Modulation of neural activity in auditory cortex and midbrain by sensory and behavioral context

By

Daniela Saderi

A DISSERTATION

Presented to the Neuroscience Graduate Program

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

March 6, 2019

Table of Contents

Acknowledgments

It was October 2011 when I sent Dr. Larry Trussell an email asking if I could come visit his laboratory at Oregon Health & Science University. I had just finished my Master's degree in Neuroscience and was visiting the States for the very first time. My husband and I were on our honeymoon, a road trip across the Pacific coast. I was supposed to relax and enjoy the vacation. But how could I pass through Portland and not visit the Vollum Institute? I knew that I wanted to study the auditory system, and OHSU was packed with world-renowned auditory scientists. Dr. Trussell not only showed me his laboratory, but organized a series of interviews with other faculty members in the Neuroscience Graduate Program. If I am here today, it is because of people like Dr. Trussell and Dr. Gary Westbrook, who have seen some value in me and have encouraged me to apply to the program. I thank them for believing I could do it. I also thank the rest of my dissertation committee, including Dr. Erick Gallun, for patiently listening to the many experimental challenges I encountered throughout my Ph.D. years and for giving me some wise advice that I should have been more diligent in following.

Dr. Stephen David arrived at OHSU only a few months before me and welcomed me in his new lab as his first rotating graduate student. Even though I was intimidated by words like 'STRF' and 'multiple linear regression' and I had no experience programming, his open and welcoming attitude convinced me that joining his laboratory would be the right choice for me. I became his first Ph.D. student. I thank him for always been so incredibly supportive of my many passions, allowing me to nurture them, even if sometimes that meant being away from the lab. There is no doubt that I would not be where I am today without his support.

One of the many passions I discovered halfway through my Ph.D. journey is advocacy for a more open way of doing science. Robin Champieux gave me countless opportunities to learn and grow as an organizer and facilitator, giving me the confidence I needed to become a leader in a global open movement. I thank her for inviting me to this space, for believing in me and encouraging me, for listening to my rumbles, and for providing thoughtful feedback to my applications. I also thank her deeply for her friendship and for being what effectively is my "open" guru.

But Robin was not alone in doing all of that. With her was another woman of great strength, compassion, character, and intelligence. I thank Danielle Robinson for showing me that even as a student I could make the difference. I followed Danielle's footsteps and I could have not been luckier to have had her come before me. Her wisdom and passion were contagious. I am so grateful I have the honor to continue to learn from her in what is now a friendship and a professional relationship.

Danielle was a founding member of Women in Science Portland (WIS PDX), a group of women students and postdocs who fought for a better parental leave and created a space for women in the sciences to commune and learn from one another. I want to thank all of the incredible women I met at WIS PDX, particularly Lilly, Allison, Astrid, and Jolanda. They have taught me so much about organization, compassion, and self-care.

During this journey there have been many other people who have influenced my path in ways that go beyond making it easier to be at work. I have learned so much from each one of my lab mates. I want to thank Sean for teaching me what being a scientist with integrity means. I thank Henry for being the first person to show me around the lab

and for a friendship that continues to this day. I thank Zack for engaging in conversations around science, life, and social justice, and for always being there when I needed help and moral support. I thank Brad for embarking in the “Chapter 2” project with me, for all the advice, help, kindness and wisdom he always shows me. I admire Brad as a scientist and, most importantly, as a superb human being. I thank Jean and Marlene for showing me what happens when you mix kindness and outstanding intelligence. I thank Jean, Mateo, Itallia and Jacob for playing D&D with me and not laughing too much at me when I brought a Mind Flyer to the game. I also thank Luke, Jesyin, and Charlie for making the lab fun, for their thoughtful feedback on my work, and for always helping me when I needed.

More friends that made this journey memorable are Carolina, Rachelle, Rita, Katrina, Luci, Glynis, and Heather. I thank them for their friendship and constant support, for the chocolate, and for cheering me up during some of the toughest times. I also thank Brian for not letting me be the only one asking questions in class and for waiting to graduate so that I would not feel alone as a 7th year graduate student.

We got to that point in which I have to acknowledge and thank the people without whom I truly would have never made it. First, my mom and dad, who have sacrificed so much to get me where I am today, and have supported me in all my “unusual” life choices even if they not always understood the why. I thank my mom for raising me to be an independent and strong woman, but also a kind and a caring one (well, she tried, at least!). And my grandparents, whose hard work and sacrifice are what enabled me to dedicate my life to science, as a first-generation college student. I’m so proud to make them proud. They are the stars that will never fade in my sky. And my best friend,

Tiziana. I thank her for being who she is, a never-ending source of light, love, and compassion. Her friendship over the past 15 years has been everything. I'm so proud to be her best friend and cannot wait to do 'all things USA' with her now that she moved to Oregon for her postdoc. Her talent, tenacity, and kindness are always there to show me the way.

Last but certainly not least is my husband, Ivar. I have so much to thank him for. He was there through it all. GRE, English test, grad school application, qualifying exam, NRSA application, Mozilla application, thesis writing, and all the crises in between. He has been the pillar of my life, never letting me break or give up, always encouraging and believing I could do anything. If I made it, it is because he helped me see I could. He is and will always be my sunshine.

Abstract

The overall focus of my doctoral research was to characterize how sensory and behavioral context affect how the auditory brain perceives and interprets sound. In Chapter 1, I overview the anatomical and physiological background for the experimental work presented in the following chapters. In Chapter 2, we explored the neural bases of auditory streaming in the ferret cued by the repetition of a complex sound in the presence of a changing mixture. In humans, sound repetition evokes a strong pop-out effect that leads to the separation of the repeated sound in the foreground in the context of a changing background composed of a mixture of noise. We first confirmed that ferrets were able to perceive complex sound repetitions as distinct sensory objects. We then recorded single- and multiunit extracellular activity in response to these stimuli in core (A1) and belt (PEG) regions of the ferret auditory cortex. We found that auditory responses were reduced in response to the repetition of any given sound, likely as a consequence of neural adaptation. However, activity evoked by the repeating, foreground stream was selectively enhanced compared to the background, an effect that was more prominent in PEG and for units whose responses were tuned to the repeating sound. Taken together these results provide evidence for stream segregation that emerges in A1 and is refined in PEG. In Chapter 3, we combined behavior, electrophysiological and pupil size recordings, and modeling to dissociate the modulatory effects of two state variables, task engagement and arousal, on auditory processing at the level of the auditory midbrain and cortex. We found that arousal as indexed by pupil size accounted for a large component of the activity modulation between behavioral contexts. Because task- and arousal-related effects on neural activity were correlated, we found that many of

the units – particularly in the IC – that would have been counted as modulated by task in a more traditional analysis, were in fact modulated either solely by arousal or by both. Furthermore, in IC, but not in A1, units with weaker auditory responses showed stronger state modulation. Taken together, these results demonstrate that task engagement and arousal can be dissociated in most neurons. This approach provides a general method for dissociating the influence of continuous and discrete behavioral state variables on sensory representation.

Chapter 1: Introduction

The auditory system detects sound derived from the mechanical disturbance of particles in a medium (for humans, usually the Earth's atmosphere), propagating as alternating compression and rarefaction pressure waves generated by the vibration of an object, *i.e.* a sound source. The anatomical structures involved in sound processing have evolved over millions of years across different animal lineages to optimize hearing as an active process. Hearing requires the integration of the physical properties of sound, such as amplitude and frequency, with information relevant to its interpretation within the specific context in which auditory perception occurs. In this introduction, I provide an anatomical and physiological overview of auditory processing to frame the experimental work presented in the following chapters.

The auditory brain coordinates the meaningful integration of auditory information with information from other senses, and uses internal and external cues to inform the correct interpretation of the incoming sound to guide behavior. Essential to this process is separating and distinguishing sound sources. In Chapter 2, using ferrets as an experimental model, we present evidence to support a hierarchical theory in which repetition facilitates sound source segregation. This process appears to begin in primary auditory cortex and be refined in secondary auditory cortical regions.

Another remarkable ability of the auditory brain is its adaptability to an ever-changing environment. It is well-established that in higher levels of the auditory pathway, auditory perception relies extensively on past experience, attention, and emotional state (Cherry 1953; Dai et al. 1991; Shinn-Cunningham & Best 2008). These state variables

likely do not operate in isolation, making it difficult to parse out the unique contributions of each to auditory processing. Therefore, state-dependent modulation of the auditory brain has not yet been integrated into a coherent model of auditory representation. In Chapter 3, we present an approach that allowed us to investigate the simultaneous contributions to firing rate modulations of two state variables, task engagement and arousal, in the inferior colliculus and primary auditory cortex.

Overview of the auditory system: structure and connectivity

Sensory systems are classically described in terms of a hierarchical organization of the brain regions that comprise them. In the case of the auditory system, we observe a progressive transformation of sound representation from the cochlea to the auditory cortex. Physical attributes of sound are encoded more faithfully in "lower level" regions (auditory nerve and brainstem nuclei), whereas abstract properties related to the perception and meaning of the stimulus are encoded in "higher level" regions (auditory cortex and above) (Bizley & Cohen 2013). The anatomical bases of such organization are composed of intricately organized networks of ascending and descending auditory projections (Winer & Lee 2007) (see Figure 1.1 for a schematic representation of this circuit).

Ascending auditory projections start in the ear and run in parallel all the way up to the cortex. In mammals, sounds waves are funneled into the ear canal by the auricle (also known as pinna), and are mechanically transmitted to the sensory organ for hearing, the organ of Corti, situated in the cochlear partition of the inner ear. It is here that vibrations are transduced into electrical and chemical signals. In response to receptor potentials

generated by the cochlear hair cells, auditory nerve fibers (collectively forming the VIII cranial nerve) transmit information to the brainstem in the form of a pattern of action potentials. Starting at the first brain auditory region, the cochlear nucleus, sound information diverges into a number of tracts converging at the obligatory midbrain auditory structure, the inferior colliculus (IC). From there, auditory information is transmitted to the auditory thalamus, and relayed to the auditory cortex (AC).

Starting in the midbrain, ascending pathways are subdivided anatomically into two systems known as the primary (lemniscal) and secondary (non-lemniscal) projections (Figure 1.2). Physiological studies have shown that primary and secondary neurons across auditory regions are different in many ways, and potentially play different roles in auditory perception (Elgueda et al. 2019; Lee 2015; Rauschecker & Tian 2000).

Descending projections from the AC connect with virtually every nucleus in the midbrain, brainstem, all the way to the periphery. This feedback network is less understood, but it is thought to be tasked with selecting the salient information that can then guide behavior (Winer & Lee 2007).

Because my dissertation focuses on the IC and primary and secondary regions of the auditory cortex, I will now review the known anatomical and physiological properties of these areas.

The inferior colliculus

Although sound frequency and intensity are directly encoded in the ear, some aspects of sound localization require the integration of cues that are extracted from the

spectral, temporal, and level differences collected from both ears. This integration starts in the IC (Winer & Schreiner 2005).

The IC sends projections to several portions of the auditory thalamus (*i.e.*, medial geniculate body or MGB) (Andersen et al. 1980, Huffman & Henson), as well as feedback projections to almost all brainstem nuclei from which it receives input (Huffman & Henson). The IC receives ascending projections from almost all parts of the brainstem auditory nuclei (Masterton et al. 1992; Morest & Oliver 1984; Oliver 1987; Saint Marie et al. 1997), and descending input from the MGB. It also receives a large efferent projection directly from the auditory cortex (Winer et al. 1998). This information is integrated locally by a complex network of intrinsic and commissural connections (Aitkin & Phillips 1984; Malmierca et al. 1995; Oliver et al. 1991; Saldaña & Merchán 1992).

Anatomical and cytoarchitectonic properties of the IC have been used to assign three main sub-regions: the central nucleus (CN), the lateral nucleus (LN), and dorsal cortex (DC) (Figure 1.2), each of which has different neuronal structure (Morest & Oliver 1984), connections (Rockel & Jones 1973a,b), and functional properties (Semple & Aitkin 1979). The CN is the largest division of the IC, and it is tonotopically organized into distinct iso-frequency laminae along a dorso-medial axis (Huang & Fex 1986). Neurons in the CN receive direct input from auditory brainstem nuclei (lemniscal pathway) and tend to have characteristic short response latency and narrow bandwidth tuning (Aitkin & Moore 1975; Aitkin et al. 1975; Moore et al. 1983; Slee & David 2015a).

A classical model describes the LN and DC as areas predominantly contacted by auditory fibers descending from the thalamus and cortex, as well as the main destination of neuromodulatory and other sensory input (Winer 2005). However, thanks to new techniques of cell-specific monosynaptic tracing, we now know that external cortex IC regions do in fact receive a large ascending projection (Chen et al. 2018). Neurons in these regions typically have longer response latencies, considerably broader auditory tuning, and, in general, lack consistent tonotopic organization (Slee & Young 2013; Syka et al. 2000).

Converging anatomical and physiological evidence has shown that, in addition to auditory information, the IC processes multisensory information, such as visual, oculomotor, and somatosensory, as well as signals relating to behavioral context, motivation, and reward contingencies (Gruters & Groh 2012). The major sources of neuromodulatory input targeting the IC are the locus coeruleus and the dorsal raphe nucleus (Coleman & Clerici 1987; Klepper & Herbert 1991). Furthermore, the large corticofugal system has been hypothesized to be implicated in context-dependent modulation of IC activity either directly via the cortico-collicular (Winer et al. 1998) and thalamo-tectal (Kuwabara & Zook 2000) systems, or indirectly via midbrain modulatory centers, such as the pedunculopontine and latero-dorsal tegmental nuclei (Chen et al. 2018; Motts & Schofield 2009). Interestingly, top-down projections from other auditory areas, multisensory input, and most neuromodulatory fibers tend to overlap in the same non-lemniscal regions of IC, suggesting that neuromodulators play a key role in modulating the convergence of ascending auditory information with descending auditory and non-auditory information.

The auditory cortex

The organization of the auditory cortex is relatively well conserved from insectivores to primates, despite vast differences in terms of number of areas, relative position, connections, and topographic organization (Hackett 2011; Winer 1992) (Figure 1.3). The regional boundaries are based on cytoarchitectonic, physiological, and connective properties, and often reflect methodological artifacts rather than actual functional differences (Oliver 2005). A common theme across auditory cortices in different species, is the presence of a core region, referred to as the primary auditory cortex (A1) (Luethke et al. 1988) (Figure 1.3). This region receives direct auditory input from lemniscal areas of the MGB (mainly ventral divisions). A1 neurons tend to respond reliably to pure tones, and do so with relatively short latencies and a frequency preference clearly organized along a tonotopic axis (Ehret & Schreiner 1997; Heil et al. 1994; Schreiner & Mendelson 1990), the orientation of which varies across species (Figure 1.3).

Systematic variations in the response properties of AC have been used as a basis for segregating different auditory fields into distinct processing streams (Griffiths & Warren 2004; Rauschecker & Tian 2000; Read et al. 2002). Primary *core* regions are surrounded by secondary *belt* regions, in which sound-evoked onset activity of neurons is delayed with respect to core regions, less reliable across repetitions, and have broader bandwidth (Rauschecker 1998). Secondary auditory areas are predominantly innervated by primary auditory areas (A1), and also receive input from non-lemniscal regions of the auditory thalamus (mainly medial and dorsal divisions).

In human and non-human primates, there is also a tertiary region referred to as the *parabelt* (Kaas & Hackett 2000; Moerel et al. 2014). Tertiary regions are primarily innervated by secondary areas and project to auditory association areas, which have been related to decision-making, such as the dorsolateral prefrontal cortex (dlPFC). Neurons in these areas tend not to respond to pure tones but rather preferring complex stimuli, such as vocalizations and other natural sounds (Hackett et al. 1998a,b). Furthermore, a recent study showed that neurons in the rostral ventral posterior auditory field (VPr) of the tertiary cortex of the ferret exhibited context-dependent changes in auditory responses, encoding non-acoustical sound features such as associated behavioral meaning and task timing (Elgueda et al. 2019).

The ferret as an animal model for sensory processing

In this thesis we used the ferret as an animal model. The ferret provides an ideal preparation for studying high-order auditory representations because of its rich behavioral repertoire (Bajo et al. 2010; Bizley et al. 2013; David et al. 2012; Fritz et al. 2003) and well-characterized auditory network (Atiani et al. 2014; Bizley et al. 2005, 2015). The well-defined auditory network provides potential advantages over mice, where the cortical organization is less clearly hierarchical (Bandyopadhyay et al. 2010; Hackett 2011). At the same time, the ferret allows for substantially lower costs and handling complexity than non-human primates, providing a valuable, complement for studies of brain networks.

The ferret IC is located under the cerebellum, about 7-9mm deep below the surface of the cortex. It receives bottom up projections, mostly contralateral, from

virtually every regions of the auditory midbrain (Moore 1988) and top down projection from the auditory thalamus and cortex (Bajo et al. 2007).The ferret auditory cortex is located in the temporal lobe and occupies a large portion of the ectosylvian gyrus (Bizley et al. 2005) (Figure 1.3). Core regions, A1 and the anterior auditor field (AAF), are located in the medial part of ectosylvian gyrus. Belt regions are located more posteriorly, and are named posterior pseudosylvian field (PPF), and posterior supra-sylvian field (PSF). Parabelt regions are positioned more anteriorly, and comprise the anterior dorsal field (ADF) and the anterior ventral field (AVF).

