal inputs.

1.4 Attentional selection across the modalities

Sensory cortical receptive fields are known to be shaped by behavioral experience and task goals (Gilbert, Sigman, and Crist 2001; Fritz et al. 2003; Winkowski et al. 2013; Yin,

Fritz, and Shamma 2014). While crossmodally-evoked neural activity is observed in the primary sensory areas of anesthetized animals, suggesting that at least some of this convergence is due to pre-attentive, bottom-up mechanisms (Kayser et al. 2005), crossmodal sensory signals are also rapidly modulated by behavioral context and attention. A striking example of this is the finding that during an auditory categorization task, the visual signal of a cue light and the somatosensory experience of a lever press can both elicit widespread and short-latency activation in monkey auditory cortex (Brosch, Selezneva, and Scheich 2005). This crossmodal activity was context-dependent, disappearing when the monkeys instead performed a visually-guided task. These findings demonstrate that the influence of one sensory modality on the cortical processing of another can be determined by behavioral salience and attentional processes.

Sensory processing in service of a behavioral task can bring together signals from different modalities, but it can also selectively parse and discard them to meet behavioral demands. Chapter 3 of this dissertation presents an experiment in which we test the neural mechanisms of modality-specific attention in auditory cortex. What follows is a brief review of attentional mechanisms and how they may interact with multimodal processing.

Attention, broadly defined, is the allocation of resources by the brain to a subset of features of perception or thought at the expense of others (Desimone and Duncan 1995; Hromádka and Zador 2007; Harris and Thiele 2011). The sensory systems, like any information processing system, are throughput-limited (Reynolds and Chelazzi 2004); attentional selection weights inputs by their behavioral relevance or salience, while still maintaining representation of less pressing sensory signals. Researchers have long noted that attention critically modulates activity even in early sensory cortex (Hubel et al. 1959). Since this time, work on attentional mechanisms in extrastriate visual cortical areas of the

monkey has made great progress on elucidating neural mechanisms of feature-based and spatial selective attention (Moran and Desimone 1985; Seidemann and Newsome 1999; McAdams and Maunsell 2000; Williford and Maunsell 2006; Reynolds and Chelazzi 2004), showing that effects differ based on task demands, such as the presence or absence of distractor stimuli, and that stimulus contrast thresholds for visually-evoked activity can be reduced by attention. Work on selective attention in the auditory domain has been more limited and has typically focused on electroencephalographic signals during human dichotic listening tasks (Hillyard et al. 1973; Giard et al. 2000; Bidet-Caulet et al. 2007), demonstrating that attention can both increase magnitudes of auditory cortical responses to attended stimuli, but also suppress responses to unattended stimuli.

The neural processes that support attentional modulation are varied, ranging from those with local effects on receptive fields to more general correlates of behavioral state. The term “task engagement” has been used to describe the brain states of arousal or vigilance employed when an animal is working toward a behavioral goal (Otazu et al. 2009), in contrast to the focused effects of selective attention on a particular space or feature of the sensory field. Studies of task engagement often compare processing of sensory stimuli within a task to passive exposure when the same set of stimuli are presented but are not behaviorally useful. In animal models, this research has typically been carried out in the auditory rather than visual domain, perhaps owing to the difficulty of maintaining the requisite visual fixation during the passive period without providing a behavioral incentive. Auditory cortical work suggests that task engagement may recruit mechanisms that act broadly on auditory processing, shaping auditory receptive fields (Fritz et al. 2003; Yin, Fritz, and Shamma 2014) and affecting stimulus-evoked (Bagur et al. 2018; Otazu et al. 2009; Niwa et al. 2012) and spontaneous rates (Buran, von Trapp, and Sanes 2014). A potential

confound of this research is that the transition from states of somnolence to arousal is marked by differences in cortical sensory processing that are unrelated to engagement in a task (Niell and Stryker 2010; McGinley, David, and McCormick 2015; Bigelow et al. 2019).

Chapter 3 of this dissertation addresses what some have termed "modality-specific attention" (Helbig and Ernst 2008; Mozolic et al. 2008) or "sensory selection" (Wimmer et al. 2015): the allocation of attentional resources to one domain at the expense of another. Much prior work on the cortical bases of this phenomenon comes from human imaging studies, which showed that attention to the attended modality increases activity in the sensory systems devoted to processing that modality (Grady et al. 1997; Johnson and Zatorre 2005; Johnson and Zatorre 2006). Perhaps more surprising is the finding that attending to a unisensory stimulus in one modality tends to decrease the baseline activity of sensory cortical regions that do not preferentially process that attended modality (Laurienti et al. 2002; Kawashima and O'Sullivan 1995). This work suggests that modality-specific attention may operate at a brain-wide level, allocating resources to the system most relevant for a task and simultaneously quieting activity in others. This presents an interesting parallel to the attentional enhancement of responses to attended space and the suppression of competing stimuli in surrounding locations (Desimone and Duncan 1995; Kastner and Ungerleider 2000).

Attention at both the fine scale of within-modality processing and the broader scale acting across modalities appears to act through both enhancement and suppression. However, to suggest that attentional mechanisms simply turn up activity levels associated with attended stimuli and turn down those associated with unattended stimuli would likely be an oversimplification. Recordings from the auditory cortex in task-engaged animals, including those presented in Chapter 3, suggests that the predominant effect of engaging

in an auditory task is suppression of both task-relevant stimulus-evoked (Otazu et al. 2009; Bagur et al. 2018) and pre-stimulus activity (Buran, von Trapp, and Sanes 2014), although many reports (including ours) also show that a fraction of recorded cells also increase stimulus-evoked activity (Kuchibhotla et al. 2017; Carcea, Insanally, and Froemke 2017; Buran, von Trapp, and Sanes 2014). These results suggest that mechanisms of sensory modulation by task engagement in the auditory cortex are varied but often employ suppression of activity, a process likely to be shaped by inhibitory networks (Kuchibhotla et al. 2017). Nonetheless, transmission of information relevant to a task may be increased even when neural responses to task stimuli are suppressed on average. This could be achieved through multiple mechanisms, such as greater suppression of background activity or increased temporal precision of stimulus responses. This relationship between information transmission and task engagement is explored in Chapter 3.

Chapter 2. Visual information present in infragranular layers of mouse auditory cortex

2.1 Introduction

The cerebral cortex enables dynamic, flexible responses to the sensory environment. To achieve this, signals from a variety of sources must come together, often across the sensory modalities. For example, the acoustic signal (auditory domain) and the lip movements of speech (visual domain) are integrated to fundamentally influence perception of speech (McGurk and MacDonald 1976), a process that occurs, at least in part, in the cortex (Skipper et al. 2007). Such sensory information is often viewed as traveling through a cortical hierarchy, from primary regions through secondary regions and eventually to association cortex. Traditionally, the earliest stages of this processing were thought to be exclusively unimodal, exhibiting responses to only one sensory modality (Felleman and Van Essen 1991). Recently this notion has been challenged by evidence that even the primary regions of visual, auditory, and somatosensory cortex all receive and integrate information from other sensory modalities (Calvert et al. 1997; Foxe et al. 2000; Ghazanfar et al. 2005; Schaefer et al. 2006; Iurilli et al. 2012).

Evidence for multisensory convergence in early sensory cortex comes from two complementary lines of research. First, anatomical tracing has revealed direct connections between cortical or thalamic regions of different sensory systems (e.g., Falchier et al. 2010; Banks et al. 2011; Henschke et al. 2015). Second, physiological studies have shown that neural activity, as measured by firing rate or field potential response, can be altered in unimodal versus multimodal stimulus conditions. For example, neural responses to sound in both core and belt regions of monkey auditory cortex can be modulated by the presence

of concurrent visual stimulation (Ghazanfar et al. 2005; Kayser, Petkov, and Logothetis 2008). Less common is the finding that a neuron would respond to a different sensory modality in absence of stimulation in the modality preferred by the surrounding cortex. In the auditory cortex, spiking responses to visual stimuli have been shown to develop after behavioral training; specifically, neurons within the auditory cortex of primates trained on an auditory categorization task exhibit responses to the onset of a cue light (Brosch, Selezneva, and Scheich 2005). In the untrained context, auditory cortex responses to visual stimulation have been reported in rats (Wallace, Ramachandran, and Stein 2004), ferrets (Bizley et al. 2007) and gerbils (Kobayasi, Suwa, and Riquimaroux 2013), but are generally reported to represent a small fraction of responses. Furthermore, the stimulus preferences and cortical organization of these visually-responsive neurons remain poorly understood. Such information is critical for understanding the role that these responses play in auditory processing.

Here we sought to determine if and how the auditory cortex of the mouse (*Mus musculus*) responds to visual stimulation. We performed acute, awake recordings in mouse auditory and visual cortices during auditory and visual stimulation. In a subset of recording sites, neurons in mouse auditory cortex responded to visual stimulation even without concurrent auditory stimulus. These neurons reside almost exclusively in layers 5 and layer 6 of the cortex and may signal the presence and timing of a salient visual stimulus to the local circuitry of the auditory cortex.

2.2 Materials and Methods

Animals. All experiments were approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco. For optogenetic identification of

layer 6 (L6), we employed the Ntsr1-Cre knock-in line (GENSAT GN220), in which Cre recombinase is expressed specifically in L6 corticothalamic cells (Olsen et al. 2012). To achieve targeted activation of L6, this line was crossed with the Ai32 strain (JAX Stock Nr. 012569), which encodes the light-gated depolarizing cation channel channelrhodopsin-2 (ChR2) conjugated to eYFP, after a floxed stop cassette under the CAG promoter. For all other experiments characterizing the visual response in auditory cortex, we used mouse strains on a C57BL/6 background that were not expressing optogenetic effector proteins. Mice in all experiments were between 6-12 weeks old. All adult mice were housed in groups of 2-5 under a 12 hr/12 hr light/dark cycle. Both female (5/19) and male mice were included in this study.

In-vivo awake recordings. A surgery to implant a custom steel headplate over the temporal skull using dental cement was conducted 2-7 days before each recording. The headplate was positioned to allow access to a point putatively centered on primary auditory cortex, 2.5 mm posterior to bregma and under the squamosal ridge. On the day of the recording, the animal was anesthetized using isoflurane and a ~2 mm diameter opening was made in the skull over auditory cortex using a dental drill. This opening was promptly covered with silicone elastomer (Kwik-Cast, World Precision Instruments, Sarasota, FL), and the animal was allowed to recover from anesthesia for 1-2 hrs. The animal was then affixed by its headplate over a free-spinning spherical treadmill and the silicone plug over the craniotomy was removed. A 16-channel linear probe (50 μm site spacing; Neuronexus, Ann Arbor, MI) was covered in the lipophilic dye Di-I (2.5 mg/mL in EtOH) using a needle and syringe and then slowly inserted in the brain using a motorized microdrive (FHC, Bowdoin, ME). After reaching the desired depth, the brain was allowed to settle for ~10

min, after which neural recording and stimulus presentation commenced. Typically, 3-5 acute penetrations were performed per animal.

The signal acquisition system consisted of an Intan RHD2000 recording board and an RHD2132 amplifier (Intan Technologies, Los Angeles, CA), sampling at 30 kHz. Auditory stimuli were presented with a free-field electrostatic speaker (ES1, Tucker-Davis Technologies, Alachua, FL) driven by a Roland Quad Capture external soundcard (Roland, Los Angeles, CA) at a sampling rate of 192 kHz. Visual stimuli were presented on a 19 in LCD monitor with a 60 Hz refresh rate (VW199, ASUS USA, Fremont, CA). Auditory and visual stimuli were both generated in MATLAB using the Psychophysics Toolbox Version 3 (Kleiner, Brainard, and Pelli 2007).

Sound stimuli were either click trains or pure tone sequences. Click trains consisted of broadband 5 ms white noise pulses, presented at 20 Hz for 500 ms duration, and were used as a search stimulus to determine auditory responsiveness; they were not analyzed further. Pure tone stimuli consisted of 100 ms tones of varied frequency (4 to 64 kHz, 0.2 octave spacing) and sound attenuation levels (30 to 60 dB in 5 dB linear steps), with an interstimulus interval of 500ms to construct a frequency-response area (FRA). Between 6 and 10 trials were presented at each frequency-attenuation level.

Visual stimuli consisted of either flash stimuli or drifting grating stimuli on a monitor centered in front of and 25 cm away from the mouse. Monitor luminance was calibrated to 25 cd/m^2 for a gray screen, measured at the approximate location of the mouse's eyes. The flash stimulus was a white square (32° horizontal x 32° vertical) on a black background, 150 ms in duration, with a peak brightness of 95 cd/m^2. Typically, 150 flash presentations were used per block, and the inter-stimulus interval (ISI) was randomly

varied between 650 and 2850 ms. Drifting gratings were presented full-screen (79° horizontal x 50° vertical) for 1 s using parameters optimal for driving mouse visual cortex: 4 Hz temporal frequency, 0.02 cycles/deg spatial frequency, 100% contrast (Niell and Stryker 2008). Gratings were presented in 12 orientations from 0 to 330 degrees in a randomly varied sequence, with 50 presentations per orientation and a randomly varied ISI between 500 and 1600 ms. To verify that the monitor did not produce sound, we recorded and analyzed sounds during visual stimulus presentation using an ultrasonic microphone and recording device sampling at 250 kHz (UltraSoundGate 416H, Avisoft Bioacoustics, Glienicke, Germany).

For L6 optogenetic identification experiments, we activated ChR2 by illuminating the cortex with a blue 470 nm LED (Mightex, Pleasanton, CA) coupled to a 400 μm-diameter optical fiber, NA = 0.39 (Thorlabs, Newton, NJ). A micromanipulator was used to position the fiber tip just above the cortical surface immediately adjacent to the probe penetration site. Light powers were ranged between 0.2 and 2.2 mW; trials with light powers of 1.6 to 2.2 mW were used for later analysis. Light duration was 500 ms, with a 50 ms linear ramp to reach full power anda recovery time randomly varied between 1600 and 2600 ms.

Data analysis. After recordings, the raw voltage trace was bandpassed between 600 Hz and 6 kHz and events were extracted using a moving-window 4.5 SD threshold. For single unit (SU) analysis, event waveforms were sorted using custom software in MATLAB (KFMMAutoSorter, written by Mathew Fellows). Multi-unit analysis was performed on all events captured by the aforementioned 4.5 SD threshold; as such, this analysis includes all units recorded on a channel, as well as events that could not be attributed to a single-unit

neural source due to the absence of a uniquely identifiable waveform shape. Such analysis is typically thought to capture spiking activity from tens of neurons in the vicinity of the recording electrode (Buzsáki 2004).

For all MU and SU analyses, auditory responsiveness was defined as a significant difference in firing rate between the 100 ms prior to stimulus onset and the 100 ms post-stimulus period (paired t-test, Benjamini-Hochberg corrected for false discovery rate, $q = 0.001$; (Benjamini and Hochberg 1995). The Benjamini-Hochberg procedure, used here to correct for multiple comparisons in determining significance of auditory and visual responses, is a method for controlling false positives (Type I errors) that has increased power relative to more common family-wise error rate control procedures, such as the Bonferroni correction; the latter class of procedures attempts to control the probability of including one false positive, typically at the expense of a high false negative (Type II error) rate. On the other hand, false discovery rate methods set an acceptable rate of Type I errors (Benjamini and Hochberg 1995). Here, we have set our false discovery rate to $q = 0.001$. Auditory FRAs were generated using firing rate in the 100 ms post sound-onset. Significant tuning to frequency, used in auditory response classification, was defined as a modulation of firing rate by frequency using a one-way ANOVA ($\alpha = 0.05$).

Visual responsiveness was defined as a significant difference in firing rate between the 200 ms preceding stimulus onset and 200 ms post-stimulus onset (paired t-test, Benjamini-Hochberg corrected for false discovery rate; $q = 0.001$). Sites and units considered in this dataset were from recordings determined to be in auditory cortex based on multi-unit responses to pure tones and only included sites within cortex, as measured by our electrode track tracing procedure described below (n = 676 MU sites; n = 223 SUs).

For both auditory and visual responses, latency to onset was defined as the time-point at which the post-stimulus firing rate exceeded the pre-stimulus firing rate by 4 SDs. Likewise, response offset was defined as the point after onset at which firing rate dropped back down below 4 SDs above pre-stimulus firing rate.

To identify the Ntsr1-Cre-positive L6 band in our Ai32/Ntsr1-Cre optogenetic activation experiments, we analyzed recordings for a group of three or more adjacent light-activated channels. A channel was defined as light-activated if it showed sustained activation throughout the light-on period. To remove transient onset effects that were often observed throughout the cortical column, our analysis focused on the last 200 ms of the light-on period. We converted firing rate during this period to a z-score using the baseline (200 ms prior to stimulus onset) mean and standard deviation. Degree of firing rate modulation differed greatly between recordings, presumably because of factors such as light penetrance in cortical tissue. As such, we defined significant activation as any period that surpassed half of the peak z-score firing rate observed across all light levels in a given recording. A channel was considered light responsive if at least half of all 20 ms time bins showed such activation, and an Ntsr1-Cre band was defined as three adjacent channels (i.e. activation spanning 150 μm) Using this method, the Ntsr1-Cre-positive band of layer 6 was readily identifiable in 5 out of 8 experiments with visually responsive single- or multi-units.

To determine the degree of tuning to orientation of drifting grating stimuli, we calculated an orientation selectivity index (OSI):

$$OSI = \frac{R_{pref} - R_{orth}}{R_{pref} + R_{orth}}$$

In the above equation, R_{pref} is the mean response to the stimulus of the preferred orientation (that which elicited the response with the highest firing rate), and R_{orth} is the mean response to the two orientations orthogonal to the preferred orientation. All responses were baseline-corrected by subtracting the mean pre-stimulus firing rate averaged over all trials. For units in which R_{orth} was suppressed relative to baseline, OSI will be greater than 1. Analyses were performed on the 500 ms after stimulus onset.

Histological verification of electrode track depth. To visualize recording site locations, we used the fluorescent lipophilic dye Di-I, which has been shown to reliably mark the full extent of electrode tracks in extracellular recordings (DiCarlo et al. 1996), and has been used to visualize silicon probe tracks in the mouse brain (Lee et al. 2015). After completion of physiological recordings, the animal was euthanized and the brain was removed and placed into a solution of 4% paraformaldehyde in phosphate-buffered saline (PBS; 0.1 M, 7.4 pH) for 12 hours, followed by a 30% sucrose in PBS solution for several days. The brain was then frozen and sections were cut on a sliding microtome (SM2000R, Leica Biosystems, Buffalo Grove, IL). Slices were then mounted and imaged on a fluorescence microscope with a red filter cube (Eclipse 90i, Nikon, Melville, NY). Fluorescent marks on slices were mapped to each recording with the aid of a penetration site map drawn based on the exposed cortical surface during the experiment. Roughly half the brains were fully imaged and mapped onto the Paxinos and Franklin mouse brain atlas (2004) to identify the locations of recorded sites; in the other half, histology was used only to identify the depth of recorded sites.

Classification of auditory sites. Auditory cortex contains subfields with characteristic neural responses to sound stimuli (Stiebler et al. 1997; Joachimsthaler et al. 2014). We

classified recording sites as either primary-like in their responses (putatively primary auditory cortex, A1, or anterior auditory field, AAF) or secondary-like (for example, secondary auditory cortex, A2, or dorsal-posterior field, DP). Primary and secondary regions are most differentiable by latency to response onset, with primary regions exhibiting shorter onset latencies than secondary ones (Joachimsthaler et al. 2014). As such, the classification procedure we used was as follows. A channel-wise automatic classification was made for multi-unit pure tone responses based on the following criteria: primary-like responses were those with significant firing rate tone responses and onset latencies of <14ms; secondary-like responses were those with significant tone responses and onset latencies of >14ms (Joachimsthaler et al. 2014). Channels not significantly responsive to sound were coded as non-auditory. Next, channels whose classification differed from that of their neighbors were examined in the context of all auditory responses on the probe and coded by eye. This dealt with two problematic cases resulting from automatic channel-wise classification. First, this allowed us to correct for "one-off" inconsistencies in classification within a probe. Second, this increased labeling accuracy in border cases where the probe track did not enter orthogonal to the brain surface, and both primary- and secondary-like responses were recorded from the same auditory experiment. All auditory response classification was performed blind to results of the visual response experiments conducted at the same site.

2.3 Results

Visual responses in mouse auditory cortex. To measure visual responses in mouse auditory cortical neurons, we performed acute extracellular recordings in awake mice using a linear probe to simultaneously measure responses in different layers of cortex (Fig. 2.1.a). In separate blocks, mice were either presented with 100 ms pure tones of varied

frequencies and attenuations or 150 ms flashes of a white square on a black background (Fig. 2.1.b). We then analyzed multi- and single-unit activity evoked by these auditory and visual stimuli (Fig. 2.1.c). Recording site depth and location were determined from post-hoc histological visualization of the lipophilic dye Di-I, which was applied to the probe shank before each penetration (Fig. 2.1.d, 2.1.e). Probe placement was targeted to the auditory cortex based on surface vasculature (Stiebler *et al.* 1997) and confirmed by robust responsiveness to sound stimulation at the recording sites (see Methods; Fig. 2.1.f, left and center). After identifying auditory cortex, we presented flashes and determined whether firing rate was modulated by these purely visual stimuli (Fig. 2.1.f, right). In 28 out of 48 auditory cortex laminar recordings from 16 mice, at least one multi-unit showed a statistically significant increase in firing rate in response to the visual stimulus (paired t-test, Benjamini-Hochberg corrected for false discovery rate, $q = 0.001$, see Methods). Visually-evoked spiking responses were found in both multi-units (Fig. 2.1.g) and single-units (SUs; Fig. 2.1.h). Habituation of responses to repeated stimulus presentations is a well-established feature of auditory cortical processing (Cook, Ellinwood and Wilson 1968). We tested whether auditory cortical SU visual responses found here also habituated over the course of stimulus presentation blocks by testing for a systematic increase or decrease in response magnitude using Spearman's correlation analysis. Wefound 3/15 units from nine mice whose stimulus-evoked firing rates changed (with no corresponding change in baseline firing; $p<0.05$). Of these, two decreased in firing rate and one increased. Additionally, we found no trend in stimulus-evoked firing rate on the population level (Jonckheere-Terpstra trend test, J-T stat. = 0.84, $p = 0.20$). Together, these tests show that there is little evidence for systematic increase or decrease of response magnitude over time with our recording protocol.

Recordings were performed in a blocked manner, with auditory stimuli presented together followed by visual stimuli. Not all units identified as visually-responsive were also identifiable in the auditory blocks, due either to sparse firing in response to auditory stimulation or electrode drift over time. There were nine well-isolated visually-responsive SUs also identified in auditory recordings (n=7 recordings from 7 mice); of these, seven units (7/9), exhibited significant auditory responses (example: Fig. 2.1.i), all of which were also tuned to auditory frequency. This result shows that some neurons in mouse auditory cortex multiplex auditory and visual stimuli.

Deep layer bias of visual responses in ACtx. The laminar location of a neural response can be strongly suggestive of its computational role. We mapped the geometry of the probe electrode sites onto brain slice images marked by Di-I (Fig. 2.1.e, right and Fig. 2.2.a), to determine the cortical depths of visually-responsive electrode sites (example: Fig. 2.2.b). We then measured the distance from white matter (WM) of each recording site and normalized this by the WM-pia distance of the corresponding brain slice to correct for any tissue distortion. This yielded a fractional cortical depth measurement for each recorded channel, which was then assigned to a cortical layer (from 16 mice: L6, n = 160 sites; L5, n = 254; L4 = 129; L2/3, n = 120; L1, n = 13; Fig. 2.2.d). Analysis of these data reveals that the majority of multi-unit and single-unit visual responses occur in layer 6 (L6), with the bulk of the remainder occurring in layer 5 (L5; Fig. 2.2.c; 2.2.d).To provide physiological confirmation of our depth measurements, we used the Ntsr1-Cre mouse strain in which Cre recombinase is expressed specifically in L6 corticothalamic (CT) cells (Fig. 2.2.e; Gong et al. 2007; Olsen et al. 2012). This mouse line was crossed with mice of the Ai32 strain, which expresses channelrhodopsin-2 (ChR2) conjugated to eYFP in a Cre-dependent manner, so that ChR2 was restricted to L6 (Fig. 2.2.e). Illumination of the cortical surface

with blue light resulted in strong multi-unit activation of a distinct band of channels deep on the probe (Fig. 2.2.f, 2.2.g), thereby providing an optogenetically-induced physiological marker of L6. A band of three or more adjacent channels with sustained optogenetic activation was identified in five out of eight recordings (n = 4 of 6 mice). We determined the depth of visual responses relative to this band of activation, and found that all of them occurred <= 200 μm from its lower border (Fig. 2.2.h). This "photo-tagging" approach to identify the band of L6 CT cells further confirmed the deep layer bias of visual responses.

Visual responses in primary and secondary regions of ACtx. The neural signatures of multisensory integration are more commonly observed in secondary or "higher-order" areas of sensory cortex compared with primary regions (Ghazanfar et al. 2005; Bizley et al. 2007). To test whether this finding holds for visual responses in mouse auditory cortex, we classified our recording sites into primary or secondary regions using temporal dynamics and frequency tuning of evoked multi-unit responses to pure tones of varied frequencies and attenuations (see Methods). Recordings classified as primary-like typically showed robust frequency-attenuation tuning (example: Fig. 2.3.a), while many secondary regions did not (example: Fig. 2.3.b). When reconstructed, most primary-like sites were localized to within A1/AAF on the Paxinos and Franklin atlas (2004; Fig. 2.3.c). Mean normalized FRAs from primary and secondary sites centered on BF show that, on average, MUs from both classification exhibit tuning, but BF- and off-BF responses were closer in magnitude in secondary than in primary sites (Fig. 2.3.e). Multi-unit onset and peak response latencies to BF sound stimuli in primary areas were lower than those in secondary areas (primary onset: 10 ± 4 ms, mean ± STD; secondary onset: 17 ± 11 ms; Wilcoxon rank-sum Z = 11.7, p = 1.71e-31; primary peak: 18 ± 8 ms; secondary peak: 30 ± 16 ms; rank sum Z = 9.76, p = 1.67e-22; n=396, 167 MUs in n=14, 13 mice for primary and secondary recordings, respectively; Fig. 2.3.d).

While distributions of recorded auditory BFs were biased to the middle of the frequency band tested (approx 8-30 kHz), visual responses were more prominent at those sites with BFs near 64 kHz (Fig. 2.3.f). The fraction of visual responses was also slightly higher at secondary than primary recording sites in both L5 and L6, the only two layers which exhibited any substantial visual responsiveness (Fig. 2.3.g). Thus, our results indicate that visual MU responses are slightly biased toward secondary sites and toward sites with high frequency tuning in primary areas, but are nevertheless present at sites with a variety of auditory BFs.

Visual response latencies compared in auditory and visual cortices. Latencies to response onset for tones at BF in the auditory cortex vary from 8 to 30 ms, depending on auditory field (Fig. 2.3.d). What are the temporal dynamics of the visual response in auditory cortex? Visually-responsive SUs in primary auditory regions exhibit onset latencies of 75 ± 10 ms (mean ± STD; n=7 SUs from 6 mice), while those in secondary regions have onsets of 92 ± 25 ms (n=8 SUs from 5 mice; Fig. 2.4.a); this difference did not reach statistical significance (Wilcoxon rank-sum = 47.5, p =0.34). MU visual response onset latencies were 85 ± 37 ms in primary sites (n=45 MUs from 12 mice) and 95 ± 25 ms (n=28 MUs from 8 mice) in secondary (Fig. 2.4.b); primary MU visual responses occurred significantly earlier than those in secondary (Wilcoxon rank-sum Z = 2.3, p = 0.021).

Anatomical tracing work has shown that mouse auditory cortex receives direct inputs from several visual cortical regions, and that these inputs show a preference for L1 and L6 (Banks *et al.* 2011). If these projections are carrying the visual information to the auditory cortex, visual stimulation should elicit earlier responses in visual cortex than in auditory cortex. To test this, we recorded from awake mouse visual cortex using flash

stimuli with the same parameters as used to elicit responses in auditory cortex (Fig. 2.4c, 2.4.d). Analysis of MU data shows that responses to visual stimuli in visual cortex are significantly earlier in onset than in auditory cortex (VCtx: 40 ± 11 ms, n=78 MUs from 8 recordings in 3 mice; ACtx: 90 ± 32 ms, n=73 MUs from 30 recordings in 15 mice; Wilcoxon rank-sum Z = 9.06, p = 1.26e-19) and peak (VCtx: 70 ± 33 ms; ACtx: 115 ± 39 ms; Wilcoxon rank-sum Z = 7.60, p = 2.95e-14). Median latencies to response offset in visual cortex and auditory cortex were not significantly different (VCtx: 160 ± 73 ms; ACtx: 131 ± 40 ms; Wilcoxon rank-sum Z = 1.4 p = 0.16; Fig. 2.4.e, 2.4.f), although visual cortex sites showed a wider distribution of offset latencies (Fig. 2.4.f). These dynamics show that visual cortex begins processing the visual flash stimulus before it arrives in the auditory cortex.

Visual orientation tuning in the auditory cortex. A hallmark of visual cortical processing is tuning of neurons to edges of particular orientations. We sought to test whether visual responses in auditory cortex also carry specific information about the visual scene such as edge orientation. While recording in auditory cortex, we presented full-screen 1 s drifting gratings of 12 orientations and found strong responses in a subset of our flash-response SUs (example, Fig. 2.5.a). Comparison of response PSTHs and firing rate histograms typically revealed only moderate orientation tuning in the auditory cortex (examples: Fig. 2.5.b, 2.5.c). For reference, we also recorded drifting grating responses from visual cortex units. Side-by-side comparison of the most orientation-selective auditory cortex and visual cortex units shows a much higher degree of orientation selectivity in the visual cortex (Fig. 2.5.d, 2.5e). We calculated the orientation-selectivity index (OSI; see Methods) for all ACtx and VCtx SUs. We find strongly orientation-selective units in the visual cortex (41% [7/17] of OSIs>0.75, n=7 recordings in 3 mice), along with weakly tuned units, but only weakly tuned units in auditory cortex (0% [0/7] of OSIs>0.75; n=5 recordings in 4 mice; Fig. 2.5.f);

orientation-selectivity differs significantly between these two populations (one-tailed Kolmogorov-Smirnov test, K-S stat. = 0.512; p = 0.049). These results suggest that visual responses in auditory cortex do not carry fundamental visual information about edge orientation, but instead may represent a more general signal indicating the timing and presence of a salient visual stimulus.

2.4 Discussion

To determine whether mouse auditory cortex responds to visual stimulation, we presented awake mice with unimodal visual and auditory stimuli under passive conditions while performing acute recordings from auditory or visual cortices. In both primary and secondary auditory cortex, we found single-unit and multi-unit activity that responded directly to visual flash and drifting grating stimuli in the absence of sound. These responses were almost entirely restricted to layer 6 (L6) and, to a lesser degree, layer 5 (L5). In L6 of auditory cortex, approximately 25% of MUs were visually responsive; in L5, this value was 10%; yet, fewer than 2% of MUs in layers 2-4 were visually responsive. Visually-responsive units in auditory cortex have longer latencies than those in visual cortex and, unlike visual cortex neurons, are not strongly tuned to drifting grating orientation. Taken together, these results suggest that the deep layers of cortex may represent a locus for cortical multisensory integration in the mouse.

These findings are supported by anatomical tracing work that shows that mouse primary auditory cortex receives inputs from visual cortex (Banks et al. 2011). Anterograde tracers injected into secondary regions both lateral (V2L) and medial (V2M) of V1 reveal robust labeling of terminals in A1. Since this work, much has recently been done to further parcellate the fields of mouse secondary visual cortex (Garrett et al. 2014); it remains to be

tested if projections to auditory areas vary further by visual cortical subfield. Particularly relevant here is the finding that neurons from V2 primarily send projections to L1 and L6 (Banks et al. 2011). This points to a potential anatomical pathway for visual signals to elicit spiking responses in the deep layers of mouse auditory cortex via monosynaptic connections from V2M and V2L. Although we did not observe visually-responsive cells in L1, this layer of cortex was not well represented in this dataset (n=13 MU sites recorded). Note that projections arriving at L1 do not necessarily target the sparse population of cells that reside there: the superficial visual projection to auditory cortex may terminate on the relatively large mass of L1 apical dendrites from pyramidal cells in L5 and L2/3 (Larkum, Zhu, and Sakmann 1999; Larkum and Zhu 2002), and thus may produce some of the deep layer visual responses we observed. Further work involving methods such as trans-synaptic tracing must be employed to resolve questions of this nature.

Beyond direct connections from secondary visual cortices, there are several other potential pathways by which visual signals may drive spiking responses in the auditory cortex. Other areas of association cortex may send feedback projections to modulate processing in the auditory cortex. The gerbil primary auditory cortex receives direct input from multisensory cortical regions such as posterior parietal cortex (Budinger et al. 2006). In addition, several thalamic regions with projections to auditory cortex show multimodal responses; the medial aspect of the medial geniculate body exhibits multisensory responses (Wepsic 1966) and sends a dense projection to L6 of rat auditory cortex (Linke and Schwegler 2000). Furthermore, the suprageniculate nucleus, another highly multimodal thalamic region, projects to L5 and L6 of rat auditory cortex (Smith, Manning, and Uhlrich 2010). The termination patterns of these projections are also consistent with deep layer visual spiking responses; none of these anatomical pathways can be ruled out

based on our results. While neurons in many visual stations, including primary and secondary visual cortices as well as visual thalamus, exhibit tuning to orientation, the untuned responses we observe in auditory cortex could result from the pooling of such tuned inputs.

The layer specificity of visually-evoked spiking responses in auditory cortex may inform the role of such responses in modulating activity within the cortical column (Douglas, Martin, and Whitteridge 1989). The deep or "infragranular" layers of cortex are considered the primary subcortical output layers, but also send collaterals to cortical targets, including local circuitry. Historically, the role of L6, in particular, has been considered enigmatic due in part to its high degree of morphological and physiological heterogeneity (Briggs 2010) and atypical sensory responses (Zhou et al. 2010). Recent work has shown that layer 6, through synapses onto local inhibitory interneurons, may play a role in gain control of sensory responses (Bortone, Olsen, and Scanziani 2014). Furthermore, L6 neurons are known to influence cortical receptive field structure (Bolz and Gilbert 1986) and gate sensory input through corticothalamic connections (Farran Briggs and Usrey 2008). Cells in these layers appear to be strategically located for sculpting and modulating sensory processing. The restriction of visually-evoked spiking responses to the infragranular layers suggests that such modulation of sensory activity may be controlled, in part, by crossmodal inputs.

Previous work has shown that visually-evoked spiking responses are rare in the auditory cortex. Kobayasi *et al.* (Kobayasi, Suwa, and Riquimaroux 2013) recorded from A1 of the Mongolian gerbil, and concluded that 2 of their 128 units exhibited responses to a visual stimulus alone. In the auditory cortex of the ferret, reported percentages of visually-

responsive neurons are much higher, with ~15% of primary auditory neurons showing responses to the light flash of an LED (Bizley et al. 2007). In the rat auditory cortex, ~6% of units showed responses to the visual stimulus alone (Wallace, Ramachandran, and Stein 2004). While factors such as cross-species differences likely explain some of these discrepancies in visual responsiveness, our work brings up the possibility that they may also be due, in part, to differences in laminar sampling. Our work also shows that visual responses are more prominent in secondary auditory cortex, consistent with findings in the ferret and monkey (Kayser, Petkov, and Logothetis 2008; Bizley et al. 2007).

The findings presented here extend this literature by revealing that visual stimuli evoke spiking responses in the mouse auditory cortex and showing conclusively that visually-responsive units show a strong laminar bias. Previous non-primate work on audiovisual integration in the auditory cortex has been performed in anesthetized animals (Bizley et al. 2007; Wallace, Ramachandran, and Stein 2004). Given that many anesthesias preferentially inhibit corticocortical connections (Raz et al. 2014) and that at least some visual information likely arrives at the auditory cortex through such connections (Banks et al. 2011), recordings in the awake animal may uncover responses otherwise obscured in anesthetized recordings. Furthermore, our examination of the orientation tuning properties of auditory responses begins to answer questions about the type of information visual responses convey to the local circuitry.

This work must be considered in the broader context of the literature on multisensory integration, much of which has found effects of crossmodal stimulation not in firing rate changes, but in evoked field potential responses and oscillatory changes (Ghazanfar et al. 2005; Lakatos et al. 2009). For example, Lakatos and colleagues found that

visual stimulus onset resulted in phase reset of ongoing oscillations in the monkey auditory cortex with no change in multiunit spiking (Lakatos et al. 2007). It remains to be seen whether such a mechanism is also present in the mouse. Furthermore, there are additional aspects of the spiking signal in response to crossmodal stimulation that remain to be examined. The quenching of trial-to-trial response variability, for example, is a widespread phenomenon related to stimulus onset and may be an additional mechanism by which visual signals affect auditory cortical processing (Churchland et al. 2010).

This work is motivated partly by the utility of the mouse as a mammalian model organism for cell-type specific microcircuit dissection. Tools such as fluorescent cell labeling, optogenetics and chemogenetics, when applied to the problems of multisensory integration, may help elucidate the microcircuitry that integrates crossmodal signals. We hope that this study of visual responses in the auditory cortex of a genetic model organism will further the use of cell-type specific tools for microcircuit dissection of multisensory phenomena.

2.5 Figures

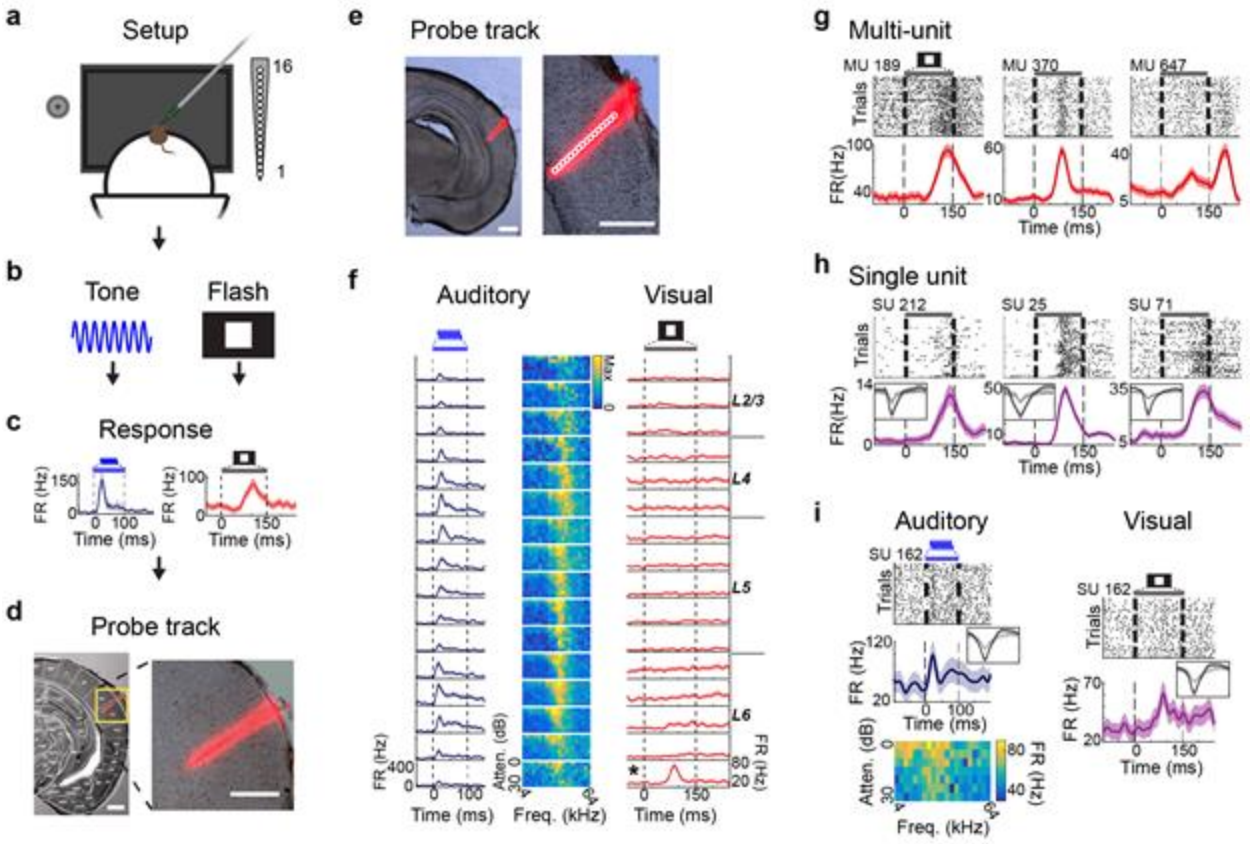

Figure 2.1. Visual stimuli elicit spiking responses in mouse auditory cortex. a. Awake, acute recordings in right mouse auditory cortex using a 16 channel multisite probe (right), experimental sequence: **b-d**. **b.** Tone stimuli used for identification of characteristic auditory cortical responses (left) and visual flash stimuli to test for visual responsiveness (right). **c.** Auditory and visual multi-unit (MU) responses on the same recording channel. **d.** Probe location visualized using fluorescent dye Di-I, which was applied to the probe prior to recording. Images at 1x magnification (left) and 4x (right) used to determine location and laminar depth. Scale bar = 500 μm. **e.** Probe track marked by Di-I, marked with electrode site locations (white circles indicate locations of channels shown in panel **f**). **f.** MU responses on all 16 channels: responses to tones of best frequency (left); frequency response area (middle); response to visual stimulus (right). Asterisk indicates response significant at q=0.001 after Benjamini-Hochberg false discovery rate correction (see **Methods**). **g.** Additional examples of statistically significant MU responses to visual stimuli from different mice. **h.** Examples of statistically significant single-unit responses to visual stimuli (waveforms inset; SU in black, MU from corresponding channel in gray). **i.** Example SU showing both auditory and visual responses.

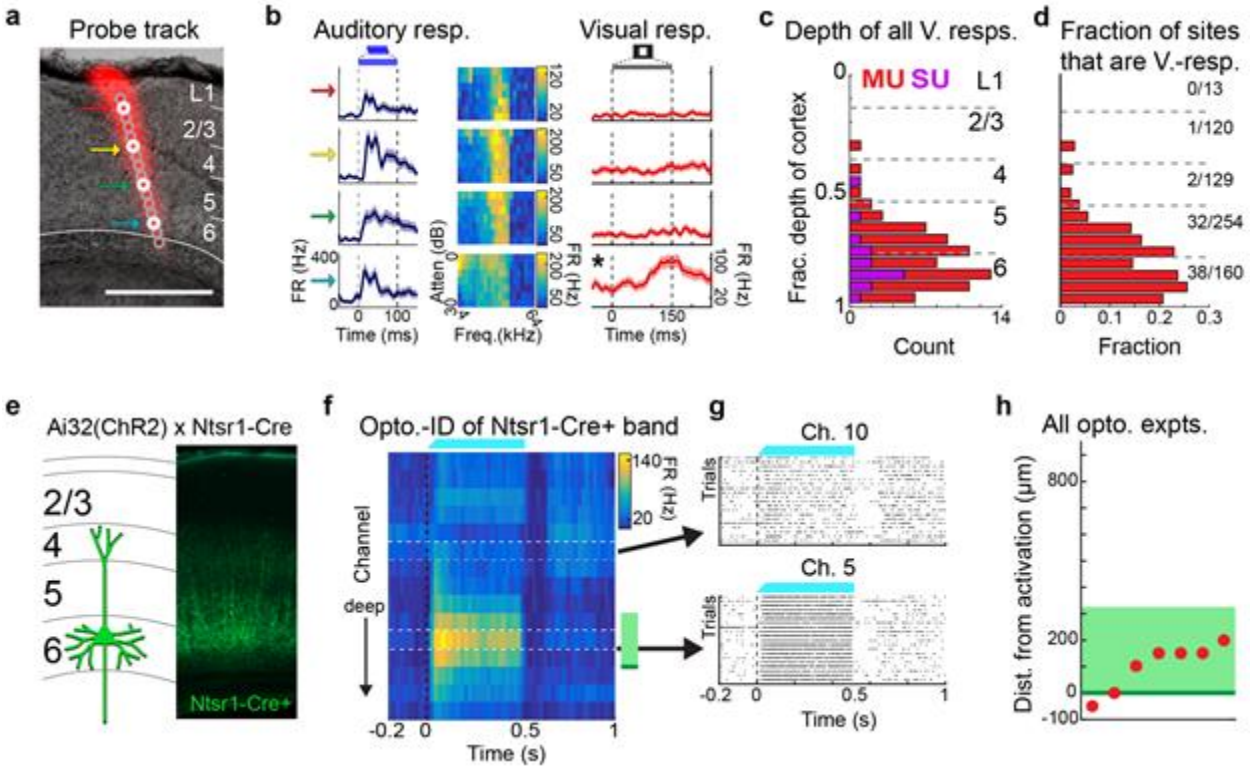

Figure 2.2. Visual responses are primarily found in the deepest layers of auditory cortex. **a.** Probe track in auditory cortex visualized with Di-I, electrode sites marked by circles. Channels shown in panel **b** indicated by colored arrows. Scale bar = 500 µm. **b.** Multi-unit (MU) auditory responses (left), frequency response area (middle), and visual responses (right) at sites from cortical depths indicated by colored arrows in **a**. Visually responsive MU, indicated by asterisk, is located in layer 6 (L6). **c.** All visual responses shown as a function of depth, MUs in red; single-units in purple (n=73 MUs and 15 SUs from 30 recordings in 15 mice). **d.** Fraction of recorded sites that are visually responsive by depth. Numbers at left indicate total number of MU visual responses over total number of sites recorded in each layer. **e.** Ai32/Ntsr1-Cre mice express YFP-tagged channelrhodopsin (ChR2) in L6; histology showing YFP on right. **f.** Representative example of optogenetic identification of L6 through activation of Ntsr1-Cre-positive cells. MU activity (red) shows a band activated strongly during light-on period (blue); significantly modulated channels indicated by green band at right. **g.** Superficial channel shows minimal light-related MU activity (top); deep channel MU activity shows strong effect of light activation (bottom). **h.** Summary plot of all visually-responsive MUs (red dots) from all recordings with an identifiable light-activated L6 band (n = 7 MUs; n = 5 recordings in 4 mice) plotted by depth relative to the lower border of this band (green).

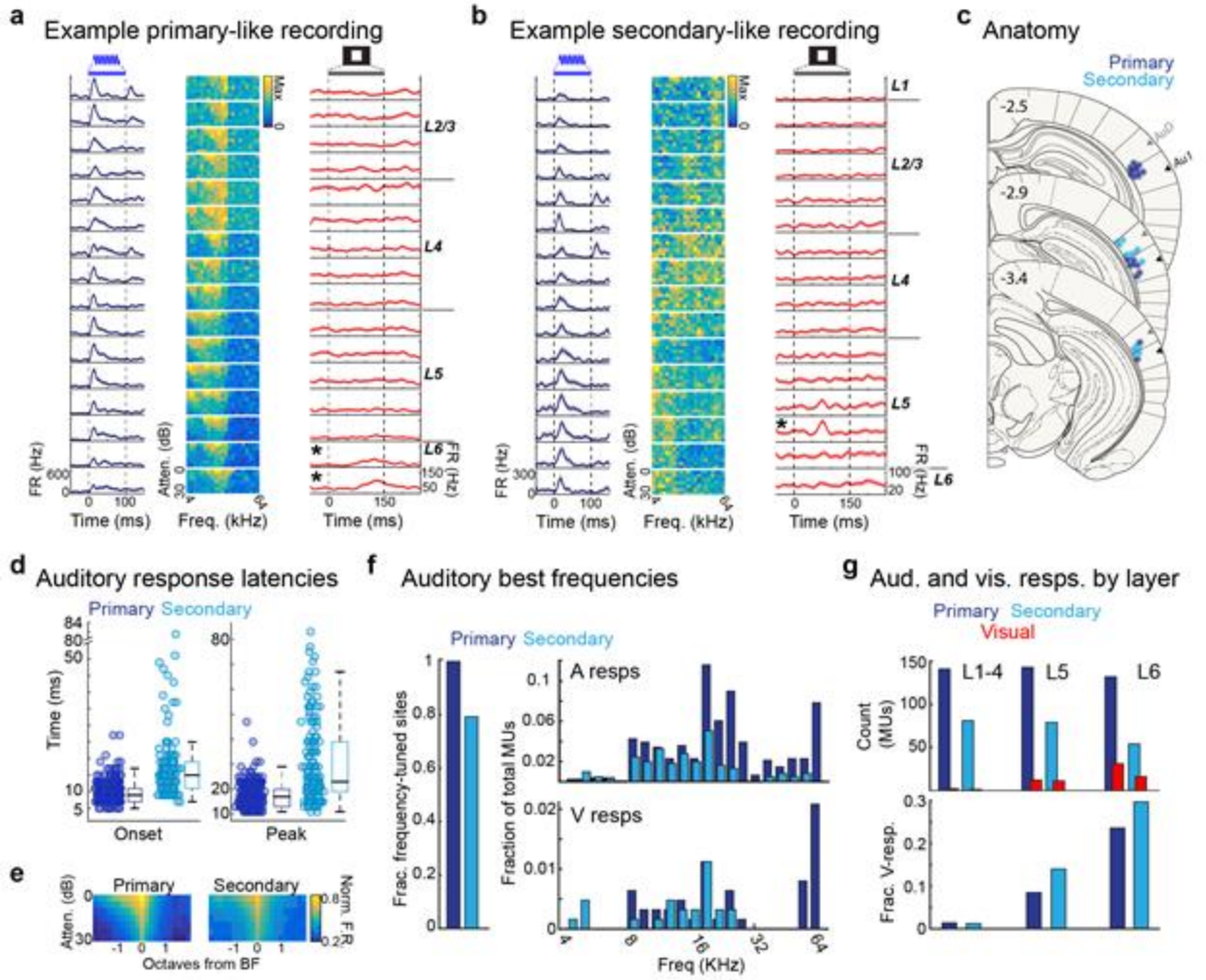

Figure 2.3. Secondary-like regions of auditory cortex are more visually responsive than primary-like regions. a. Example recording classified as primary-like, based on short latency of responses to tones at best frequency (BF; left), all 16 channels shown. Primary-like sites show a high degree of tuning to frequency and attenuation (middle). Response to visual stimulation at same site (right); responses on deepest two channels were significant at q = 0.001 (asterisk; see **Methods**). **b.** Recording classified at secondary-like, based on longer latency of responses to tones of BF, panels as in **a.** Significant MU visual response (q = 0.001) recorded on channel third from bottom. **c.** Anatomical locations of visually responsive reconstructed sites (n=13 mice), color-coded blue (n=22 sites) or cyan (n=24 sites) for primary or secondary classification, respectively (section drawings from Paxinos and Franklin 2004). Depth and anterior-posterior distance determined from electrode track histology (see Figure 2.1d). Distance behind the mouse skull landmark bregma and putative positions of primary auditory cortex (Au1, black) and dorsal auditory cortex (AuD, gray) are marked on each section. **d.** MU auditory response latencies by auditory classification; latencies at primary-like regions are shorter both in onset (left; Wilcoxon rank-sum Z = 11.75, p = 6.50e-32) and peak (right; Wilcoxon rank-sum Z = 10.34, p = 4.20e-25). **e.** Average MU FRA showing tuning from all sites classified as primary (left) or secondary (right), centered on BF. Before averaging, all FRAs were normalized to peak

response. **f.** Auditory best frequency of all auditory-responsive MUs (top) and visually-responsive MUs (bottom). **g.** Total counts of all recorded auditory- and visual-responsive MUs by layer (top). Fraction of visually responsive MUs (bottom), showing biases toward deeper layers and sites with secondary-like responses.

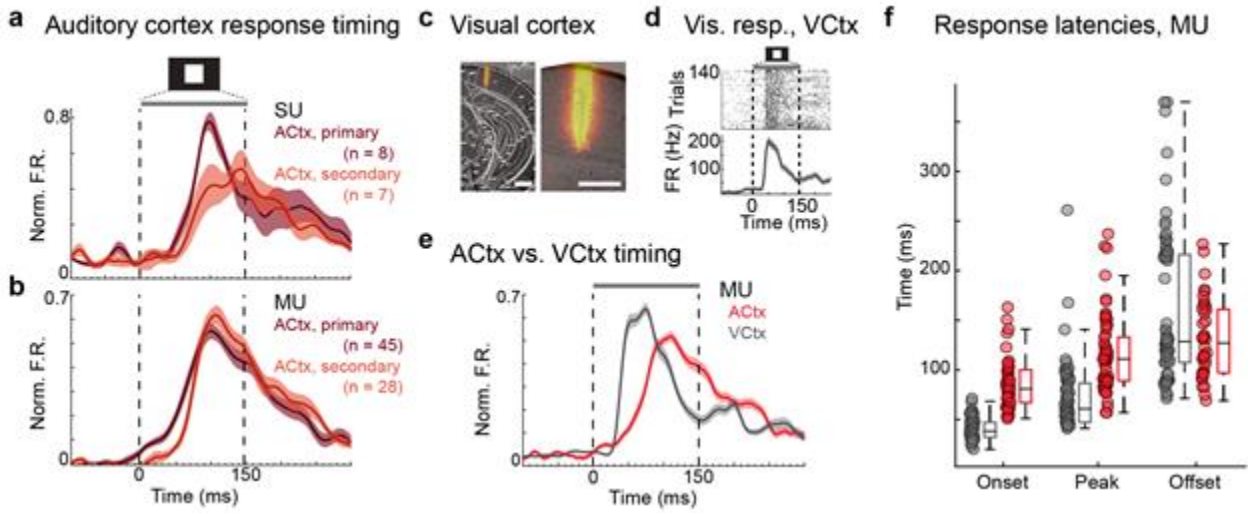

Figure 2.4. Responses to visual stimuli exhibit longer latencies in auditory cortex compared to visual cortex. a. Mean normalized firing rate of visually responsive single-units ($q<0.001$ following Benjamini-Hochberg false discovery rate correction) in primary (dark red: n=45 MUs from 19 recordings in 12 mice) or secondary (light red: n=28 MUs from 11 recordings in 8 mice) auditory cortex (ACtx) following a flash stimulus. **b.** Mean normalized firing rate of visually responsive multi-unit sites in primary (dark red: n=45 MUs from 19 recordings from 12 mice) or secondary (light red: n=28 MUs from 11 recordings from 8 mice) auditory cortex following a flash stimulus. **c.** Example probe track of recording in in visual cortex (VCtx). Scale bar = 500 μm. **d.** Example MU response to flash stimulus in VCtx. **e.** Comparison of response dynamics between ACtx and VCtx, including all significantly visually-responsive MUs from both regions (n=73 ACtx sites from 30 recordings in 15 mice; n=78 VCtx sites from 5 recordings from 3 mice). **f.** Latencies to onset, peak and offset recorded from VCtx and ACtx, showing VCtx peak and onset responses occur earlier than those in ACtx, while offsets in VCtx occur later (see Results).

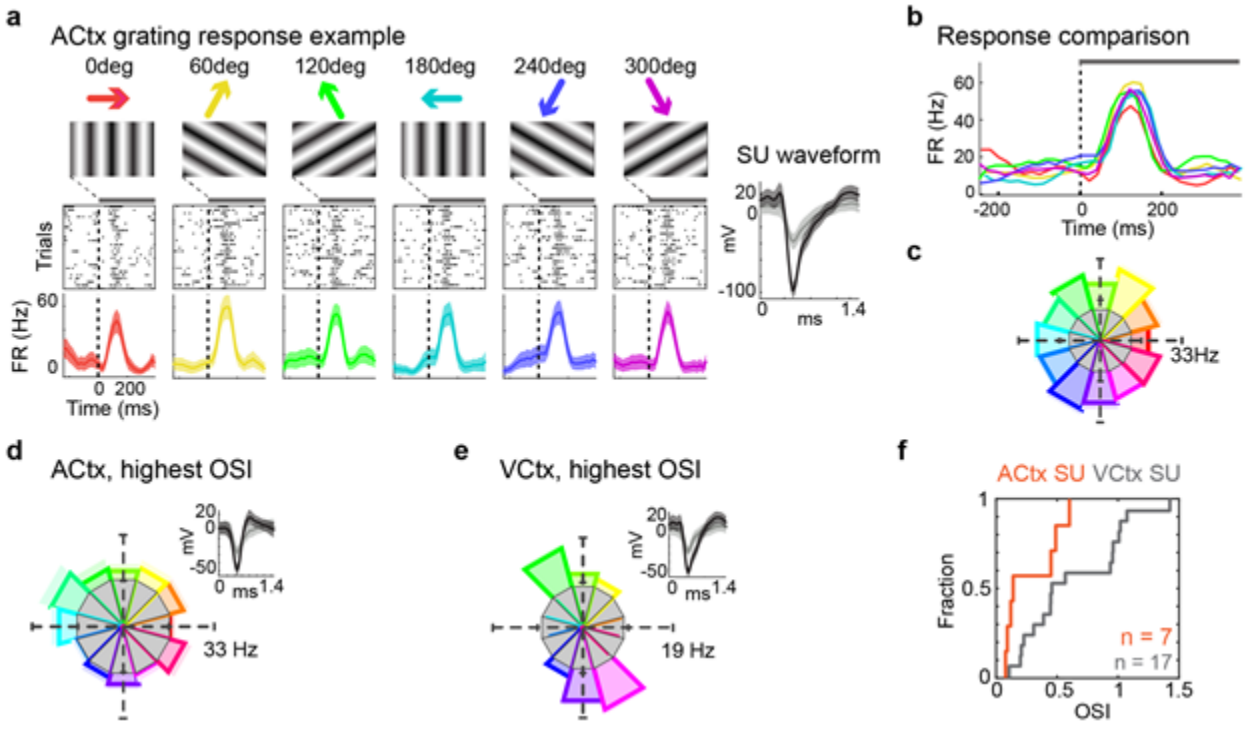

Figure 2.5. Auditory cortex shows minimal preference for grating orientation when compared to visual cortex. a. Example ACtx SU response to drifting gratings of varied orientations (colors); SU waveform shown at right. **b.** All PSTH responses from **a** plotted together, revealing little preference for grating orientation. **c.** Results in **b**, visualized as baseline-normalized circular histogram; baseline firing rate shown in gray. **d.** ACtx unit with highest orientation selectivity index (OSI). **e.** For comparison, VCtx unit with highest OSI. **f.** Cumulative distribution plot of all OSIs, showing a higher degree of tuning in VCtx SUs (n=17 units from 7 recordings in 3 mice) than those in ACtx (n=7 units from 5 recordings in 3 mice).

Chapter 3. Modality-specific attentional state modulates sensory processing in task-engaged animals

3.1 Introduction

To make effective behavioral responses in a dynamic environment, the brain must employ selective attention, highlighting some sensory features while filtering out others. This process is modulated by behavioral context and must occur across the senses. For example, imagine you are driving a car while listening to your favorite song on the radio and you hear a siren. You immediately start a visual scan for the flashing lights of an emergency vehicle, and you begin to ignore the song. On the other hand, if you are at home and hear a siren through an open bedroom window while listening to the same song, you can likely continue to enjoy the music. In these cases, behavioral demands necessitate allocating attention to either the visual domain, ignoring the song while looking for the emergency lights, or the auditory domain, continuing to listen. The perceptual experience of the music in these two scenarios is qualitatively different. How does the sensory cortex differentially encode stimuli when they are attended versus when they should be ignored?