Sensory- and behavior-dependent modulation of sound representation in the auditory system

In the past 20 years, evidence from human and non-human studies has revealed that the neural encoding of sound, particularly in cortical regions, does not solely depend on the sensory input. Rather, it is heavily influenced by the *context* in which such input is presented and perceived. With the term "context", I here refer to both the *sensory* conditions in which a given sound is presented, and the *behavioral* state of the animal receiving sensory stimulation. Many of these studies have challenged the classical and simplistic model that views primary sensory areas as static encoders of the attributes of the physical stimulus. In order to understand how the brain allows humans and other animals to successfully experience, react, and adapt to complex, ever-changing sensory environments, it is important to explore mechanisms of sensory processing that include and control for context variables that are likely to affect sound representation.

Sensory context effects have been associated with the ability of auditory neurons, both cortically and subcortically, to rapidly adapt to the statistics (*e.g.*, spectral and temporal components) of sound (Bendixen et al. 2012; Winkler et al. 2009). The effects of behavioral context reflect relatively slower changes in sound-evoked activity occurring as a consequence of changes in the task structure and reward/penalty association, or changes in the internal state of the listener, *e.g.*, arousal, sleep, and state of anesthesia (Osmanski & Wang 2015).

Sensory context: Streaming of complex sounds in the auditory cortex

Separating and distinguishing sound sources are essential processes of hearing. Complex mixtures of sounds generated by different sources often arrive at the ears simultaneously or in close succession, and yet are effortlessly decomposed by the auditory brain into distinct sound sources (Griffiths & Warren 2004). The computational difficulty of this task is emphasized by the fact that for subjects with hearing impairments and for speech recognition devices, it is extremely hard to decompose sound signals and correctly estimate the sound of interest embedded in a mixture (Divenyi 2004; Marrone et al. 2008; McDermott 2009; Wang & Brown 2006).

Psychologist Albert S. Bregman was the first to extensively describe these processes and to popularize them under the name of *auditory scene analysis* (Bregman 1990). The principles of auditory scene analysis are simple and based on the statistical analysis of streams of sounds; auditory streams that come from the same sound source share common spectro-temporal properties that make them separable (*i.e.*, “stream

segregation") from other sources, and perceivable as unique acoustic events (Bregman 1990; Schnupp et al. 2012; Winkler et al. 2009).

The principles of stream segregation have been extensively investigated using sequences of pure tones, denoted "A" and "B", presented in alternating patterns (Bregman 1990; Moore & Gockel 2012; van Noorden 1975). One common pattern is the "ABAB" pattern, which generates a different percept based on the frequency separation (ΔF) and presentation rate (PR) of the two tones (Figure 1.4A). If the ΔF is small (<10%) and the PR is slow (<10Hz), a human subject would perceive the sound as a coherent alternating sequence, a single stream. However, if the ΔF is large or the PR fast, the stimulus would be perceived as two separate streams (Figure 1.4C). Intermediate cases are ambiguous as they can be perceived as either one or two streams, in a so-called bistable percept (Pressnitzer & Hupé 2006; Schnupp 2008; Schnupp et al. 2012). Furthermore, the perception of two streams does not occur instantly following the alternated presentation of two successive sounds. Instead, the probability of segregation to occur increases with time of sequence exposure (Anstis & Saida 1985; Bregman 1978a). This "buildup" effect suggests that the auditory brain accumulates information over time, and the final percept depends on the information available in a specific sensory context.

Early models of sound segregation describe the perceptual organization of sequential sounds as based on spectral differences, because two tones at different frequencies lead to two or more non-overlapping (or weakly overlapping) activations of different "channels" in the cochlea (Beauvois & Meddis 1996; Hartmann & Johnson 1991; McCabe & Denham 1997). This place-based model of segregation is supported by

a series of seminal findings by Fishman and colleagues. The authors presented un-anesthetized macaque monkeys with alternating A and B tones at different PRs, while recording multiunit activity from A1 regions tuned to the frequency of A tones but not to the B tones (Fishman et al. 2001, 2004). As the PR and pitch separation of A and B tones increased, the firing rate increased in response to the A tones, while it decreased in response to the B tones. These findings have since been replicated (Bee & Klump 2004, 2005; Bidet-Caulet & Bertrand 2009; Kanwal et al. 2003; Micheyl et al. 2007a) and found a possible explanation in the phenomenon of "forward masking", *i.e.*, the reduction in the neural response to a stimulus due to a preceding stimulus (Brosch & Schreiner 1997; Calford & Semple 1995). These results are consistent with what is sometimes referred to as the "peripheral channeling" or "population separation" hypothesis, in which peripheral and pre-attentive neural mechanisms underlying frequency selectivity and differential forward masking are described as the main contributors to the perceptual segregation of sequential acoustic events (Beauvois & Meddis 1996; McCabe & Denham 1997).

Although most models based on the population separation hypothesis have successfully accounted for many important aspects of perceptual integration and separation of sequential streams of pure tones, they fail to account for the observed influence of time in synchronous presentations of the same sound sequences (Figure 1.4B). Elhilali *et al*. (2009) showed that sequences of tones separated by an octave or more – which would be perceived as two streams if presented in alternation – are perceived as a single stream when presented synchronously and fully coherently in time (Elhilali et al. 2009a; Shamma et al. 2011). In addition, psychophysical and physiological

studies conducted in humans using simultaneous, multi-talker paradigms, have demonstrated that even if the frequency spectra overlap, one stream of speech can be perceptually segregated from another as long as they differ on any other perceptual dimension, such as pitch (Akeroyd et al. 2005; Mesgarani & Chang 2012), timbre (Cusack & Roberts 2000; Roberts et al. 2002; Singh & Bregman 1997), or spatial location (Carlyon 2004; Micheyl et al. 2007a). These results led Elhilali and collaborators to formulate an alternative model for sound segregation based on the *temporal coherence* between responses that encode various features of a sound source (Elhilali et al. 2009a; Shamma et al. 2011). In addition to spectral separation, the model proposes that temporal relationships between sound elements are key factors used by the auditory system to perceptually organize acoustic scenes (Elhilali et al. 2009a; Micheyl et al. 2013b,a; Shamma et al. 2011). Psychophysical (Teki et al. 2013) and physiological (Christiansen et al. 2014; O'Sullivan et al. 2015; Teki et al. 2016) studies have since provided experimental evidence in support of this model.

The discussed streaming models considered here are primarily based on pre-attentive processes. With pre-attentive or "bottom-up" processes, I here refer to auditory processes that are driven solely by stimulus physical characteristics, and therefore presumably do not require "top-down" processes, such as prior knowledge or focused attention to the sounds. However, it is well established that auditory experience is influenced by central and top-down mechanisms of learning and attention (Bee & Micheyl 2008; Elhilali et al. 2009b; Shinn-Cunningham & Best 2008). Support for a theory of sound segregation that includes central mechanisms in addition to bottom-up mechanisms, is provided by investigations of the role of attention in auditory streaming.

For example, upon presentation of ambiguous combinations of sequential streams, human listeners are able to intentionally bias their perception toward grouping or streaming (Pressnitzer & Hupé 2006). Another example is that the buildup of stream segregation can be reset by changes in the ear of stimulus presentation or by experimentally controlled shifts in attention (Cusack et al. 2004; Snyder et al. 2006).

In summary, although stream segregation is promoted by several acoustic cues whose contribution can be explained by simple mechanisms of peripheral activation of different frequency channels, it is likely that central mechanisms involving auditory and non-auditory brain regions are tasked with selectively enhancing these attributes (Alain & Arnott 2000; Alain & Woods 1994; Snyder et al. 2012; Sussman & Steinschneider 2006).

Repetition as a cue for stream segregation

Complex auditory scenes with sound sources overlapping in both spectral and temporal dimensions, represent a common everyday experience for hearing animals, including humans. Sounds emanating from animate and inanimate objects in the environment tend to be statistically regular and often repetitive (*e.g.*, water falling, flapping wings, locomotion sounds). Thus, the ability to detect regularities within the incoming sensory input is a critical aspect of scene analysis.

The temporal coherence model highlights the role of time in the formation of separate streams. Teki *et al.* (2016) demonstrated that human listeners are highly sensitive to the *repetition* of a sound presented in the context of a changing mixture of chords. Perceptually, the repeating sounds are fused together into a "foreground" that pops out from a randomly changing "background" (Teki et al. 2011, 2013). McDermott,

Wrobleski, and Oxenham (2011) demonstrated that for human listeners repetition alone is sufficient to induce segregation of complex sounds (McDermott et al. 2011). The authors used a set of artificially-generated, naturalistic stimuli that lacked first-order statistics of natural sounds carrying major grouping cues for auditory streaming (*e.g.*, common onset and harmonicity), while retaining second-order correlations of natural sound statistics (*e.g.*, high nearby spectro-temporal correlations) (Figure 1.5 A-B). Human subjects presented with sequences of temporally overlapping mixtures of these sounds were able to successfully report the occurrence of a repeating novel "target" sound if it appeared in the context of non-repeating, spectrally diverse mixture of other sounds. Similar to what was shown by Teki and coworkers with sequences of tone pips, perceptually, the repeated sequence popped out as a foreground from a simultaneously presented changing background. These findings are consistent with the observation that, in natural environments, sounds that come from the same source tend to repeat with consistency (*i.e.*, are co-modulated) and therefore are grouped by the brain into a single object (Bizley & Cohen 2013).

Listeners can use prior knowledge of specific features of sounds. For example, in the case of speech, humans can isolate the presence of a word or a phoneme presented simultaneously with another sound (Warren 1970). Listeners also assume that frequency components that are regularly spaced (Cheveigné et al. 1998; Roberts & Brunstrom 1998) start and end at the same time (Darwin & Ciocca 1992), come from the same spatial location (Best et al. 2005), or belong to the same source. However, because the complex naturalistic sounds used by McDermott *et al.* were artificially generated and could not have mapped into any meaningful object in the memory of their listeners, the authors

concluded that prior knowledge of sounds is not required to stream sources using repetition as the sole cue for stream segregation (McDermott et al. 2011). In a subsequent experiment, the authors showed that streaming sound sources via repetition is likely a pre-attentive effect. Human listeners engaged in a decoy visual task were also able to correctly report the identity of a sound when repeated in the context of a changing background (Masutomi et al. 2015).

In Chapter 2 of this dissertation, we used a very similar paradigm to McDermott *et al*. (2011) to investigate the neurophysiological underpinnings of streaming cued by repetition using the ferret model. Our findings suggest that streaming of repeated sequences might engage mechanisms of adaptation, similar to those described below.

A dominant idea in studies of sensory context is that the brain's ability to detect regularities is not only crucial for identifying an auditory object embedded in a busy scene, but also allows subjects to form predictions about the environment, thereby making the system sensitive to deviance. Deviance detection has been extensively explored in humans, with experiments often combining behavior with electroencephalography (EEG) recordings and event-related brain potentials (ERPs). ERPs are time-locked events extracted from the ongoing EEG recording. The composite ERP component mismatch negativity (MMN) has been widely used as an index of deviance detection and to assess processes of auditory streaming. MMN is elicited by sound input violations in a regularly presented pattern of sounds (Näätänen 2001).

At the level of single unit activity, a parallel phenomenon to MMN has been described as stimulus-specific adaptation (SSA) (Pérez-González & Malmierca 2014;

Ulanovsky et al. 2003). SSA is a particular case of response adaptation; while the response to a repeated sound adapts, a new "oddball" sound presented at random times in the sequence elicits a rapid and strong response. This adaptation specific to the stimulus is thought to arise from a combination of feedforward synaptic depression and local cortical inhibition (Ayala & Malmierca 2013; Natan et al. 2015; Yarden & Nelken 2017). Although the exact correlation between SSA and MMN remains debated, some have argued that SSA is not the direct substrate of MMN because the two phenomena differ in latencies, NMDA receptor-dependence, and sensitivity to certain regularities (Khouri & Nelken 2015).

SSA was first observed in single-unit activity in auditory cortex, but has since been described also in the IC (Ayala & Malmierca 2013; Duque & Malmierca 2015; Malmierca et al. 2009) and in the auditory thalamus (Anderson et al. 2009; Antunes et al. 2010; Malmierca et al. 2015). Interestingly, it appears to be stronger in non-lemniscal divisions of these auditory nuclei (Pérez-González & Malmierca 2014).

Given that the majority of sound stimuli that have been used to investigate deviance detection have been very simple (*i.e.*, sequences of repeated tones), it has been hypothesized that the adaptation of neural response is likely a major contributor to the observed deviant responses (Briley & Krumbholz 2013; Grill-Spector et al. 2006; Nelken 2014). Another hypothesis sees deviant response as a phenomenon arising from the neural processes associated with a circuit primed to report any mismatch detected between incoming sensory input and an expected template of what that input might look like given some learned expectations (Daikhin & Ahissar 2012; Khouri & Nelken 2015; Taaseh et al. 2011).

The results presented in Chapter 2 indicate that in A1 and even more so in PEG, there are units that may hold the signature of streaming sounds as a virtue of its repetition in the context of changing backgrounds. We found that streams containing the repetition were often less suppressed or even enhanced compared to the background streams of random mixtures. Even though our findings contradict what we would predict from SSA – responses to the repeating stream were relatively enhanced with respect to the background stream – mechanisms and circuits of adaptation similar to those underlying SSA may be used to stream sequences of repeating objects. These brain processes would make it possible for the brain to selectively enhance a stream of repeating sound, exploiting its predictable statistical structure to extract it as a separate source (Rubin et al. 2016; Winkler et al. 2009).

Behavioral context: state-dependent modulation in the AC and IC

Over the past 20 years, a large number of studies have shown that the information represented in primary cortical areas is strongly influenced by the behavioral state of the animal and by the association between a given sound and contingencies of the task. Non-sensory variables such as motor activity, arousal, learning, and task engagement have been found to strongly modulate responses in primary visual (Niell & Stryker 2010; Shuler & Bear 2006), somatosensory (Petreanu et al. 2012; Sachidhanandam et al. 2013), and auditory cortices (Downer et al. 2015; Niwa et al. 2012; Otazu et al. 2009; Rodgers & DeWeese 2014; Schneider et al. 2014). These studies have challenged the traditional assumption that sensory primary cortical regions are simple sensory analyzers, and that comprehension and behavioral relevance are confined to non-sensory areas farther away

from the periphery (Campbell 1905). Moreover, these studies have helped reshape our understanding of the adult brain and its components from that of a static organ, to an ensemble of cells infused with possibility. Given that the brain is highly plastic, it should not be surprising that the activity of sensory neurons can be altered by a number of non-sensory variables. In the next section, I review some of the key studies that have led scientists to these conclusions, highlighting findings that are relevant to the questions investigated in this dissertation.

Task-related plasticity in primary auditory cortex and inferior colliculus

The first controlled evidence of adult plasticity in A1 was provided by Weinberger and colleagues in a series of pioneering experiments in the mid-1990s. They showed that the tuning of A1 neurons shifted to match the frequency of pure tones presented as conditioned stimuli in a classical conditioning paradigm (Bakin & Weinberger 1990; Bakin et al. 1992; Weinberger 1997). These results demonstrated that the plasticity of receptive fields of A1 neurons is *associative*, as it requires pairing of a conditioned acoustic stimulus, a pure tone, and an aversive stimulus, a mild shock; and it is *specific*, as the maximum increase in response happened at the conditioned stimulus frequency, while neighboring frequencies remained unchanged or decreased their firing rates (Bakin et al. 1992).

A few years later, Fritz and colleagues used instrumental conditioning to train ferrets on an avoidance go/no-go tone detection task (Figure 1.6A). Ferrets were instructed to freely lick from a water spout during the presentation of broadband reference stimuli, temporally orthogonal ripple combinations (TORCs) (Klein et al.

2000), but to promptly refrain from licking upon presentation of a target tone. Failure to lick during the target window would lead to a mild shock to the posterior paw, which acted as a negative reinforcer (Fritz et al. 2003, 2005b). To investigate stimulus-specific changes related to task engagement, the authors measured spectro-temporal receptive fields (STRFs) of A1 units derived from the responses to the TORCs, and compared them between active engagement and passive listening (Figure 1.6A). An STRF is a linear encoding model that uses a weighted version of the stimulus spectrogram (a measure of the energy in the stimulus at each frequency and time) to predict the neuronal discharge rate. The multiunit data showed a population enhancement of the responses to the target tone and suppression at adjacent frequencies (Figure 1.6B, left).

These results corroborated Weinberger and colleagues' findings that auditory cortex plasticity can be highly specific in the adult brain. However, the findings differed on a few important points. In Fritz *et al*., animals were not anesthetized, and receptive field plasticity occurred rapidly (after as few as five trials) as the ferrets switched from passive listening to the task stimuli to active engagement in which the same stimuli had a learned association with reward and penalty. Moreover, while in the case of Weinberger *et al*., receptive field changes in the direction of the conditioned frequency persisted for days and even weeks after a single 30-trial conditioning (Weinberger et al. 1993), in Fritz *et al*. (2003), receptive field changes for many neurons tended to rapidly revert back to the passive listening (baseline) condition.

These results and results from similar experiments by the same group (Fritz et al. 2005b, 2007b) led to a model of real-time adaptive plasticity, in which the auditory system changes its filter properties to optimize the ability to discriminate task relevant

sounds by enhancing the target (in this paradigm paired with a shock) and suppressing the references (paired with reward) (Fritz et al. 2007a). However, when the same group attempted to replicate the findings in a positive reinforcement, approach version of the task – one in which animals had to refrain from licking a water spout during the TORC reference stimuli and lick only to report the occurrence of a target tone to get a water reward (Figure 1.6A) – they found the opposite pattern of plasticity: for several recorded units in A1, active behavior responses were suppressed near the target frequency of neurons whose BF matched the one of the target compared to the passive condition (David et al. 2012) (Figure 1.6B,C). The results were in agreement with previous results from Otazu *et al*., who found task engagement effects in rat A1 was also suppressive of sounds-evoked activity (Otazu et al. 2009). David *et al.* (2012) concluded that reward contingencies were likely responsible for this inversion of the sign of plasticity.

Another aspect of task-related plasticity investigated by the Fritz, Shamma research group, is the influence of task difficulty on the magnitude of these changes. Atiani and coworkers modified the avoidance task used in Fritz *et al*. (2003) such that the target was no longer presented alone, but embedded in one of the TORCs that acted as a masker (Atiani et al. 2009). By changing the signal-to-noise ratio (SNR) of the tone/masker, they were able to titrate the difficulty of the task. Under these circumstances, they found that A1 units tuned near the target frequency displayed an enhanced sensitivity at their BF, while those far from it experienced suppressed activity. Furthermore, these effects were stronger in experiments when behavioral performance was best, indicating that the strength of plasticity might be tightly related to motivation and reward. These results suggest that as perceptual demand increases, greater effort is

required to perform the task, and the increased effort might be what is reflected in cortical activity.

So far, I have presented an overview of task-related plasticity focused on work in primary auditory cortical regions. A few studies have also explored how auditory representation emerges and transforms along the auditory pathway. Slee and David (2015) conducted an experiment with stimuli, animal model, and behavioral paradigm matching the experiment in David *et al*. (2012) to allow for a direct comparison (Slee & David 2015a). The authors computed STRFs from single-unit responses recorded in the IC while animals performed the approach version of the tone discrimination task, and compared them to the STRFs computed with responses during the passive condition. During training the target tone frequency was selected randomly – but spanned a broad range of the ferret's frequency range, whereas during recording it was selected to either match (on-BF) or be at or above 0.5 octaves (off-BF) from neurons' best frequency (BF). Just as in A1, local gain changes measured by the STRF at the target frequency were suppressed in both the central nucleus (CN) and external cortex areas of the IC (LN and DC) (Figure 1.6C). Moreover, these effects were stronger when the target frequency was on-BF. However, global gain changes measured across frequencies tended to be also suppressive in IC, while the same analysis performed on the A1 data showed that global gain changes were equally likely to be suppressed or enhanced (Slee & David 2015a).

To investigate task-related modulation in higher level regions of auditory processing, similar experiments have been performed in ferret belt areas and in frontal cortex (Atiani et al. 2014; Fritz et al. 2010; Niwa et al. 2013). Here behavioral

modulation was found to be stronger, gradually moving towards a more abstract representation of task-related categories and behavioral outcome. Taken together these results support a hierarchical model of auditory representation in which a selective encoding of an attended object begins subcortically and emerges gradually as information ascends along the cortical pathway.

State-dependent modulation of auditory activity

The rapid changes in sound-evoked activity described in the previous section are strong evidence in favor of an active mechanism of re-tuning that develops within minutes as listeners engage in learned associative behaviors. Such modulations are often described using encoding models of neural activity that use both stimulus and context variables as model parameters (David 2018). However, several studies, many of which are cited above, have only focused on a single aspect of behavioral context, such as task engagement, and explored its influence on neural activity as a binary variable (*e.g.*, passive/active). Although this discretization is convenient as it increases the statistical power and can capture changes across trials in each condition, it likely fails to capture the granularity of other changes that might be associated with fluctuations of factors related to the internal state of the animal (Brody et al. 2003; Ecker & Tolias 2014; McGinley et al. 2015b).

It is increasingly clear that factors that are not purely sensory can be responsible for some of the variability in sensory neurons' excitability (Kato et al. 2012; Luck et al. 2013; McGinley et al. 2015b; Schwartz & David 2015). Factors like arousal, indexed by changes in pupil size, locomotion, and whisking, for example, have been shown to

change smoothly in non-anesthetized animals over the course of an experiment, and to track fluctuations of sensory neurons' membrane potentials and behavioral performance (Cohen & Maunsell 2010; Goris et al. 2014; McGinley et al. 2015a; Niell & Stryker 2010; Poulet & Petersen 2008; Reimer et al. 2014; Vinck et al. 2015).

In our laboratory, we found that during passive recordings of single neurons in A1 of non-anesthetized ferrets presented with repetition of two 3s-long ferret vocalizations, pupil size varied considerably under constant luminance (Figure 1.7A). In some neurons, this variation in pupil size was tightly coupled with trial-by-trial variability in spontaneous and sound-evoked activity (Figure 1.7B-E). These fluctuations are likely to contribute to the variability in experimental results observed across studies, cortical areas, and even across different time points within a single experiment. Therefore, in order to formulate coherent theories of sensory representation, it is important that experiments control for known variables and incorporate such variables into traditional auditory encoding models (David 2018).

In Chapter 3 of this dissertation, we show that accounting for pupil size variability accounts for some but not all changes in neural activity between active and passive behavior in A1 and more so in IC.

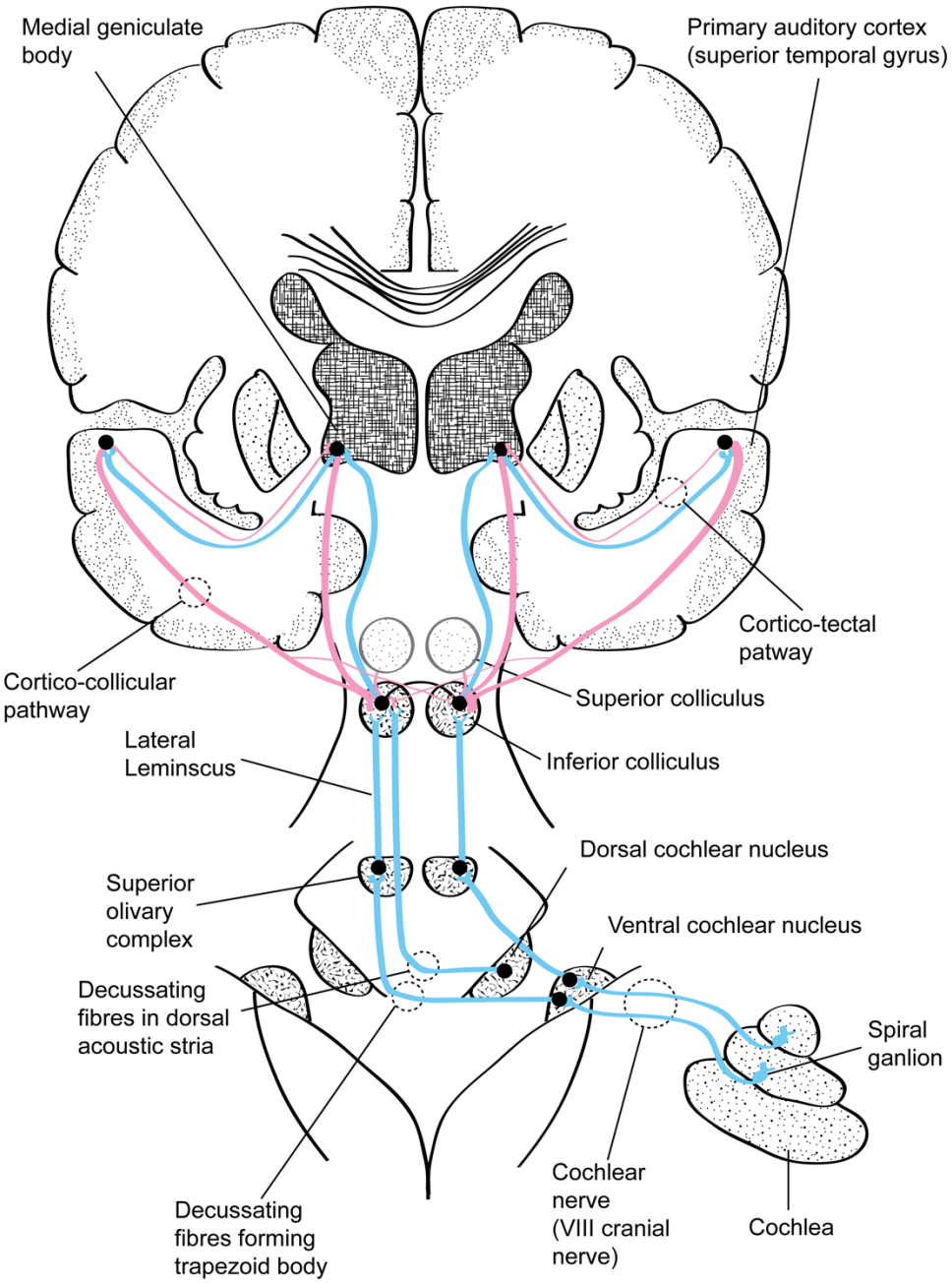

Figure 1.1: Auditory pathway from the cochlea to primary auditory cortex

Schematic of ascending (cyan) and descending (pink) auditory pathway with main auditory structures. Descending fibers below the inferior colliculus are not shown.

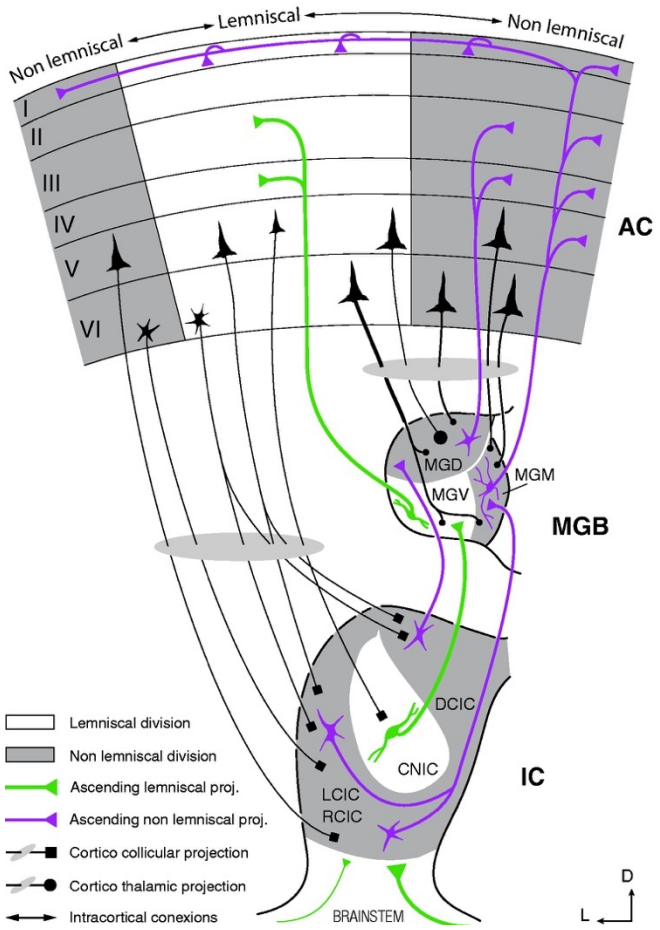

Figure 1.2: Lemniscal and non-lemniscal auditory pathways.

Schematic diagram of the auditory pathway, showing the major stations and projections that constitute the lemniscal and non-lemniscal pathways. Adapted from (Malmierca et al. 2015).

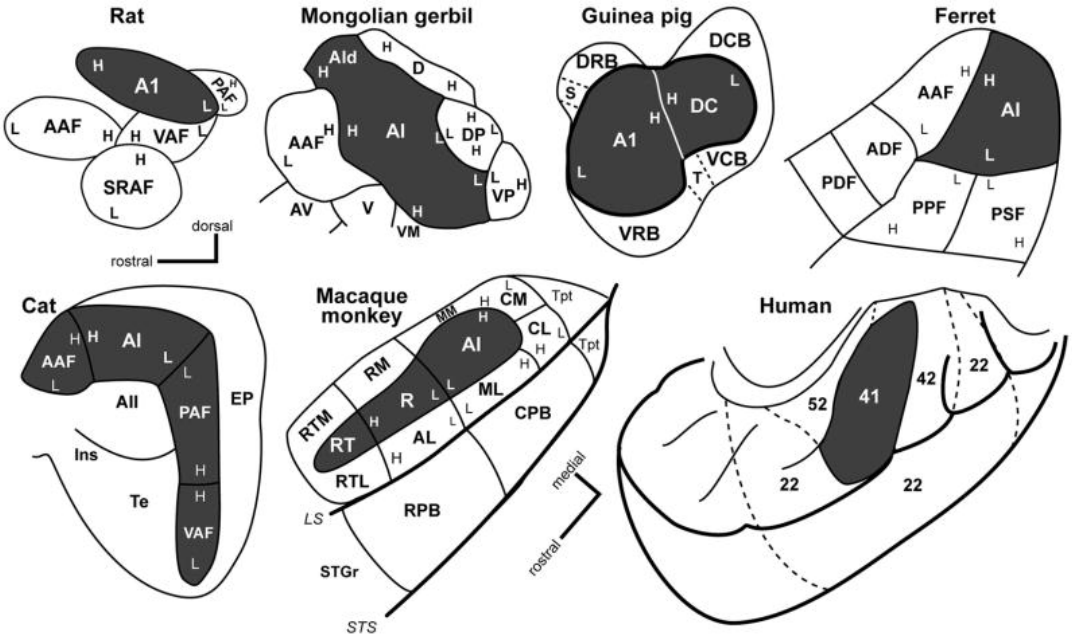

Figure 1.3: Auditory cortex anatomy across species

Schematics of the auditory cortex in selected mammals. Primary core auditory areas are darkly shaded. Belt and parabelt areas are unshaded. Tonotopic gradients are indicated by H (high) and L (low) frequency. Dorsal-rostral axis marker applies to all panels except macaque and human. Adapted from (Hackett 2011).

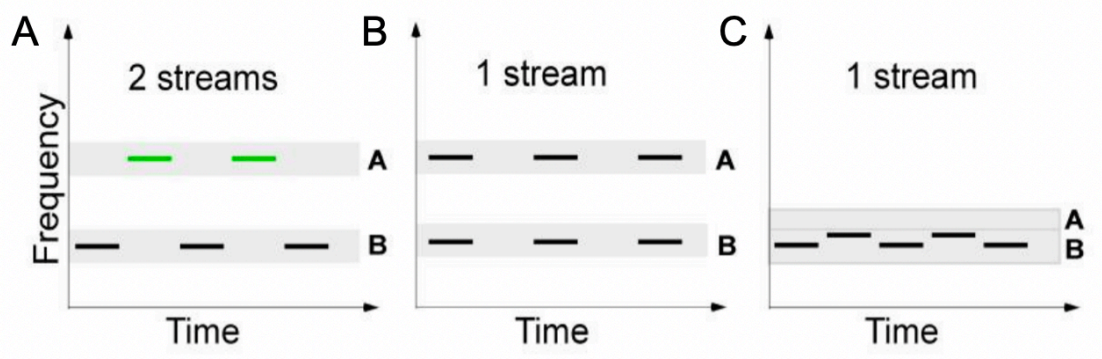

Figure 1.4: Examples of sequential and simultaneous organization of pure-tone sequences

A. Two tones, A and B, presented sequentially (incoherently), with frequency difference $\Delta F > 10\%$ are usually perceived as two separate streams. The green color indicates a separate stream. Shaded regions indicate two hypothetical neural auditory channels activated by the tones. **B.** Same sequence as in A, but presented simultaneously (coherently). In this case the two sequences are perceived as a single stream, despite the large ΔF. **C.** Alternating tones of nearby frequencies ($\Delta F < 10\%$) are usually heard as a single perceptual stream that oscillates in frequency regardless of tone presentation rates. Figure adapted from (Shamma et al. 2011).

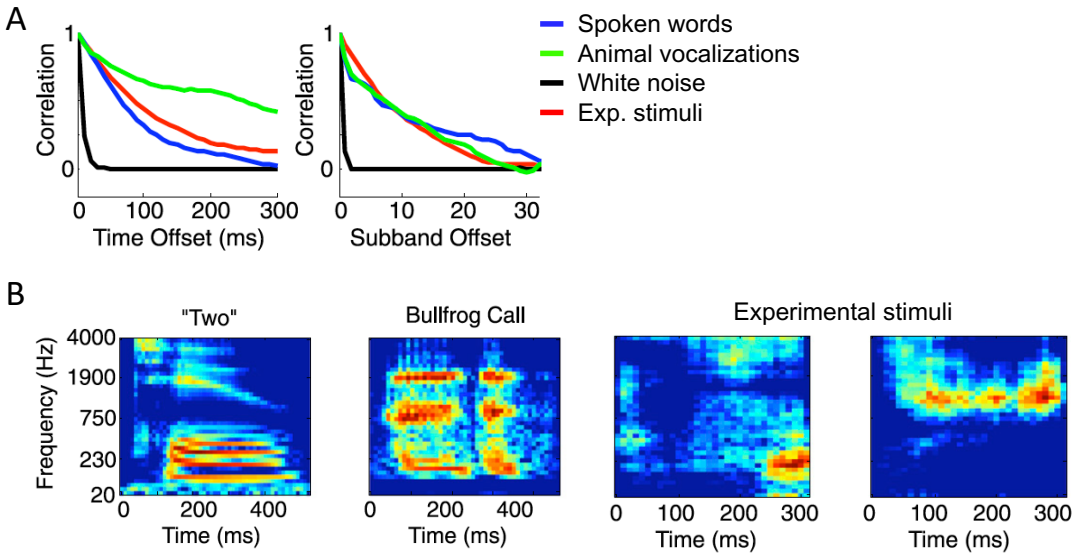

Figure 1.5: Naturalistic sounds used in McDermott *et al.*, 2011

A. Correlation between nearby time-frequency cells as a function of their temporal (left) and spectral (right) separation. **B.** Spectrogram of a spoken word, a bullfrog vocalization, and two synthetically generated stimuli. Adapted from (McDermott et al. 2011).