The auditory cortex (ACtx) is a locus for the convergence of sound-driven sensory signals and modulatory signals driven by task-engagement and attention. It receives its main sensory drive from the medial geniculate body, the auditory thalamus, but also receives a wealth of inputs from the frontal, parietal, cingulate, non-auditory sensory and other networks (Budinger et al. 2008; Budinger and Scheich 2009) that send feedback signals to rapidly alter sensory processing to meet behavioral demands (Winkowski et al. 2013; Rodgers and DeWeese 2014; Park et al. 2015). During a behavioral task, ACtx stimulus-evoked and pre-stimulus spiking responses are modulated when compared to passive

processing (Otazu et al. 2009; Buran, von Trapp, and Sanes 2014; Carcea, Insanally, and Froemke 2017; Kuchibhotla et al. 2017; Bagur et al. 2018), as are population-level dynamics such as signal and noise correlations (Downer, Niwa, and Sutter 2015; Downer et al. 2017). One limitation of this approach is that the passive state, to which the task-engaged state is typically compared, may comprise multiple states of arousal, ranging from somnolent to active and vigilant. These arousal state transitions have widely been reported to affect sensory cortical stimulus processing in multiple cortical regions (Niell and Stryker 2010; Fu et al. 2014; McGinley, David, and McCormick 2015; Dadarlat and Stryker 2017; Bigelow et al. 2019). Furthermore, paradigms comparing passive to task engaged processing poorly approximate the rapid and dynamic shifts in context-dependent behavioral responses required by real-world behavior.

To address this, we have created a novel audiovisual rule-switching task for mice that employs distinct modality-specific attentional modes while maintaining task engagement. In one part of the task, mice attended to auditory stimuli while ignoring visual stimuli. In the other, mice attended to visual stimuli while ignoring sound. The same set of audiovisual stimuli was used for both parts of the task to facilitate direct comparisons of stimulus responses across the rules and isolate the modulatory effects of modality-specific attention. We recorded single-unit activity while mice performed the task and analyzed the effects of task rule on firing rate responses, pre-stimulus activity levels, and information encoding of both task-relevant decision stimuli and task-irrelevant auditory receptive field mapping stimuli. Our findings suggest that switching between tasks induces a shift in auditory cortical stimulus processing, affecting response levels for task-relevant target and distractor stimuli as well as task-irrelevant stimuli. Pre-stimulus activity decreases when the animals are guided by the auditory rule, which appears to account for decreases in

absolute stimulus-evoked activity. We show that despite the reduced spike counts in the auditory rule, spike train-stimulus decoding and raw mutual information is not degraded. Instead, auditory attention appears to increase encoding efficiency, specifically in the deep cells of the auditory cortex.

3.2 Materials and Methods

Animals. All experiments were approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco. Twenty-seven (27) C57BL/6 background male mice were surgically implanted with a headpost and began behavioral training in this study, of which 14 eventually achieved sufficient task expertise for advancement to physiology recordings. Of these, 12 successfully performed the behavior during physiology recording sessions. All mice began the experiment between ages P56 and P84. Mice used in this report expressed optogenetic effectors in a subset of interneurons, which we intended to use for optogenetic identification of cells (Lima et al. 2009; analysis not included here). These mice were generated by crossing an interneuron subpopulation-specific Cre driver line (PV-Cre JAX Stock Nr. 012358; Sst-Cre: JAX Stock Nr. 013044) with either the Ai32 strain (JAX Stock Nr. 012569), expressing Cre-dependent eYFP-tagged channelrhodopsin-2, or the Ai40 strain (JAX Stock Nr. 021188), expressing Cre-dependent eGFP-tagged archaerhodopsin-3. Of the 12 mice included in this report, 7 were Ai32/Sst-Cre, 4 were Ai32/PV-Cre and Ai40/Sst-Cre. In some experiments, brief, low-level optogenetic pulses during the inter-trial interval of the task were used to identify opsin-expressing neurons (<0.3 mW light; 5 light pulses of 10 ms duration, every ~1.5 min); analysis of these results are outside of the scope of this report.

All mice were housed in groups of 2-5 for the duration of the behavioral training until the craniotomy. Post-craniotomy and during physiology recordings mice were housed singly (up to 6 days) to protect the surgical site. Mice were kept in a 12 hr/12 hr reversed dark/light cycle. All training occurred during the dark period, when mice showed increased activity and behavioral task performance (Roedel et al. 2006).

Audiovisual rule-switching behavior task. Adult mice were trained on an audiovisual go/no-go rule-switching behavior task. In this task, mice were positioned on a floating spherical treadmill in front of a monitor and a speaker, and an optical computer mouse recorded treadmill movement. Mice licked to receive a reward depending on auditory, visual or audiovisual stimulus presentation (here, referred to collectively as "decision" stimuli, a term we use to include both target and distractor), but the modality predictive of the reward changed partway through the behavioral session. Each session would start with a unimodal go/no-go block, in which a series of auditory (A1, A0; high or low tone clouds [TC]) or visual (V1, V0; upward or rightward moving gratings) stimuli was presented. After stimulus presentation, mice signaled choice by either licking a spout positioned in front of the mouth or withholding licking. Licking at the go unimodal stimulus would trigger a water reward, while licking at the no-go would trigger a short timeout. After a fixed number of unimodal trials, the stimuli would become audiovisual, but the rule for which stimulus predicted reward would carry over from the unimodal block. All four stimulus combinations (A1V1, A1V0, A0V1, A0V0) would be presented in the audiovisual block, such that two audiovisual combinations would be go stimuli and two would be no-go. Then, after completing a fixed number of trials in the audiovisual block, the task using the rule of the opposite modality would begin; a unimodal block with the other modality would start, followed by a second audiovisual block using the rule from the preceding unimodal block.

For any mouse, the stimuli predictive of the reward in each rule was kept constant across days and training sessions (e.g., a high TC would always predict a reward in the auditory rule, and a rightward grating would always predict a reward in the visual rule). In most cases, blocks proceed without interruption, although in some cases, the task is paused and resumed momentarily if the mouse fails to correctly transition between rules.

The task was self-paced using a virtual foraging approach, in which mouse locomotion (measured as rotation of the trackball) would cause a track of randomly placed dots on the screen to move down. After a randomly varied virtual distance, a decision stimulus would be presented, at which point the mouse would lick or withhold licking to signal choice. For receptive field mapping during physiology experiments, a random double-sweep (RDS) stimulus was presented in between decision stimuli, during the inter-trial "track" portion.

Behavior training and apparatus. Before beginning task training, mice were anesthetized using isoflurane and surgically implanted with a stainless steel headplate for head fixation during the task, and later, for physiology recordings. Three days post-implant, mice began a water restriction protocol based on previously published guidelines (Guo et al. 2014). Throughout the course of training, mice received a minimum water amount of 25 mL/kg/day, based on weight at time of surgical implant. After recovery from surgery, mice were given ~7 days to adjust to water restriction. Then, mice were head fixed and habituated to the floating treadmill for 15-30 min daily sessions with no stimulus presentation for 2-3 days. After mice appeared comfortable on the treadmill, a phased behavioral task training regimen was started, during which mice were trained once daily for ~6 days per week. On day 1, mice were introduced to an auditory-only (A-only)

"stimulus training" version of the task in which A1 ("go"/"rewarded") or A0 ("no-go"/"unrewarded") stimuli were presented, and a reward would be automatically administered shortly after the onset of A1. Next, the mice were put on an operant version of the A-only task, which required licking any time after the onset of A1 to receive a reward and withholding of licking during A0 to avoid a timeout punishment. Mice achieved proficiency, defined as two or more consecutive days of sensitivity index d' > 1.5 (see 'Data analysis' for calculation), on the A-only task after 9.5 ± 3.0 d after start of training (median ± SD, n = 12 successful mice). Then, a similar training structure was repeated for the visual task; V-only stimulus training with automatic rewards for V1, but not V0, followed an operant version of the visual task in which licks after V1 were rewarded and licks after V0 were punished (median time to proficiency: 24.0 ± 4.5 d after start). When mice had learned the tasks for each modality separately, they were introduced to an auditory-audiovisual (A-AV) version, in which the rule from the auditory stimulus carried over to the audiovisual block. This was intermixed with training days on a visual-audiovisual (V-AV) version of the task. Number of training days on A-AV or V-AV were decided based on prior performance, with extra training given where individual mice showed deficits. Mice were considered proficient at this stage after performing with d'>1.5 on each rule (A-AV; V-AV) on two consecutive days (median time to proficiency: 43.5 ± 7.0 d after start). Finally, the full rule-switching task was introduced (Fig. 3.1.d), generally alternating between days of visual rule-first (V-AV-A-AV) and the auditory rule-first (A-AV-V-AV) task sequences but allocating more training days to task configurations where individual mice showed deficits. Because physiology recordings were acute and strictly limited to 6 days after craniotomy, we set a greater threshold for expert-level performance on the full task before considering a mouse ready for physiology: three consecutive days of d'>2.5 (median time to expertise: 90.5 ± 25.8

d). Care was taken to train each mouse at a roughly consistent time of day (no more than ~1-2 hrs day-to-day variation). During expert-level task performance, mice typically completed 260-300 trials in a daily session (30 A-only; 100 to 120 A-AV; 30 V-only; 100 to 120 V-AV).

The behavior training setup was controlled by two computers: a behavior monitoring and reward control PC (OptiPlex 7040 MT, Dell) and a dedicated stimulus presentation machine running Mac OS X (Mac Mini, Apple). Both machines ran custom software built with MATLAB, and inter-machine communication used the ZeroMQ messaging protocol. Auditory and visual stimuli during the task and in passive sessions were generated and presented using the Psychophysics Toolbox Version 3 (Kleiner, Brainard, and Pelli 2007). Water rewards were administered using a programmable syringe pump (NE-500, New Era Pump Systems, Farmingdale, NY), positioned outside of the sound-attenuating recording chamber. Early in training, water reward volume was set at 0.01 mL per correct response, but over training the reward volume was gradually decreased to 0.006 mL to achieve greater trial counts before the mouse reached satiety. Licking events were recorded using a custom photobeam-based lickometer circuit based on plans by Evan Remington (Xiaoqin Wang Lab, Johns Hopkins University). Licks were registered when an IR photobeam positioned in front of the lick tube was broken. Lickometer output was read by an Arduino Uno microcontroller (Arduino, LLC) that interfaced with the computer controlling the behavior system. The lickometer was queried for photobeam breaks at a sampling rate of 100 Hz.

***In vivo* awake recordings during behavior.** Animals in this experiment underwent two surgeries: first, as noted, before training a surgery to implant a custom steel headplate over

the temporal skull using dental cement was conducted. The animal was anesthetized using isoflurane and a headplate was positioned over auditory cortex, ~2.5 mm posterior to bregma and under the squamosal ridge, to allow for physiology recordings after achieving task expertise. When mice successfully completed the training regimen outlined above, a craniotomy was performed. The animal was again anesthetized using isoflurane and an elliptical opening (0.75 mm wide x 1.5 mm long) was made in the skull over ACtx using a dental drill. This opening was promptly covered with silicone elastomer (Kwik-Cast, World Precision Instruments), and the animal was allowed to recover from anesthesia overnight. The following day, the animal was affixed by its headplate over the spherical treadmill inside of a sound-attenuating recording chamber, the silicone plug over the craniotomy was removed and the craniotomy was carefully flushed with saline. A silver-chloride ground wire was placed into the craniotomy well at a safe distance from the exposed brain. A 64-channel linear probe (20 μm site spacing; Cambridge Neurotech, Cambridge, UK) was slowly inserted in the brain using a motorized microdrive (FHC, Bowdoin, ME) at an approximate rate of ~1 μm / s (Fiáth et al. 2019). After reaching the desired depth, the brain was allowed to settle for 10 min, after which the water spout, lickometer, visual stimulus delivery monitor and speaker were positioned in front of the mouse, and the behavior session commenced. Behavior sessions were sometimes stopped early and restarted due to poor performance. In approximately half of behavior-physiology sessions (15 of 27 successful recordings), the task was stopped due to low performance after the rule transition and restarted at the beginning (unimodal block) of the second rule. To control for possible effects of task order, attempts were made to counterbalance recordings from A-rule first (18) and V-rule first (9) behavior sessions.

After completion of the behavior task, the water spout and lickometer were removed, and a series of auditory and/or visual passive experiments were conducted in order to characterize the response properties of the recording site. All stimuli were presented with the auditory and visual stimulation apparati described above. Following completion of these experiments, the probe was slowly removed, and the brain was covered with a thin layer of freshly mixed 2.5 % agarose in saline, followed by a layer of silicone elastomer. The animal was returned to its home cage, and the following day the physiological recording process was repeated. Recordings were made for up to 6 days after the craniotomy. The neural signal acquisition system consisted of an Intan RHD2000 recording board and an RHD2164 amplifier (Intan Technologies), sampling at 30 kHz.

Auditory and visual stimuli. In-task auditory decision stimuli were 1 s tone clouds (TCs), consisting of 50 ms tone pips overlapping by 25 ms, with frequencies in a 1 octave band around either 17 kHz or 8 kHz. TCs were frozen for the duration of the task, so that each mouse always heard the same 17 kHz TC and the same 8 kHz TC, allowing for direct comparisons of sound-evoked neural responses across rules without concern that stimulus peculiarities may be driving observed differences. Visual decision stimuli consisted of a circular moving grating stimulus (33° diameter subtended visual space), which appeared at the center of the screen for 1 s (coincident with TC stimulus, during bimodal presentation). The gratings moved with a 4 Hz temporal frequency, 0.09 cycles/degree spatial frequency at 50% contrast. Grating movement was either upward or rightward. In between decision stimulus presentations during the behavior task, a random double-sweep (RDS) stimulus was presented for receptive field mapping (Gourévitch et al. 2015). The RDS consisted of two uncorrelated random sweeps that vary continuously and smoothly between 4 and 64 kHz, with a maximum sweep modulation frequency of 20 Hz.

After the behavior task, a number of passive auditory and/or visual stimulus experiments were conducted to characterize response properties of the site. Responses to click trains and pure tones were used to determine the span of sound responsive sites on the probe for depth measurement. Click trains consisted of broadband 5 ms white noise pulses, presented at 20 Hz for 500 ms duration. Pure tone stimuli consisted of 100 ms tones of varied frequencies (4 – 64 kHz, 0.2 octave spacing) and sound attenuation levels (30 – 60 dB in 5 dB linear steps), with an interstimulus interval of 500 ms.

Auditory stimuli were presented from a free-field electrostatic speaker (ES1, Tucker-Davis Technologies) driven by an external soundcard (Quad-Capture or Octa-Capture, Roland) sampling at 192 kHz. Sound levels were calibrated using a Brüel & Kjær model 2209 meter and a model 4939 microphone. Visual stimuli were presented on a 19-inch LCD monitor with a 60Hz refresh rate (VW199), positioned 25 cm in front of the mouse and centered horizontally and vertically on the eyes of the mouse. Monitor luminance was calibrated to 25 cd/m2 for a gray screen, measured at approximate eye level for the mouse.

Data analysis. *Behavioral performance.* Task performance was evaluated by calculation of the *d*' sensitivity index:

$$d' = Z(H) - Z(F)$$

where *H* is hit rate and *F* is false alarm rate, and *Z* is the inverse normal transform. Because this transform is undefined for values of 0 or 1 and hit rates of 1 commonly occurred in this study, we employed the log-linear transformation, a standard method for correction of extreme proportions, for all calculations of *d*' (Hautus 1995). In this correction, a value of

0.5 is added to all elements of the 2 x 2 contingency table that defines performance such that

$$H = (hits + 0.5)/(hits + misses + 1)$$

$$F = (FA + 0.5)/(FA + CR + 1)$$

where FA is the false alarm count and CR is the correct reject count. To ensure that mice properly transitioned between task rules, *d*' values were calculated separately for responses in the A-rule and the V-rule. Behavioral sessions during physiological recording with *d*' < 1.5 in either rule were excluded from analyses (n = 28 successful sessions, n = 12 mice; 1 session excluded due to recording artifact, see below).

Spike sorting and unit stability evaluation. To assign spikes to unit clusters, we used the free, open-source automatic spike sorting software KiloSort2 (KS2; Pachitariu et al. 2016), which is explicitly designed for detection of unit drift. KiloSort2 has been benchmarked against competing automatic spike sorting algorithms and performs well on ground-truth intracellular/extracellular tests (Magland et al. 2020). KS2 waveform clusters were evaluated in a multi-step process. First, after cluster generation, KS2 recommends candidate "good" clusters using a score based on the event rate in the central 2 ms bin of the histogram relative to the baseline of the auto-correlogram. These clusters were then evaluated using a custom MATLAB interface. Waveforms that appeared non-neuronal in shape, likely attributable to electrical or light artifact, were labeled as noise and excluded from analysis. Additionally, clusters with 2 ms refractory period violations of >0.5% were labeled as multi-unit and not analyzed further (Laboy-Juárez, Ahn, and Feldman 2019; Sukiban et al. 2019). Activity for each included unit plotted for the duration of the recording as a spike raster and binned spike counts (2 min bins) and manually examined for periods where the unit showed a substantial decrease in activity (periods flagged for instability

exhibited 88 ± 10 % decrease in firing rate, mean ± SD). Time periods where the unit no longer exhibited meaningful activity were marked and removed from analysis (117/873 SUs with flagged durations >10% of recording time). Of the 28 physiology recording sessions with successful behavior, one was excluded due to a high degree of electrical noise contamination.

Classification of clusters by depth and waveform shape. Probes with electrode spans of 1275 μm were used, allowing for channels to record below and above ACtx. During recording, the probe was intentionally lowered such that at least several channels showed a prominent drop in field potential amplitude and spiking activity, indicating penetration into the white matter (Land et al. 2013). After behavior sessions, a set of auditory and visual stimulation protocols was used to map response properties of each electrode site, and multi-unit activity (MUA) responses were analyzed. Here, we define MUA as threshold crossings of 4.5 SD above a moving window threshold applied to each channel. Analysis of MUA was restricted to site characterization and is not included in the main results. Multi-unit responses to 100 ms pure tones (see Auditory and visual stimuli) were used to generate a PSTH for all tones and frequency response area (FRA) heatmap for all recorded channels, organized by depth (Fig. 3.2.b). We analyzed each tone or click PSTH for reliable responses, which we defined has trial-to-trial similarity of $p<0.01$ (Escabí et al. 2014). Based on the deepest and most superficial channel with a reliable MUA sound response of any magnitude, we designated a span of responsive channels that were putatively located in cortex, as well as a span of unresponsive channels above and below the cortical depth. In line with histologically measured mouse ACtx depths (Christianson, Sahani, and Linden 2011; Morrill and Hasenstaub 2018), cortical responsive channels spanned a length of 940 ± 121 μm along the probe (median ± SD; range: 740 - 1140 μm; n = 12 mice, 27 penetrations).

This span was then used to divide channels within the cortex into depth groups, based on measurements of the fraction of cortex attributed to supragranular (layers 1-3), granular (layer 4) and infragranular (layers 5-6) in the mouse ACtx, as shown in the Allen Institute Mouse Brain Atlas (https://mouse.brain-map.org/). Because our method lacks additional physiological or histological confirmation of these laminar divisions, we refer to these depth groupings as "superficial", "middle", and "deep." Single units were given the fractional depth and the corresponding depth grouping of the channel on which the largest magnitude waveform was recorded, which is most likely to be the channel closest to the soma (Humphrey and Schmidt 1990).

Clusters were also classified into broad-spiking (BS; putatively excitatory) and narrow-spiking (NS; putatively fast-spiking inhibitory) units, as is common in extracellular recording studies (Nandy, Nassi, and Reynolds 2017; Bigelow et al. 2019; Cardin, Palmer, and Contreras 2007; Phillips, Schreiner, and Hasenstaub 2017). Classification into BS/NS was made on the basis of the bimodal distribution of waveform peak-trough durations (Fig. 3.2.c, NS/BS transition boundary = 0.6 ms). From sessions with successful behavior, we recorded 873 SUs from cortical depths, comprising 17.2% (150) NS units and 82.8% (723) BS units.

Firing rate analysis and trial filters. To compare firing rate (FR) responses to identical stimuli across task rules and to the receptive field mapping stimulus, we measured FR in the first 300 ms post-stimulus onset. Only units with nonzero FRs in both rules were included. To ensure that our measurements were capturing periods of engagement in the task, all trials with incorrect responses (both misses and false alarms) were excluded from this and all analyses. We also excluded trials with recorded licks earlier than the 300 ms

post-stimulus onset, or in the 500 ms pre-stimulus onset. Given these filters, analyses were restricted to units present in the recording during at least 10 trials (correct and without "early licks") for each stimulus type (see Results for number of units included for each analysis). For analysis of pre-stimulus FRs (-300 to 0 ms), we included correct trials and those without early licks for all bimodal trials in each rule. Significance of individual SU FR difference across rules was assigned based on an unpaired t-test between responses in A-rule and V-rule trials ($p<0.01$).

PSTH-based Euclidean distance decoding. A peristimulus time histogram(PSTH)-based decoder was used to compare how much information A-rule and V-rule spike trains provided about stimulus identity (Fig. 3.6.a; see Foffani and Moxon 2004 for complete details). In this method, two or more neuronal responses are compared by generating template PSTHs with the removal of one test trial. This test trial response is also binned into a single-trial PSTH, and then classified as belonging to the nearest template in n-dimensional Euclidean space, where n is the number of PSTH bins. More formally, the nearest template is that which minimizes the Euclidean norm between test and template vectors (PSTHs), which is the square root of the sum of squared differences between each element of the vector. This process is then repeated for all trials comprising the template PSTHs. Decoding accuracy is the percentage of trial responses that are correctly assigned to the stimuli that elicited them. Mutual information (MI) is calculated from a confusion matrix of classifications as follows:

$$MI = \sum_i \sum_j P(X_i Y_j) log_2 (\frac{P(X_i Y_j)}{P(X_i) * P(Y_j)})$$

where X is the decoder prediction, Y is the actual, $P(X_i Y_j)$ represents the value of the (i, j) element of the confusion matrix, and $P(X_i)$ and $P(Y_j)$ represents the sums on the marginals.

This yields a value of mutual information in bits. This PSTH-based decoding method and its MI calculation both depend on spike count. To measure encoding efficiency (bits per spike), we normalized mutual information by the joint mean spikes per trial of the responses submitted to the decoder (Zador 1998; Buracas et al. 1998; Bigelow et al. 2019).

For consistency with FR analyses, a time window of 0 to 300 ms, where stimulus onset is 0, was chosen for decoding analysis. A PSTH binwidth of 30 ms was chosen based on optimal binwidth calculations for mouse ACtx using the same decoding method (Hoglen et al. 2018). To filter out units with low responsiveness to any of the stimuli in a given decoding analysis, we required a minimum FR of 5 Hz during the 0-300 ms window in at least one of the stimulus conditions. As such, unit sets may differ between each decoding analysis due to units that were responsive to one set of stimuli but unresponsive to others. Included unit counts for all analyses are noted in text.

Spectrotemporal receptive field (STRF) analysis. To test whether task rule modulates auditory receptive fields, we presented a random-double sweep stimulus (RDS, described in *Auditory and visual stimuli*) in between trials. The RDS consists of two uncorrelated random sweeps that vary continuously between 4 and 64 kHz (Gourévitch et al. 2015) and was presented for durations between ~1 and 15 s, depending on rate of task progression (determined by running speed). Different randomly generated RDS segments were presented in each inter-trial interval. Because total RDS duration varied between the A-rule and the V-rule in a single session, we equated presentation time across rules by truncating the segments of the rule with greater RDS time (RDS presentation duration: 7.58 ± 2.50 min, mean ± SD; n = 27 sessions). This was done to ensure that longer stimulus presentation time did not bias receptive field property measurements. The first 200 ms of

RDS response was dropped from all STRF analyses to minimize bias reflecting strong onset transients. SU activity during these short RDS segments was used to generate spectrotemporal fields (STRFs) for each segment using standard reverse-correlation techniques (Aertsen and Johannesma 1981). In brief, the spike-triggered average (STA) was calculated by adding all stimulus segments that preceded spikes using a window of 200 ms before and 50 ms after each spike. The choice of 200 ms prior to each spike reflects the upper limit of temporal integration times (or "memory") of auditory cortical neurons (Atencio and Schreiner 2013), and the 50 ms post-spike time was included to estimate acausal values, i.e., those that would be expected by chance given the stimulus and spike train statistics (Gourévitch et al. 2015). For a more straightforward interpretation of stimulus-driven changes in activity, STRFs were transformed into units of firing rate (Hz) using standard methods discussed elsewhere (Rutkowski et al. 2002). To compare across rule conditions, STRFs were calculated separately from RDS segments presented during the A-rule and those presented during the V-rule. To ensure robust analysis of STRF properties, we filtered STRFs using a trial-to-trial correlation metric (Escabí et al. 2014). First, STRF segments were randomly divided into two halves, re-averaged separately, and a correlation value was calculated for the two STRFs. This process was then repeated 1000 times, and the mean of correlations defined the reliability value for each STRF. Because such reliability metrics are influenced by spike statistics such as count and interspike interval, we compared the mean observed STRF reliability to a null distribution of reliabilities. This distribution was generated by repeating the procedure on null STRFs made from circularly-shuffled spike trains, thereby preserving spike count and interspike interval but breaking the timing relationship between spikes and stimulus. A p-value was calculated as the fraction of the null STRF reliabilities greater than the mean observed

STRF reliability, and significance was set at $p<0.01$. STRFs with significant reliability in either rule were included in subsequent analyses.

The STRF provides a linear approximation of the temporal and spectral preferences of a neuron, but it does not fully describe the relationship between stimulus and stimulus-evoked activity. For instance, it does not provide an estimate of stimulus selectivity, whether a unit fires exclusively in response to an ideal stimulus (a perfect match between stimulus and receptive field), or whether it instead fires in graded approximations to the ideal stimulus. The nonlinear input/output function ("nonlinearity") provides an estimate of this relationship (Escabí and Schreiner 2002; Atencio, Sharpee, and Schreiner 2008; Atencio and Schreiner 2012). The nonlinearity is a scaled ratio of two distributions: one reflecting the similarity of the windowed stimulus segments (RDS) preceding a spike and the STRF, and the other reflecting the similarity of all possible windowed stimulus segments and the STRF, regardless of whether a spike occurred (Fig. 3.9.a). Stimulus-STRF similarity is operationally defined as the inner product between the STRF and the stimulus of equivalent dimensions, with relatively high values reflecting closer matches between the STRF and stimulus. To calculate the nonlinearity, projections between RDS stimulus segments that elicited a spike (s) and the STRF were made by calculating the inner product $z = s \cdot STRF$, which defines the distribution $P(z|spike)$. Then the distribution $P(z)$ was made from similarity calculations of all possible windowed RDS segments and the STRF. The mean μ and the standard deviation (SD) σ of $P(z)$ were calculated, and the distributions were transformed into units of SD: $x = (z - \mu)/\sigma$, yielding distributions of $P(x|spike)$ and $P(x)$ expressed in units of SD. The nonlinearity was then calculated as:

$$P(spike|x) = P(spike)\frac{P(x|spike)}{P(x)}$$

where *P*(*spike*) is the mean FR. The nonlinearity thus provides an estimate of the firing rate from the range of similarity values (*x*) between the STRF and the stimulus. If the nonlinearity is high when *x* is near 0, the neuron fires indiscriminately during the RDS stimulus. Instead, if nonlinearity values are only high as *x* increases, the neuron preferentially fires when the stimulus and STRF are similar.

Using the distributions described above, a spike count-normalized measure of mutual information between the calculated STRF and the spike train can be calculated as:

$$MI = \sum P(x|spike) log_2(\frac{P(x|spike)}{P(x)})$$

To compare nonlinearities from the same neuron calculated during different task rules, we calculated two statistics to characterize the nonlinearity shape: the asymmetry index and skewness (Atencio and Schreiner 2012; Atencio, Sharpee, and Schreiner 2008; Atencio and Schreiner 2008). The asymmetry index (ASI) is defined as:

$$ASI = (R - L)/(R + L)$$

where *R* is the sum value of nonlinearity values greater than 0, and *L* is the sum value of the nonlinearity less than 0. The ASI ranges from -1 to 1, with 1 indicating a unit that spikes exclusively during positive correlations with the filter (STRF), and -1 indicating exclusive spiking during anti-correlation.

The skewness of the nonlinearity describes the amount of information in the tail. Formally, it is defined as the third central moment of the nonlinearity divided by its cubed SD (Atencio and Schreiner 2008):

$$skew = \sum_i (Y_i - \mu)^3/(N\sigma^3)$$

where Y_i are values of the nonlinearity, μ is the nonlinearity mean, N is the number of points in the nonlinearity and σ is the SD of the nonlinearity. Skewness values near 1 indicate a gradual transition of activity as the correlation of the stimulus and the filter increase, whereas higher values indicate a sharper transition (Fig. 3.9.c for examples). We used these values to compare how well STRFs from A-rule and V-rule inter-trial intervals predict activity levels, and thus whether activity is well described by this standard filter model or whether activity is likely influenced by other sources unrelated to the stimulus.

Statistics. All statistical calculations were performed in MATLAB r2019a and its Statistics and Machine Learning Toolbox, v11.5. For comparisons of SU responses across task rules at the group level, a paired Wilcoxon signed-rank (WSR) test was used. To determine if individual SUs were significantly modulated by rule, an unpaired Student's t-test on FR was used with a threshold of $p<0.01$. Descriptive statistics reported in text are mean ± standard deviation (SD), unless otherwise noted. Fractional change values between task rules are reported as median of the A-rule/V-rule. All other statistical tests are described in Results. Sample sizes (n) are indicated for each comparison in Results.

3.3 Results

A novel audiovisual rule-switching task for mice and extracellular recordings from the auditory cortex. We created an audiovisual rule-switching behavioral task, in which mice made decisions using auditory stimuli while ignoring simultaneously presented visual stimuli or made decisions based on visual stimuli while ignoring auditory stimuli. Mice were head-fixed and running on a trackball suspended on air (Fig. 3.1). This was a self-paced virtual foraging task in which the movement of the visual stimuli on a monitor in front of the mouse and the presentation of auditory stimuli was determined by locomotion

speed. At each stimulus presentation, a go or no-go stimulus was presented, and the mouse was required to lick a spout to get a water reward (for 'go' stimuli) or withhold licking (for 'no-go' stimuli). Decision stimuli were 1 s long tone clouds [TCs] centered at 8 kHz or 17 kHz paired with 1 s long drifting gratings, which were vertically- or horizontally-oriented (Fig. 3.1.b; see Methods for stimulus details). In between trials, a random double sweep stimulus was presented for mapping of receptive fields during the two attentional states (Fig 3.1.c).

In the task, a short block of unimodal (auditory or visual) stimuli was first presented, which served to cue the rule for the subsequent audiovisual block. Next, a block of the other unimodal stimulus was presented, which cued the rule for a final block of audiovisual stimuli (Fig. 3.1.d). For example, if the first cue block were unimodal auditory, the mouse would lick when presented with the go A1 stimulus, and withhold licking for the no-go A0 stimulus. The auditory rule would then carry over to the following audiovisual block, so that the mouse should lick for A1V1 and A1V0, but not A0V1 or A0V0 stimuli (decision-relevant stimulus feature underlined). Next, a visual-only cueing block would be presented, with V1 as the go stimulus and V0 as no-go. In the subsequent visual-rule audiovisual block, the go stimuli would be A1V1 and A0V1, while the no go stimuli would be A1V0 and A0V0.

The set of audiovisual stimuli in each of the rules was physically identical, although their behavioral significance sometimes differed. For example, A1V0 in the auditory rule, a go stimulus, was the same tone-cloud and grating combination as A1V0 in the visual rule, in which it was a no-go stimulus. (Note the naming convention that a "1" or "0" indicates a go or no-go stimulus, respectively, only in the rule indicated by the preceding character, "A" or "V".) For each animal, the rewarded and unrewarded stimuli were held constant for the duration of the experiment. Between animals, we attempted to counterbalance stimuli; for

the auditory rule, 15 animals were 8 kHz-go, while 13 animals were 17 kHz-go. For the visual rule, 17 animals were vertical-go while 11 were horizontal-go.

After achieving expertise in the task, a craniotomy was performed over the right auditory cortex and acute recordings were made daily for up to six days. During these 1-3 hr recordings, mice performed the audiovisual rule-switching behavior task. Subsequently a set of passive auditory and/or visual stimuli was presented for characterization of the recording site (Fig. 3.2.b). Using multisite silicon probes to span the full cortical depth (Fig. 3.2.a), we recorded single units from the auditory cortices of 12 mice during 27 successful behavioral sessions (Fig 3.1.f**)**. Units were assigned depth based on the position of the electrode channel that recorded the largest amplitude waveform from the unit, presumed to be closest to the soma, relative to the span of channels responsive to pure-tone auditory search stimuli (Fig. 3.2.b). Based on this depth measurement, units were assigned to either superficial, middle, or deep groups. Our method approximates the division of cortex into supragranular, granular and infragranular laminae, although without physiological or histological markers for their boundaries we cannot make conclusive laminar assignments. We further divided SU waveforms into narrow-spiking (NS) or broad-spiking (BS) populations based on peak-to-trough time. The NS/BS distinction provides a coarse classification into putative fast-spiking interneurons and regular-spiking neurons, the latter of which are dominated by excitatory pyramidal cells (Cardin, Palmer, and Contreras 2007; Nandy, Nassi, and Reynolds 2017; Bigelow et al. 2019; Phillips, Schreiner, and Hasenstaub 2017). Our dataset consisted of 17% (150) NS and 83% (723) BS units. We used these divisions to determine whether subpopulations of putatively inhibitory and putatively excitatory cells are selectively modulated by the modality-specific attention switching required in the task.

Auditory-specific attention generally suppresses response firing rates relative to visual attention. We recorded SUs during the audiovisual task-switching behavior, focusing on comparing responses to stimuli in the bimodal blocks across the rules. A wide variety of responses to stimuli and modulatory effects of task rule were observed (Fig. 3.3), with many cells showing clear firing rate changes in response to stimulus onset. We first asked how modality-specific attentional state modulates firing rates during responses to task decision stimuli. To do this, we compared SU FR responses to AV stimuli in the A-rule with responses to the same AV stimuli in the V-rule. To capture a predominantly sensory-driven component of the response, we restricted our analysis to the first 300 ms post-stimulus onset (Fig. 3.4.a.i and Fig. 3.4.a.ii). This window was chosen in part to avoid the onset of the licking behavioral response, which typically occurred at ~600 ms post-stimulus onset (median: 583 ms; 5th-95th percentiles: 275-1064 ms; 5.8% of licks <300 ms, n = 3,342 lick trials across dataset). Trials with licking activity earlier than 300 ms were removed from analysis. We first compared A-rule and V-rule responses to the stimuli that were rewarded in the A-rule (A1V1 and A1V0 responses combined, indicated by A1*). We found that, on average, SU responses in the middle and deep layers were suppressed in the A-rule relative to the V-rule, but no significant change was found in the superficial units (Fig. 3.4.b.i). For all layers, the NS and BS populations were similarly modulated. For middle and deep groups, population firing rate was suppressed by 11-14% in the A-rule relative to the V-rule (Super. BS: p = 0.68, Z = -0.41, V-rule FR: 4.52 ± 7.34, A-rule: 4.43 ± 6.94, n= 101; Super. NS: p = 0.13, Z = 1.51, V-rule: 17.37 ± 17.33, A-rule: 15.77 ± 14.61, n= 40; Mid. BS: p = 0.011, Z = 2.54, median fold change: 0.85, V-rule: 5.68 ± 8.45, A-rule: 5.89 ± 13.15, n = 153; Mid. NS: p = 0.04, Z = 2.06, median fold change: 0.86, V-rule: 17.13 ± 14.68, A-rule: 13.85 ± 11.01, n = 32; Deep BS: p = 6.4e-06, Z = 4.51, median fold change: 0.89, V-rule: 6.22 ± 8.92, A-rule: 5.32 ± 7.92, n

= 370; Deep NS: p = 0.0011, Z = 3.25, median fold change: 0.86, V-rule: 25.39 ± 20.27, A-rule: 21.79 ± 19.27, n = 74; paired Wilcoxon signed rank [WSR]; fold change: A-rule/V-rule; mean Hz ± SDs). While suppression was the dominant trend on the population level, many cells had enhanced rates during the A-rule as well. When we examined significantly modulated units ($p<0.01$ FR difference between rules, unpaired t-test), we found a roughly equal proportion of A-rule suppressed and A-rule enhanced BS units in the superficial and middle layers. In the deep layer, the fraction of suppressed cells was roughly twice that of enhanced (see figure for counts: Fig. 3.4.b.i). For NS cells, there were more significantly suppressed units than enhanced in all layers.

We found that this pattern of FR modulation was not unique to the stimulus A1*, which is rewarded in the auditory rule. A0* (A0V1 and A0V0 combined) stimuli, those unrewarded in the auditory rule, showed a strikingly similar pattern of modulation between the A-rule and V-rule; at a population level, the superficial cells did not show any change, while the middle and deep NS and BS activity was suppressed in the A-rule by 18-20% (Super. BS: p = 0.79, Z = 0.26, V-rule: 4.35 ± 5.60, A-rule: 4.49 ± 6.22, n= 96; Super. NS: p = 0.087, Z = 1.71, V-rule: 17.74 ± 13.62, A-rule: 15.73 ± 12.10, n= 40; Mid. BS: p = 0.04, Z = 2.05, median fold change: 0.81, V-rule: 5.74 ± 7.74, A-rule: 5.64 ± 10.77, n = 148; Mid. NS: p = 0.047, Z = 1.98, median fold change: 0.80, V-rule: 20.96 ± 20.18, A-rule: 17.37 ± 16.05, n = 32; Deep BS: p = 5.3e-08, Z = 5.44, median fold change: 0.82, V-rule: 6.12 ± 10.22, A-rule: 4.98 ± 8.15, n = 354; Deep NS: p = 0.0058, Z = 2.76, median fold change: 0.80, V-rule: 21.29 ± 15.29, A-rule: 18.81 ± 16.66, n = 74; paired WSR; fold change: A-rule/V-rule; mean Hz ± SD). Relative fractions of significantly enhanced and suppressed units for A0* stimuli were similar to those described above for A1* stimuli (Fig. 3.4.b.ii).

To determine if attentionally-induced suppression and enhancement acts similarly across stimuli with different behavioral valences and acoustic properties (8 kHz or 17 kHz TC frequency centers), we asked whether the units show the same sign of modulation for A1* as they do for A0*. For significantly task-rule modulated BS units, 79% show the same sign of modulation (52% suppressed for both A1* and A0*, 27% enhanced for both; Fig. 3.4.c). For significantly modulated NS units, the 96% showed consistent modulation (67% suppressed for both, 29% enhanced for both). These results suggest that FR modulation by attentional state influences rewarded and unrewarded auditory stimuli in similar ways.

Does the modulation by modality-specific attention we observed act only on the task-relevant decision stimuli, or do attentional effects act on other sound stimuli as well? The latter would suggest that the switch between task rules alters the mode in which the ACtx is processing sensory stimuli in a general manner, rather than acting only on stimuli useful for the task. In between trials containing decision stimuli, mice were presented with a random-double sweep (RDS) stimulus for receptive field mapping (see Methods for details). This allowed us to compare the effects of modality-specific attention between task-relevant and task-irrelevant stimuli (Fig. 3.4.d). We analyzed the firing rate response to the inter-trial interval RDS stimulus during the 0-300 ms post-stimulus onset and found that rule modulation for the RDS stimulus resembled that for decision stimuli. Middle and deep layer BS and NS populations exhibited a population-level FR suppression during the A-rule. The superficial layer units were not significantly modulated, although both NS and BS cells exhibited a trend toward suppression in the A-rule (Super. BS: $p = 0.18$, $Z = 1.33$, V-rule: 3.00 ± 4.00, A-rule: 3.15 ± 4.82, n= 92; Super. NS: $p = 0.17$, $Z = 1.37$, V-rule: 12.83 ± 8.67, A-rule: 11.60 ± 9.29, n= 36; Mid. BS: $p = 0.0011$, $Z = 3.27$, median fold change: 0.89, V-rule: 4.81 ± 5.89, A-rule: 4.25 ± 6.08, n = 138; Mid. NS: $p = 0.0099$, $Z = 2.58$, median fold change:

0.70, V-rule: 14.26 ± 14.77, A-rule: 10.98 ± 10.76, n = 32; Deep BS: p = 1.1e-11, Z = 6.79, median fold change: 0.81, V-rule: 4.12 ± 5.35, A-rule: 3.26 ± 4.25, n = 350; Deep NS: p = 0.0026, Z = 3.01, median fold change: 0.77, V-rule: 16.48 ± 12.08, A-rule: 13.52 ± 10.77, n = 71; paired WSR; fold change: A-rule/V-rule; mean Hz ± SD). In all subpopulations we found both units that were significantly suppressed in the A-rule and units that were enhanced ($p<0.01$ FR difference between rules, unpaired t-test; Fig. 3.4.e, right). Suppressed units represented a greater share of the subpopulation than enhanced for all groups, but this difference was particularly large for the deep layers (BS: 8% sig. enhanced, 28% sig. suppressed; NS: 15% sig. enhanced, 52% sig. suppressed). These results suggest that task rule modulates stimulus-evoked activity levels, and that the dominant trend is suppression of activity in the BS and NS populations of the middle and deep layers in the A-rule when compared to the V-rule. However, we also observed units at all depth and BS/NS subgroups that showed a significant increase in activity in the A-rule when compared to the V-rule. The sign of modulation was largely consistent across the 8 kHz and 17 kHz TC stimuli, suggesting that neither these frequency differences nor their associated behavioral valences explain the bulk of the FR effects. Task rule FR modulation is also not specific to the task-relevant stimuli, affecting both tone-cloud/grating stimuli used for decision making and the RDS stimuli.

Modulation in pre-stimulus firing rates partially accounts for changes in stimulus-evoked firing rates across rules. Ongoing activity in the sensory cortex is known to affect subsequent sensory-evoked responses (Arieli et al. 1996; Haider and McCormick 2009). The reduction in stimulus-evoked activity during auditory attention could be explained by a decrease in stimulus-related drive or by a general decrease in auditory cortical activity. To address these possibilities, we determined if pre-stimulus baseline activity levels were also

modulated by rule. In this analysis, we used a window 300 ms prior to the onset of a decision stimulus during which no auditory stimulus was presented (Fig. 3.5.a). An important caveat is that this pre-stimulus period immediately follows the offset of the RDS inter-trial interval stimulus and precedes the onset of the decision stimulus, such that the animal is cued for expectation of a trial. As such, there may be anticipatory activity that influences the estimation of the baseline, and this analysis window does not represent a purely spontaneous rate.

We found that pre-stimulus firing rate decreased across the rules with similar patterns to those found for stimulus-evoked activity; middle and deep subpopulations (both BS and NS cells) were both suppressed in the A-rule relative to V-rule. Again, we observed no significant group-level modulation of superficial unit FRs (Fig. 3.5.b; Super. BS: $p = 0.43$, $Z = -0.79$, V-rule: 2.90 ± 4.40, A-rule: 3.21 ± 4.88, n= 104; Super. NS: $p = 0.94$, $Z = 0.08$, V-rule: 12.33 ± 10.40, A-rule: 12.12 ± 9.87, n= 40; Mid. BS: $p = 0.0077$, $Z = 2.66$, median fold change: 0.90, V-rule: 4.57 ± 5.63, A-rule: 4.23 ± 6.90, n = 153; Mid. NS: $p = 0.04$, $Z = 2.06$, median fold change: 0.80, V-rule: 15.29 ± 14.68, A-rule: 12.84 ± 12.17, n = 32; Deep BS: $p = 1.2e-08$, $Z = 5.70$, median fold change: 0.84, V-rule: 4.30 ± 5.65, A-rule: 3.47 ± 4.61, n = 382; Deep NS: $p = 0.013$, $Z = 2.48$, median fold change: 0.85, V-rule: 15.81 ± 12.07, A-rule: 14.00 ± 11.06, n = 74; paired WSR; fold change: A-rule/V-rule; mean Hz ± SD).

To test whether this observed pre-stimulus reduction in FR accounts for stimulus-evoked changes, we recalculated response FRs as fold change over the 300 ms pre-stimulus FRs (Fig. 3.5.a). Adjusting for the pre-stimulus reduction in activity accounted for most of the stimulus-evoked FR changes, and most subpopulations no longer showed significant modulation by task rule. Interestingly, adjusting for spontaneous rate did show a small reduction in the superficial putatively-inhibitory NS population, with median FR

reductions of 16% for A1* (Super. BS: p = 0.24, Z = 1.18, V-rule: 4.18 ± 8.56, A-rule: 3.44 ± 8.09, n= 96; Super. NS: p = 0.012, Z = 2.50, median fold change: 0.84, V-rule: 2.65 ± 4.67, A-rule: 2.01 ± 2.59, n = 40; Mid. BS: p = 0.85, Z = 0.19, V-rule: 2.36 ± 5.96, A-rule: 2.09 ± 3.49, n= 151; Mid. NS: p = 0.68, Z = 0.41, V-rule: 3.27 ± 6.93, A-rule: 2.52 ± 4.05, n= 32; Deep BS: p = 0.11, Z = -1.61, V-rule: 8.24 ± 50.57, A-rule: 7.82 ± 42.63, n= 347; Deep NS: p = 0.83, Z = 0.22, V-rule: 2.14 ± 2.20, A-rule: 2.39 ± 3.97, n= 74; Wilcoxon signed rank, paired; fold change: A-rule/V-rule; mean FR fold change ± SD) and 12% for A0* (Super. BS: p = 0.69, Z = 0.40, V-rule: 2.33 ± 2.76, A-rule: 2.20 ± 2.10, n= 90; Super. NS: p = 0.0027, Z = 3.00, median fold change: 0.88, V-rule: 3.60 ± 10.45, A-rule: 2.17 ± 2.95, n = 40; Mid. BS: p = 0.093, Z = -1.68, V-rule: 1.47 ± 2.06, A-rule: 1.97 ± 3.69, n= 144; Mid. NS: p = 1, Z = 0.00, V-rule: 2.54 ± 4.38, A-rule: 2.23 ± 3.06, n= 32; Deep BS: p = 0.11, Z = 1.59, V-rule: 4.19 ± 20.30, A-rule: 4.72 ± 23.02, n= 340; Deep NS: p = 0.6, Z = 0.53, V-rule: 2.14 ± 4.03, A-rule: 1.92 ± 2.39, n= 74; paired WSR; fold change: A-rule/V-rule; mean ± SDs) . Taken together, these results suggest that at the group level, a general A-rule reduction in ACtx activity, both pre- and post-stimulus, may account for changes in stimulus evoked rates.

Task rule can be decoded from all layers using a PSTH pattern classifier. We next tested whether stimulus responses in one task rule can be reliably discriminated from those in the other rule using a Euclidean distance-based PSTH pattern classifier (Foffani and Moxon 2004; Malone, Scott, and Semple 2007; Hoglen et al. 2018). For each unit, this method generates single-trial test PSTHs and then compares these to two or more template PSTHs (Fig. 3.6.a). Template PSTHs are generated without the test trial, and then the test trial is assigned to the template that is closest in n-dimensional Euclidean distance space, where n is the number of bins in the PSTH. This is repeated for all trials recorded, generating new templates for each classifier run. After all trials have been classified, a confusion matrix is

generated. From this, accuracy of classification (expressed as a percentage) can be calculated, as well as mutual information (bits), and a spike-rate normalized mutual information which we refer to as encoding efficiency (bits/spk; see Methods and Fig. 3.6.a). As for FR analyses, the 0-300ms post-stimulus onset window was used in this method to restrict decoding to a predominantly sensory-driven component of the response. The binwidth for generating PSTHs was 30 ms. Only trials with correct responses (hits and correct rejects) and units with a minimum stimulus response FR of 5 Hz to at least one of the stimuli in the decoder comparison were included.

We first used this method to determine if rule identity could be decoded from PSTHs, by comparing responses to bimodal stimuli in the A-rule versus the V-rule. For all bimodal stimuli (A1V1, A1V0, A0V1 and A0V0), classification accuracy significantly above the 50% chance level for both BS and NS units (Fig. 3.6.b; A1V1 BS: 59.99 ± 12.95, NS: 61.19 ± 13.63; A1V0 BS: 60.68 ± 12.79, NS: 62.16 ± 12.45; A0V1 BS: 61.85 ± 12.48, NS: 63.71 ± 13.12; A0V0 BS: 60.60 ± 12.70, NS: 60.38 ± 12.35; mean % correct ± SD; all $p<10^{-13}$, all $z > 7.4$, one-way WSR vs chance [50%]). Using a repeated-measures ANOVA, we found no difference of stimulus or BS/NS group in decoding (stimulus [within-subject]: $F(2.59,454.65) = 0.71$, $p = 0.53$; BS/NS [between-subject]: $F(2.59,454.65) = 1.21$, $p = 0.31$; Greenhouse-Geisser-corrected). We then asked whether task rule might be differentially encoded across the cortical depth (Fig. 3.6.c). Task rule could be decoded from all depth groupings of BS and NS cells (Super. BS: 58.53 ± 9.84; Super. NS: 58.35 ± 7.06; Mid. BS: 60.51 ± 9.24; Mid. NS: 60.74 ± 10.22; Deep BS: 61.03 ± 10.69; Deep NS: 63.77 ± 11.21; mean % correct ± SD, decoding averaged across stim.; all $p<10^{-4}$, all $z > 4.3$, one-way WSR vs chance [50%]). When we compared these means with a two-way ANOVA, we found that there was a significant effect of depth ($F(2,440) = 4.7$, $p = 0.0097$) but not BS/NS $F(1,440) = 0.67$, $p = 0.41$) or their

interaction ($F(2,440) = 0.84$, $p = 0.43$). Post-hoc comparison by depth showed that decoding was significantly better in the deep group (62.4 ± 10.8, mean % correct ± SD, n = 86) than the shallow group (58.4 ± 8.6 % corr., n= 248; super. vs. deep: $p = 0.0071$, Tukey's HSD). These results show that, in aggregate, task rule modulates SU responses at all depths and BS/NS groupings in a manner that is discriminable by our PSTH-based pattern classifier, and the decoder performs particularly well on the responses from the subpopulation of deep units. Because this decoder utilizes both rate and timing, this is consistent with our finding that FR is most strongly modulated by task rule in the deep group.

Attention to sound increases information efficiency in deep units of auditory cortex. While discrimination of task rule suggests that auditory cortical responses differ under conditions of auditory versus visual attention, it does not determine if and how information processing is changing between the two rules. To address this question, we used the PSTH neural pattern classifier to achieve what is ostensibly the same behavioral goal of the animal performing the auditory rule task: discrimination of the rewarded (A1*) and unrewarded (A0*) bimodal stimuli. Our hypothesis at the outset was that during the A-rule, decoder accuracy for the A1*/A0* discrimination would increase relative to the V-rule, when the A1*/A0* discrimination is not behaviorally relevant. We used the decoder to discriminate between these stimuli in both rules, and compared accuracy, mutual information (MI) and MI rate. In both rules, the decoder performed at greater than chance accuracy for A1V1 vs A0V1 across all depth x BS/NS groups (Fig. 3.7.a; Super. BS: V-rule % correct: 69.43 ± 16.10, A-rule: 70.95 ± 12.15, n= 29; Super. NS: V-rule: 72.27 ± 16.06, A-rule: 70.92 ± 15.89, n= 32; Mid. BS: V-rule: 68.83 ± 16.51, A-rule: 68.75 ± 15.77, n= 58; Mid. NS: V-rule: 71.54 ± 14.52, A-rule: 72.35 ± 15.14, n= 27; Deep BS: V-rule: 72.81 ± 16.30, A-rule: 73.38 ± 15.37, n= 147; Deep NS: V-rule: 72.87 ± 17.32, A-rule: 72.99 ± 15.16, n= 56; all $p<2.1e-05$, all $z >$

4.2, one-way WSR vs chance [50%]). The same was true for the A1V0 vs A0V0 decode (Fig. 3.7.b; Super. BS: V-rule % correct: 69.80 ± 14.62, A-rule: 67.07 ± 14.10, n= 41; Super. NS: V-rule: 72.61 ± 13.34, A-rule: 70.36 ± 15.17, n= 38; Mid. BS: V-rule: 69.15 ± 15.51, A-rule: 67.68 ± 15.26, n= 73; Mid. NS: V-rule: 72.43 ± 14.67, A-rule: 72.95 ± 15.08, n= 31; Deep BS: V-rule: 71.65 ± 17.40, A-rule: 72.17 ± 16.53, n= 168; Deep NS: V-rule: 72.18 ± 15.29, A-rule: 71.22 ± 17.12, n= 70; all p<2.3e-06, all z > 4.7, one-way WSR vs chance [50%]). This shows that low and high TC stimulus identity can be discriminated from SU activity regardless of task rule. Do accuracy results differ across the rules? We did not find a significant difference in decoding between A-rule and V-rule for either A1* vs A0* discrimination (Fig. 3.7.a, ***A1V1 vs A0V1 comparison across rules:*** Super. BS: p = 0.38, Z = -0.88; Super. NS: p = 0.48, Z = 0.71; Mid. BS: p = 0.76, Z = 0.31; Mid. NS: p = 0.6, Z = -0.53; Deep BS: p = 0.62, Z = -0.49; Deep NS: p = 0.61, Z = 0.51; Fig. 3.7.b; ***A1V0 vs A0V0:*** Super. BS: p = 0.16, Z = 1.42; Super. NS: p = 0.14, Z = 1.46; Mid. BS: p = 0.31, Z = 1.01; Mid. NS: p = 0.48, Z = -0.71; Deep BS: p = 0.67, Z = -0.42; Deep NS: p = 0.92, Z = 0.10; paired WSR). MI is typically correlated with decoder accuracy, and we found the same patterns for these results: no significant MI difference across rules (Fig. 3.7.c, ***A1V1 vs A0V1:*** Super. BS: p = 0.48, Z = -0.70, V-rule bits: 0.25 ± 0.28, A-rule: 0.23 ± 0.18, n= 29; Super. NS: p = 0.77, Z = 0.29, V-rule: 0.28 ± 0.23, A-rule: 0.25 ± 0.25, n= 32; Mid. BS: p = 0.5, Z = 0.68, V-rule: 0.25 ± 0.27, A-rule: 0.24 ± 0.24, n= 58; Mid. NS: p = 0.65, Z = -0.46, V-rule: 0.24 ± 0.21, A-rule: 0.26 ± 0.25, n= 27; Deep BS: p = 0.72, Z = -0.36, V-rule: 0.29 ± 0.27, A-rule: 0.30 ± 0.28, n= 147; Deep NS: p = 0.26, Z = 1.12, V-rule: 0.30 ± 0.27, A-rule: 0.27 ± 0.25, n= 56; Fig. 3.7.d, ***A1V0 vs A0V0:*** Super. BS: p = 0.077, Z = 1.77, V-rule: 0.24 ± 0.24, A-rule: 0.20 ± 0.19, n= 41; Super. NS: p = 0.5, Z = 0.67, V-rule: 0.25 ± 0.23, A-rule: 0.25 ± 0.25, n= 38; Mid. BS: p = 0.42, Z = 0.80, V-rule: 0.23 ± 0.25, A-rule: 0.22 ± 0.25, n= 73; Mid. NS: p = 0.8, Z = -0.25, V-rule: 0.27 ± 0.24, A-rule: 0.27 ± 0.26, n= 31; Deep BS: p = 0.7, Z = -0.38, V-

rule: 0.30 ± 0.29, A-rule: 0.29 ± 0.28, n= 168; Deep NS: p = 0.88, Z = -0.15, V-rule: 0.27 ± 0.26, A-rule: 0.27 ± 0.28, n= 70; paired WSR; mean bits ± SDs).

Because decoder accuracy and MI tend to increase as firing rate increases (accuracy: r(3001) = 0.49, p = 2.3e-180; MI: r(3001) = 0.41, p = 1.5e-123; Pearson's correlation, all A1* vs A0* decoder runs), normalization by activity rate provides crucial insight into how efficiently spikes are being used for decoding. As such, we compared encoding efficiency across rules, which is MI normalized by the mean joint per-trial spike rate for the two stimulus responses in each decode. We found a ~20% A-rule increase in encoding efficiency of the deep BS units, but no other depth x BS/NS group was significantly modulated. This increased deep-layer efficiency was consistent for both the A1V1 vs A0V1 and A1V0 vs A0V0 decodes (Fig. 3.7.e, ***A1V1 vs A0V1:*** Super. BS: p = 0.15, Z = -1.44, V-rule bits/spk: 0.12 ± 0.11, A-rule: 0.15 ± 0.13, n= 29; Super. NS: p = 0.41, Z = -0.82, V-rule: 0.08 ± 0.07, A-rule: 0.09 ± 0.09, n= 32; Mid. BS: p = 0.79, Z = -0.27, V-rule: 0.12 ± 0.13, A-rule: 0.12 ± 0.13, n= 58; Mid. NS: p = 0.68, Z = -0.41, V-rule: 0.07 ± 0.08, A-rule: 0.11 ± 0.24, n= 27; Deep BS: p = 9.2e-05, Z = -3.91, median fold change: 1.26, V-rule: 0.11 ± 0.10, A-rule: 0.14 ± 0.15, n = 147; Deep NS: p = 0.57, Z = -0.56, V-rule: 0.05 ± 0.05, A-rule: 0.06 ± 0.06, n= 56; Fig. 3.7.f, ***A1V0 vs A0V0:*** Super. BS: p = 0.37, Z = 0.89, V-rule: 0.12 ± 0.13, A-rule: 0.11 ± 0.13, n= 41; Super. NS: p = 0.87, Z = 0.17, V-rule: 0.06 ± 0.06, A-rule: 0.08 ± 0.09, n= 38; Mid. BS: p = 0.76, Z = -0.31, V-rule: 0.09 ± 0.09, A-rule: 0.10 ± 0.09, n= 73; Mid. NS: p = 0.43, Z = -0.78, V-rule: 0.06 ± 0.08, A-rule: 0.07 ± 0.07, n= 31; Deep BS: p = 0.0086, Z = -2.63, median fold change: 1.19, V-rule: 0.13 ± 0.22, A-rule: 0.14 ± 0.13, n = 168; Deep NS: p = 0.22, Z = -1.23, V-rule: 0.05 ± 0.05, A-rule: 0.06 ± 0.06, n= 70; paired WSR; fold change: A-rule/V-rule; mean bits/spk ± SDs). Taken together, these results suggest that raw decoding accuracy is not modulated by task rule. Instead, the predominant effect is an increase in A-rule encoding

efficiency in the deep layer units, putatively caused by a reduction in A-rule FR without a concomitant reduction in decoding accuracy.

Information encoding efficiency changes are driven by suppressed units. The finding that A-rule information rate increases in the deep BS unit group, which also shows a decrease in A-rule FR, led us to explore the relationship between activity levels and PSTH pattern decoding results. Specifically, we sought to answer two questions. First, is the A-rule increase in information efficiency driven by the units with suppressed responses? Second, how do units with increased A-rule FRs perform in the decoder? We restricted this analysis to the deep BS units, and divided this group into those with increased FRs in the A-rule (Fig. 3.8.a) and those with decreased FRs (Fig. 3.8.b). Here we again decoded auditory stimulus identity (A1* vs A0*), but given the similar patterns observed for the two comparisons presented in the previous section (A1V1 vs A0V1 and A1V0 vs A0V0), we collapsed visual stimulus for clarity (A1* = A1V1 + A1V0; A0* = A0V1 + A0V0). We found that deep BS units with increased A-rule firing rates (36%; n = 65) exhibited a small but significant increase in A-rule decoding accuracy and MI (***accuracy:*** p = 0.0042, Z = -2.87, median fold change: 1.05, V-rule % correct: 73.14 ± 17.00, A-rule: 77.01 ± 14.25, n = 66; **MI:** p = 0.04, Z = -2.06, median fold change: 1.18, V-rule bits: 0.33 ± 0.30, A-rule: 0.36 ± 0.29, paired WSR; fold change: A-rule/V-rule; mean ± SDs). There was no significant change in MI rate in this group (p = 0.31, Z = 1.02, V-rule bits/spk: 0.17 ± 0.20, A-rule: 0.14 ± 0.12). For deep BS units with suppressed firing rates (64%; n = 114), there was no significant change in decoding accuracy or mutual information (***accuracy:*** p = 0.13, Z = 1.53, V-rule % correct: 71.64 ± 15.71, A-rule: 70.13 ± 15.17; **MI:** p = 0.42, Z = 0.80, V-rule bits: 0.27 ± 0.27, A-rule: 0.26 ± 0.24), but there was a 45% increase in A-rule encoding efficiency (p = 3.7e-08, Z = -5.50,

median fold change: 1.45, V-rule bits/spk: 0.09 ± 0.09, A-rule: 0.14 ± 0.13; paired WSR). These results suggest that the minority of units that do increase FR in the A-rule perform marginally better at decoding the auditory stimulus, and that the units that decrease FR do not lose accuracy, but instead show strongly increased information efficiency.