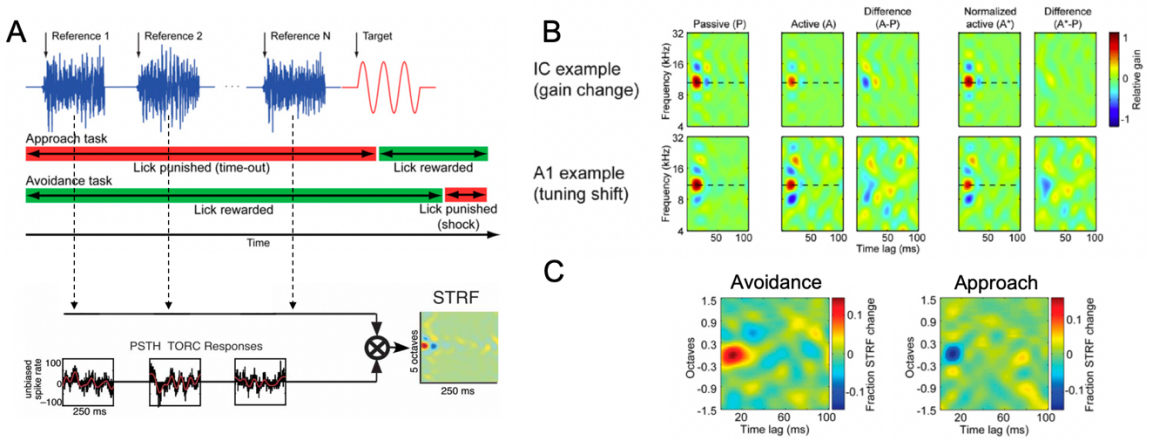

Figure 1.6: STRF measurement and differences observed in A1 and IC

A. Approach *versus* avoidance behaviors. In both tasks, animals were required to detect a pure tone target (red) after a random number of reference noise sounds (blue). During the approach behavior (timeline), subjects were positively rewarded with water for licking a water spout 0.1–1.0 s after target onset (green bar) and punished with a timeout for licking earlier (red bar). During avoidance, subjects were rewarded by licking a continuously flowing stream of water during the references and punished with a mild tail shock if they failed to stop licking 0.4 s after target offset. Reference stimuli were used to compute a spectro-temporal receptive field (STRF), a linear model that shows the time-varying receptive field of a unit. Red regions indicate frequencies and time lags with increased responsiveness, and blue regions indicate a decrease. Green regions indicate no change. **B.** STRF examples from IC (top) and A1 (bottom) neurons that show how STRF differences between passive listening and active behavior are computed. Both examples are from approach task. **C.** Average STRF difference between approach behavior and passive listening, aligned at the target frequency and averaged across units for approach (left) and avoidance (right) tasks. It the approach behavior, the average STRF difference showed a selective local decrease in sound-evoked responses near the target frequency. The opposite was observed in the avoidance task. Figure adapted with authors' permission from (David et al. 2012) panel A, top, and panel C; (Fritz et al. 2003) panel A, bottom; and (David 2018) panel B.

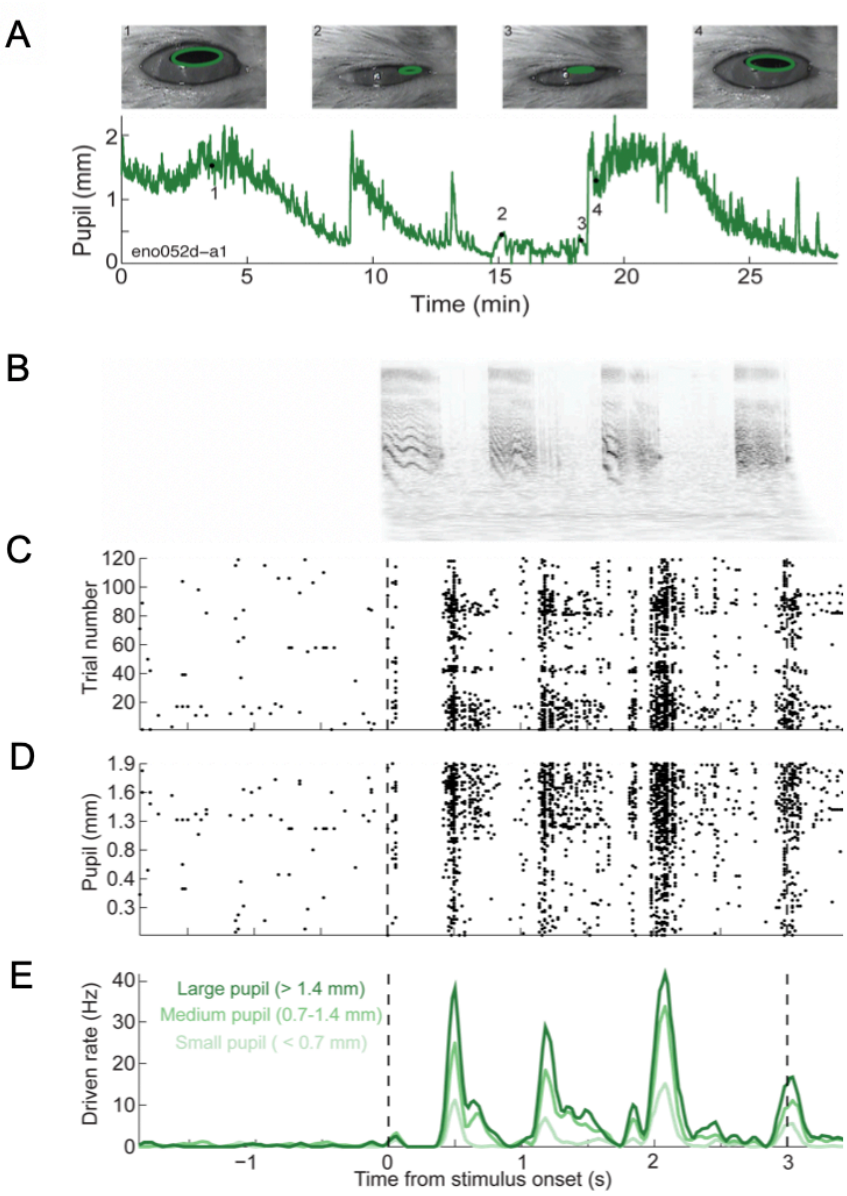

Figure 1.7: Pupil size tracks neural activity in A1

A. Pupil size recorded under constant illumination over the course of an experimental session where the animal was passively listening to ferret vocalizations. **B.** Spectrogram of one of two ferret vocalizations presented to the animal. **C.** Raster plot of the activity of an A1 single unit. Each row is a trial organized according to presentation time. This neuron presented a strong inhibition of its activity in response to sound, followed by a rebound excitatory response whose strength varied across trials. **D.** Same as in C., but trials are ordered here according to pupil size. **E.** Peri-stimulus time histogram (PSTH) of

unit's firing rate for trials in which pupil size was small (< 0.7 mm), medium (0.7-1.4 mm), or large (> 1.4 mm). Data and figure kindly provided by Zachary Schwartz (unpublished).

Chapter 2

Streaming of repeated noise in ferret primary and secondary fields of the auditory cortex

Daniela Saderi, Brad N. Buran, Stephen V. David

Oregon Hearing Research Center, Oregon Health & Science University,

Portland, OR 97239, USA

Acknowledgements: This study was supported by grants from the NIH (SVD, R00 DC 010439, R01 DC 014950) and the Hearing Health Foundation (BNB).

Abstract

Statistical regularities in natural sounds facilitate the perceptual segregation of auditory sources, or streams. Repetition is one cue that drives stream segregation in humans, but the neural bases of the perceptual phenomenon remain unknown. We trained ferrets to detect a stream of repeating noise samples (foreground) embedded in a stream of random noise samples (background). While ferrets were listening passively, we recorded neural activity in primary (A1) and secondary (PEG) fields of the auditory cortex. We used context-dependent encoding models to assess whether evidence for streaming of the repeating stimulus could be observed in these brain areas. Separate models tested whether the strength of the neural response was better predicted by scaling the prediction to both streams equally (global response gain), or by scaling the predicted response of one stream relative to another (stream-specific response gain). Consistent with adaptation, we found a reduction in global response gain when the stimulus was repeated. However, when we measured stream-specific changes in gain, neural responses to the foreground stream were enhanced relative to the background stream. This relative enhancement was stronger in PEG than in A1. In A1, the degree of enhancement depended on auditory tuning. It was strongest in units that displayed low sparseness (*i.e.*, broad sensory tuning), and were tuned preferentially to the repeated sample. Thus, while overall auditory responses were reduced by the repeating sound, a relative enhancement of the foreground stream relative to the background provides evidence for stream segregation that emerges in A1 and is refined in PEG.

Significance Statement

To interact with the world successfully, the brain must parse behaviorally important information from a complex sensory environment. Essential to this process is separating and distinguishing sounds. Complex mixtures of sounds often arrive at the ears simultaneously or in close succession, yet they are effortlessly segregated into distinct sound sources. This process breaks down in hearing-impaired individuals and speech recognition devices. By identifying the underlying neural mechanisms that facilitate streaming, we can develop strategies for ameliorating hearing loss and improving speech recognition technology in the presence of background noise. Here, we present evidence to support a hierarchical model in which sound repetition facilitates sound segregation. This process begins in primary auditory cortex and is refined in secondary auditory cortical fields.

Introduction

Sounds generated by different auditory objects (sound sources) impinge on the ear as a complex mixture, with acoustic energy generated by each source overlapping in both time and frequency. The auditory system has the remarkable ability to group these dynamically changing spectro-temporal sound features into percepts of their distinct sources, in a process known as auditory streaming (Bregman 1990; Griffiths & Warren 2004). Streaming requires statistical analysis of sound sources: streams that come from the same sound source share statistical properties, and the brain uses these properties as cues for stream integration or segregation (Bregman 1990; Carlyon 2004; Darwin 1997; McDermott 2009; Winkler et al. 2009).

Basic acoustic features, such as separation in frequency and time, are key perceptual cues for segregating simple, alternating sequences of pure tones (Bregman 1978b; Bregman et al. 2000; Oberfeld 2014; van Noorden 1975). However, segregating more complex, spectrally overlapping sounds require perceptual dimensions such as pitch (Akeroyd et al. 2005; Mesgarani & Chang 2012), timbre (Cusack & Roberts 2000; Roberts et al. 2002; Singh & Bregman 1997), spatial location (Carlyon 2004; Cusack & Roberts 2000; Mesgarani & Chang 2012; Micheyl et al. 2007b; Roberts et al. 2002; Singh & Bregman 1997), and temporal regularity (Agus et al. 2010; Andreou et al. 2011; Bendixen et al. 2010; Szalárdy et al. 2014). McDermott *et al*. (2011) tested specifically for the benefit of temporal regularity with a set of artificially-generated naturalistic noise samples that lacked other cues for streaming, such has harmonicity and common onset (McDermott et al. 2011). The authors found that simple repetition was sufficient for humans to recover the identity of individual noise samples from mixtures, while the same

samples could not be segregated from background noise outside of the context of repetition. The neural bases of this perceptual pop-out remain unknown.

In contrast to the robust perceptual enhancement reported for a repeating foreground stream, studies of auditory cortical activity during presentation of temporally regular stimuli have emphasized a suppressive effect of repetition (Pérez-González & Malmierca 2014). Single neurons undergo stimulus-specific adaptation (SSA), where responses to repeated tones adapt, but responses to an oddball stimulus, such as a tone at a different frequency, are less adapted or even facilitated, reflecting perceptual pop-out of the oddball sound (Nelken 2014; Ulanovsky et al. 2003). In recordings of human electroencephalography (EEG), a possibly related phenomenon is observed in a late event-related component, called the mismatch negativity (MMN). Although the dynamics differ from SSA, MMN is also elicited by rare deviant sounds randomly interspersed among frequent standard sounds (Näätänen 2001). While there is no evidence that links SSA or MMN with repetition-based grouping, it is possible that these processes share some of the same circuits. How the brain might use adaptation to a repeating sound to enhance its perception is not known.

In this study, we investigated the neuronal correlates of streaming induced by repetition of complex sounds in primary (A1) and secondary (PEG) fields of the auditory cortex of ferrets. We first established the ferret as an animal model for streaming of repeating noise sounds, by designing a behavioral paradigm that assessed the animal's ability to detect repetitions embedded in mixtures. We then recorded single- and multi-unit activity in A1 and PEG of un-anesthetized, passively listening ferrets either trained or naïve to the detection task. We tested the prediction that auditory cortical neurons

facilitate stream segregation by selectively enhancing their response to the repeating (*i.e.*, foreground) stream. We used context-dependent sound encoding models to quantify the relative contribution of the two overlapping streams to the evoked neural response. We found that neural responses to the repeated stimuli were reduced overall in both areas. Yet, some neurons in both brain regions had foreground-specific responses that were enhanced with respect to responses to the simultaneous background stream. These results provide evidence for a model of streaming cued by repetition that starts in primary and is refined in secondary fields of the auditory cortex.

Materials and Methods

All procedures were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee and conform to the United States Department of Agriculture standards.

Surgical procedure

Animal care and procedures were similar to those described previously for neurophysiological recordings from awake ferrets (Slee & David 2015b). Five young adult ferrets (two females, three males) were obtained from an animal supplier (Marshall Farms, New York). Normal auditory thresholds were confirmed by measuring auditory brainstem responses. A sterile surgery was then performed under isoflurane anesthesia to mount a post for subsequent head fixation and to expose a 10-mm^2 portion of the skull over the auditory cortex where the craniotomy would be subsequently opened. A light-cured composite (Charisma, Heraeus Kulzer) anchored a custom stainless-steel head-post

on the midline in the posterior region of the skull. The stability of the implant was also supported by 8-10 stainless self-tapping set screws mounted in the skull (Synthes). The whole implant was then built up to its final shape with layers of Charisma and acrylic pink cement (AM Systems).

During the first week post-surgery, the animal was treated prophylactically with broad-spectrum antibiotics (10 mg/kg Baytril). For the first two weeks the wound was cleaned with antiseptics (Betadine and Chlorexidine) and bandaged daily. After the wound margin was healed, cleaning and bandaging occurred every 2-3 days through the life of the animal. This method revealed to be effective in minimizing infections of the wound margin.

Stimuli and Acoustics

Repeated embedded noise stimuli used in the present study were generated using the algorithm from McDermott *et al*. (2011) (McDermott et al. 2011). Brief, 250- or 300-ms duration samples of broadband Gaussian noise were filtered to have spectro-temporal correlations matched to natural sounds but without common grouping cues, such as harmonic regularities and common onsets (McDermott et al. 2011). The spectral range of the noise (125-16,000 Hz or 250-20,000 Hz) was chosen to span the tuning of the current recording site. An experimental trial consisted of continuous sequences of 10-12 noise samples (0 ms inter-sample interval) drawn randomly from a pool of twenty distinct samples (Figure 1). The order of samples varied between trials. Either one stream of samples was presented (single stream trial) or two streams were overlaid and presented simultaneously (dual stream trial). At a random time (after 3-11 samples), the sample in

one stream (target sample) began to repeat. In dual stream trials, this repetition occurred only in one of the two streams, while samples in the other stream continued to be drawn randomly. In human studies, the repeating sample has been shown to pop out perceptually as a salient stream (McDermott et al. 2011). Thus, the stream containing the repeated sample is referred to here as the *foreground*, and the non-repeating stream as the *background* (Figure 1B). The period of the trial containing only random samples is referred to as the *random phase*, and the segment starting with the first repetition of the target sample, where the two streams perceptually diverge, is referred to as the *repeating phase* (Figure 2.1B).

All behavioral and physiological experiments were conducted inside a custom double-walled sound-isolating chamber with inside dimensions of 8' × 8' × 6' (L × W × H). A custom second wall was added to a single-walled factory chamber (Professional Model, Gretch-Ken Inc.) with a wooden frame and an inner wall composed of ¾" MDF board. The air gap between the outer and inner walls was 1.5". The inside wall was lined with 3" sound absorbing foam (Pinta Acoustics). The chamber attenuated sounds above 2 kHz by more than 60dB. Sounds from 0.2-2 kHz were attenuated 30-60 dB, falling off approximately linearly with log-frequency.