Receptive fields mapped during the inter-trial interval are modulated by task-rule. In the inter-trial interval, we presented an auditory receptive field mapping stimulus to test whether switching between task rules affects both the encoding of task-relevant stimuli and non-task relevant stimuli. We used the random-double sweep (RDS) stimulus, which consists of two uncorrelated random frequency sweeps (Gourévitch et al. 2015), ranging between 4 and 64 kHz. Randomly varied segments of RDS stimuli (ranging from ~1 to 15 s, median 3.0 s ± 2.1 SD) were presented between all trials in the task. SU responses to the RDS stimuli were then separated into those occurring the A-rule and those in the V-rule. Reverse correlation was then used to generate two spectrotemporal receptive fields (STRF) for each neuron: one from RDS stimuli presented during the A-rule and one from the V-rule. The mean-subtracted STRF, conventionally plotted on a reversed time axis, shows the time-frequency fields that both excite (Fig 3.9.a, red fields) and inhibit (blue fields) spiking activity. As expected, this procedure did not yield STRFs from all recorded units. To remove STRFs whose fields could not be measured reliably, we filtered for significant STRF trial-to-trial similarity relative to a null STRF ($p<0.01$; see Methods) in at least one condition (A-rule or V-rule). This yielded 277 BS units and 82 NS units STRFs for subsequent analysis.

Previous research has shown that neurons in the auditory cortex can rapidly alter receptive fields between periods of task engagement and passive listening (Fritz et al. 2003; Yin, Fritz, and Shamma 2014). We sought to test whether receptive field properties are

similarly altered between conditions of auditory attention and visual attention. First, we measured the characteristic frequency (CF) of significant STRF excitatory (E) and inhibitory (I) fields during the two task rules. To achieve this, we determined significant regions of the STRF by comparing observed pixel values to the distribution of pixel values from a null STRF, derived from a time-shifted spike train. STRF pixels with absolute values beyond the 1st percentile ($p<0.01$) were considered significant. To measure CF, we summed the significant E and I fields separately along the frequency axis. The peak absolute value of each sum was considered the CF, calculated separately for the E and I fields (Fig. 3.9.c). We tested whether CFs change between the A- and V-rules and found no significant difference between the distributions for the E CFs (Fig. 3.9.d; BS units: $p = 0.58$, KS stat = 0.19; NS units: $p = 0.07$, KS = 0.31; Kolmogorov-Smirnov [KS] test) or I CFs (BS units: $p = 0.38$, KS = 0.22; NS units: $p = 0.23$, KS = 0.25; KS test). When CF distributions were divided between mice for which the go auditory stimulus was the 8 kHz TC and those for which go was 17 kHz TC (Fig. 3.9.d), we found a significant difference between the distributions of the E field CFs for BS units (BS units: $p = 0.035$, K-S stat = 0.31; NS: $p = 0.43$, K-S stat = 0.16 ; one-tailed KS test), but not I field CFs for either unit type (BS: $p = 0.87$, K-S stat = 0.06; NS: $p = 0.43$, K-S stat = 0.16; one-tailed KS test). E field CFs for the 8 kHz TC-rewarded mice tended to lie within (or at the upper edge) of the 8 kHz TC frequency band. A similar preference for the rewarded TC frequency band was observed for CFs from the 17 kHz TC-rewarded mice.

The classical STRF represents a linear fit between the spectrotemporal properties of a sound stimulus and the spike train it evokes, but auditory cortical neurons are known to vary widely in their response linearity (Atencio, Sharpee, and Schreiner 2008). To better characterize the relationship between the observed reduction in stimulus activity and

information processing, we calculated a nonlinear input/output function for each A-rule and V-rule STRF. The nonlinearity represents the neuronal response as a function of the similarity between the STRF and the stimulus from which it is derived (Fig. 3.10.a; see Methods). It is generated by comparing the distribution of STRF-stimulus projections (similarity values) from stimulus segments that elicited a spike ($P(x|spike)$) to a distribution of STRF-stimulus projections using randomly chosen stimulus segments ($P(x)$). These distributions also form the basis for calculation of a spike-rate normalized mutual information (encoding efficiency) measure between the STRF and the spike train:

$$MI = \sum P(x|spike) log_2(\frac{P(x|spike)}{P(x)})$$

We compared this encoding efficiency measure derived from the inter-trial interval across rules by depth and BS/NS classification. There was no difference between encoding efficiency in the superficial or middle BS/NS groups, or the deep NS group. Instead, consistent with our earlier findings for decision stimuli, encoding efficiency shows a small but significant increase in the deep broad-spiking subpopulation (Fig. 3.10.b; Super. BS: p = 0.56, Z = 0.59, V-rule bits/spk: 0.15 ± 0.15, A-rule: 0.15 ± 0.14, n= 20; Super. NS: p = 0.21, Z = -1.24, V-rule: 0.03 ± 0.03, A-rule: 0.04 ± 0.04, n= 16; Mid. BS: p = 0.72, Z = 0.36, V-rule: 0.06 ± 0.07, A-rule: 0.06 ± 0.08, n= 52; Mid. NS: p = 0.52, Z = -0.64, V-rule: 0.04 ± 0.05, A-rule: 0.04 ± 0.04, n= 18; Deep BS: p = 0.042, Z = -2.04, median fold change: 1.11, V-rule: 0.10 ± 0.17, A-rule: 0.11 ± 0.18, n = 79; Deep NS: p = 0.75, Z = -0.32, V-rule: 0.02 ± 0.02, A-rule: 0.02 ± 0.04, n= 26; paired WSR; fold change: A-rule/V-rule; mean bits/spk ± SDs). This measure of MI rate is different from that used for decision stimuli; the current measure shows mutual dependence between the STRF and the spike train, while that for decision stimuli shows mutual dependence between stimulus identity classification and the binned spike

train. Nevertheless, both metrics indicate a modulation of information consistent in sign and specific to BS cells from the deep layers of cortex.

To better understand the shift in MI rate, we analyzed two properties of STRF empirical nonlinearity: the asymmetry index (ASI) and the skewness. ASI quantifies the difference in likelihood of spiking for positive and negative values of the nonlinearity, defined as $(R-L)/(R+L)$ where R is the sum of positive nonlinearity values (right of 0), and L is the sum of negative values. ASI values range from -1 to 1, and values close to 1 indicate that spiking activity only increases when the stimulus is highly similar to the STRF (Atencio and Schreiner 2012). Skewness quantifies the amount of information in the tail of the nonlinearity (see Fig. 3.10.c for examples). It describes how quickly the spike rate changes with increasing STRF-stimulus similarity. High skewness values indicate that a cell's activity is restricted to instances when the STRF closely matches the structure of the stimulus. Given the trend of decreased FR and increased encoding efficiency in the A-rule, we hypothesized that these measures would increase in the A-rule. We compared ASIs calculated from A-rule and V-rule STRFs and found that during the A-rule, asymmetry increases slightly in the deep BS subpopulation (Fig. 3.10.d.; Super. BS: p = 0.63, Z = -0.32, V-rule: 0.65 ± 0.27, A-rule: 0.63 ± 0.28, n = 42; Super. NS: p = 0.29, Z = 0.56, V-rule: 0.42 ± 0.29, A-rule: 0.43 ± 0.32, n = 23; Mid. BS: p = 0.26, Z = 0.63, V-rule: 0.53 ± 0.31, A-rule: 0.55 ± 0.28, n = 83; Mid. NS: p = 0.056, Z = 1.59, V-rule: 0.39 ± 0.27, A-rule: 0.45 ± 0.26, n = 22; Deep BS: p = 0.025, Z = 1.96, mean ASI difference [A-V] = 0.06, V-rule: 0.56 ± 0.29, A-rule: 0.63 ± 0.29, n = 152; Deep NS: p = 0.53, Z = -0.09, V-rule: 0.40 ± 0.30, A-rule: 0.40 ± 0.22, n = 37; paired WSR; mean ASI ± SDs). We found no significant change in ASI in any of the NS depth groups. Similarly, when we compared skewness across rules, we found that only the deep BS group exhibited a difference (Fig. 3.10.e.; Super. BS: p = 0.33, Z = 0.45, V-rule: 1.36 ±

1.15, A-rule: 1.42 ± 1.05, n = 42; Super. NS: p = 0.94, Z = -1.60, V-rule: 0.99 ± 1.03, A-rule: 0.59 ± 0.97, n = 23; Mid. BS: p = 0.079, Z = 1.41, V-rule: 0.86 ± 1.06, A-rule: 1.03 ± 1.15, n = 83; Mid. NS: p = 0.2, Z = 0.84, V-rule: 0.82 ± 1.06, A-rule: 1.08 ± 1.36, n = 22; Deep BS: p = 0.014, Z = 2.20, mean skew difference [A-V] = 0.31, V-rule: 1.05 ± 1.07, A-rule: 1.38 ± 1.18, n = 152; Deep NS: p = 0.69, Z = -0.50, V-rule: 0.61 ± 1.21, A-rule: 0.50 ± 1.15, n = 37; paired WSR; mean skew ± SDs). These measures of the bias in the nonlinearity provide a link between the observed reduction in firing rate and the change in information processing deep in the cortex; the V-rule tends to elicit more spiking activity during the RDS that is not well described by the STRF, while activity during the RDS in the A-rule is more restricted to moments when the STRF matches the stimulus closely.

3.4 Discussion

Here we presented a novel audiovisual rule-switching behavior task in which mice had to either attend to and discriminate sound stimuli while ignoring a visual stimulus, or ignore sound stimuli while discriminating a visual stimulus. During the task, we recorded 873 single units (SUs) from auditory cortex (ACtx) using a multisite probe to span the cortical laminae. We compared SU responses to identical audiovisual stimuli in the auditory rule (A-rule) of the task versus the visual rule (V-rule). On average, deep and middle layer responses to audiovisual task-relevant stimuli were suppressed in the A-rule relative to the V-rule. This suppression was evident not only during task-relevant stimuli, but also for inter-trial interval receptive field mapping stimuli. The effect of suppression across rules was abolished when adjusting for baseline activity, which also showed an A-rule decrease. This suggests that, on the whole, the auditory selective attention required by our task quiets the auditory cortex, generally reducing the background activity over which cortical sensory responses occur. We then used a PSTH-based pattern classifier to determine how

effectively ACtx sensory responses can be used to decode task rule, and whether task rule changes the amount of information present in spike trains. While a reduction in decoding might be expected from decreased activity levels, we observed no such change: despite reduced spike count, both decoding accuracy and mutual information were preserved, and, as a result, encoding efficiency in bits per spike increased during the A-rule. This effect was restricted to the deep broad-spiking cells, suggesting that attentional modulation varies by cortical depth. Analysis of spectrotemporal receptive fields (STRFs) recorded during the inter-trial interval confirmed the increased A-rule information rate in the deep broad-spiking cells. We also found, through analysis of STRF nonlinearities, that activity levels are better described by the linear STRF in the A-rule than in the V-rule, providing a link between the reduction in activity levels and the changes in stimulus encoding.

Past research, largely from extrastriate visual cortex, has suggested that attention often increases neural response gain to a space or feature (Moran and Desimone 1985; Desimone and Duncan 1995; Reynolds and Chelazzi 2004). Increasing gain should in turn increase the reliability of the sensory cortical readout. Another putative mechanism for attentional enhancement is the reduction of noise, which has been shown to occur at the level of single neurons (Mitchell, Sundberg, and Reynolds 2007) and in correlations at the population level (Cohen and Maunsell 2009; Downer, Niwa, and Sutter 2015). Our finding that auditory attention increases encoding efficiency and STRF-stimulus selectivity suggest mechanisms of noise reduction. Previous work comparing ACtx responses during task engagement and passive listening has shown that engagement increases signal-to-noise ratios through decreased pre-stimulus rates without a change in sensory-evoked rates (Buran, von Trapp, and Sanes 2014). At the group level, we did not observe this phenomenon, finding that accounting for pre-stimulus rates generally abolished the

group-level firing rate change across rules. However, the timing of activity in ACtx is known to carry substantial information (Krishna and Semple 2000; Malone, Scott, and Semple 2007; Hoglen et al. 2018), which would not be captured by coarse estimations of rate. By accounting for temporal patterns at a finer scale using a PSTH-based pattern classifier, we were able to uncover another mechanism for attentional modulation: increased efficiency of information encoding.

Of course, mechanisms of gain increase and noise reduction likely act in tandem. For example, in human ACtx response gain is increased for attended signals and decreased for unattended signals (Hillyard et al. 1973; Giard et al. 2000; Bidet-Caulet et al. 2007). The latter effect may be considered a form of noise reduction through decreased representation of competing but behaviorally irrelevant stimuli. While in our study, the predominant effect of auditory attention was to reduce spike counts, in all depth and excitatory/inhibitory subgroups we did find that a fraction of units significantly increased their firing rates. Even in the deep putative-excitatory group, where we saw the most robust activity suppression, ~⅓ of units increased their activity in the A-rule. This result is in line with several recent mouse ACtx SU studies (Carcea, Insanally, and Froemke 2017; Kuchibhotla et al. 2017), suggesting that a similar fraction of units show increased activity during task engagement when compared to passive listening. Past research on selective attention in monkey visual cortex shows that stimulus-evoked responses for attended features are generally enhanced but can also be suppressed for features outside of the receptive field (Reynolds and Chelazzi 2004). As such, one might be tempted to conclude that the smaller fraction of cells with attentionally-increased activity is simply the group that is selectively tuned for task-relevant features. Our finding that units typically show the same sign of modulation for both target (A1*) and distractor (A0*) stimuli (79% of broad-

spiking units; 96% of narrow-spiking units) complicates this interpretation, suggesting enhancement or suppression is not selective for at least the features that distinguish these stimuli. When we divided the decoding analysis between units with increased and decreased FRs in the auditory rule (FR+, FR-; Fig. 3.8), we found that the FR+ group showed increased decoder accuracy and mutual information in the auditory rule, while the FR- group showed strongly increased encoding efficiency. This suggests that the effects of attention on ACtx are nuanced and unlikely to be fully captured by analyses of coarse single-unit groupings. For example, downstream stations might "listen" more to the smaller share of FR+ neurons with increased stimulus information, while the FR- neurons reduce activity that does not represent the stimulus. Such hypotheses are in line with "efficient pooling" mechanism of attention, in which downstream stations selectively utilize the inputs that optimally discriminate behaviorally-relevant features (Serences 2011; Pestilli et al. 2011), and need to be further explored in the context of ACtx attentional mechanisms.

We used the PSTH-based pattern decoder for two purposes: first, we sought to determine whether task rule could be decoded from responses and found strong evidence for this at all depth and BS/NS groups studied. Second, we sought to determine whether information processing in ACtx changes between the auditory and the visual rule. The decoder was used to discriminate between spike trains evoked by target (A1*) and distractor (A0*) stimuli, tone clouds centered around 8 kHz or 17 kHz, ostensibly the same task faced by the mice in our task during the A-rule. These highly discriminable stimuli were chosen intentionally to decrease the difficulty of a cognitively demanding task, but this presents a caveat: as we expected, ACtx neurons often respond with highly distinct patterns to these different tone clouds (e.g., Fig. 3.3). While our results suggest that changes in information processing do occur between the task rules, one must consider

whether observed changes would be different when comparing stimuli that evoke less inherently distinguishable ACtx responses. We speculate that decoding of responses to more similar stimuli would yield more pronounced effects between attended and ignored states.

Several studies have addressed differences between task-engaged and passive sensory processing in the ACtx at the SU level. Even though our study explicitly compares two task-engaged states, this literature may provide the most relevant background for understanding our results, and many of our findings are consistent with this work. Engagement in a behavioral task is known to modulate correlations at the population level (Downer, Niwa, and Sutter 2015; Downer et al. 2017) as well as stimulus-evoked and pre-stimulus single-unit activity. Several groups have reported that mouse and ferret ACtx stimulus-evoked spiking responses are predominantly suppressed during self-initiated task engagement when compared to passive listening (Otazu et al. 2009; Kuchibhotla et al. 2017; Carcea, Insanally, and Froemke 2017; Bagur et al. 2018). Mouse and gerbil ACtx activity levels preceding a stimulus have also been shown to decrease during self-initiated tasks in several studies (Buran, von Trapp, and Sanes 2014; Carcea, Insanally, and Froemke 2017), but not others (Otazu et al. 2009). A similar finding of reduced pre-stimulus activity during task engagement has been observed in rat gustatory cortex, leading to sharpened taste selectivity and reduced responses to suboptimal stimuli (Yoshida and Katz 2011). Recent work has shown pre-stimulus suppression of activity in monkey primary visual cortex immediately after a task cue (Cox et al. 2019), a finding consistent with earlier studies in monkey extrastriate areas (Sato and Schall 2001; Bisley and Goldberg 2006; Herrington and Assad 2010). In the present study, we show a median 10-20% reduction in SU spiking activity in the period immediately prior to the decision stimulus onset when comparing A-

rule to V-rule. As in the above cited studies, the pre-stimulus period we use should be considered an anticipatory period, as it immediately follows the offset of the inter-trial interval RDS stimulus, which reliably predicts imminent presentation of a decision stimulus (Fig. 3.1.c). Unlike the studies cited above, we are comparing rates across two states of task-engagement. Our finding suggests that the task-engaged reduction of pre-stimulus activity may not be a cortical-wide phenomenon that occurs during any period of expectation, but instead may represent a mechanism of attentional highlighting specific to the sensory system used in the task (here, the auditory cortex during auditory engagement). Nevertheless, due to task design we cannot rule out the possibility that there is also a reduction in pre-stimulus ACtx activity during the anticipation of stimuli in the visual rule, which might be revealed when compared to a passive condition.

How might suppression of both stimulus-evoked and spontaneous activity subserve attention? As others have noted (Otazu et al. 2009), it seems counterintuitive that a process whose goal is perceptual highlighting would decrease responses to behaviorally-relevant stimuli. The work presented here shows that decreasing activity does not necessarily decrease stimulus information. Speculatively, this may occur by removal of the noise background on which external stimulus-encoding activity sits, a proposition in line with an influential theory of attention. This theory (Harris and Thiele 2011) posits that spontaneous fluctuations in cortical activity partly reflect internal processes such as mental imagery or memory recall. These processes are contrasted with activity that arises from external sensory stimulation. One potential mechanism of attention is the suppression of internally-generated spontaneous activity to favor the processing of behaviorally-relevant external stimulation. Such a theory is consistent with attentional suppression which tends to affect both pre-stimulus and stimulus-evoked activity. It is also

consistent with our finding that suppression of activity does not degrade stimulus-spike train decoding accuracy or mutual information, and instead increases stimulus encoding efficiency.

One way that the present study extends the current literature is the focus of this work on analysis by cortical depth. Our use of multisite probes to span the cortical laminae allowed us to group units in an approximation of supragranular, granular and infragranular divisions. These divisions are likely to be highly relevant to attentional modulation. Such modulatory feedback signals are traditionally thought to arrive in the supragranular and infragranular layers (Felleman and Van Essen 1991), and depth divisions in task-related columnar modulation have been found by groups examining current-source density (Zempeltzi et al. 2020) or multi-unit activity (O'Connell et al. 2014). We find that auditory attention both suppressed firing rates (middle and deep groups) and increased encoding efficiency (deep group) in a depth-dependent manner. Recent findings on the columnar circuitry involving L2/3 (superficial) and L5 (deep) cells suggest a possible cortical circuit at play. Pluta *et al.* showed that optogenetic activation of L2/3 both suppressed the activity and increased the stimulus selectivity of L5 pyramidal neurons, mediated by somatostatin-positive interneurons (Pluta et al. 2019). As the main cortical output layer, shifts in information processing in the infragranular population would have ramifications at subcortical sites, as well as other cortical regions through its corticocortical connectivity (Salin and Bullier 1995).

The present work should also be considered in the context of previous rule-switching behavioral tasks. This research, in both humans and non-human animals, has implicated frontal, parietal and striatal networks in rule switching and flexibility (Toth and Assad 2002; Roshan Cools, Clark, and Robbins 2004; Rougier et al. 2005; Crone et al. 2006;

R. Cools, Ivry, and D'Esposito 2006; Licata et al. 2017; Wimmer et al. 2015; Rikhye, Gilra, and Halassa 2018). These networks may act as a context-dependent switch, routing attentional modulatory feedback to the appropriate sensory systems. Because our focus was on the role of task rule in modulating sensory processing rather than the executive function of rule-switching, here we took care to present physically identical audiovisual stimuli across the rules of the task. This allowed us to isolate the effects of task rule on stimulus processing, demonstrating that attending to sound suppresses both baseline and stimulus-evoked activity and increases encoding efficiency in a cortical depth-dependent manner.

3.5 Figures

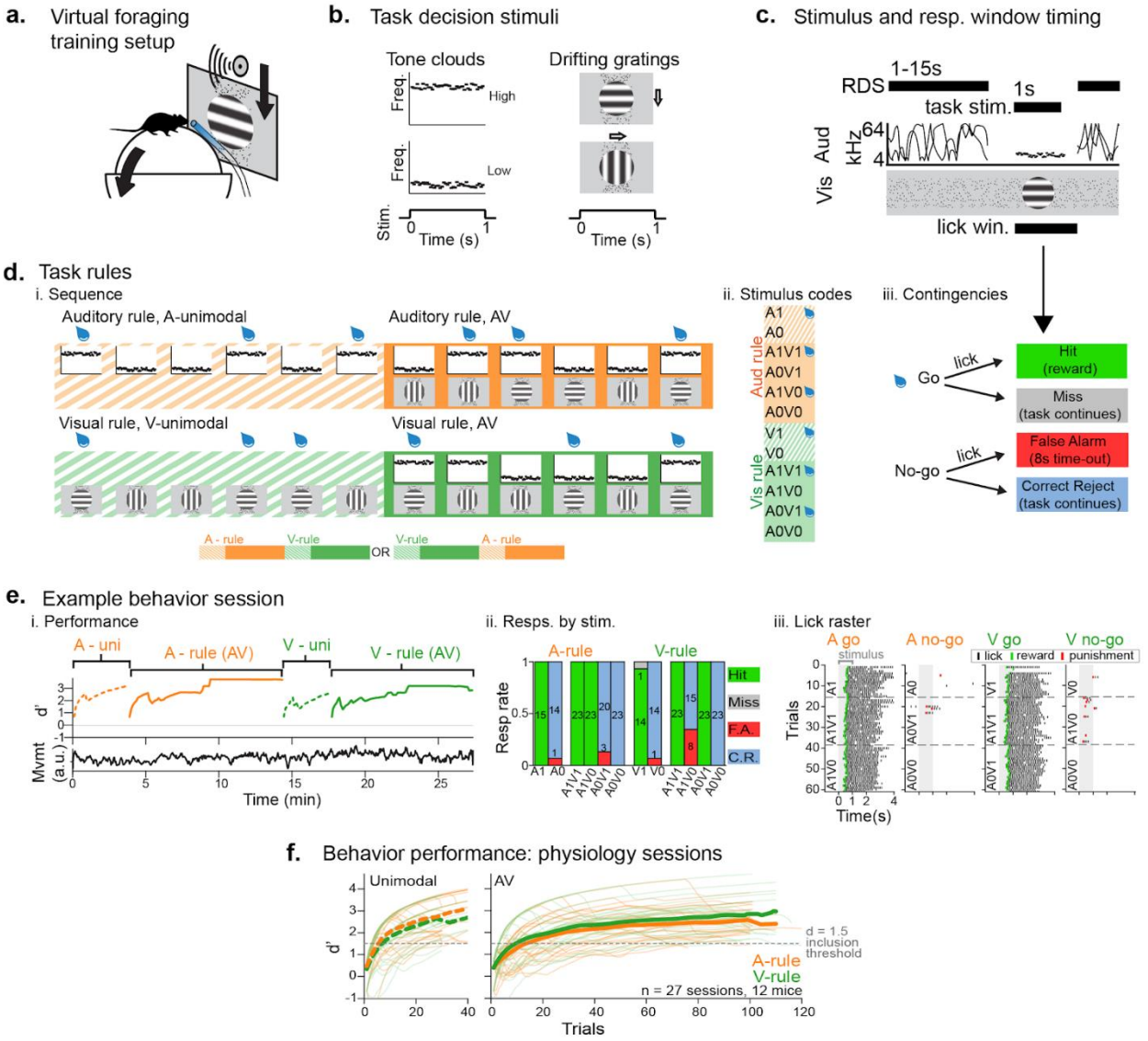

Figure 3.1 A novel audiovisual rule-switching task. a. Virtual foraging training setup: a head-fixed mouse runs on a floating trackball. Locomotion measured by trackball movement controls auditory and visual stimulus presentation. A water spout in front of the mouse provides reward administration, and an attached lickometer records licks, which determine reward or punishment administration. **b.** Decision stimuli are either 1 s tone-clouds (TC; 8-kHz centered or 17 kHz-centered) or drifting gratings (horizontal or vertical orientation). **c.** In-task stimulus sequence. Mice run while a 'track' of moving dots provides visual feedback. During the inter-trial track period, a random double sweep (RDS) auditory receptive field mapping stimulus is also presented. This stimulus has no behavioral task relevance. When the RDS stimulus ends, there is a brief pause in auditory stimulation, followed by presentation of a go or no-go stimulus. The animal can make a behavioral choice (lick) anytime after decision stimulus onset, although trials with early licks (>0.3 s post-stimulus onset) were removed from subsequent analysis. **d.i.** Behavioral sessions

begin with a unimodal block (in this example: A-only, hatched orange hatched background; typically ~30 trs). Water drops represent go stimuli, during which mice have an opportunity for reward. Next an AV block is presented (solid orange; typically 120 trs), during which the mouse must continue to use the A-rule from the previous unimodal block. Next, a V-only block is presented (hatched green; ~30 trs), followed by the final AV block, in which the mouse continues to use the V-rule (solid green; ~120 trials). Each mouse is trained to always lick for only one auditory stimulus (8kHz or 17kHz TC) and only one visual stimulus (vertical or horizontal), and go stimulus identity was counterbalanced across mice. Task order sometimes begins with the A-rule, and sometimes begins with the V-rule. **d.ii.** Stimulus codes, for reference. **d.iii.** Contingencies for reward (typically 0.006 mL water), punishment (time-out, with a darkened screen for 4-8 s), or task continuation. **e.i.** Performance on an example session, showing sensitivity index (d') calculated by task block. Mouse locomotion is shown below. **e.ii.** Response rate by stimulus code for same example session. (Colors as in d.iii). **e.iii.** Stimulus-onset aligned lick rasters for example session, organized by rule and go/no-go. Green ticks indicate water rewards, while red ticks indicate time-out punishments. **f.** Behavioral performance (moving window d', log-linear corrected, see Methods) for all sessions used in physiology analysis (n = 27 sessions from 12 mice).

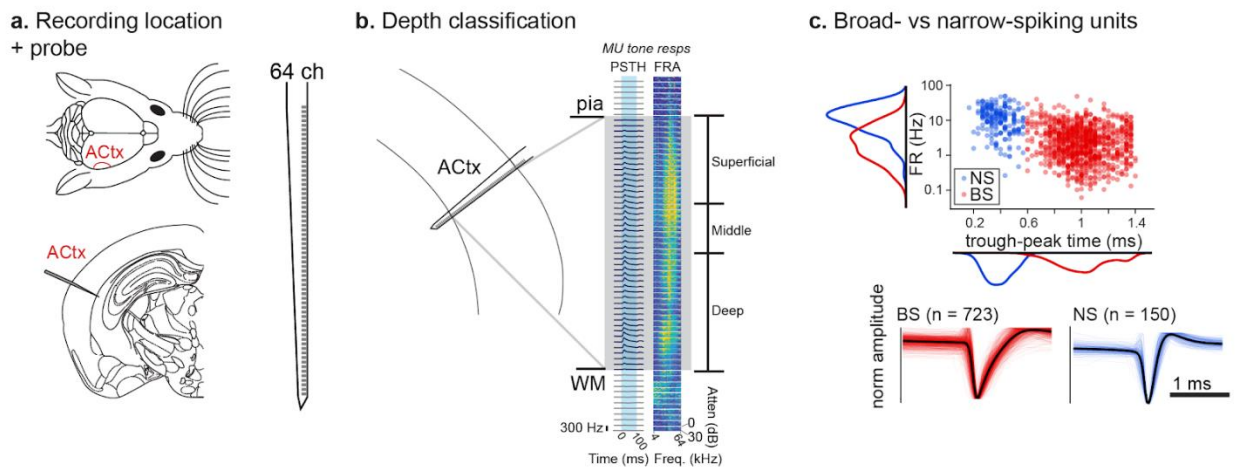

Figure 3.2 In-task extracellular physiology and measurement of recording site depth. a. After mice achieved task-expertise, a craniotomy was made over the right auditory cortex (ACtx; red), and acute extracellular recordings were made daily for up to 6 days. Right, multi-site probe to record from all layers of auditory cortex. **b.** Multi-unit event responses to 0.1 s pure tones were recorded after each behavioral session to confirm that recording site was in ACtx and to determine depth of the probe within cortex. The topmost and bottommost channels with any significant MU tone response were considered to demarcate the putative pia-WM span, and channels were categorized as 'Superficial', 'Middle', or 'Deep'. Single-unit (SU) waveforms were typically recorded across multiple channels, so SU depth was assigned based on the channel with the largest amplitude deflections. **c.** After spike sorting, SU waveforms were divided into narrow-spiking (putative fast-spiking inhibitory) and broad-spiking (putative excitatory) populations based on a waveform trough-peak time boundary of 0.6 ms.

Figure 3.3 Varied single-unit responses to audiovisual decision stimuli across task rules. **a.** Seven example broad-spiking (BS) units (**i-vii**) recorded during the behavioral task, organized by depth category. Each row of raster plots (top) and corresponding PSTHs (bottom) represents responses to indicated stimuli in the visual rule (orange) and auditory rule (green). Note that for each indicated stimulus, tone-clouds and grating orientations are physically identical across rules. Stimulus codes correspond to those in Fig. 3.1.d; A1V1 is rewarded in both A- and V-rule; A1V0 is rewarded in A- but unrewarded in V-rule, etc. Some units are minimally modulated by rule, while others show clear firing rate modulation. While many units respond strongly prefer A1* stimuli (includes A1V1 and A1V0; **ii, iii, v, vi**), others exhibit strong response preference for A0* stimuli (collapsed A0V1 and

A0V0; **vii**) . Some respond robustly to both A1* and A0* stimuli, but may be modulated by task rule (**ii, iv**). **b.** Six example narrow-spiking (NS) units (**i-vi**). Conventions as in **a.** As with BS units, NS units exhibit a variety of response PSTH profiles. Many are highly modulated by task rule (**ii-v**), while others are minimally modulated (**i, vi**). Consistent with putative classification as fast-spiking interneurons, NS units frequently exhibit high sustained firing rates.

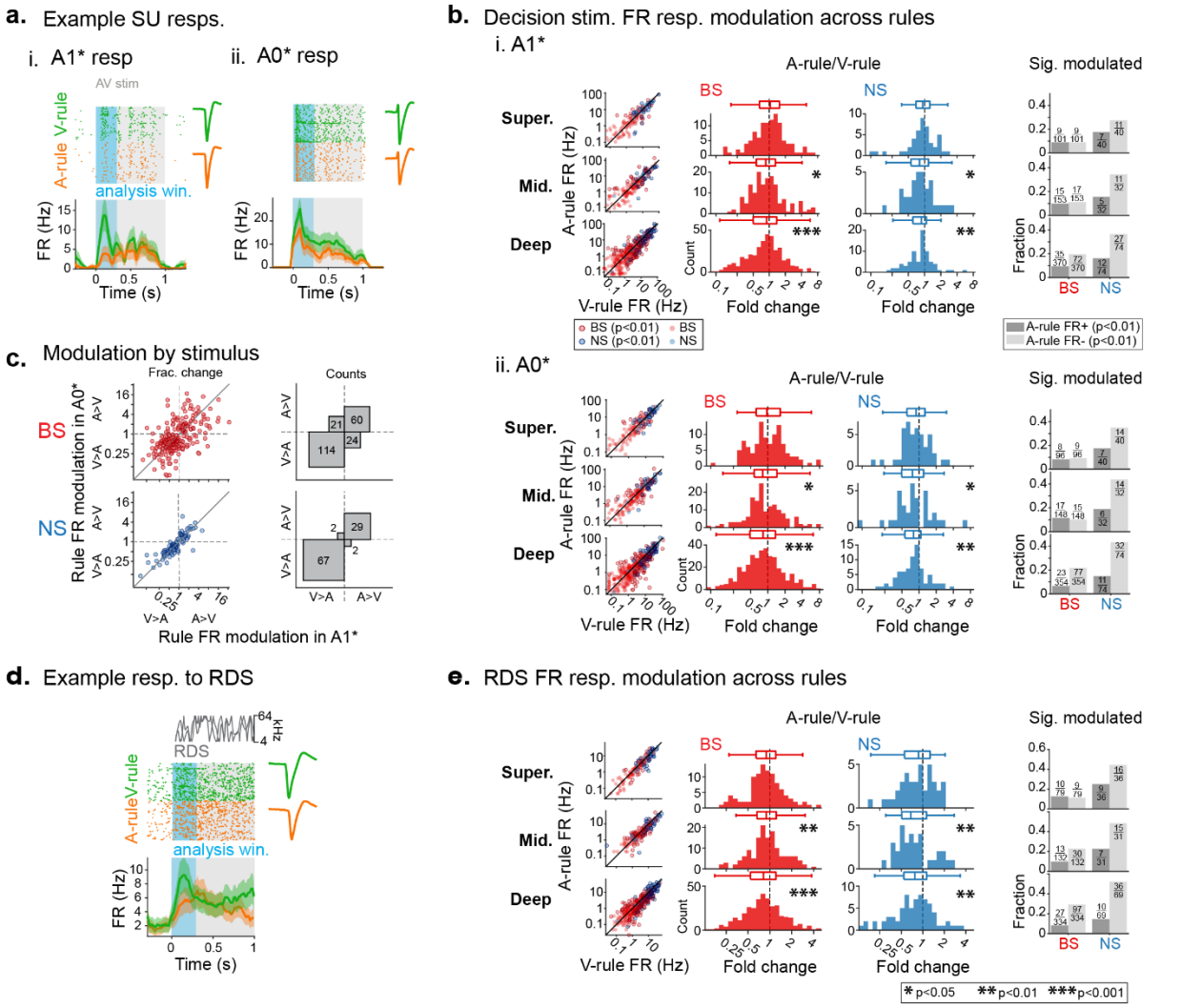

Figure 3.4 Stimulus-evoked firing rates are generally suppressed during the auditory rule compared to the visual rule. a. Responses to physically identical AV stimuli across the two task rules. Blue shaded window represents the analysis period of 0-0.3 s for calculating firing rate (FR), chosen to capture a predominantly sensory-driven component of the response. **i.** Example unit response to A1* stimuli (A1V1 and A1V0, collapsed) across rules. **ii.** Different example unit response to A0* stimuli (A0V1 and A0V0, collapsed) across rules. **b. i.** Group data showing A1* FR response across rules by depth grouping. A1* stimuli are those that are rewarded in A-rule. Scatter plots (left) show SU V-rule FR on abscissa and A-rule FR on ordinate axes. BS units shown in red (red outline: significantly modulated BS, $p<0.01$; paired t-test). NS units shown in blue (blue outline: significantly modulated NS, $p<0.01$; paired t-test). Fold change histograms show A-rule FR divided by V-rule FR for all units shown in scatters; bins to the left of 1 (dashed line) indicate suppression in the A-rule. Histograms in red show BS units; blue show NS units. Box plots above histograms: central line indicates median, box edges indicate 25th and 75th percentiles, whiskers extend to data points not considered outliers. Right: fractions of significantly modulated units ($p<0.01$) over total, by depth group and NS/BS classification. Dark grey indicates units with significantly increased FRs in the A-rule compared to V-rule, and light grey indicates significantly suppressed FRs in A-rule. **ii.** As in **i.**, but for A0* (unrewarded in A-rule)

responses. **c.** Comparison of unit FR rule modulation between A1* (abscissa) and A0* (ordinate). Top: BS units, bottom: NS units. Scatter plots (left) show all overlapping units included in **b.i** and **b.ii.** As in fold change histograms (**b.**), modulation values >1 indicate greater FR response in A-rule, while <1 indicates greater FR response in V-rule. Note the increased density of units below 1 on both axes. Right: counts of units by direction of FR rule modulation by stimulus, including units with significant FR changes across rule for A1*, A0* or both. The bulk of units lie in lower left and upper right quadrants, indicating similar direction of modulation across stimuli. **d.** Example unit showing firing rate modulation to the onset of the random double sweep (RDS) stimulus, a task-irrelevant RF mapping stimulus. The same analysis window for calculating FR was used: 0-0.3 s (blue). **e.** Group data for RDS FR modulation across rules by depth and BS/NS classifications. Conventions as described in **b.**

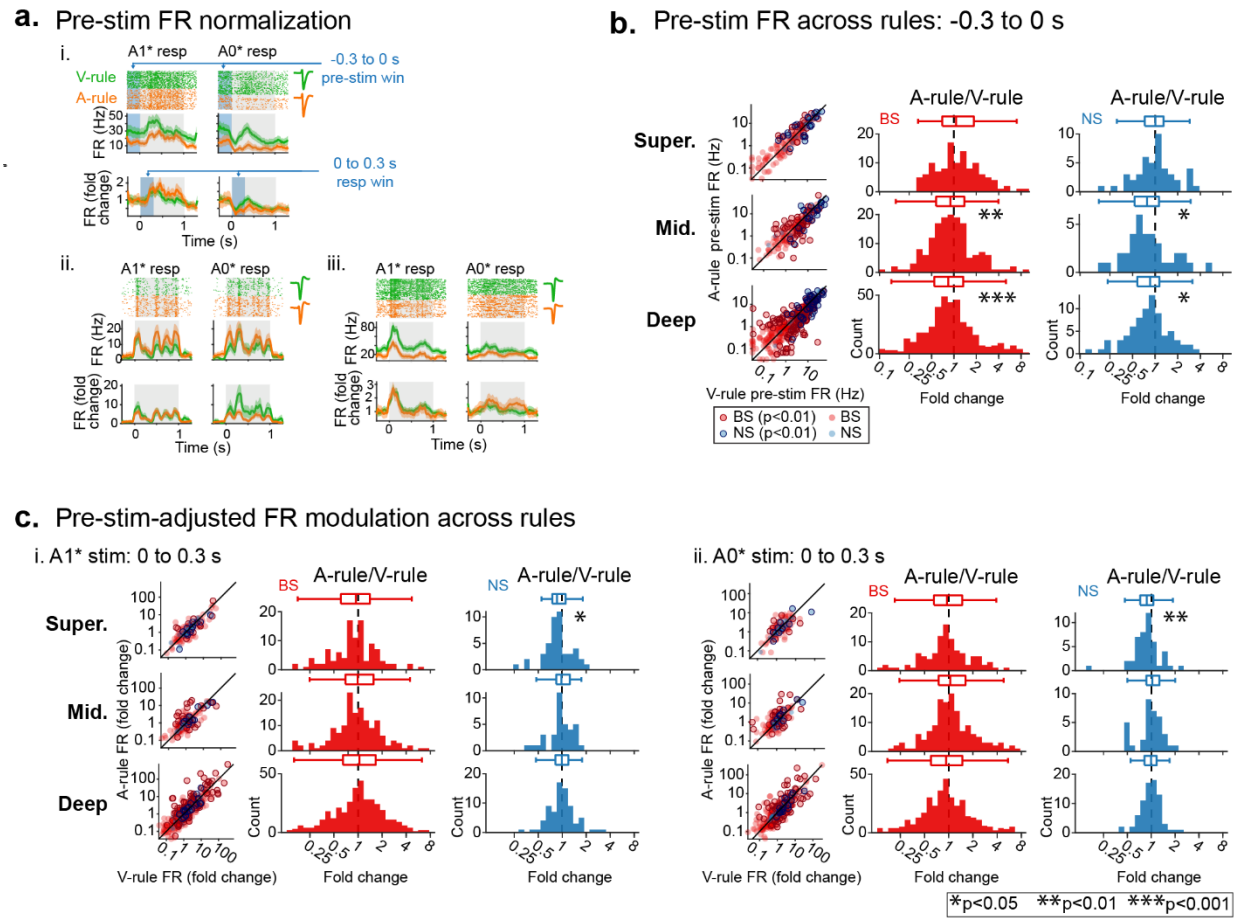

Figure 3.5 Modulation in pre-stimulus firing rates partially accounts for changes in stimulus-evoked firing rates across rules. a. Pre-stimulus (-0.3 to 0) onset firing rate measurement, and normalization of FR response. **i.** Example broad-spiking unit, showing raw FR (upper PSTH) and fold-change FR (lower PSTH), which is the raw FR divided by pre-stimulus FR. Pre-stimulus FR window (group data shown in **b.**, upper raster, PSTH) and post-stimulus FR window (group data shown in **c.**, lower PSTH) marked in blue. **ii.** Example superficial NS unit showing increased V-rule FR response after normalization. **iii.** Example NS unit showing similar FR responses across rules after normalization. **b.** Pre-stimulus onset FR compared across rules, with data organized by depth and BS/NS (red/blue). Conventions as in Fig. 3.4. Scatter plots (left) show individual units, with significantly modulated units outlined ($p<0.01$; paired t-test). Difference histograms show A-rule/V-rule for all units shown in scatters; fold change <1 indicates suppression during the A-rule. Note that data comes from -0.3 to 0 s preceding all stimuli in the bimodal (but not unimodal) portions of each rule. **c.** AV stimulus FR response as fold change, normalized by pre-stimulus FR. **i.** A1* fold change responses across rules, conventions as above. **ii.** A0* fold change responses, conventions as above.

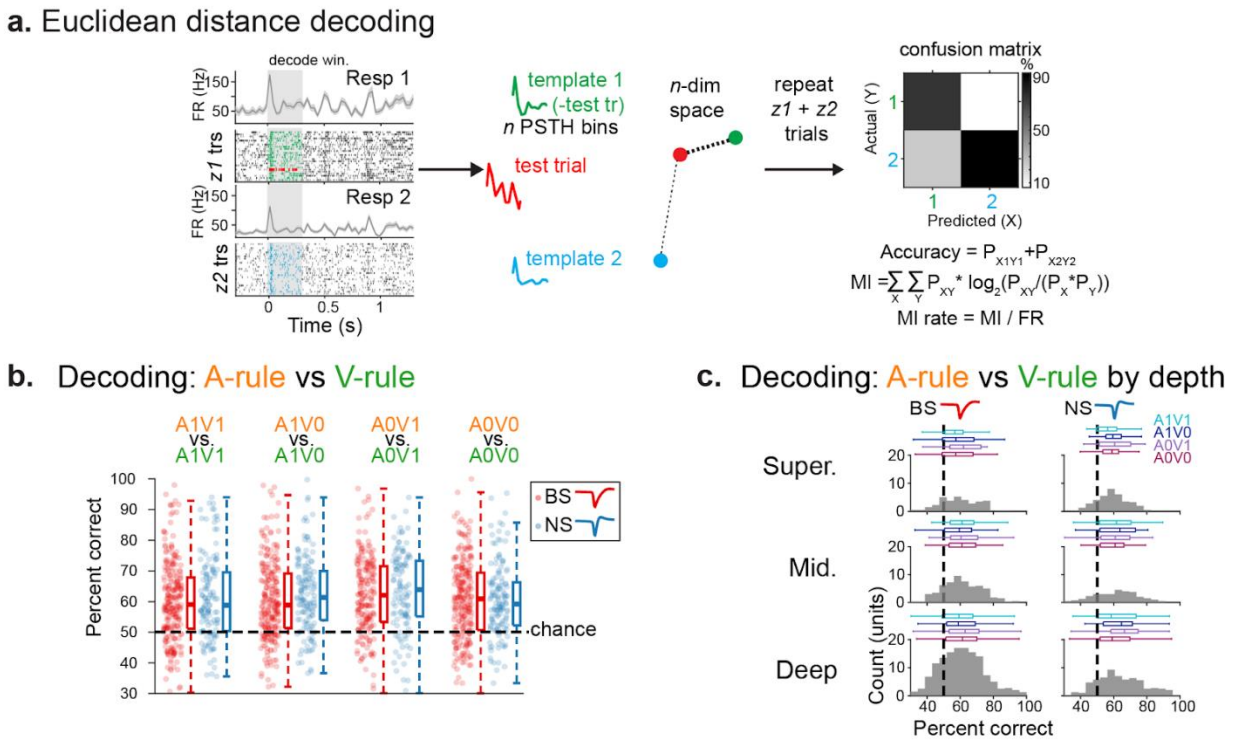

Figure 3.6 Task rule can be decoded from audiovisual SU responses in auditory cortex. a. PSTH-based decoding of stimulus identity. We use this method to compare two response PSTHs from the same unit (Resp 1, Resp 2). The decoder is run *z1*+*z2* times, where *z1* = nr. of trials in PSTH for Resp 1 and *z2* = nr. of trials in PSTH for Resp 2. For each decoder run, a trial is removed; this is the test trial (red). Template PSTHs for the two responses are generated with the exclusion of the test trial. The test trial is classified as belonging to the template that is closest in Euclidean-distance *n*-dimensional space, where *n* = number of bins in PSTHs. Based on classifications from all runs, a confusion matrix (right) is generated, and accuracy, mutual information (MI) and encoding efficiency (MI rate, bits/spk) are calculated as indicated. **b.** Decoding of rule identity. Responses to each of the four AV stimuli in the A-rule vs responses to the same stimuli in the V-rule were decoded based on 0-0.3 s stimulus onset window; PSTHs were constructed with 30 ms bins. Each dot represents SU decoding accuracy (% correct); BS units in red, NS in blue. Chance decoding is indicated by the dashed line. A minimum 5 Hz FR response to at least one stimulus in the decode was required for inclusion. Box plots as before: central line indicates median, box edges indicate 25th and 75th percentiles, whiskers extend to data points not considered outliers. **c.** Decoding by depth group. Median decoding by stimulus indicated as colored horizontal box plots (as described above). Gray distributions represent decoding by depth, each SU averaged across stimulus codes. Left: broad-spiking units, right: narrow-spiking. See text for statistics.

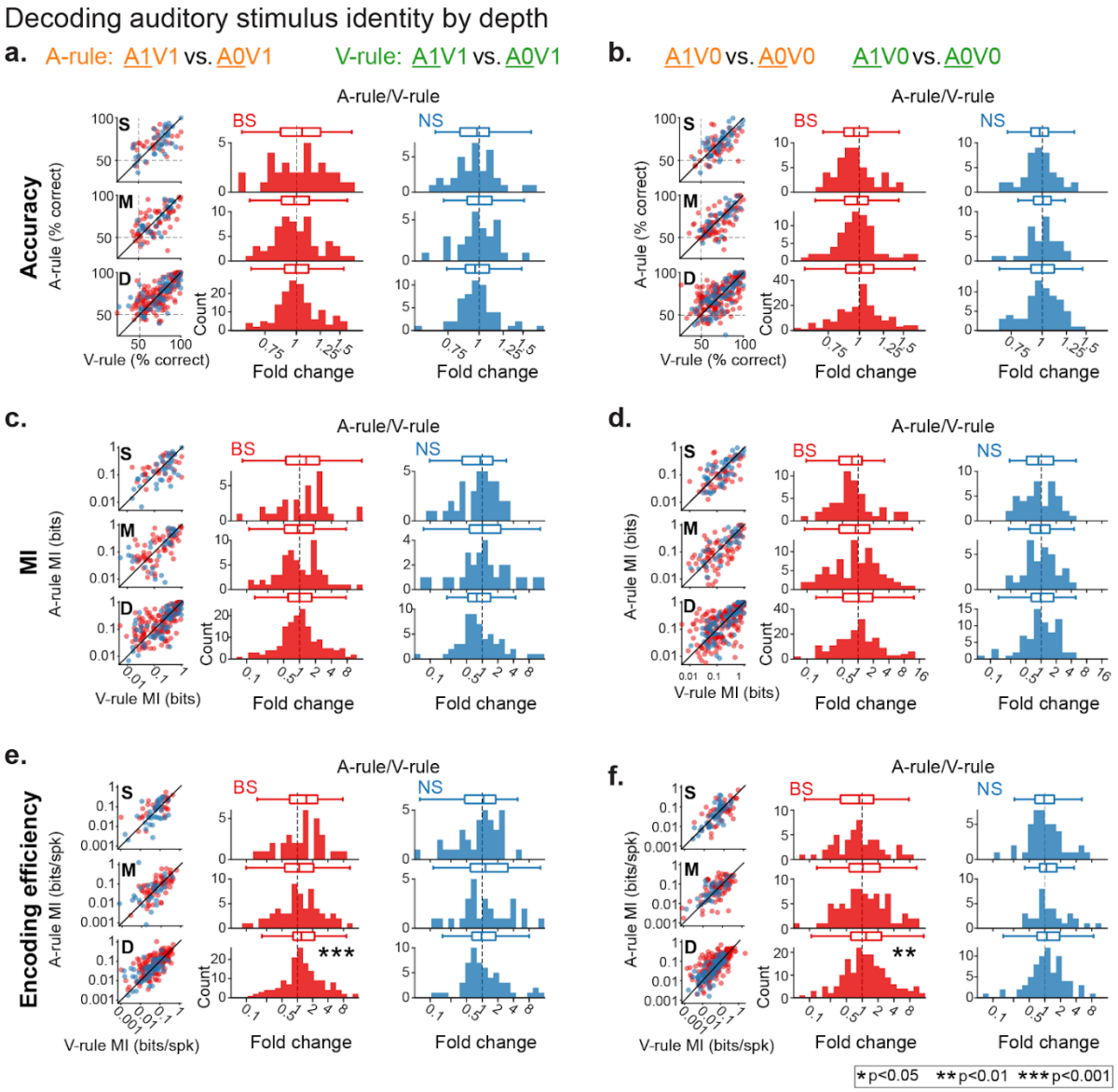

Figure 3.7 Stimulus encoding efficiency increases during the auditory rule. Decoding of auditory stimulus identity, to compare decoder accuracy and information across rules using the PSTH-based Euclidean distance pattern decoder (see Methods, Fig. 3.6). Comparisons were A1V1 vs. A0V1 (**a,c,e**) and A1V0 vs. A0V0 (**b,d,f**). Panels show indicated measure (Accuracy: **a,b**; MI: **c,d** encoding efficiency: **e,f**) by depth and BS/NS classification, conventions as before. Insets on scatter plots indicate depth group for the row: S = superficial; M = middle; D = deep. In scatters, red dots represent BS units and blue dots represent NS units. Histograms show raw unit counts for A-rule/V-rule fold change for BS (red) and NS (blue); fold change values >1 indicate increase in A-rule. Box plots as described elsewhere.

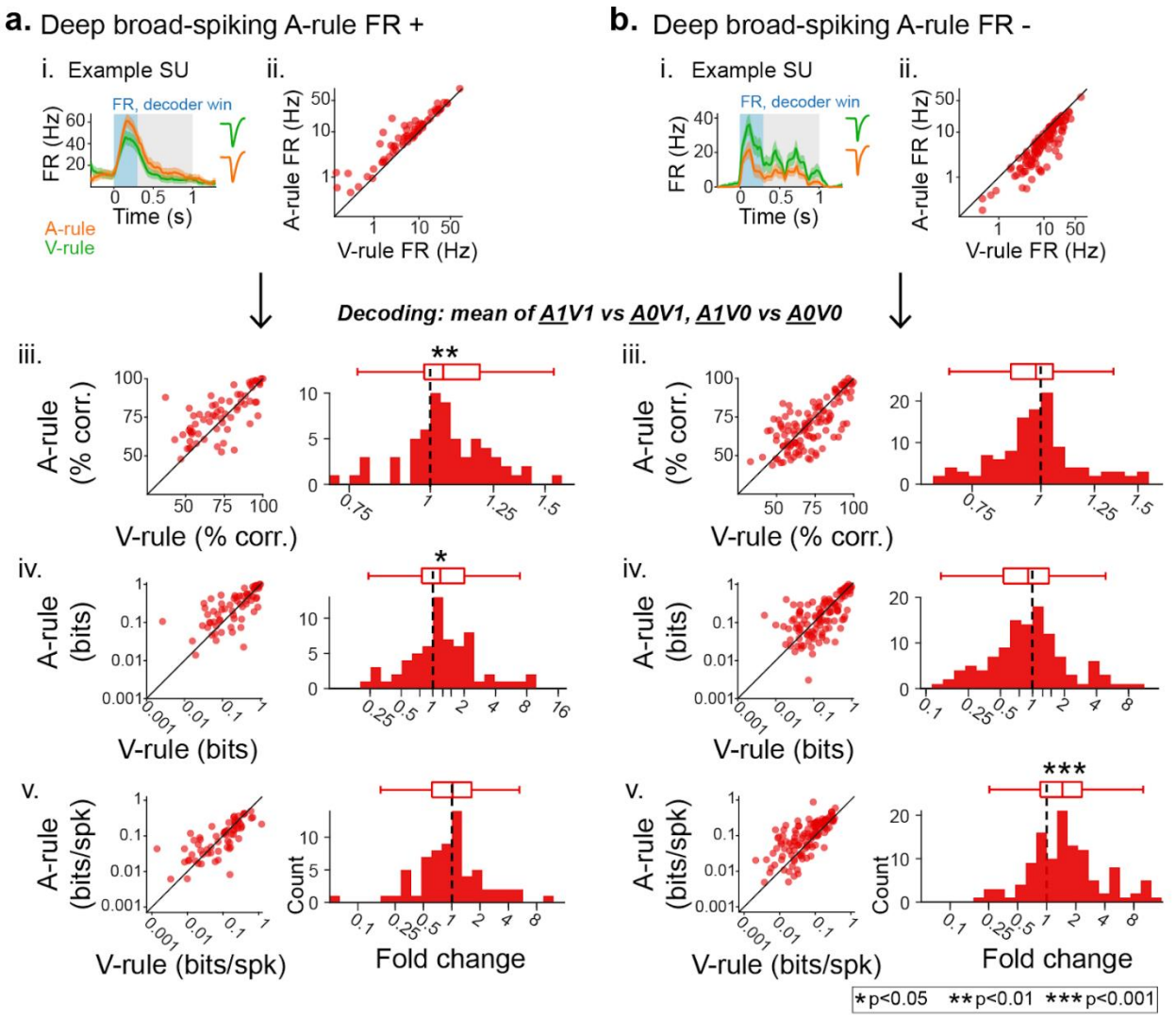

Figure 3.8 Deep ACtx A-rule information efficiency increases are driven by suppressed units. Deep broad-spiking units were split into those with increased A-rule FRs compared with V-rule (n= 65; **a.**) and those with decreased A-rule FRs (n = 114; **b.**), based on raw FR, and mean decoding of auditory stimulus identity (A1* vs A0*) was compared. See Fig. 3.6 for decoding procedure. **a.** Auditory stimulus identity decoding from A-rule FR increased deep BS units. **i.** Example deep BS SU showing increased FR resp in A-rule to A1* stimuli. **ii.** Mean FRs to A1* and A0* stimuli across rules, showing only A-rule FR+ units. **iii.** Decoding accuracy of the A-rule FR+ group (values for each SU are averages of A1V1 vs A1V0 and A1V0 vs A0V0 decodes, collapsed for clarity). **iv.** Mutual information (MI). **v.** MI rate. **b.** Auditory stimulus decoding from A-rule FR decreased deep BS units. Conventions for **i-v.** as in **a.** Box plots as described elsewhere.

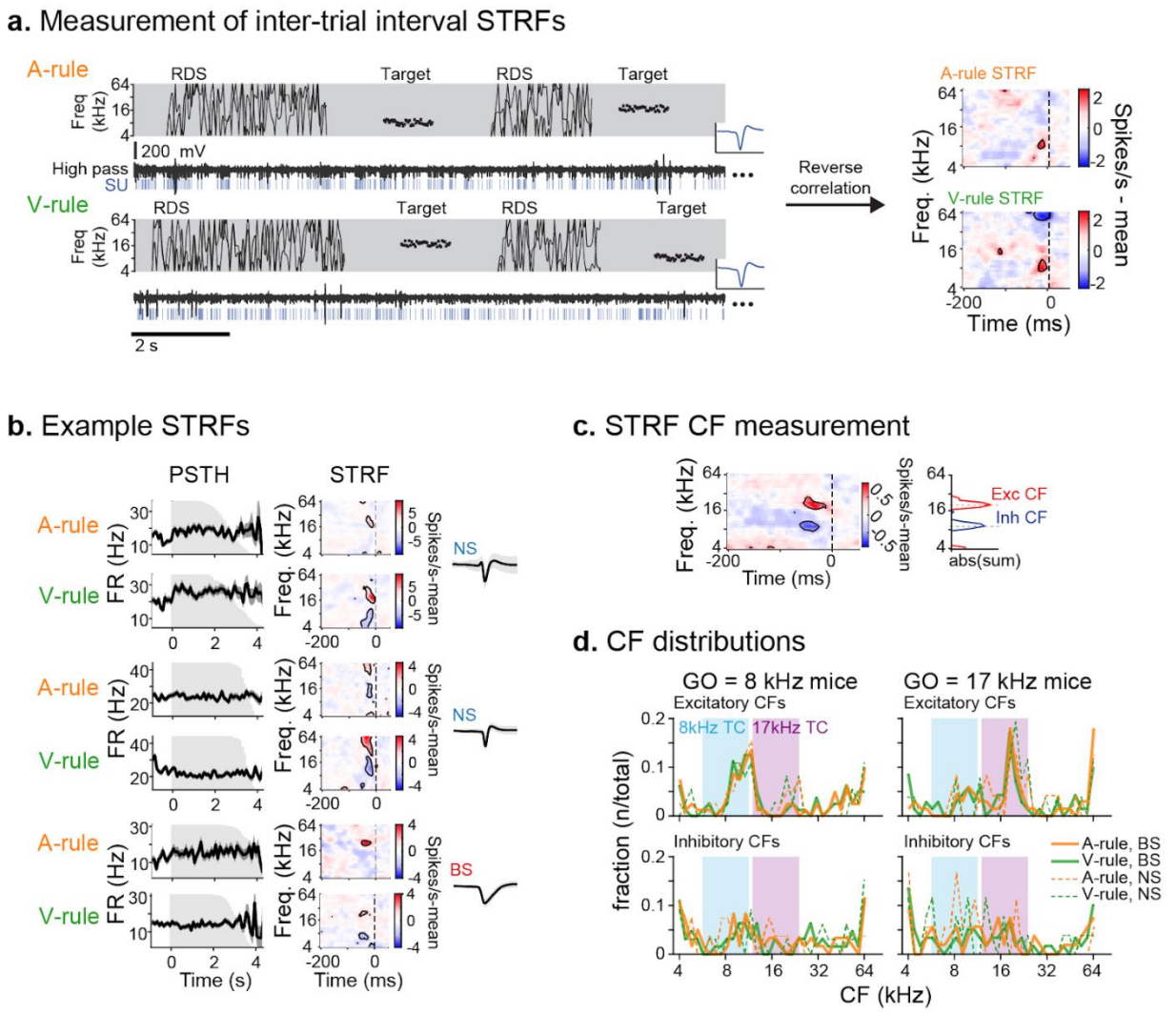

Figure 3.9. Measuring auditory receptive fields during the inter-trial interval. a. A random-double sweep (RDS) stimulus was presented between trials. RDS length varied from 1-15 s (3.0 ± 2.1 s, median ± SD), depending on rate of task progression (determined by mouse locomotion). Responses to each RDS stimulus presentation (segment) were divided into those from the A-rule and those from the V-rule. Duration of RDS presentation time was equated between the two rules. Spectrotemporal receptive fields (STRFs) from each segment were calculated by reverse correlation, and combined by weighted averaging, resulting in a STRF from the A-rule (top) and a STRF from the V-rule (bottom). To remove units whose STRFs that could not be measured reliably, STRFs were filtered for significant trial-to-trial similarity ($p<0.01$ compared to null STRFs). Properties of these STRFs were then compared to determine the effects of task rule on task-irrelevant stimulus processing. **b.** Example RDS response PSTHs and STRFs from two NS units (upper) and one BS unit (lower). PSTHs are triggered on RDS onset, and grey indicates the durations of presented RDS stimuli. **c.** Measurement of characteristic frequency (CF) from the STRF. Significant excitatory (red) and inhibitory (blue) STRF fields were determined by comparing the measured STRF to a null STRF generated from shuffled spike data (see Methods). To measure CF, significant excitatory and inhibitory fields were summed separately along the

frequency axis, and the peak absolute value of each sum was considered the CF. **d.** CF distribution for excitatory (top) and inhibitory (bottom) fields. Left: CFs from SUs recorded from mice for which 8 kHz was the go stimulus. Right: CFs from 17 kHz go stimulus mice. For reference, the frequency bandwidth for the two tone cloud stimuli is shown in blue (8 kHz) and purple (17 kHz).