Stimulus presentation and behavioral control were provided by custom MATLAB software (Mathworks Inc.). Digitally generated sounds were D/A converted (100 kHz, National Instruments PCI-6229), and presented through a sound transducer (Manger W05) driven with a power amplifier (Crown D-75A). The speaker was placed one meter from the animal's head, 30° contralateral to the cortical hemisphere under study. Sound

level was calibrated using a ½" microphone (Bruel & Kjaer 4191). Stimuli were presented with 10ms $\cos^2$ onset and offset ramps.

Behavior

Two ferrets (one male, ferret H, and a female, ferret O) were trained to report the occurrence of repeated target noise samples in the repeated embedded noise stimuli using a go/no-go paradigm (David et al. 2012). Two weeks after the implant surgery, each ferret was gradually habituated to head fixation by a custom stereotaxic apparatus in a plexiglass tube. Habituation sessions initially lasted for 5 minutes and increased by increments of 5-10 minutes until the ferret lay comfortably for at least one hour. At this time the ferret was placed on a restricted water schedule and began behavioral training. During training and physiological recording sessions that involved behavior, the ferret was kept in water restriction for five days/week, and almost all the daily water intake (40-80 ml) was delivered through behavior. Their diet was supplemented with 20 ml/day of high protein Ensure (Abbott). Water restriction was to be discontinued if weight dropped below 20% of the initial weight, but this did not happen with either ferret. Water rewards were delivered through a spout positioned close to the ferret's nose. Delivery was controlled electronically with a solenoid valve. Each time the ferret licked the waterspout, it caused a beam formed by an infrared LED and photo-diode placed across the spout to be discontinued (Figure 2.1A). This system allowed us to precisely record the timing of each lick relative to stimulus presentation.

After trial onset, animals were required to refrain from licking until the onset of the repeating phase, *i.e.*, after the occurrence of a repeated sample. Licks during the

random phase were recorded as false alarms and punished with a 4-6 sec time-out. Licks that occurred in the repeating phase were recorded as hits and always rewarded with one-two drops of water (Figure 2.1B). Each behavioral session had two target samples whose identity varied from session to session to avoid ferret overexposure to a given target spectro-temporal features.

To shape the animal's behavior, training started with a high signal-to-noise ratio (SNR) between random and repeating phases. SNR was then slowly decreased until 0dB SNR was reached. Parameters such as spectral modulation depth of the two streams and length of the random phase/false alarm window were also adjusted over the training period. Performance was assessed by a discrimination index (DI) computed from the area under the receiver operating characteristic (ROC) curve for detection of the target in the repeating phase (David et al. 2012; Yin et al. 2010). DI combines information about hit rate, false alarm rate, and reaction time, and has a higher value for the higher, lower, and faster these scores are, respectively. A DI greater than 0.5 indicates above-chance performance. Criterion was reached as the ferret performed at DI > 0.5, with 0 SNR and 0 modulation depth difference for four consecutive days.

Electrophysiology

Single- and multi-unit neural recordings were performed in the two trained animals and in three additional task-naïve animals. A small (~1-2 mm diameter) craniotomy was opened over the auditory cortex, in a location chosen based on stereotaxic coordinates and superficial landmarks on the skull marked during surgery. Initial recordings targeted primary regions of the ferret auditory cortex (A1), and

recording location was confirmed by characteristic short-latency responses to tone stimuli and by tonotopic organization of frequency selectivity (Bizley et al. 2005). Recordings in secondary auditory cortex (PEG) were then performed in the field ventral to A1, identified by a reversal in the tonotopic gradient.

On each recording day, 1 to 4 high-impedance tungsten microelectrodes (FHC or A-M Systems, impedance 1-5 MΩ) were slowly advanced into cortex with independent motorized microdrives (Alpha-Omega). The electrodes were positioned (Kopf Instruments) such that the angle was roughly normal to the surface of the brain (~28-40°). Stimulus presentation and electrode advancement were controlled from outside the sound booth, and animals were monitored through a video camera. Neural signals were recorded using open-source data acquisition software (MANTA, (Englitz et al. 2013)) Raw traces were bandpass-filtered (0.3-10 kHz), amplified (10k, A-M Systems 1800 or 3600 AC amplifier), digitized (20 kHz, National Instruments PCI-6052E) and stored for subsequent offline analysis. Putative spikes were extracted from the continuous signal by collecting all events ≥4 standard deviations from zero. Different spike waves were separated from each other and from noise using principle component analysis and *k*-means clustering (David et al. 2009). Single units (>95% isolation) and stable multiunits (>70% isolation) were included in this study, resulting in a total of 141 A1 and 136 PEG units.

Between recording sessions, the exposed recording chamber surrounding the craniotomy was covered with polysiloxane impression material (GC America). After several electrophysiological penetrations (usually about 5-10), the craniotomy was expanded or a new craniotomy was opened to expose new regions of auditory cortex.

When possible, old craniotomies were covered with a layer of bone wax and allowed to heal. Multiple craniotomies were performed on both hemispheres.

Analysis

Effect of repetition on target responses

To assess the effect of repetition on overall responsiveness, we first measured changes in the response to the target sample between random and repeating trial phases. We computed the peristimulus time histogram (PSTH) response to each occurrence of target sample in the stimulus separately for the random phase and repeating phase, using data from dual-stream trials only. Spontaneous rate was subtracted from the PSTH to ensure the fraction term reflected changes in the evoked response. We then computed the gain term that minimized the least squares difference between evoked responses in the two phases. Log of the measured gain is reported to allow for direct comparison with the results of subsequent modeling analysis (see below).

PSTH-based models

Auditory cortical neurons could support segregation of the repeated stream either by changing the overall gain of their response to the repeating stream (stream-independent) or by differentially enhancing responses to one or the other stream (stream-dependent). To test these alternative predictions, we fit the data using *stream-independent* and *stream-dependent* models. In both models, responses were predicted using a weighted sum of time-varying responses to each noise sample. During the random phase,

the time-varying response was the linear sum of a response to the foreground stream, response to the background stream, and spontaneous spike rate:

$$r_{rand}(S_{fg}, S_{bg}, i) = \bar{r}(S_{fg}, i) + \bar{r}(S_{bg}, i) + r_0 \qquad (Eqn.\ 2.1).$$

Here, S_{fg} and S_{bg} are the identity of samples in the foreground and background streams, respectively, and r_0 is the spontaneous rate. $\bar{r}(S, i)$ is the contribution of sample S to the evoked spike count in *i*-th time bin following sample onset.

For the stream-independent model, responses during the repeated phase were computed,

$$r_{rep\_ind}(S_{fg}, S_{bg}, i) = [\bar{r}(S_{fg}, i) + \bar{r}(S_{bg}, i)] \times e^{RGglobal} + r_0 \qquad (Eqn.2.2),$$

where RG_{global} scales responses to both streams. For the stream-dependent model, responses during the repeated phase were computed,

$$r_{rep\_dep}(S_{fg}, S_{bg}, i) = \bar{r}(S_{fg}, i) \times e^{RGfg} + \bar{r}(S_{bg}, i)e^{RGbg} + r_0 \qquad (Eqn.\ 2.3),$$

where RG_{fg} and RG_{bg} modulate the respective stream responses separately before they are summed. The use of an exponent simplifies interpretation of gain changes such that values of $RG > 0$ indicate enhancement and values of $RG < 0$ indicate suppression. $\bar{r}(S, i)$ can be negative, which allows for suppressed responses relative to the spontaneous rate. In this case, if a unit has both enhanced and suppressed responses, RG will scale both responses equally (*e.g.*, if $RG > 0$, there will be a decrease in spike rate during negative responses and an increase in spike rate during enhanced responses). The difference $RG_{fg} - RG_{bg}$ is the relative enhancement between streams, here referred to as *foreground enhancement*. If $RG_{fg} > RG_{bg}$, then the neural response to the foreground stream is stronger than to the background stream.

Models were fit to maximize Poisson likelihood of free parameters using Bayesian regression. A normal prior with mean 0 and standard deviation 10 was set on both r_0 and $\bar{r}$. A normal prior with mean 0 and standard deviation 1 was set on all RG parameters. The model was fit three times using a different set of random starting values for each coefficient. Two thousand samples for each fit were acquired with a No-U-Turn Sampler, an extension to Hamiltonian Monte Carlo that eliminates the need to set a number of steps (Hoffman & Gelman 2011). Gelman-Rubin statistics were computed for each fit to ensure all the fits converged to the same final estimate ($\overline{r} < 1.1$).

The posteriors for RG_{global}, RG_{fg} and RG_{bg} were extracted from the Bayes model. An RG parameter for which the 95% credible interval (as derived from the posterior) was less than 0 were considered to have significant suppression. Parameters with an interval greater than 0 were considered to have significant enhancement. For all data points shown, the means of the relevant posterior are plotted.

Lifetime sparseness and target preference

We quantified sparseness (*S*), a measure of unit selectivity for a given sample relative to the others in the collection (adapted from (Vinje & Gallant 2000)),

$$S = \left[1 - \frac{\left(\frac{1}{n}\sum_{i=1}^{n} r_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} r_i^2}\right] \Big/ \left[1 - \frac{1}{n}\right] \qquad (Eqn.\ 2.4),$$

where r_i is the standard deviation of the PSTH (computed using the average of response to the token in the random phase of single-stream trials) for the i^{th} sample and n is the total number of noise samples. We quantified target preference (*TP*), a measure of how well the target sample modulates the unit's response,

$$TP = {r_{tar}}\Big/{\left(\frac{1}{n}\sum_{i=1}^{n} r_i\right)} \quad \textit{(Eqn. 2.5)},$$

where r_{tar} is the standard deviation of the target PSTH and the other terms are defined as for sparseness. The use of standard deviation to measure response magnitude means that strong suppression or enhancement yield similar response strength.

To assess whether there was a significant effect of cortical area, sparseness and/or target preference, we used a mixed linear model with area (A1 or PEG), sparseness (S), and target preference (TP) as the fixed effects and unit as the random effect. All two- and three-way interactions were included according to the following model design:

$$\beta_0 + \beta_1\, PEG + \beta_2\, S + \beta_3\, PEG\, S + \beta_4\, TP + \beta_5\, PEG\, TP +$$
$$+\beta_6\, S \;+\; \beta_7\, PEG\, S\, TP \quad \textit{(Eqn.2.6)}.$$

Spectro-temporal receptive field models

In addition to the PSTH-based models, which fit responses to individual noise samples, we confirmed that the same streaming effects were captured by a context-dependent spectro-temporal receptive field (STRF) model (David 2018). The classic linear-nonlinear (LN) STRF models neural activity as the linear weighted sum of the preceding stimulus spectrogram, the output of which passes through a static nonlinearity to predict the time-varying spike rate response (Aertsen & Johannesma 1981; deCharms et al. 1998). The STRF, $h(x,u)$ is as a linear weight matrix that is convolved with the logarithm of the stimulus spectrogram, $s(x,t)$:

$$r_{LIN}(t) = \sum_{x=1}^{X}\sum_{u=1}^{U} h(x,u)\; s(x, t-u) \quad \textit{(Eqn. 2.7)},$$

where x = 1…X are the frequency channels, t = 1…T is time, and u is the time lag of the convolution kernel. Taking the log of the stimulus spectrogram accounts for nonlinear gain in the cochlea. Free parameters in the weight matrix, h, indicate the gain applied to frequency channel x at time lag u to produce the predicted response. Positive values indicate components of the stimulus correlated with increased firing, and negative values indicate components correlated with decreased firing.

The output of the linear STRF is passed through a static nonlinear sigmoid function to account for spike threshold and saturation (Thorson et al. 2015),

$$r(t) = F[r_{LIN}(t)] \qquad (Eqn.\ 2.8),$$

$$F(x) = r_0 + A\ \exp\left[-\exp\left(\kappa(x - x_0)\right)\right] \qquad (Eqn.\ 2.9).$$

Free parameters here are x_0, inflection point of the sigmoid, r_0, spontaneous spike rate, A, maximum spike rate, and κ, the slope of the sigmoid.

We developed a modified LN STRF to account for stream-dependent changes in gain. The input spectrogram for each stream was scaled by a gain term that depended on stream identity (foreground or background) and trial phase (random or repeating). We refer to this model as the *phase+stream model*. The stimulus was modeled as the sum of two log spectrograms, computed separately for the foreground and background streams, s_1 and s_2, respectively. In the random phase, the total stimulus, $s(x,t)$, was modeled as the linear sum of these two stimuli:

$$s(x,t) = s_1(x,t) + s_2(x,t) \qquad (Eqn.\ 2.10).$$

In the repeating phase, each stimulus was scaled by a repetition gain for the respective stream,

$$s(x,t) = RG_{fg}\ s_1(x,t) + RG_{bg}\ s_2(x,t) \qquad (Eqn.\ 2.11).$$

All model parameters were estimated by gradient descent (Byrd et al. 1995; David 2018; Thorson et al. 2015). STRF parameters were initialized to have flat tuning (*i.e.*, uniform initial values of h) and were iteratively updated using small steps in the direction that optimally reduced the mean squared error between the time-varying spike rate of the neuron and the model prediction. To maximize statistical power with the available data, the STRF was fit using both single- and dual-stream data. For single-stream trials, the second stimulus spectrogram was fixed at zero, $s_2(x,t) = 0$, and a separate gain term was fit for those trials to prevent bias in estimates of RG_{fg} and RG_{bg}. Measurements of prediction accuracy were obtained by 20-fold cross validation, in which a separate model was fit to 95% of the data then used to predict the remaining 5%. Fit and test data were taken from interleaved trials. This procedure was repeated 20 times with non-overlapping test sets, so that the final result was a prediction of the entire time-varying response. Prediction accuracy was then measured as the correlation coefficient (Pearson's *r*) between the predicted and actual response. Standard error on prediction correlation was measured by jackknifing (Efron & Tibshirani 1986), and only units with prediction error significantly greater than zero were included in model comparisons ($p < 0.05$, jackknife *t*-test).

To quantify effects of phase- and stream-dependent gain, we also fit models using the same data and fitting procedure, but where stream identity (phase-only model) or both phase and stream (baseline model) were shuffled in time. An improvement in prediction accuracy for a model with a non-shuffled over shuffled variable indicated a beneficial effect of the corresponding gain parameter on model performance, and thus of a stream-dependent change in sound encoding. Significant differences in model performance were

assessed by a Wilcoxon rank sum test between prediction correlations for the set of units fit with each model.

Results

Ferrets perceive repeated patterns embedded in noise

To investigate the physiological underpinnings of repetition-based streaming in an animal model, we first developed a behavioral paradigm to assess ferrets' ability to detect repetitions embedded in noise. The repeated embedded noise stimuli were composed of two overlapping continuous streams of brief (250- or 300-ms) noise samples (McDermott et al. 2011). The noise samples had second-order statistics (*i.e.*, spectral and temporal correlations) matched to natural sounds. Consistent with the goal of this study, the only streaming cue was repetition. These stimuli lacked other conventional streaming cues such as harmonicity and onset time.

During the initial phase of each trial, samples for both streams were drawn randomly from a pool of 20 distinct noise samples. We refer to this initial phase of the trial as the *random phase* (1-2.5 sec, Figure 2.1B). When the samples are drawn randomly for both streams, they are perceived as a single object. At the end of the random phase, a target noise sample, different for each behavioral block, started to repeat in one sequence but not in the other. In humans, this stimulus structure leads to perceptual separation of the two sequences into discrete streams (McDermott et al. 2011). We refer to this phase of the trial as the *repeating phase*. In addition, we refer to the sequence that contains the repeating target sample as the *foreground stream*, and the sequence with no repetition as the *background stream* (Figure 2.1B).

Two ferrets (O and H) were trained to report when they detected the repetition of the target using a go/no-go detection paradigm. Head-fixed animals were required to withhold from licking a waterspout during the random phase and to lick shortly after the onset of the repeating phase (Figure 2.1A-B). In each behavioral block (~50-100 trials), two noise samples were chosen as targets from a pool of 20, each with 50% chance of occurring in a trial. Changing the identity of the targets between blocks to help avoid overtraining on a specific target. To measure behavioral performance in a task with continuous distractors and variable target times, we used a discrimination index (DI, see *Materials and Methods*). This metric uses hit rate, false alarm rate, and reaction time to compute the area under the receiver operating characteristic (ROC) curve for target detection (David et al. 2012; Yin et al. 2010). A DI greater than 0.5 indicates above-chance behavior. Both ferrets were able to learn the task and perform above chance within two months of training, indicating that they were able to perceive the repeating noise stream (Ferret O: mean DI = 0.61±0.004 SEM, n = 327; Ferret H: mean DI = 0.55±0.005 SEM, n = 171; Figure 2.1C).

Neuronal responses are suppressed during the repeating phase

We recorded multi- and single-unit neural activity in primary (A1, n = 152) and secondary (PEG, n = 138) regions of the auditory cortex of five ferrets passively listening to the task stimuli. Two were trained on the repetition embedded noise task (behavior described above), and three were naïve to the task. Although all physiological data presented here were recorded in passively-listening ferrets, for consistency we refer to the same trial structure terminology described in the previous section. During

electrophysiological recordings, one of the noise samples selected as target, was chosen to match each unit's best response while the other was chosen at random (see *Materials and Methods*). Both targets were presented with the same probability of occurrence across trials.