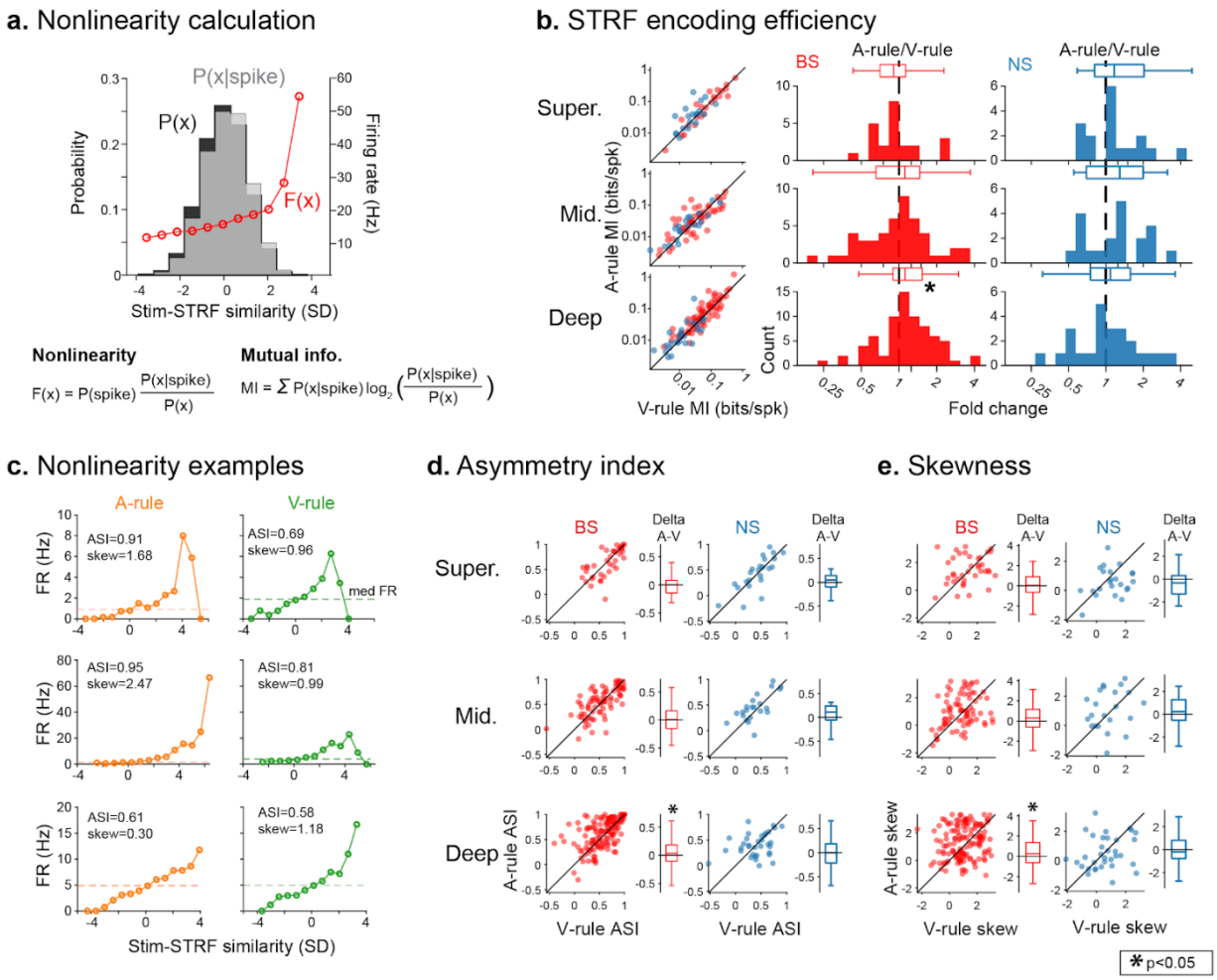

Figure 3.10. Auditory attention modulates inter-trial stimulus response information and nonlinearity-based measures of selectivity. a. Calculation of nonlinear input/output function relating STRF to stimulus evoked activity (see Methods). P(x|spike) is the distribution of similarities between the STRF and the RDS segments that elicited a spike. P(x) represents a null distribution of similarity values between the STRF and randomly chosen segments of the RDS stimulus. The nonlinearity is the ratio of the two, scaled by likelihood of spiking (P(spike)). Mutual information (MI) is also calculated from these distributions. **b.** MI rate (bits/spk) between spiking activity and RDS stimulus by depth, compared across rules. Conventions as in previous figures. In scatters, BS units are red dots. NS are blue dots. Difference histograms for BS (red) and NS (blue) represent A-rule/V-rule; values >1 indicate increased MI rate in A-rule. **c.** Example nonlinearities from three units in the A-rule (left, orange) and V-rule (right, green). Shown in text are two values that characterize the relationship between stim-STRF similarity and spiking activity: asymmetry index (ASI) and skewness (skew). ASI is the difference in likelihood of spiking for positive and negative values of the nonlinearity. Skewness is the amount of information present in the tail of the nonlinearity (see text). **d.** Comparison of ASI across rules. Top: BS and NS values for all depths combined, split by BS units (L, red) and NS units (R, blue). Below: units divided by depth group. * $p<0.05$ sign-rank test comparison of values across rules. **e.** Skewness across rules, conventions as in **d**.

Chapter 4. Concluding remarks

This dissertation comprises findings from two projects on the interplay between the auditory and visual modalities in mouse auditory cortex (ACtx). In Chapter 2, we investigated visually-evoked spiking responses in ACtx. We showed that a visual flash stimulus evoked multi- and single-unit spiking responses in the absence of any sound stimulation. Most striking was the depth distribution of these responses; they were largely restricted to layer (L) 6 and, to a lesser degree, L5. Visual responses in ACtx occurred at longer latencies than auditory responses and also at longer latencies than visual responses in visual cortex. Presentation of drifting gratings showed that ACtx responses were minimally tuned to orientation. This work suggests that the infragranular layers may be a locus of crossmodal convergence in AC, conveying the presence and timing of salient visual stimuli. There are several ways in which this finding should be extended through future research. This report did not address the role of visual stimulation in modulating responses to simultaneously presented auditory stimuli. While previous work has addressed this question and found modulation in of auditory spiking responses in the ferret and monkey ACtx (e.g., Kayser, Petkov, and Logothetis 2008; Bizley and King 2009) as well as field potential responses in the monkey (e.g., Ghazanfar et al. 2005), future work should attempt to relate the specific finding of infragranular visual responses presented in Chapter 2 to modulation of sound representation. However, the possibility that such visual inputs to the auditory cortex are latent and provide a mechanism for audiovisual learning should not be ruled out. Finally, experiments leveraging the repertoire of genetic tools in the mouse could aid in circuit dissection of this visual input. First, intersectional tracing work using modified viral vectors could be used to identify and label cells receiving inputs from visual stations, cortical or otherwise (Finke, Conzelmann, and Callaway 2007). Second, the ever-

growing list of cell-type specific *Cre* lines that label subsets of interneurons or pyramidal cells could be used for optogenetic identification of the cell populations that respond to crossmodal stimuli (Lima et al. 2009). These techniques would help provide genetic handles for the circuits bridging the sensory systems.

In Chapter 3, we investigated the role of auditory and visual modality-specific attention on audiovisual sensory responses in mouse ACtx. We designed a novel audiovisual rule-switching task in which mice had to either attend to sound while ignoring visual stimuli or ignore sound while attending to visual stimuli. After mice achieved expertise in this task, we recorded single-unit responses while mice performed the behavior and compared neural responses to identical audiovisual task stimuli under different attentional demands. We found distinct patterns of activity associated with the attend-auditory condition relative to the attend-visual condition. The predominant effect was suppression of firing rate, however, a smaller subset of cells also showed enhanced firing rates during the auditory rule. Similar trends for firing rate modulation were seen for task-irrelevant receptive field mapping stimuli. When corrected for pre-stimulus firing rate, much of the modulation of stimulus-evoked activity observed at the group level was abolished, suggesting that auditory attention may cause a general reduction in levels of activity rather than a specific reduction stimulus-related drive. Using a PSTH-based pattern decoder, we found no difference in decoding ability or mutual information across attentional states, but instead observed increased encoding efficiency in the deep putative-excitatory cells. These results show that suppression of firing rates and increased encoding efficiency are neural correlates of auditory modality-specific attention and that cells in the deepest layers of auditory cortex show the greatest degree of modulation by task rule. Several important caveats for this study must be noted. First, care has been taken to qualify

the use of the term “attention” to emphasize that the mechanisms we are studying may be distinct from forms that highlight local features of a perceptual scene. Such forms of attention have largely been studied in humans or nonhuman primates (Desimone and Duncan 1995; Kastner and Ungerleider 2000; Reynolds and Chelazzi 2004), and the tasks required for their study may not be well-suited to rodent models. Second, although probe penetrations were made using coordinates centered on the mouse primary ACtx, in this study we have not made any distinction between units from primary and secondary regions. Past research suggests that attention modulates cortical processing throughout the auditory cortical hierarchy in humans (Grady et al. 1997; Jäncke, Mirzazade, and Shah 1999), although work at the single-unit level in animals has not explicitly compared attentional effects across subregions (Otazu et al. 2009; Buran, von Trapp, and Sanes 2014; Carcea, Insanally, and Froemke 2017). Future work on crossmodal attention should deliberately target distinct sensory cortical subregions to determine, at the fine-scale resolution provided by animal models, how attention affects sensory processing at different levels of the cortical hierarchy. Third, this work has not examined the role of field potential oscillations in crossmodal attention. Findings that attentionally-induced increases in alpha band oscillations are related to inhibitory function (Klimesch 2012) provide an interesting avenue of exploration given the suppressive effects of modality-specific attention we observed. Finally, while this study suggests that a change in encoding efficiency may be a neural correlate of auditory-specific attention, we have yet to demonstrate that this has behavioral ramifications. Future analysis of changes in encoding efficiency during incorrect trials or on low-performance blocks may better support a causal link between this neural feature and behavior.

Unexpectedly, both of these projects implicated the deepest cells in the ACtx as mediators of crossmodal-related phenomena. Given the different nature and goals of these two projects, overlapping cell populations or mechanisms should not be assumed. Instead, together these results suggest that the cells in the infragranular layer may be important sites of convergence for a variety of extramodal signals. The anatomical structure of the infragranular population makes it well-suited for this role. The large pyramidal cells in L5 have dendritic tufts that extend to L1, but also receive inputs from all other layers, suggesting that they integrate both driving and modulatory information (Larkum, Zhu, and Sakmann 1999). L6 contains a diversity of pyramidal cells, the tallest of which span the entire cortical column (L6b to L1; Katz 1987; Thomson 2010). Other L6 pyramidal cells have apical tufts that reach L4 and L5 (Thomson 2010; Briggs 2010). Their wide array of columnar inputs suggests that cells in both L5 and L6 may be important integrators. Finally, the infragranular layer is traditionally considered the cortical output layer, sending processed information back to subcortical stations and to other cortical areas (Felleman and Van Essen 1991; Douglas and Martin 2004). It may be expected that modulatory effects from extramodal processes such as attention are most readily measurable at the putatively final stage of columnar processing.

Several caveats apply to the work presented in both Chapters 2 and 3. First, this dissertation has focused on effects at the single-unit level and has not examined population-level phenomena (Averbeck, Latham, and Pouget 2006). Attention appears to modulate the degree of correlations between recorded neurons, an effect observed in both visual (Cohen and Maunsell 2009) and auditory (Downer, Niwa, and Sutter 2015; Downer et al. 2017) cortices. Furthermore, recent work on the rat ACtx suggests that when simultaneously recorded neurons are grouped into ensembles based on coordinated

spiking, the encoding of auditory sensory information is enhanced (See et al. 2018). While previous work has studied the population-level effects of attentional modulation, far less research has explored the ways in which multisensory effects may emerge at the population level. With advances in imaging (Stirman et al. 2016) and electrophysiological (Jun et al. 2017) techniques that now allow for simultaneous recording of hundreds to thousands of neurons, this is likely to be a fruitful avenue for future research.

On a technical level, while the use of multisite silicon probes allows for recording from the entire cortical depth, this technique likely does not capture single-unit activity at all depths equally well. In particular, neurons in the superficial layers may sustain greater damage from probe insertion than those in the deeper layers and could exhibit decreased stability or altered physiological response properties. This is of concern because of the anatomically identified visual inputs to superficial auditory cortex (Banks et al. 2011) as well as examples of attentional-related current-source density modulations in supragranular layers (Mehta, Ulbert, and Schroeder 2000; Lakatos et al. 2009). While we were able to record isolated and stable single-units from superficial cortex in Chapter 3, we cannot fully rule out effects that were undetected due to uneven tissue damage inherent to the technique. Future work aimed at recording the full laminae should use a combination of partial- and full-cortical depth probe insertions to rule out this potential confound.

Taken together, findings from these projects underscore two main principles: first, the ACtx is dynamically modulated by non-auditory inputs to shape sound processing according to behavioral demands. This integration from across the modalities as well as attentional and other networks is central to the function of ACtx, altering how it encodes information. Second, the laminar structure of cortex provides an important architecture

for the integration of extramodal information. Explicitly addressing neural effects by depth in the awake, behaving animal, whenever possible, will continue to yield new insight into the circuitry underlying cortical information processing.

Bibliography

Adesnik, Hillel, and Alexander Naka. 2018. “Cracking the Function of Layers in the Sensory Cortex.” *Neuron* 100 (5): 1028–43.

Aertsen, A. M., and P. I. Johannesma. 1981. “The Spectro-Temporal Receptive Field. A Functional Characteristic of Auditory Neurons.” *Biological Cybernetics* 42 (2): 133–43.

Allman, Brian L., Ruben E. Bittencourt-Navarrete, Leslie P. Keniston, Alexandre E. Medina, Meng Y. Wang, and M. Alex Meredith. 2008. “Do Cross-Modal Projections Always Result in Multisensory Integration?” *Cerebral Cortex* 18 (9): 2066–76.

Arieli, A., A. Sterkin, A. Grinvald, and A. Aertsen. 1996. “Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses.” *Science* 273 (5283): 1868–71.

Atencio, C. A., and C. E. Schreiner. 2013. “Stimulus Choices for Spike-Triggered Receptive Field Analysis.” *Handbook of Modern Techniques in Auditory Cortex. New York: Nova Biomedical*, 61–100.

Atencio, Craig A., and Christoph E. Schreiner. 2008. “Spectrotemporal Processing Differences between Auditory Cortical Fast-Spiking and Regular-Spiking Neurons.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28 (15): 3897–3910.

Atencio, Craig A., and Christoph E. Schreiner. 2012. “Spectrotemporal Processing in Spectral Tuning Modules of Cat Primary Auditory Cortex.” *PloS One* 7 (2): e31537.

Atencio, Craig A., Tatyana O. Sharpee, and Christoph E. Schreiner. 2008. “Cooperative Nonlinearities in Auditory Cortical Neurons.” *Neuron* 58 (6): 956–66.

Averbeck, Bruno B., Peter E. Latham, and Alexandre Pouget. 2006. “Neural Correlations, Population Coding and Computation.” *Nature Reviews. Neuroscience* 7 (5): 358–66.

Bagur, Sophie, Martin Averseng, Diego Elgueda, Stephen David, Jonathan Fritz, Pingbo Yin, Shihab Shamma, Yves Boubenec, and Srdjan Ostojic. 2018. “Go/No-Go Task Engagement Enhances Population Representation of Target Stimuli in Primary Auditory Cortex.” *Nature Communications* 9 (1): 2529.

Banks, Matthew I., Daniel J. Uhlrich, Philip H. Smith, Bryan M. Krause, and Karen A. Manning. 2011. “Descending Projections from Extrastriate Visual Cortex Modulate Responses of Cells in Primary Auditory Cortex.” *Cerebral Cortex* 21 (11): 2620–38.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 289–300.

Bental, E., N. Dafny, and S. Feldman. 1968. “Convergence of Auditory and Visual Stimuli on Single Cells in the Primary Visual Cortex of Unanesthetized Unrestrained Cats.” *Experimental Neurology* 20 (3): 341–51.

Bidet-Caulet, Aurélie, Catherine Fischer, Julien Besle, Pierre-Emmanuel Aguera, Marie-Helene Giard, and Olivier Bertrand. 2007. “Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27 (35): 9252–61.

Bigelow, James, Ryan J. Morrill, Jefferson Dekloe, and Andrea R. Hasenstaub. 2019. “Movement and VIP Interneuron Activation Differentially Modulate Encoding in Mouse Auditory Cortex.” *eNeuro* 6 (5).

Bisley, James W., and Michael E. Goldberg. 2006. “Neural Correlates of Attention and Distractibility in the Lateral Intraparietal Area.” *Journal of Neurophysiology* 95 (3): 1696–1717.

Bizley, Jennifer K., and Andrew J. King. 2008. “Visual-Auditory Spatial Processing in Auditory Cortical Neurons.” *Brain Research* 1242 (November): 24–36.

Bizley, Jennifer K., and Andrew J. King. 2009. “Visual Influences on Ferret Auditory Cortex.” *Hearing Research* 258 (1-2): 55–63.

Bizley, Jennifer K., Fernando R. Nodal, Victoria M. Bajo, Israel Nelken, and Andrew J. King. 2007. “Physiological and Anatomical Evidence for Multisensory Interactions in Auditory Cortex.” *Cerebral Cortex* 17 (9): 2172–89.

Bolz, J., and C. D. Gilbert. 1986. “Generation of End-Inhibition in the Visual Cortex via Interlaminar Connections.” *Nature* 320 (6060): 362–65.

Bortone, Dante S., Shawn R. Olsen, and Massimo Scanziani. 2014. “Translaminar Inhibitory Cells Recruited by Layer 6 Corticothalamic Neurons Suppress Visual Cortex.” *Neuron* 82 (2): 474–85.

Briggs, Faran. 2010. “Organizing Principles of Cortical Layer 6.” *Frontiers in Neural Circuits* 4 (3). https://doi.org/10.3389/neuro.04.003.2010.

Briggs, Farran, and W. Martin Usrey. 2008. “Emerging Views of Corticothalamic Function.” *Current Opinion in Neurobiology* 18 (4): 403–7.

Brosch, Michael, Elena Selezneva, and Henning Scheich. 2005. “Nonauditory Events of a Behavioral Procedure Activate Auditory Cortex of Highly Trained Monkeys.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25 (29): 6797–6806.

Budinger, E., P. Heil, A. Hess, and H. Scheich. 2006. “Multisensory Processing via Early Cortical Stages: Connections of the Primary Auditory Cortical Field with Other Sensory Systems.” *Neuroscience* 143 (4): 1065–83.

Budinger, Eike, Anna Laszcz, Holger Lison, Henning Scheich, and Frank W. Ohl. 2008. “Non-Sensory Cortical and Subcortical Connections of the Primary Auditory Cortex in Mongolian Gerbils: Bottom-up and Top-down Processing of Neuronal Information via Field AI.” *Brain Research* 1220 (July): 2–32.

Budinger, Eike, and Henning Scheich. 2009. “Anatomical Connections Suitable for the Direct Processing of Neuronal Information of Different Modalities via the Rodent Primary Auditory Cortex.” *Hearing Research* 258 (1-2): 16–27.

Buracas, G. T., A. M. Zador, M. R. DeWeese, and T. D. Albright. 1998. “Efficient Discrimination of Temporal Patterns by Motion-Sensitive Neurons in Primate Visual Cortex.” *Neuron* 20 (5): 959–69.

Buran, Bradley N., Gardiner von Trapp, and Dan H. Sanes. 2014. “Behaviorally Gated Reduction of Spontaneous Discharge Can Improve Detection Thresholds in Auditory Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34 (11): 4076–81.

Callaway, Edward M. 2005. “A Molecular and Genetic Arsenal for Systems Neuroscience.” *Trends in Neurosciences* 28 (4): 196–201.

Calvert, G. A., E. T. Bullmore, M. J. Brammer, R. Campbell, S. C. Williams, P. K. McGuire, P. W. Woodruff, S. D. Iversen, and A. S. David. 1997. “Activation of Auditory Cortex during Silent Lipreading.” *Science* 276 (5312): 593–96.

Carcea, Ioana, Michele N. Insanally, and Robert C. Froemke. 2017. “Dynamics of Auditory Cortical Activity during Behavioural Engagement and Auditory Perception.” *Nature Communications* 8 (February): 14412.

Cardin, Jessica A., Larry A. Palmer, and Diego Contreras. 2007. “Stimulus Feature Selectivity in Excitatory and Inhibitory Neurons in Primary Visual Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27 (39): 10333–44.

Chandrasekaran, Chandramouli, Andrea Trubanova, Sébastien Stillittano, Alice Caplier, and Asif A. Ghazanfar. 2009. “The Natural Statistics of Audiovisual Speech.” *PLoS Computational Biology* 5 (7): e1000436.

Christianson, G. Björn, Maneesh Sahani, and Jennifer F. Linden. 2011. “Depth-Dependent Temporal Response Properties in Core Auditory Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31 (36): 12837–48.

Cohen, Marlene R., and John H. R. Maunsell. 2009. “Attention Improves Performance Primarily by Reducing Interneuronal Correlations.” *Nature Neuroscience* 12 (12): 1594–1600.

Constantinople, Christine M., and Randy M. Bruno. 2013. “Deep Cortical Layers Are Activated Directly by Thalamus.” *Science* 340 (6140): 1591–94.

Cools, R., R. B. Ivry, and M. D’Esposito. 2006. “The Human Striatum Is Necessary for Responding to Changes in Stimulus Relevance.” *Journal of Cognitive Neuroscience* 18 (12): 1973–83.

Cools, Roshan, Luke Clark, and Trevor W. Robbins. 2004. “Differential Responses in Human Striatum and Prefrontal Cortex to Changes in Object and Rule Relevance.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 24 (5): 1129–35.

Cox, Michele A., Kacie Dougherty, Geoffrey K. Adams, Eric A. Reavis, Jacob A. Westerberg, Brandon S. Moore, David A. Leopold, and Alexander Maier. 2019. “Spiking Suppression Precedes Cued Attentional Enhancement of Neural Responses in Primary Visual Cortex.” *Cerebral Cortex* 29 (1): 77–90.

Crone, Eveline A., Carter Wendelken, Sarah E. Donohue, and Silvia A. Bunge. 2006. “Neural Evidence for Dissociable Components of Task-Switching.” *Cerebral Cortex* 16 (4): 475–86.

Dadarlat, Maria C., and Michael P. Stryker. 2017. “Locomotion Enhances Neural Encoding of Visual Stimuli in Mouse V1.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37 (14): 3764–75.

Desimone, R., and J. Duncan. 1995. “Neural Mechanisms of Selective Visual Attention.” *Annual Review of Neuroscience* 18: 193–222.

DiCarlo, J. J., J. W. Lane, S. S. Hsiao, and K. O. Johnson. 1996. “Marking Microelectrode Penetrations with Fluorescent Dyes.” *Journal of Neuroscience Methods* 64 (1): 75–81.

Diederich, Adele, and Hans Colonius. 2004. “Bimodal and Trimodal Multisensory Enhancement: Effects of Stimulus Onset and Intensity on Reaction Time.” *Perception & Psychophysics* 66 (8): 1388–1404.

Douglas, Rodney J., and Kevan A. C. Martin. 2004. “Neuronal Circuits of the Neocortex.” *Annual Review of Neuroscience* 27: 419–51.

Downer, Joshua D., Mamiko Niwa, and Mitchell L. Sutter. 2015. “Task Engagement Selectively Modulates Neural Correlations in Primary Auditory Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35 (19): 7565–74.

Downer, Joshua D., Brittany Rapone, Jessica Verhein, Kevin N. O’Connor, and Mitchell L. Sutter. 2017. “Feature-Selective Attention Adaptively Shifts Noise Correlations in

Primary Auditory Cortex." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37 (21): 5378–92.

Driver, Jon, and Charles Spence. 1998. "Cross-Modal Links in Spatial Attention." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 353 (1373): 1319–31.

Ernst, Marc O., and Martin S. Banks. 2002. "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion." *Nature* 415 (6870): 429–33.

Escabí, Monty A., Heather L. Read, Jonathan Viventi, Dae-Hyeong Kim, Nathan C. Higgins, Douglas A. Storace, Andrew S. K. Liu, et al. 2014. "A High-Density, High-Channel Count, Multiplexed μECoG Array for Auditory-Cortex Recordings." *Journal of Neurophysiology* 112 (6): 1566–83.

Escabí, Monty A., and Christoph E. Schreiner. 2002. "Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22 (10): 4114–31.

Fairhall, S. L., and E. Macaluso. 2009. "Spatial Attention Can Modulate Audiovisual Integration at Multiple Cortical and Subcortical Sites." *The European Journal of Neuroscience* 29 (6): 1247–57.

Falchier, Arnaud, Charles E. Schroeder, Troy A. Hackett, Peter Lakatos, Sheila Nascimento-Silva, Istvan Ulbert, Gyorgi Karmos, and John F. Smiley. 2010. "Projection from Visual Areas V2 and Prostriata to Caudal Auditory Cortex in the Monkey." *Cerebral Cortex* 20 (7): 1529–38.

Felleman, D. J., and D. C. Van Essen. 1991. "Distributed Hierarchical Processing in the Primate Cerebral Cortex." *Cerebral Cortex* 1 (1): 1–47.

Fiáth, Richárd, Adrienn Lilla Márton, Ferenc Mátyás, Domonkos Pinke, Gergely Márton, Kinga Tóth, and István Ulbert. 2019. “Slow Insertion of Silicon Probes Improves the Quality of Acute Neuronal Recordings.” *Scientific Reports* 9 (1): 111.

Finke, S., K. K. Conzelmann, and E. M. Callaway. 2007. “Retrograde Neuronal Tracing with a Deletion-Mutant Rabies Virus.” *Nature*. https://www.nature.com/articles/nmeth999.

Foffani, Guglielmo, and Karen Anne Moxon. 2004. “PSTH-Based Classification of Sensory Stimuli Using Ensembles of Single Neurons.” *Journal of Neuroscience Methods* 135 (1-2): 107–20.

Foxe, J. J., I. A. Morocz, M. M. Murray, B. A. Higgins, D. C. Javitt, and C. E. Schroeder. 2000. “Multisensory Auditory-Somatosensory Interactions in Early Cortical Processing Revealed by High-Density Electrical Mapping.” *Brain Research. Cognitive Brain Research* 10 (1-2): 77–83.

Foxe, John J., and Charles E. Schroeder. 2005. “The Case for Feedforward Multisensory Convergence during Early Cortical Processing.” *Neuroreport* 16 (5): 419–23.

Fritz, Jonathan, Shihab Shamma, Mounya Elhilali, and David Klein. 2003. “Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in Primary Auditory Cortex.” *Nature Neuroscience* 6 (11): 1216–23.

Fu, Yu, Jason M. Tucciarone, J. Sebastian Espinosa, Nengyin Sheng, Daniel P. Darcy, Roger A. Nicoll, Z. Josh Huang, and Michael P. Stryker. 2014. “A Cortical Circuit for Gain Control by Behavioral State.” *Cell* 156 (6): 1139–52.

Gau, Remi, Pierre-Louis Bazin, Robert Trampel, Robert Turner, and Uta Noppeney. 2020. “Resolving Multisensory and Attentional Influences across Cortical Depth in Sensory Cortices.” *eLife* 9 (January). https://doi.org/10.7554/eLife.46856.

Ghazanfar, Asif A., Joost X. Maier, Kari L. Hoffman, and Nikos K. Logothetis. 2005. "Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25 (20): 5004–12.

Ghazanfar, Asif A., and Charles E. Schroeder. 2006. "Is Neocortex Essentially Multisensory?" *Trends in Cognitive Sciences* 10 (6): 278–85.

Giard, M. H., A. Fort, Y. Mouchetant-Rostaing, and J. Pernier. 2000. "Neurophysiological Mechanisms of Auditory Selective Attention in Humans." *Frontiers in Bioscience: A Journal and Virtual Library* 5 (January): D84–94.

Gilbert, C. D., M. Sigman, and R. E. Crist. 2001. "The Neural Basis of Perceptual Learning." *Neuron* 31 (5): 681–97.

Goldman-Rakic, P. S. 1988. "Topography of Cognition: Parallel Distributed Networks in Primate Association Cortex." *Annual Review of Neuroscience* 11: 137–56.

Gong, Shiaoching, Martin Doughty, Carroll R. Harbaugh, Alexander Cummins, Mary E. Hatten, Nathaniel Heintz, and Charles R. Gerfen. 2007. "Targeting Cre Recombinase to Specific Neuron Populations with Bacterial Artificial Chromosome Constructs." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27 (37): 9817–23.

Gourévitch, Boris, Florian Occelli, Quentin Gaucher, Yonane Aushana, and Jean-Marc Edeline. 2015. "A New and Fast Characterization of Multiple Encoding Properties of Auditory Neurons." *Brain Topography* 28 (3): 379–400.

Grady, C. L., J. W. Van Meter, J. M. Maisog, P. Pietrini, J. Krasuski, and J. P. Rauschecker. 1997. "Attention-Related Modulation of Activity in Primary and Secondary Auditory Cortex." *Neuroreport* 8 (11): 2511–16.

Grant, K. W., and P. F. Seitz. 2000. “The Use of Visible Speech Cues for Improving Auditory Detection of Spoken Sentences.” *The Journal of the Acoustical Society of America* 108 (3 Pt 1): 1197–1208.

Guo, Zengcai V., S. Andrew Hires, Nuo Li, Daniel H. O’Connor, Takaki Komiyama, Eran Ophir, Daniel Huber, et al. 2014. “Procedures for Behavioral Experiments in Head-Fixed Mice.” *PloS One* 9 (2): e88678.

Haider, Bilal, and David A. McCormick. 2009. “Rapid Neocortical Dynamics: Cellular and Network Mechanisms.” *Neuron* 62 (2): 171–89.

Harris, Kenneth D., and Alexander Thiele. 2011. “Cortical State and Attention.” *Nature Reviews. Neuroscience* 12 (9): 509–23.

Hautus, Michael J. 1995. “Corrections for Extreme Proportions and Their Biasing Effects on Estimated Values Of d’.” *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc* 27 (1): 46–51.

Helbig, Hannah B., and Marc O. Ernst. 2008. “Visual-Haptic Cue Weighting Is Independent of Modality-Specific Attention.” *Journal of Vision* 8 (1): 21.1–16.

Henschke, Julia U., Tömme Noesselt, Henning Scheich, and Eike Budinger. 2015. “Possible Anatomical Pathways for Short-Latency Multisensory Integration Processes in Primary Sensory Cortices.” *Brain Structure & Function* 220 (2): 955–77.

Herrington, Todd M., and John A. Assad. 2010. “Temporal Sequence of Attentional Modulation in the Lateral Intraparietal Area and Middle Temporal Area during Rapid Covert Shifts of Attention.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (9): 3287–96.

Hillyard, S. A., R. F. Hink, V. L. Schwent, and T. W. Picton. 1973. “Electrical Signs of Selective Attention in the Human Brain.” *Science* 182 (4108): 177–80.

Hoglen, Nerissa E. G., Phillip Larimer, Elizabeth A. K. Phillips, Brian J. Malone, and Andrea R. Hasenstaub. 2018. “Amplitude Modulation Coding in Awake Mice and Squirrel Monkeys.” *Journal of Neurophysiology* 119 (5): 1753–66.

Horn, G. 1965. “The Effect of Somaesthetic and Photic Stimuli on the Activity of Units in the Striate Cortex of Unanaesthetized, Unrestrained Cats.” *The Journal of Physiology* 179 (2): 263–77.

Hromádka, Tomás, and Anthony M. Zador. 2007. “Toward the Mechanisms of Auditory Attention.” *Hearing Research* 229 (1-2): 180–85.

Hubel, David H., Calvin O. Henson, Allen Rupert, and Robert Galambos. 1959. “‘Attention’ Units in the Auditory Cortex.” *Science* 129 (3358): 1279–80.

Humphrey, Donald R., and Edward M. Schmidt. 1990. “Extracellular Single-Unit Recording Methods.” In *Neurophysiological Techniques: Applications to Neural Systems*, edited by Alan A. Boulton, Glen B. Baker, and Case H. Vanderwolf, 1–64. Totowa, NJ: Humana Press.

Iurilli, Giuliano, Diego Ghezzi, Umberto Olcese, Glenda Lassi, Cristiano Nazzaro, Raffaella Tonini, Valter Tucci, Fabio Benfenati, and Paolo Medini. 2012. “Sound-Driven Synaptic Inhibition in Primary Visual Cortex.” *Neuron* 73 (4): 814–28.

Jäncke, L., S. Mirzazade, and N. J. Shah. 1999. “Attention Modulates Activity in the Primary and the Secondary Auditory Cortex: A Functional Magnetic Resonance Imaging Study in Human Subjects.” *Neuroscience Letters*.

Jiang, W., M. T. Wallace, H. Jiang, J. W. Vaughan, and B. E. Stein. 2001. “Two Cortical Areas Mediate Multisensory Integration in Superior Colliculus Neurons.” *Journal of Neurophysiology* 85 (2): 506–22.

Joachimsthaler, Bettina, Michaela Uhlmann, Frank Miller, Günter Ehret, and Simone Kurt. 2014. “Quantitative Analysis of Neuronal Response Properties in Primary and Higher-Order Auditory Cortical Fields of Awake House Mice (Mus Musculus).” *The European Journal of Neuroscience* 39 (6): 904–18.

Johnson, Jennifer Adrienne, and Robert J. Zatorre. 2006. “Neural Substrates for Dividing and Focusing Attention between Simultaneous Auditory and Visual Events.” *NeuroImage* 31 (4): 1673–81.

Johnson, Jennifer A., and Robert J. Zatorre. 2005. “Attention to Simultaneous Unrelated Auditory and Visual Events: Behavioral and Neural Correlates.” *Cerebral Cortex* 15 (10): 1609–20.

Jun, James J., Nicholas A. Steinmetz, Joshua H. Siegle, Daniel J. Denman, Marius Bauza, Brian Barbarits, Albert K. Lee, et al. 2017. “Fully Integrated Silicon Probes for High-Density Recording of Neural Activity.” *Nature* 551 (7679): 232–36.

Kastner, S., and L. G. Ungerleider. 2000. “Mechanisms of Visual Attention in the Human Cortex.” *Annual Review of Neuroscience* 23: 315–41.

Katz, L. C. 1987. “Local Circuitry of Identified Projection Neurons in Cat Visual Cortex Brain Slices.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 7 (4): 1223–49.

Kawashima, Ryuta, Brendan T. O’Sullivan, and Per E. Roland. 1995. “Positron-Emission Tomography Studies of Cross-Modality Inhibition in Selective Attentional Tasks: Closing the ‘Mind’s Eye.’” Proceedings of the National Academy of Sciences 92 (13): 5969–72.

Kayser, Christoph, and Nikos K. Logothetis. 2007. “Do Early Sensory Cortices Integrate Cross-Modal Information?” *Brain Structure & Function* 212 (2): 121–32.

Kayser, Christoph, Christopher I. Petkov, Mark Augath, and Nikos K. Logothetis. 2005. "Integration of Touch and Sound in Auditory Cortex." *Neuron* 48 (2): 373–84.

Kayser, Christoph, Christopher I. Petkov, and Nikos K. Logothetis. 2008. "Visual Modulation of Neurons in Auditory Cortex." *Cerebral Cortex* 18 (7): 1560–74.

Kleiner, M., D. Brainard, and D. Pelli. 2007. "'What's New in Psychtoolbox-3?'" In *ECVP Abstract Supplement*.

Klimesch, Wolfgang. 2012. "Alpha-Band Oscillations, Attention, and Controlled Access to Stored Information." *Trends in Cognitive Sciences* 16 (12): 606–17.

Kobayasi, Kohta I., Yoichi Suwa, and Hiroshi Riquimaroux. 2013. "Audiovisual Integration in the Primary Auditory Cortex of an Awake Rodent." *Neuroscience Letters* 534 (February): 24–29.

Krishna, B. S., and M. N. Semple. 2000. "Auditory Temporal Processing: Responses to Sinusoidally Amplitude-Modulated Tones in the Inferior Colliculus." *Journal of Neurophysiology* 84 (1): 255–73.

Kuchibhotla, Kishore V., Jonathan V. Gill, Grace W. Lindsay, Eleni S. Papadoyannis, Rachel E. Field, Tom A. Hindmarsh Sten, Kenneth D. Miller, and Robert C. Froemke. 2017. "Parallel Processing by Cortical Inhibition Enables Context-Dependent Behavior." *Nature Neuroscience* 20 (1): 62–71.

Laboy-Juárez, Keven J., Seoiyoung Ahn, and Daniel E. Feldman. 2019. "A Normalized Template Matching Method for Improving Spike Detection in Extracellular Voltage Recordings." *Scientific Reports* 9 (1): 12087.

Lakatos, Peter, Chi-Ming Chen, Monica N. O'Connell, Aimee Mills, and Charles E. Schroeder. 2007. "Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex." *Neuron* 53 (2): 279–92.

Lakatos, Peter, Monica N. O'Connell, Annamaria Barczak, Aimee Mills, Daniel C. Javitt, and Charles E. Schroeder. 2009. "The Leading Sense: Supramodal Control of Neurophysiological Context by Attention." *Neuron* 64 (3): 419–30.

Land, R., G. Engler, A. Kral, and A. K. Engel. 2013. "Response Properties of Local Field Potentials and Multiunit Activity in the Mouse Visual Cortex." *Neuroscience* 254 (December): 141–51.

Larkum, Matthew E., and J. Julius Zhu. 2002. "Signaling of Layer 1 and Whisker-Evoked Ca2+ and Na+ Action Potentials in Distal and Terminal Dendrites of Rat Neocortical Pyramidal Neurons in Vitro and in Vivo." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22 (16): 6991–7005.

Larkum, M. E., J. J. Zhu, and B. Sakmann. 1999. "A New Cellular Mechanism for Coupling Inputs Arriving at Different Cortical Layers." *Nature* 398 (6725): 338–41.

Laurienti, Paul J., Jonathan H. Burdette, Mark T. Wallace, Yi-Fen Yen, Aaron S. Field, and Barry E. Stein. 2002. "Deactivation of Sensory-Specific Cortex by Cross-Modal Stimuli." *Journal of Cognitive Neuroscience* 14 (3): 420–29.

Lee, Eunee, Jiso Hong, Young-Gyun Park, Sujin Chae, Yong Kim, and Daesoo Kim. 2015. "Left Brain Cortical Activity Modulates Stress Effects on Social Behavior." *Scientific Reports* 5 (August): 13342.

Lemus, Luis, Adrián Hernández, Rogelio Luna, Antonio Zainos, and Ranulfo Romo. 2010. "Do Sensory Cortices Process More than One Sensory Modality during Perceptual Judgments?" *Neuron* 67 (2): 335–48.

Licata, Angela M., Matthew T. Kaufman, David Raposo, Michael B. Ryan, John P. Sheppard, and Anne K. Churchland. 2017. "Posterior Parietal Cortex Guides Visual Decisions in

Rats." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37 (19): 4954–66.

Lima, Susana Q., Tomás Hromádka, Petr Znamenskiy, and Anthony M. Zador. 2009. "PINP: A New Method of Tagging Neuronal Populations for Identification during in Vivo Electrophysiological Recording." *PloS One* 4 (7): e6099.

Linden, Jennifer F., Robert C. Liu, Maneesh Sahani, Christoph E. Schreiner, and Michael M. Merzenich. 2003. "Spectrotemporal Structure of Receptive Fields in Areas AI and AAF of Mouse Auditory Cortex." *Journal of Neurophysiology* 90 (4): 2660–75.

Linke, R., and H. Schwegler. 2000. "Convergent and Complementary Projections of the Caudal Paralaminar Thalamic Nuclei to Rat Temporal and Insular Cortex." *Cerebral Cortex* 10 (8): 753–71.

Lovelace, Christopher T., Barry E. Stein, and Mark T. Wallace. 2003. "An Irrelevant Light Enhances Auditory Detection in Humans: A Psychophysical Analysis of Multisensory Integration in Stimulus Detection." *Brain Research. Cognitive Brain Research* 17 (2): 447–53.

Macaluso, Emiliano, Chris D. Frith, and Jon Driver. 2000. "Modulation of Human Visual Cortex by Crossmodal Spatial Attention." *Science* 289 (5482): 1206–8.

Magland, Jeremy, James J. Jun, Elizabeth Lovero, Alexander J. Morley, Cole Lincoln Hurwitz, Alessio Paolo Buccino, Samuel Garcia, and Alex H. Barnett. 2020. "SpikeForest, Reproducible Web-Facing Ground-Truth Validation of Automated Neural Spike Sorters." *eLife* 9 (May). https://doi.org/10.7554/eLife.55167.

Malone, Brian J., Brian H. Scott, and Malcolm N. Semple. 2007. "Dynamic Amplitude Coding in the Auditory Cortex of Awake Rhesus Macaques." *Journal of Neurophysiology* 98 (3): 1451–74.

McAdams, C. J., and J. H. Maunsell. 2000. “Attention to Both Space and Feature Modulates Neuronal Responses in Macaque Area V4.” *Journal of Neurophysiology* 83 (3): 1751–55.

McGinley, Matthew J., Stephen V. David, and David A. McCormick. 2015. “Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection.” *Neuron* 87 (1): 179–92.

McGurk, H., and J. MacDonald. 1976. “Hearing Lips and Seeing Voices.” *Nature* 264 (5588): 746–48.

Mehta, Ashesh D., Istvan Ulbert, and Charles E. Schroeder. 2000. “Intermodal Selective Attention in Monkeys. II: Physiological Mechanisms of Modulation.” *Cerebral Cortex* 10 (4): 359–70.

Meredith, M. A., J. W. Nemitz, and B. E. Stein. 1987. “Determinants of Multisensory Integration in Superior Colliculus Neurons. I. Temporal Factors.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 7 (10): 3215–29.

Meredith, M. A., and B. E. Stein. 1983. “Interactions among Converging Sensory Inputs in the Superior Colliculus.” *Science* 221 (4608): 389–91.

Meredith, M. A., and B. E. Stein. 1986a. “Spatial Factors Determine the Activity of Multisensory Neurons in Cat Superior Colliculus.” *Brain Research* 365 (2): 350–54.

Meredith, M. A., and B. E. Stein. 1986b. “Visual, Auditory, and Somatosensory Convergence on Cells in Superior Colliculus Results in Multisensory Integration.” *Journal of Neurophysiology* 56 (3): 640–62.

Mishra, Jyoti, Antigona Martinez, Terrence J. Sejnowski, and Steven A. Hillyard. 2007. “Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27 (15): 4120–31.

Mitchell, Jude F., Kristy A. Sundberg, and John H. Reynolds. 2007. “Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4.” *Neuron* 55 (1): 131–41.

Moran, J., and R. Desimone. 1985. “Selective Attention Gates Visual Processing in the Extrastriate Cortex.” *Science* 229 (4715): 782–84.

Morrill, Ryan J., and Andrea R. Hasenstaub. 2018. “Visual Information Present in Infragranular Layers of Mouse Auditory Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 38 (11): 2854–62.

Mozolic, Jennifer L., Christina E. Hugenschmidt, Ann M. Peiffer, and Paul J. Laurienti. 2008. “Modality-Specific Selective Attention Attenuates Multisensory Integration.” *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale* 184 (1): 39–52.

Murata, K., H. Cramer, and P. Bach-y-Rita. 1965. “Neuronal Convergence of Noxious, Acoustic, and Visual Stimuli in the Visual Cortex of the Cat.” *Journal of Neurophysiology* 28 (6): 1223–39.

Nandy, Anirvan S., Jonathan J. Nassi, and John H. Reynolds. 2017. “Laminar Organization of Attentional Modulation in Macaque Visual Area V4.” *Neuron* 93 (1): 235–46.

Niell, Cristopher M., and Michael P. Stryker. 2008. “Highly Selective Receptive Fields in Mouse Visual Cortex.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28 (30): 7520–36.

Niell, Cristopher M., and Michael P. Stryker. 2010. “Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex.” *Neuron* 65 (4): 472–79.

Niwa, Mamiko, Jeffrey S. Johnson, Kevin N. O’Connor, and Mitchell L. Sutter. 2012. “Active Engagement Improves Primary Auditory Cortical Neurons’ Ability to Discriminate

Temporal Modulation." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32 (27): 9323–34.

Nozawa, G., P. A. Reuter-Lorenz, and H. C. Hughes. 1994. "Parallel and Serial Processes in the Human Oculomotor System: Bimodal Integration and Express Saccades." *Biological Cybernetics* 72 (1): 19–34.

O'Connell, Monica Noelle, Annamaria Barczak, Charles E. Schroeder, and Peter Lakatos. 2014. "Layer Specific Sharpening of Frequency Tuning by Selective Attention in Primary Auditory Cortex." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34 (49): 16496–508.

Olcese, Umberto, Giuliano Iurilli, and Paolo Medini. 2013. "Cellular and Synaptic Architecture of Multisensory Integration in the Mouse Neocortex." *Neuron* 79 (3): 579–93.

Olsen, Shawn R., Dante S. Bortone, Hillel Adesnik, and Massimo Scanziani. 2012. "Gain Control by Layer Six in Cortical Circuits of Vision." *Nature* 483 (7387): 47–52.

Otazu, Gonzalo H., Lung-Hao Tai, Yang Yang, and Anthony M. Zador. 2009. "Engaging in an Auditory Task Suppresses Responses in Auditory Cortex." *Nature Neuroscience* 12 (5): 646–54.

Pachitariu, Marius, Nicholas Steinmetz, Shabnam Kadir, Matteo Carandini, and Harris Kenneth D. 2016. "Kilosort: Realtime Spike-Sorting for Extracellular Electrophysiology with Hundreds of Channels." *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/061481.

Pandya, Deepak N., and Benjamin Seltzer. 1982. "Association Areas of the Cerebral Cortex." *Trends in Neurosciences* 5 (January): 386–90.

Park, Hyojin, Robin A. A. Ince, Philippe G. Schyns, Gregor Thut, and Joachim Gross. 2015. "Frontal Top-down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners." *Current Biology: CB* 25 (12): 1649–53.

Paxinos, George, and Keith B. J. Franklin. 2004. The Mouse Brain in Stereotaxic Coordinates. Academic Press.

Pestilli, Franco, Marisa Carrasco, David J. Heeger, and Justin L. Gardner. 2011. "Attentional Enhancement via Selection and Pooling of Early Sensory Responses in Human Visual Cortex." *Neuron* 72 (5): 832–46.

Phillips, Elizabeth A. K., Christoph E. Schreiner, and Andrea R. Hasenstaub. 2017. "Diverse Effects of Stimulus History in Waking Mouse Auditory Cortex." *Journal of Neurophysiology* 118 (2): 1376–93.

Pinsk, Mark A., Glen M. Doniger, and Sabine Kastner. 2004. "Push-Pull Mechanism of Selective Attention in Human Extrastriate Cortex." *Journal of Neurophysiology* 92 (1): 622–29.

Pluta, Scott R., Greg I. Telian, Alexander Naka, and Hillel Adesnik. 2019. "Superficial Layers Suppress the Deep Layers to Fine-Tune Cortical Coding." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 39 (11): 2052–64.

Raz, Aeyal, Sean M. Grady, Bryan M. Krause, Daniel J. Uhlrich, Karen A. Manning, and Matthew I. Banks. 2014. "Preferential Effect of Isoflurane on Top-down vs. Bottom-up Pathways in Sensory Cortex." *Frontiers in Systems Neuroscience* 8 (October): 191.

Recanzone, Gregg H. 2003. "Auditory Influences on Visual Temporal Rate Perception." *Journal of Neurophysiology* 89 (2): 1078–93.

Reynolds, John H., and Leonardo Chelazzi. 2004. "Attentional Modulation of Visual Processing." *Annual Review of Neuroscience* 27: 611–47.

Rikhye, Rajeev V., Aditya Gilra, and Michael M. Halassa. 2018. “Thalamic Regulation of Switching between Cortical Representations Enables Cognitive Flexibility.” *Nature Neuroscience* 21 (12): 1753–63.

Rockland, K. S., and D. N. Pandya. 1979. “Laminar Origins and Terminations of Cortical Connections of the Occipital Lobe in the Rhesus Monkey.” *Brain Research* 179 (1): 3–20.

Rodgers, Chris C., and Michael R. DeWeese. 2014. “Neural Correlates of Task Switching in Prefrontal Cortex and Primary Auditory Cortex in a Novel Stimulus Selection Task for Rodents.” *Neuron* 82 (5): 1157–70.

Roedel, Angelika, Corinna Storch, Florian Holsboer, and Frauke Ohl. 2006. “Effects of Light or Dark Phase Testing on Behavioural and Cognitive Performance in DBA Mice.” *Laboratory Animals* 40 (4): 371–81.

Rougier, Nicolas P., David C. Noelle, Todd S. Braver, Jonathan D. Cohen, and Randall C. O’Reilly. 2005. “Prefrontal Cortex and Flexible Cognitive Control: Rules without Symbols.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (20): 7338–43.

Rutkowski, Richard G., Trevor M. Shackleton, Jan W. H. Schnupp, Mark N. Wallace, and Alan R. Palmer. 2002. “Spectrotemporal Receptive Field Properties of Single Units in the Primary, Dorsocaudal and Ventrorostral Auditory Cortex of the Guinea Pig.” *Audiology & Neuro-Otology* 7 (4): 214–27.

Salin, P. A., and J. Bullier. 1995. “Corticocortical Connections in the Visual System: Structure and Function.” *Physiological Reviews* 75 (1): 107–54.

Sato, T., and J. D. Schall. 2001. “Pre-Excitatory Pause in Frontal Eye Field Responses.” *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale* 139 (1): 53–58.

Schaefer, Michael, Herta Flor, Hans-Jochen Heinze, and Michael Rotte. 2006. “Dynamic Modulation of the Primary Somatosensory Cortex during Seeing and Feeling a Touched Hand.” *NeuroImage* 29 (2): 587–92.

Schroeder, Charles E., and John J. Foxe. 2002. “The Timing and Laminar Profile of Converging Inputs to Multisensory Areas of the Macaque Neocortex.” *Brain Research. Cognitive Brain Research* 14 (1): 187–98.

Schwartz, Jean-Luc, Frédéric Berthommier, and Christophe Savariaux. 2004. “Seeing to Hear Better: Evidence for Early Audio-Visual Interactions in Speech Identification.” *Cognition* 93 (2): B69–78.

See, Jermyn Z., Craig A. Atencio, Vikaas S. Sohal, and Christoph E. Schreiner. 2018. “Coordinated Neuronal Ensembles in Primary Auditory Cortical Columns.” *eLife* 7 (June). https://doi.org/10.7554/eLife.35587.

Seidemann, E., and W. T. Newsome. 1999. “Effect of Spatial Attention on the Responses of Area MT Neurons.” *Journal of Neurophysiology* 81 (4): 1783–94.

Serences, John T. 2011. “Mechanisms of Selective Attention: Response Enhancement, Noise Reduction, and Efficient Pooling of Sensory Responses.” *Neuron*. Elsevier.

Shams, L., Y. Kamitani, and S. Shimojo. 2000. “Illusions. What You See Is What You Hear.” *Nature* 408 (6814): 788.

Shimojo, S., and L. Shams. 2001. “Sensory Modalities Are Not Separate Modalities: Plasticity and Interactions.” *Current Opinion in Neurobiology* 11 (4): 505–9.

Skipper, Jeremy I., Virginie van Wassenhove, Howard C. Nusbaum, and Steven L. Small. 2007. “Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception.” *Cerebral Cortex* 17 (10): 2387–99.

Smith, Philip H., Karen A. Manning, and Daniel J. Uhlrich. 2010. “Evaluation of Inputs to Rat Primary Auditory Cortex from the Suprageniculate Nucleus and Extrastriate Visual Cortex.” *The Journal of Comparative Neurology* 518 (18): 3679–3700.

Song, You-Hyang, Jae-Hyun Kim, Hye-Won Jeong, Ilsong Choi, Daun Jeong, Kwansoo Kim, and Seung-Hee Lee. 2017. “A Neural Circuit for Auditory Dominance over Visual Perception.” *Neuron* 93 (4): 940–54.e6.

Spinelli, D. N., A. Starr, and T. W. Barrett. 1968. “Auditory Specificity in Unit Recordings from Cat’s Visual Cortex.” *Experimental Neurology* 22 (1): 75–84.

Stanford, Terrence R., and Barry E. Stein. 2007. “Superadditivity in Multisensory Integration: Putting the Computation in Context.” *Neuroreport* 18 (8): 787–92.

Stein, Barry E., and Terrence R. Stanford. 2008. “Multisensory Integration: Current Issues from the Perspective of the Single Neuron.” *Nature Reviews. Neuroscience* 9 (4): 255–66.

Stein, Barry E., Mark W. Wallace, Terrence R. Stanford, and Wan Jiang. 2002. “Cortex Governs Multisensory Integration in the Midbrain.” *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry* 8 (4): 306–14.

Stein, B. E., M. A. Meredith, W. S. Huneycutt, and L. McDade. 1989. “Behavioral Indices of Multisensory Integration: Orientation to Visual Cues Is Affected by Auditory Stimuli.” *Journal of Cognitive Neuroscience* 1 (1): 12–24.

Stiebler, I., R. Neulist, I. Fichtel, and G. Ehret. 1997. “The Auditory Cortex of the House Mouse: Left-Right Differences, Tonotopic Organization and Quantitative Analysis of Frequency Representation.” *Journal of Comparative Physiology. A, Sensory, Neural, and Behavioral Physiology* 181 (6): 559–71.

Stirman, Jeffrey N., Ikuko T. Smith, Michael W. Kudenov, and Spencer L. Smith. 2016. “Wide Field-of-View, Multi-Region, Two-Photon Imaging of Neuronal Activity in the Mammalian Brain.” *Nature Biotechnology* 34 (8): 857–62.

Sukiban, Jeyathevy, Nicole Voges, Till A. Dembek, Robin Pauli, Veerle Visser-Vandewalle, Michael Denker, Immo Weber, Lars Timmermann, and Sonja Grün. 2019. “Evaluation of Spike Sorting Algorithms: Application to Human Subthalamic Nucleus Recordings and Simulations.” *Neuroscience* 414 (August): 168–85.

Sumby, W. H., and Irwin Pollack. 1954. “Visual Contribution to Speech Intelligibility in Noise.” *The Journal of the Acoustical Society of America* 26 (2): 212–15.

Talsma, Durk, Daniel Senkowski, Salvador Soto-Faraco, and Marty G. Woldorff. 2010. “The Multifaceted Interplay between Attention and Multisensory Integration.” *Trends in Cognitive Sciences* 14 (9): 400–410.

Thomson, Alex M. 2010. “Neocortical Layer 6, a Review.” *Frontiers in Neuroanatomy* 4 (March): 13.

Toth, Louis J., and John A. Assad. 2002. “Dynamic Coding of Behaviourally Relevant Stimuli in Parietal Cortex.” *Nature* 415 (6868): 165–68.

Wallace, Mark T., Ramnarayan Ramachandran, and Barry E. Stein. 2004. “A Revised View of Sensory Cortical Parcellation.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (7): 2167–72.

Wang, Ye, Simona Celebrini, Yves Trotter, and Pascal Barone. 2008. “Visuo-Auditory Interactions in the Primary Visual Cortex of the Behaving Monkey: Electrophysiological Evidence.” *BMC Neuroscience* 9 (August): 79.

Watanabe, K., and S. Shimojo. 2001. “When Sound Affects Vision: Effects of Auditory Grouping on Visual Motion Perception.” *Psychological Science* 12 (2): 109–16.

Watkins, S., L. Shams, S. Tanaka, J-D Haynes, and G. Rees. 2006. “Sound Alters Activity in Human V1 in Association with Illusory Visual Perception.” *NeuroImage* 31 (3): 1247–56.

Welch, R. B., and D. H. Warren. 1980. “Immediate Perceptual Response to Intersensory Discrepancy.” *Psychological Bulletin* 88 (3): 638–67.

Wepsic, J. G. 1966. “Multimodal Sensory Activation of Cells in the Magnocellular Medial Geniculate Nucleus.” *Experimental Neurology* 15 (3): 299–318.

Williford, Tori, and John H. R. Maunsell. 2006. “Effects of Spatial Attention on Contrast Response Functions in Macaque Area V4.” *Journal of Neurophysiology* 96 (1): 40–54.

Wimmer, Ralf D., L. Ian Schmitt, Thomas J. Davidson, Miho Nakajima, Karl Deisseroth, and Michael M. Halassa. 2015. “Thalamic Control of Sensory Selection in Divided Attention.” *Nature* 526 (7575): 705–9.

Winkowski, Daniel E., Sharba Bandyopadhyay, Shihab A. Shamma, and Patrick O. Kanold. 2013. “Frontal Cortex Activation Causes Rapid Plasticity of Auditory Cortical Processing.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 33 (46): 18134–48.

Yin, Pingbo, Jonathan B. Fritz, and Shihab A. Shamma. 2014. “Rapid Spectrotemporal Plasticity in Primary Auditory Cortex during Behavior.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34 (12): 4396–4408.

Yoshida, Takashi, and Donald B. Katz. 2011. “Control of Prestimulus Activity Related to Improved Sensory Coding within a Discrimination Task.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31 (11): 4101–12.

Zador, A. 1998. “Impact of Synaptic Unreliability on the Information Transmitted by Spiking Neurons.” *Journal of Neurophysiology* 79 (3): 1219–29.

Zempeltzi, Marina M., Martin Kisse, Michael G. K. Brunk, Claudia Glemser, Sümeyra Aksit, Katrina E. Deane, Shivam Maurya, et al. 2020. “Task Rule and Choice Are Reflected by Layer-Specific Processing in Rodent Auditory Cortical Microcircuits.” *Communications Biology* 3 (1): 345.

Zhou, Yi, Bao-Hua Liu, Guangying K. Wu, Young-Joo Kim, Zhongju Xiao, Huizhong W. Tao, and Li I. Zhang. 2010. “Preceding Inhibition Silences Layer 6 Neurons in Auditory Cortex.” *Neuron* 65 (5): 706–17.

Zhou, Yong-Di, and Joaquín M. Fuster. 2004. “Somatosensory Cell Response to an Auditory Cue in a Haptic Memory Task.” *Behavioural Brain Research* 153 (2): 573–78.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
77BC43C85723445...
Author Signature

12/17/2020
Date