To investigate the neurophysiological underpinnings of streaming due to repetition, we first looked at the raw firing rates of the recorded units in response to the repeated noise stimuli. Given the enhanced representation of repeating stimuli observed in behavioral experiments (Agus et al. 2010; Masutomi et al. 2015; McDermott et al. 2011), we reasoned that evidence for the selective enhancement of foreground representation should be found at the level of the auditory cortex. If this were true, we would expect the neural response to a target sample to change between random and repeating contexts.

To test this prediction, we first computed the average peristimulus time histogram (PSTH) response across all occurrences of the target noise samples in the random phase, and compared them to the average PSTH response to the first three repetitions of the target in the repeating phase (Figure 2.2A). Since the background sample was randomly selected for each presentation of the target, responses to the background sample were averaged out, and the PSTHs primarily reflected responses to the target. To quantify changes in the response, we computed the gain term that scaled the PSTH for the random phase to match the PSTH for the repeating phase. To allow for a direct comparison between gains generated by models (see below), all gain terms were log-transformed. Thus, negative values indicated suppressed responses during repetition and positive values indicated enhanced responses. For most units in A1 and PEG the gain term was

less than zero (Figure 2.2B), indicating that the average target response in the repeating phase was suppressed with respect to the average target response in the random phase. Considering neural adaptation to repeated stimuli in the auditory cortex (Pérez-González & Malmierca 2014; Ulanovsky et al. 2003), a decreased response to the target in the repeating phase is not unexpected.

Relative enhancement of responses to the repeating foreground stream

Simply comparing the average neural response to the target in the repeating phase relative to the random phase does not provide insight into any stream-specific effect that might emerge as a consequence of the repetition. To test for evidence of streaming in the neural response, we needed to independently assess the responses in the two streams. We reasoned that, even if the total response was suppressed, activity in the foreground stream in response to the repetition could be enhanced (or suppressed) relative to the background stream.

To test this prediction, we developed an encoding model in which the neural response was the sum of responses to samples in each stream (*stream-dependent model*, see *Materials and Methods*). In the random phase, the responses to the two concurrent samples, plus a constant term representing baseline firing rate, were added (*Eqn. 2.1*). For the repeating phase, the response to each sample was scaled according to whether it occurred in the foreground or background stream by repetition gain terms RG_{fg} and RG_{bg}, respectively, before summing (*Eqn. 2.3*). We compared this model to a *stream-independent* model in which responses to samples in both streams were scaled equally by a single term, RG_{global} (*Eqn. 2.2*). Since the predicted responses to the samples are

relative to baseline firing rate and the gain terms are log-scaled, positive gains indicate stronger modulation of the unit's response (*i.e.*, greater excitation and inhibition) and negative gains indicate weaker modulation of the unit's response relative to baseline firing rate.

Figure 2.3 plots the average responses to the target stimuli and predictions by the stream-dependent model for example units in A1 (top) and PEG (bottom). In both examples, the repetition gain for the background stream was negative (RG_{bg} = -1.0 in A1 and -1.6 in PEG). This means that neural responses to the background stream in the repeating phase were strongly suppressed (*i.e.*, on a linear scale, the neural response was scaled by 0.37 in A1 and 0.20 in PEG) relative to the random phase (Figure 2.3, black dotted line). Conversely, the foreground repetition gain term positively scaled the target sample response in the repeating phase (RG_{fg} = 1.6 in A1 and 2.2 in PEG; blue dashed line), leading to an overall enhancement of the combined response (orange solid line). That is, the neural modulation to the foreground stream was 5 (A1) and 9 (PEG) times greater in the repeating phase than in the random phase.

Across the population, repetition gain was negative in the majority of unit-target pairs for both foreground and background streams (Figure 2.4A; A1: n = 304 unit-target pairs; mean RG_{bg} = -0.610, mean RG_{fg} = -0.486; PEG: n = 276; mean RG_{bg} = -0.935; mean RG_{fg}= -0.518). Similarly, in the stream-independent mode, units in both A1 and PEG usually had a negative RG_{global} (Figure 2.4B). This global suppression was consistent with the decrease observed in the average target response described above (Figure 2.2B; Pearson's r = 0.63 between RG_{global} and target response gain, $p < 0.0001$).

To test for the relative enhancement of foreground responses in the stream-dependent model, we measured the *foreground enhancement*, the difference between foreground and background repetition gain (Figure 2.4B). A subset of unit-target pairs displayed significant foreground enhancement (41/304 in A1, 58/276 in PEG, Figure 2.4B, dark purple), meaning that in the repeating phase responses in the foreground stream were less suppressed or enhanced relative to responses in the background stream. In contrast, fewer units showed foreground suppression in either area (26/304 in A1, 12/276 in PEG, light purple). Across the set of unit-target pairs, mean foreground enhancement was significantly greater than zero in A1 (0.124, $p = 0.004$, Wilcoxon signed-rank) and PEG (0.416, $p < 0.0001$, Wilcoxon signed-rank) (Figure 2.4B). Mean foreground enhancement was stronger in PEG than in A1 ($p < 0.0001$, independent two-sample *t*-test).

Despite the overall suppression of activity during the repeating phase, these results support a model of selective enhancement of responses to the repeated foreground stream, consistent with the enhanced perception of the repeated stream relative to the random background (McDermott et al. 2011).

Auditory tuning properties predict the degree of repetition enhancement

As described above, we used two different target samples during each recording of an individual unit. Next, we wondered if the units showing significant foreground enhancement had distinct acoustic response properties or preference for the target stimuli. For each unit, we quantified lifetime sparseness, a measure of selectivity for any one sample relative to the others (see *Materials and Methods*, *Eqn. 2.4*) (Vinje & Gallant

2000). This metric is bounded between 0 and 1, where 0 indicates low sparseness (equal responses to all stimuli) and 1 indicates high sparseness (non-zero response to only one stimulus). Example responses to each noise samples in our collection of a unit with relatively high sparseness are plotted in Figure 2.5A. For each unit-target pair, we also computed target preference, the ratio of evoked response to the target sample versus the average response to all samples (see *Materials and Methods*, *Eqn. 2.5*). A target preference of 1 indicates that the modulation by the target is equivalent to the average response for all samples.

The relationship between each unit's sparseness, target preference, and auditory field (A1 or PEG) and its foreground enhancement (Figure 2.5C) was quantified by a general linear mixed model (see *Materials and Methods*, *Eqn. 2.6*) with area, target preference, and sparseness as fixed effects and unit as a random effect. All two- and three-way interactions between the fixed parameters were included, and complete results are shown in Table 2.1. This model demonstrated a significant relationship between target preference and foreground enhancement which was significantly modulated by sparseness. Foreground enhancement was stronger in units with high target preference; however, this relationship was influenced by the unit's sparseness as indicated by the negative interaction between target preference and sparseness (Table 2.1). This means that A1 units with strong responses to a target had significantly larger foreground enhancement for that target; however, this effect decreased with increasing sparseness. In contrast, in PEG there was no relationship between foreground enhancement and either sparseness, target preference, or the interaction of target preference and sparseness (Table 2.1).

Thus, in A1, units that responded to many stimuli (low sparseness) but had a relatively strong preference to a target (high target preference) tended to show the most foreground enhancement. In PEG, enhancement was stronger overall and affects responses more uniformly, regardless of auditory selectivity. These differences between PEG than A1 suggest a gradual emergence of repetition-related streaming along the cortical auditory pathway.

Foreground enhancement increases accuracy of spectro-temporal receptive field models

To validate the gain changes observed in the PSTH-based model and to quantify the effect of these changes on sound-evoked activity, we modeled the same data with a spectro-temporal receptive field (STRF). In the classic linear-nonlinear (LN) STRF (see *Material and Methods, Eqns. 2.7-2.9*), the time-varying neural response is modeled as a linear weighted sum of the stimulus spectrogram (Depireux et al. 2001; Thorson et al. 2015). We developed a context-dependent model, in which spectrograms for each stream were scaled separately by a gain term before input to the STRF (see *Material and Methods, Eqns. 2.10-2.11*). This stream-dependent scaling followed the same logic as the PSTH-based model described above. That is, each stream was scaled by free parameters that depended on stream identity (foreground or background) and phase (random or repeating). The rescaled spectrograms were summed and then provided input to a traditional LN STRF. Context gain parameters and STRF parameters were fit simultaneously (Figure 2.6A) (David 2018). In the text and figures, we refer to this model as the *phase+stream* STRF.

We used 20-fold cross validation to compare the prediction accuracy of the phase+stream STRF to two control models. A *phase-only* STRF, in which stream identity was shuffled in time before fitting; and a *baseline* STRF, in which both phase and stream identity were shuffled. The phase-only STRF accounted for changes in gain due to repetition but independent of foreground versus background stream identity, analogous to the stream-independent model above. The phase+stream STRF predicted time-varying responses more accurately than the phase-only STRF in both A1 and PEG, confirming a significant influence of stream identity on relative gain (A1: $p < 0.0001$, PEG: $p < 0.0001$, Wilcoxon signed-rank test, Figure 2.6B).

To measure the relative enhancement of the two streams, we compared the stream-specific gain terms from the model fits, equivalent to RG_{fg} and RG_{bg} discussed above. We observed a significant relative increase in foreground versus background gain in both A1 (mean increase 0.159; $p < 0.0001$, Wilcoxon signed-rank test) and PEG (mean increase 0.291; $p < 0.0001$, Wilcoxon signed-rank test, Figure 2.6C). This result provided further evidence for stream-dependent changes in gain. These changes in gain followed the same pattern as for mean foreground enhancement computed by the stream-dependent model (Figure 2.4).

The comparison of phase-only and baseline STRFs measured the effect of repetition alone on evoked activity (independent of stream identity). On average, the phase-only STRF had greater prediction accuracy than the baseline STRF in both areas (A1: $p < 0.0001$, PEG: $p < 0.0001$, Wilcoxon signed-rank test, Figure 2.6D). Overall, gain was suppressed during the repeating phase (mean A1: -0.11, PEG: -0.042), as observed in the PSTH-based models above (Figure 2.4). Thus, this approach provides

additional evidence for a model in which repetition leads to overall suppression of the neural responses, but with less prominent suppression of the foreground stream relative to the background stream.

Discussion

In natural environments, temporally co-varying sound features tend to be grouped by the brain into a single object (Bizley & Cohen 2013). Sound repetition is sufficient to induce stream segregation in human listeners (McDermott et al. 2011), and subjects are able to identify individual, previously unheard noise samples if they are presented in repeated succession simultaneously to a mixture of different non-repeating samples. The goal of the current study was to investigate the neural underpinnings of streaming cued by sound repetition. We developed an animal model for repetition detection and found evidence for enhanced representation of the repeating foreground signal in the auditory cortex. This representation appears to emerge hierarchically, as the streaming effects were stronger in secondary (PEG) than in primary (A1) auditory cortex.

Mechanisms of repetition-induced stream segregation

Previous studies that have explored the neural signature of streaming at the single-unit level have primarily used alternating sequences of pure tones (Fishman et al. 2001; Micheyl et al. 2005). Micheyl and collaborators presented "ABA_" sequences to awake macaques and examined the pattern of activity evoked in A1. Tone A was chosen to be on the best frequency of the recorded unit, while tone B was placed at a frequency of 1-9 semitones from tone A. The authors found that, even if responses to both tones decreased

relative to their presentation in isolation, responses to the non-preferred B tones decreased to a greater extent (Micheyl et al. 2005). In the current study, we observed a similar effect, that relative enhancement of the foreground stream was more pronounced in units well-tuned to the target noise sample. Thus, our results are consistent with observations based on the ABA tone paradigm. Moreover, our results provide evidence that the same principles generalize to streaming of complex, naturalistic sounds presented simultaneously, a situation that more closely relates to animals' everyday sound experience.

Sound features that belong to the same source tend to begin and end at the same time. This phenomenon has been formalized for streaming in the temporal coherence model (Elhilali et al. 2009a; Shamma et al. 2011). Teki *et al.* (2016) demonstrated that human listeners are highly sensitive to repetition of sounds presented in the context of a random mixture of chords. Similar to our findings, the authors observed that repeating sounds tend to fuse together into a "foreground" that emerges from a randomly changing background (Teki et al. 2011, 2013). Here, we propose that foreground enhancement contributes to streaming repeating sounds in the context of a random background.

However, it is important to note that an expectation of enhanced responses to "foreground" stimuli may reflect a biased expectation. There is no *a priori* requirement for sounds that perceptually pop out as a foreground to evoke an enhanced (or less suppressed) neural response. For example, Bar-Yosef and collaborators, investigated the interactions in neural activity during simultaneous presentation of bird chirps and background noise, simulating a naturalistic auditory scene (Bar-Yosef & Nelken 2007; Bar-Yosef et al. 2002). To their surprise, responses of A1 neurons in anesthetized cats

were in fact dominated by the background noise, despite it being presented at a lower intensity than the foreground bird chirp. The authors interpreted this finding in an evolutionary context that sees it as advantageous for a prey to pay attention to subtle changes in the background to avoid predators which might be using foreground sounds to mask the sound of their own approach. This example shows that the brain might enhance different components of the sound depending on the context and identity of that sound. Thus, it is important to interpret the current findings through the critical lens of our own biased interpretation of what they might mean and consider the traditional ecosystem niche of the animal model. More experiments directly comparing these variations in context will be needed to further elucidate how the brain streams repeated sound features.

Streaming analysis

To capture differences in responses to simultaneous repeated and non-repeated noise samples, we relied on model predictions. In this paradigm, the neural response is necessarily the sum of responses to two simultaneous stimuli, and the component responses cannot be separated in the raw neural firing rate. Therefore, we constructed encoding models that teased apart stream-dependent activity computationally. This analysis showed that, even though most neural responses were suppressed by repetition—likely due to the phenomenon of response adaptation (Grill-Spector et al. 2006; Pérez-González & Malmierca 2014; Ulanovsky et al. 2004), responses to the foreground stream where less suppressed than the background or even enhanced. This approach established a methodology that could be used in similar cases in which there is a need to separate effects on neural responses of simultaneously occurring inputs.

The challenge of separating responses to simultaneously occurring sounds has been previously addressed for neural population activity using a similar modeling approach (Ding & Simon 2012). Ding and Simon asked human subjects to listen to one of two competing speakers and recorded brain activity via magnetoencephalography (MEG). To investigate the neural encoding process, they fit a separate STRF (or more precisely a "TRF", given that MEG data could not resolve spectral tuning) model for each of the two simultaneously presented speech streams. Neural activity was found to selectively synchronize to the speech of the speaker to whom attention of the listener was directed. Furthermore, the latency and source location of the two components suggested a hierarchy of auditory processing in which the representation of the attended object emerges from core (primary) to posterior (secondary) auditory cortex (Ding & Simon 2012). These results are largely consistent with the foreground enhancement observed in the current study, suggesting that top-down attention and bottom-up pop-out effects could be mediated by common mechanisms.

Another approach used to investigate the neural signature of streaming is by stimulus decoding, or reconstruction (Ding & Simon 2012; Mesgarani & Chang 2012; Mesgarani et al. 2009). A decoding model describes the relationship between stimulus and response similarly to the STRF, but in the opposite direction. That is, decoding uses the neural output to reconstruct the sound stimulus input. If the reconstruction of the envelope has a higher correlation to the envelope of the attended stream rather than the non-attended stream or the two streams combined, it would suggest enhanced coding of the attended stream. In a human MEG study, Ding and Simon (2012) found that this was indeed the case. Similar results were also obtained by Mesgarani and Chang (2012) using

data collected from non-primary auditory cortex via electrocorticography (ECoG) (Mesgarani & Chang 2012).

In our study, a decoding analysis could complement the encoding approach, potentially revealing how the relative enhancement of the repeating stream allows for the separation of the two streams. Specifically, we would predict that for units with positive foreground enhancement the stimulus reconstruction would be more accurate for the foreground stream compared to the background stream, matching perception. However, in order to avoid numerical bias towards the foreground stream, an experimental design in which the background stimulus is always composed by the same set of stimuli would be necessary.

Relation of repetition enhancement to stimulus-specific adaptation

The ability of the brain to detect regularities is not only crucial for identifying an auditory object embedded in a noisy scene, but also for making predictions about the environment, thereby making the system sensitive to deviance (Bendixen et al. 2010; Winkler et al. 2009). Substantial effort has been devoted to understanding the mechanisms of deviance detection. In human encephalography (EEG), an enhanced deviant response is observed in the mismatch negativity (MMN), a pre-attentive event-related potential elicited by rare sounds randomly interspersed among frequent standard sounds (Näätänen 2001).

A similar, but arguably not overlapping (Farley et al. 2010; Nelken et al. 2013) phenomenon observed at the single-neuron level is stimulus-specific adaptation (SSA). SSA is the attenuation in the responses to a common, repeated sound that does not

generalize, or only partially generalizes, to a second, rare sound that is presented in the same sequence in alternative pattern (May & Tiitinen 2010; Pérez-González & Malmierca 2014; Ulanovsky et al. 2003). While evidence for SSA has been found in the inferior colliculus and thalamus (Anderson et al. 2009; Antunes et al. 2010; Malmierca et al. 2009), the first lemniscal region in which SSA has been shown to be prominent and strong is A1 (Malmierca et al. 2015; Nelken & Ulanovsky 2007). Mechanistically, SSA is thought to arise from a combination of feedforward synaptic depression and local cortical inhibition (Ayala & Malmierca 2013; Natan et al. 2015; Yarden & Nelken 2017).

Selective enhancement of the neural response to a repeating sound might seem like an intuitive prediction, based on behavioral studies of repetition-induced streaming (Agus et al. 2010; McDermott et al. 2011). However, this enhancement may be surprising when viewed in the context of SSA (Taaseh et al. 2011; Ulanovsky et al. 2003). If SSA affects responses to simultaneous stimuli the same way as responses to sequential stimuli, one would expect a relative suppression of responses to the foreground stream in repeated embedded noise. Additionally, response adaptation should be larger in units preferring the target over other samples. However, our results show the opposite effect, *i.e.*, a relative suppression of the non-repeating background stream, especially for target-preferring neurons. We propose that while SSA can account for the overall decreased response to both streams (Grill-Spector et al. 2006; Pérez-González & Malmierca 2014), a separate mechanism must be responsible for the additional suppression of sounds that occur simultaneously to the repeating foreground. Furthermore, the fact that foreground enhancement was more prominent in secondary auditory cortical fields (PEG) with

respect to primary areas (A1), suggests a hierarchical mechanism by which the enhancement emerges along the auditory cortical pathways.

Animal models for streaming

Most behavioral studies of auditory streaming have been performed in humans (Gutschalk & Dykstra 2014). This is not surprising as measuring auditory streaming in nonhuman species is complicated by the fact that they cannot simply be asked to attend to a given cue or report what they perceive – for recent reviews on nonhuman behavioral studies of auditory streaming see (Bee & Micheyl 2008; Fay 2008). The ferret has been previously identified as an animal model for streaming of alternating tone sequences and tone clouds (Ma et al. 2010; Micheyl et al. 2007b), and used to study its neurophysiological bases (Elhilali et al. 2009a).

Here, we developed the ferret as a model for streaming repeated sequences of simultaneously presented complex sounds. We designed an auditory task where animals had to report the occurrence of a repetition emerging from random overlapping noise samples. Ferrets were able to perform this task, suggesting that they could perceive repetition of complex sound features as a distinct component of the stimulus. Furthermore, given that the identity of the repeated sample was changed across behavioral blocks, we could exclude the possibility that the animals used specific spectro-temporal features of the target sample to perform the task. While this is not a direct proof that ferrets perceived two separate streams in the same way as humans, it confirms that they did perceive the occurrence of repetitions.

The role of attention in repetition-induced streaming

Our physiological experiments were conducted on passively listening ferrets without explicit control of attention. While attention is known to modulate sensory responses across multiple brain areas (Cohen & Maunsell 2009; Reynolds & Chelazzi 2004; Sundberg et al. 2009), the role of attention on repetition-based streaming is controversial. Masutomi, McDermott *et al*. tested this question directly by asking human subjects to perform the same task as in McDermott *et al*., 2011, but while also performing a decoy visual task (Masutomi et al. 2015). The authors found that human listeners were equally able to recover the identity of the repeating noise sample even when their attention was directed away from the sound, indicating that repetition-based streaming is a bottom-up process.

Several other studies have shown that human listeners are extremely sensitive to regular patterns rapidly emerging from complex sequences of sound (Barascud et al. 2016; Teki et al. 2011). Barascud *et al*. investigated how human listeners discover temporal patterns and statistical regularities in complex sound sequences (Barascud et al. 2016). They found that subjects' behavior matched the one of an ideal observer, even when distracted by a decoy visual task, again suggesting that detection of sound repetition might be a phenomenon that does not require attentional focus.

Streaming of more complex sounds, however, is known to be facilitated by directing attention to specific sound components that distinguish a foreground from a background. For example, Mesgarani & Chang, 2012 presented human listeners with two streams of speech. What was referred to as foreground or background changed across trials in response to a specific word that cued participants to either listen to the female or

the male voice. The authors found that listeners were much better at reporting the content of sentences that they were cued to pay attention to with respect to non-cued sentences presented simultaneously. Furthermore, the signature of this "foreground enhancement" is present at the level of neural activity measured by ECoG (Mesgarani & Chang 2012). Future experiments incorporating behavior into neurophysiological recordings may explain whether the pre-attentive foreground enhancement effects reported here are mediated by the same mechanisms as those that enhance actively attended streams.

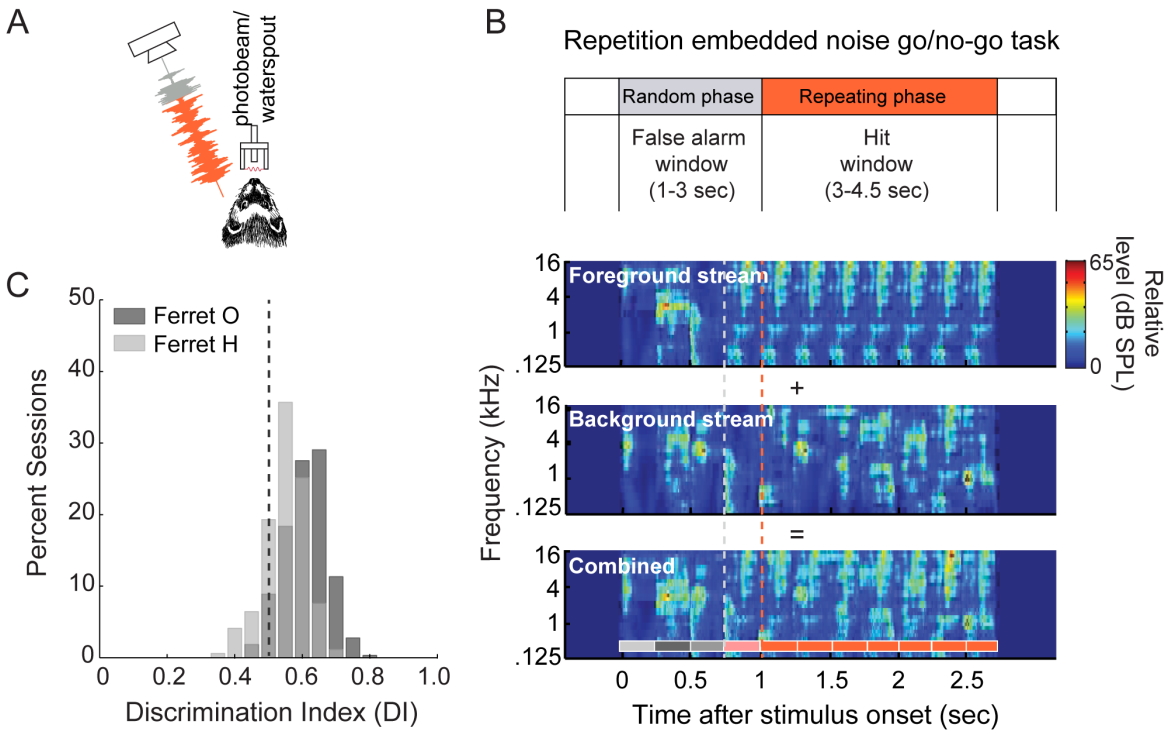

Figure 2.1: Ferrets are sensitive to repetitions embedded in mixtures.

A. Ferrets were trained to respond to sound repetition by licking a waterspout. **B.** Schematic of the go/no-go task and spectrograms of repetition embedded noise stimuli from an example behavioral trial. Animals were exposed to the combination (bottom spectrogram) of two overlapping streams: a foreground stream containing a target sample (top), and a background stream, a non-repeating sequence of noise samples (middle). In this example, the target sample (orange boxes, bottom panel) starts repeating after three random noise samples (grey boxes). The grey dashed line marks the first occurrence of the target sample (pale orange), which is included in the random phase for analysis. The transition between random and repeating phase is marked by the orange dashed line and occurs when the target sample is first repeated. Animals were trained to withhold licking from a waterspout during the random phase. To receive a water reward, they had to lick the waterspout following repetition onset. **C.** Distribution of discrimination index (DI) across behavior sessions for ferret O and ferret H after training was completed. Dashed line (0.5 DI), indicates chance performance.

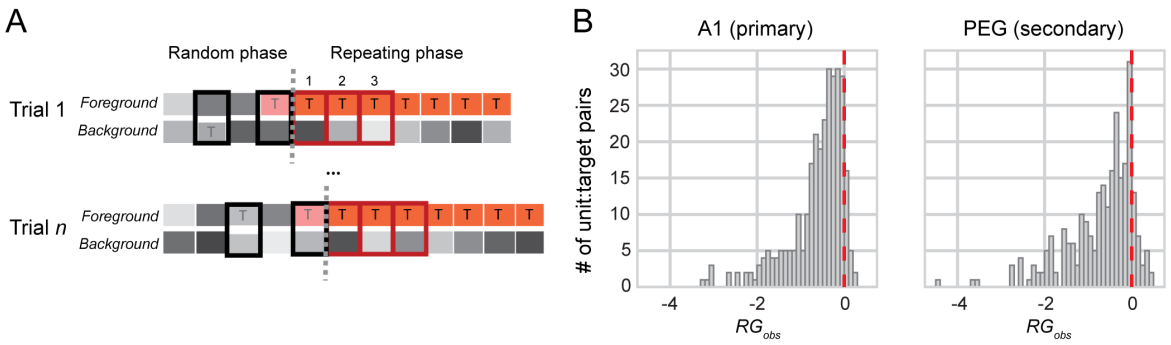

Figure 2.2: Activity in both A1 and PEG was suppressed during the repeating phase

A. Schematic of two trials. The background stream consisted entirely of random noise samples (gray). The foreground stream contained random samples during the random phase, which sometimes included the target sample (T). The final sample of the random phase is the target (light orange) as it had not yet begun to repeat. Average PSTH responses to each pair of samples that contained the target (thick rectangles) were computed separately for the random phase and repeating phase (separated by the grey dashed line in the schematic). **B.** Distribution of observed repetition gain (RG_{obs}) for A1 and PEG. The majority of target responses were suppressed ($RG_{obs} < 0$) during the repeating phase. Red dashed line indicates 0 (*i.e.*, no difference between phases). Since results may depend on how well the unit responded to the target, all analyses of neural responses were performed separately for each unique unit-target pair (n = 304 A1, n = 276 PEG). Median RG_{obs} values, A1: -0.64 (95% CI [-0.74, -0.54]); PEG: -0.70 (95% CI [-0.81, 0.60]).

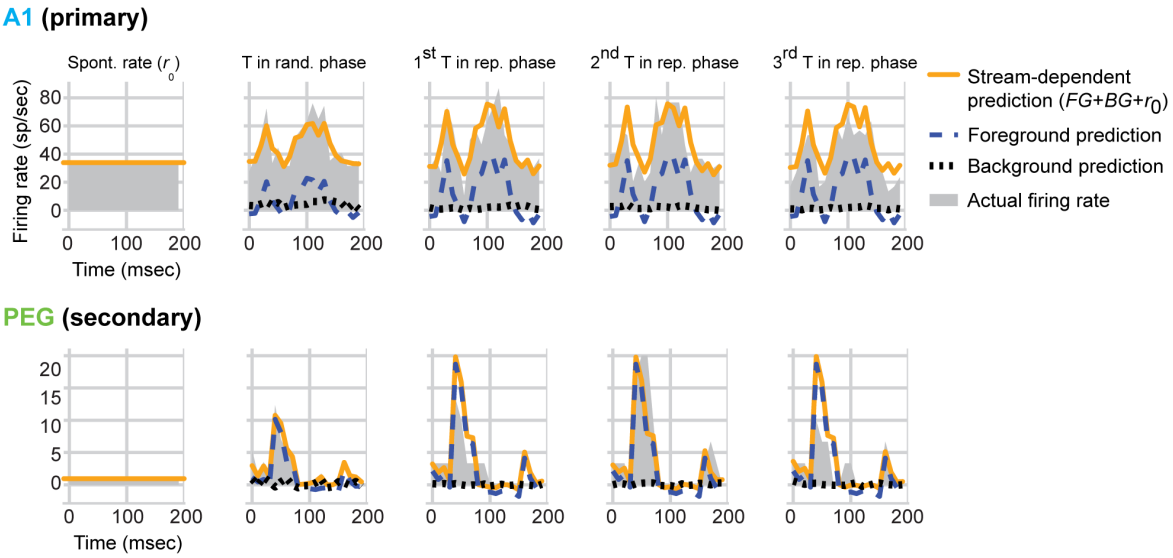

Figure 2.3: Example activity A1 and PEG

Example PSTH responses of units in A1 (top) and PEG (bottom) to the target sample (T) in the random versus repeating phase. Spontaneous rate (SR) is shown (1st column) for reference. Predictions from the stream-dependent model (orange) broken down into the contribution of the foreground (blue, dashed) and background (black, dotted) streams, shown for the average responses to the target in the random phase (2nd column), and the first three repetitions of the target sample in the repeating phase (3rd, 4th, and 5th columns).

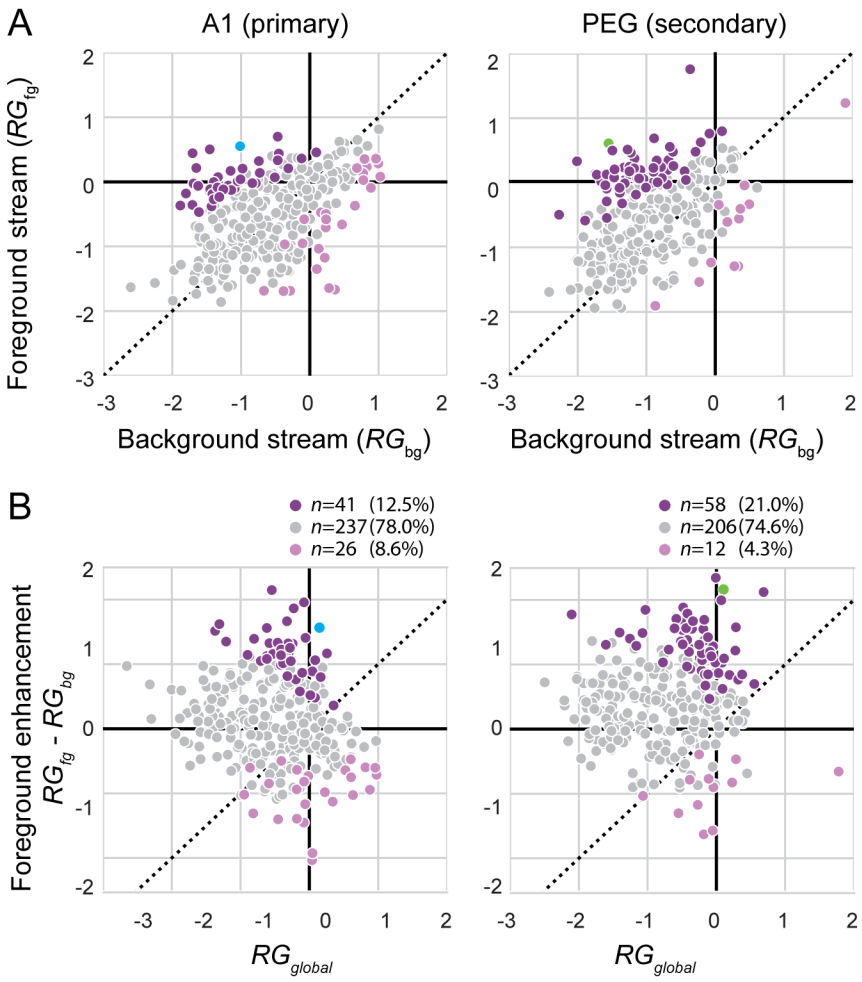

Figure 2.4: Selective foreground enhancement in A1 and PEG

A. Foreground (RG_{fg}) versus background (RG_{bg}) repetition gain measured in the stream-dependent model in A1 (n = 304 unit-target pairs; mean RG_{bg} = -0.610, mean RG_{fg} = -0.486) and PEG (n = 276; mean RG_{bg} = -0.935; mean RG_{fg}= -0.518). Color indicates unit-target pairs in which values of RG_{fg} are significantly higher (dark purple) or lower (light purple) than values of RG_{bg} (95% credible interval for the difference does not overlap with 0). Grey indicates no significant difference. Dashed line indicates equality. **B.** Foreground enhancement ($RG_{fg} - RG_{bg}$) plotted against overall gain change (RG_{global}) during the target phase for A1. Colors are as in B. Number of data points with significant foreground enhancement, significant background enhancement and no change are shown above each plot. Data for examples in Figure 3 are highlighted for A1 (cyan) and PEG (green).

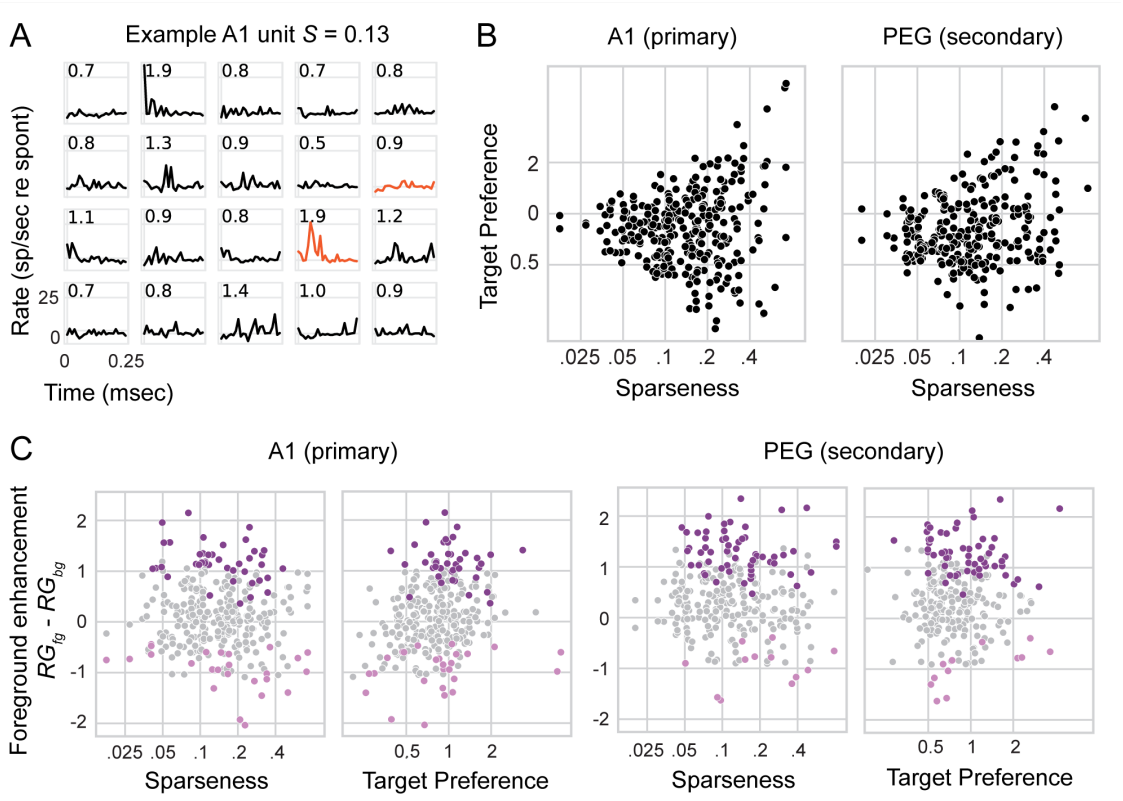

Figure 2.5: Relationship between target preference and sparseness in A1 and PEG units

A. PSTH responses to each of the 20 noise samples presented individually to a unit with relatively high sparseness ($S = 0.13$). Units such this one, responded well to only a few samples. Numbers indicate unit's preference for that particular sample with respect to the others. Responses to target samples are indicated in orange. **B.** Scatter plot of target preference versus lifetime sparseness for each unit recorded from A1 (left) and PEG (right). Target preference quantifies the response of a given unit to a target sample compared to the other 19 noise samples. Lifetime sparseness measures selectivity for the noise samples. Values of sparseness near 0 indicate units that responded similarly to all noise samples, and values near 1 indicate units that responded preferentially to a small number of samples. Units with high sparseness tended to have a greater variability in target preference. Data for examples in Figure 2.3 are highlighted for A1 (cyan) and PEG (green). **C.** Scatter plots of foreground enhancement as a function of target preference and sparseness in A1 and PEG.

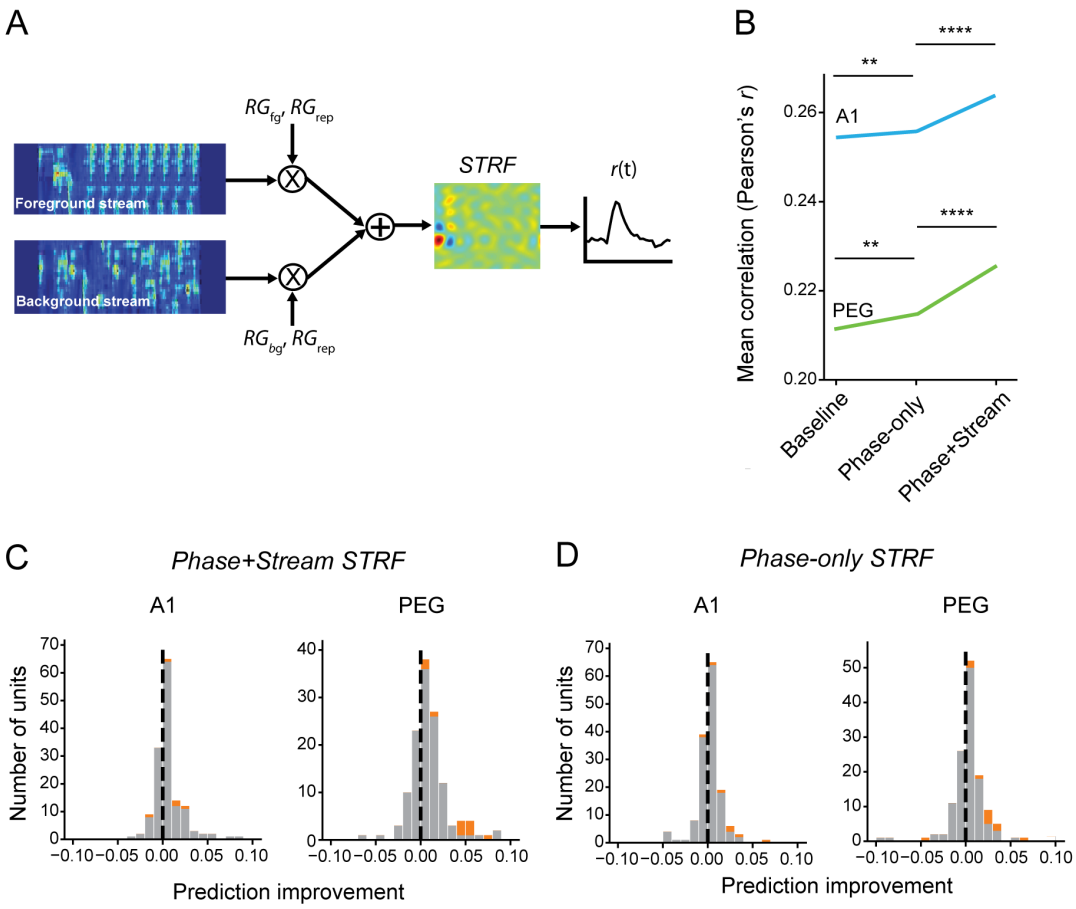

Figure 2.6: STRF-based model corroborates PSTH-based model findings of stream-specific gain changes in A1 and PEG.

A. Schematic representation of the STRF-based model. **B.** Mean prediction correlation coefficient (Pearson's *r*) for each area plotted for the *baseline* STRF-based model (both stream identity and repetition permuted in time prior to fitting), for the *phase-only* model (stream identity shuffled), and for the full *phase+stream* model. **C.** Histogram of difference in prediction accuracy between the *phase+stream* and *phase-only* STRF model for each neuron in A1 and PEG. A positive change shows a benefit of incorporating stream identity (foreground versus background) into the model. Units with a significant improvement in accuracy for the full model are plotted in orange ($p < 0.05$, jackknife *t*-test). Mean prediction correlation improved for the phase-only model in both areas (mean difference, A1: 0.003, $p < 0.0001$; PEG: 0.004, $p < 0.0001$, sign test). **D.** Histogram of difference in prediction accuracy between the *phase-only* and *baseline* STRF model for

each neuron, plotted as in C. A positive change indicates a benefit of incorporating repetition (random versus repeating phase) into the model. Mean prediction correlation improved for the *phase+stream* model over the *phase-only* model in both areas (mean difference, A1: 0.005, $p < 0.0001$; PEG: 0.007, $p < 0.0001$).

area	effect	coef.	coef	std err	z	P>|z|	Conf. Int. Low	Conf. Int. Upp.
A1	intercept	β_0	0.148814	0.041729	3.566217	3.621717e-04	0.067027	0.230601
	sparseness	β_2	-0.419478	0.315404	-1.329973	1.835271e-01	-1.037658	0.198701
	target preference	β_4	0.464842	0.106181	4.377810	1.198775e-05	0.256730	0.672954
	target preference * sparseness	β_6	-1.043268	0.266257	-3.918273	8.918570e-05	-1.565123	-0.521414
PEG	intercept	$\beta_0 + \beta_1$	0.416046	0.044176	9.417978	4.598578e-21	0.329463	0.502629
	sparseness	$\beta_2 + \beta_3$	-0.317036	0.324972	-0.975580	3.292725e-01	-0.953968	0.319897
	target preference	$\beta_4 + \beta_5$	0.049695	0.099294	0.500480	6.167374e-01	-0.144918	0.244307
	target preference * sparseness	$\beta_6 + \beta_7$	-0.106379	0.354988	-0.299671	7.644285e-01	-0.802143	0.589384

Table 2.1:

Results of the mixed linear model for foreground enhancement, with target preference (*TP*), sparseness (*S*), and area as fixed effects. Results are broken down by area. Significance was assessed using a post-hoc test of contrasts (*t*-test).

Chapter 3

Dissociation of task engagement and arousal effects in auditory cortex and midbrain

Daniela Saderi, Zachary P. Schwartz, Charlie R. Heller, Stephen David

Oregon Hearing Research Center, Oregon Health & Science University,

Portland, OR 97239, USA

Acknowledgements: This study was supported by NIH grants F31 DC014888 (DS), R010495 (SVD), and from the Tartar Trust at Oregon Health & Science University (DS).

Abstract

The brain's representation of sound is influenced my many different aspects of internal state. To isolate effects of different state variables on auditory processing, we simultaneously controlled task engagement and monitored fluctuations in arousal via pupillometry. Single- and multiunit activity was recorded in primary auditory cortex (A1) and the inferior colliculus (IC) of ferrets trained to a go/no-go tone detection task. We used a generalized linear model to isolate the contributions of task engagement and arousal on spontaneous and evoked neural activity. As expected from previous studies, fluctuations in pupil-indexed arousal were correlated with changes in task engagement, but their effects could be dissociated in most experiments. In both areas, units were modulated by task, by arousal, or by both. However, arousal effects were more prominent in IC. Engagement and arousal effects had variable sign in IC, but were mostly positive with spike rate in A1. These results indicate that some changes attributed to task engagement in previous studies should in fact be attributed to global fluctuations in arousal. Moreover, these arousal effects may explain differences in neural activity observed between passive conditions pre- and post-behavior. This same approach can be used to account for other state variables, such as selective attention and behavioral effort, providing a general method for dissociating the influence of continuous and discrete behavioral state variables.

Introduction

Hearing is a dynamic process that requires integration of the sensory evidence provided by physical attributes of sound with information about the behavioral context in which auditory perception occurs. It is well-established that sensory cortical and, more recently, subcortical regions operate as adaptive processors, making different calculations around the incoming auditory stimuli based on the immediate behavioral demand. Compared to passive listening, engaging in a task that requires subjects to associate the same set of sounds to a behavioral response motivated by a positive or negative reinforcer, leads to sizable changes in neurons' excitability, spectro-temporal and spatial receptive fields (Downer et al. 2015; Fritz et al. 2003, 2005a; Knudsen & Gentner 2013; Kuchibhotla et al. 2017; Lee & Middlebrooks 2011; Otazu et al. 2009; Ryan & Miller 1977; Yin et al. 2014). These changes are often attributed to the rapid adaptation of auditory neurons to the current behavioral demand, such as sensory representation is optimized to the relevant component of the task (Bagur et al. 2018; Fritz et al. 2007a; Mesgarani et al. 2010; Natan et al. 2017; Niwa et al. 2012). For some units, changes were reported to persist for several minutes after the active behavior, while for others, activity and tuning changes were observed to rapidly regress back to baseline (Fritz et al. 2003; Slee & David 2015a). One interpretation for such variability might be that the source of modulation is a non-auditory input carrying information about another state variable.

It is increasingly apparent from experimental studies that multiple state variables contribute to observed changes following "task engagement", acting together or in opposition to each other. For example, during behavior, the activity of auditory neurons is also affected by the anticipation of the sound (Jaramillo & Zador 2011), reward

associations (Beaton & Miller 1975; David et al. 2012), self-generated sound/stimulus interactions (Eliades & Wang 2003, 2005, 2008), and non-sound related variables, such as motor planning (Bizley et al. 2013) and decision-making (Tsunada et al. 2016). Furthermore, fluctuations are observed in neural activity throughout the forebrain, including primary sensory regions, even in the absence of external sensory input (Ringach 2009). This activity is related to cognitive states such as arousal, and might be acting to boost responses to sensory stimuli relevant to the current task demand (Fu et al. 2014; Wimmer et al. 2015). Therefore, attempting to interpret changes in neuronal activity solely in the light of a single state variable, such as task engagement, may lead to an incomplete and sometimes even incorrect understanding of how sensory information is processed under conditions that necessarily engage multiple aspects of cognition.

A straightforward and non-invasive way to measure ongoing changes in the state of cognition, is monitoring non-luminance-mediated changes in pupil size (Appen 1994; Granholm & Steinhauer 2004; Kahneman & Beatty 1966). In humans, changes in pupil size under constant light exposure have been shown to correlate with mental effort (Beatty 1982; Kahneman & Beatty 1966; Wierda et al. 2012; Winn et al. 2015), changes in states of arousal (Granholm & Steinhauer 2004; Kahneman & Beatty 1966), aspects of decision-making (Gilzenrat et al. 2010), and task performance (Jepma & Nieuwenhuis 2011; Schriver et al. 2018). Fluctuations in pupil size also closely track locomotion and evoked and spontaneous activity of different population of neurons throughout the mouse forebrain (McGinley et al. 2015b; Reimer et al. 2016; Stringer et al. 2018; Vinck et al. 2015). While both pupil size and locomotion have been used as an assay for arousal, pupil size has more accurate information as it can reveal changes in arousal even when

the animal is not moving (Shimaoka et al. 2018). Thus, pupil size can be easily measured alongside neuronal activity and incorporated as a continuous state variable in encoding models aimed at characterizing individual and interactive components of the effects of multiple state variables onto auditory activity.

To study the interaction of task engagement and arousal on auditory neural coding, we recorded extracellular firing activity and pupil size of ferrets performing an auditory task or passively listening to the task stimuli. Single- and multiunit activity was recorded from the primary auditory cortex (A1) and inferior colliculus (IC). We used a step-wise approach to fit a generalized linear model that quantified the relative contribution of pupil-indexed arousal and task-related changes in neural activity. We found that incorporating arousal state in the model significantly reduced the variability in firing rate attributed to task engagement. This result suggests that some previously reported effects of task engagement may in fact be explained by changes in arousal captured by pupil size fluctuations. Furthermore, subpopulations of recorded neurons were modulated uniquely by task or by arousal, suggesting that these effects might be mediated by distinct modulatory circuits. Finally, the magnitude of effects attributed to state changes was uncorrelated with auditory responsiveness across neurons in A1. Conversely, in IC task effects and responsiveness were negatively correlated, suggesting that top-down and bottom-up signals are initially segregated but become intermingled in cortex. This work establishes a general method for integrating continuous and discrete behavioral state variables into the architecture of encoding models aimed at describing sensory processing across behavioral contexts.

Materials and Methods

All procedures were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee and conform to the National Institutes of Health standards.

Surgical procedure

Animal care and procedures were similar to those described previously for recording neurophysiological activity from awake behaving ferrets (Slee & David 2015a). Four young adult male ferrets were obtained from an animal supplier (Marshall Farms, New York). A sterile surgery was performed under isoflurane anesthesia to mount two head-posts for subsequent head fixation and to expose auditory brain regions for recordings. A UV light-cured composite (Charisma, Heraeus Kulzer) allowed the placing of two custom-made stainless-steel head-posts spaced approximately 1 cm apart along the sagittal crest of the skull. The stability of the implant was also achieved via 6-8 stainless self-tapping set screws mounted in the skull (Synthes). The whole implant was then shaped to its final shape with the Charisma and acrylic denture material (Co-Oral-Ite). On each hemisphere, two 1.2x1.2 cm wells in which the skull was covered with only a thin layer of Charisma were built to allow access to auditory regions and to contain sterile saline during recordings.

Following the surgery, animals were treated with prophylactic antibiotics (Baytril 10 mg/kg) and analgesics (Buprenorphin 0.02 mg/kg) under the supervision of University veterinary staff. For the first two weeks the wound was cleaned with antiseptics (Betadine 1:4 in saline and Chlorhexidine 0.2%) and bandaged daily with application of topic

antibiotic ointment (Bacitracin). After the wound margin was healed, cleaning and bandaging occurred every 2-3 days through the life of the animals. This method revealed to be very effective in minimizing infections of the wound margin.

After recovery (~2 weeks), animals were habituated to a head-fixed posture inside the sound-booth chamber for about 2 weeks prior to the beginning of the training.

Behavioral paradigm and training

Four animals were trained by instrumental conditioning to perform a positively reinforced, tone *versus* noise discrimination task (Ferret L) (Slee & David 2015a) or tone-in-noise detection task (Ferrets R, B, T) (McGinley et al. 2015a). Animals under training were provided access to water *ad libitum* on weekends, but were placed on water restriction during the weekdays (Monday through Friday) to maintain 90% of their normal body weight long-term. Animals had to report the presence of a target tone either presented alone (tone *versus* noise task) or embedded last (tone-in-noise task) in a sequence of 2-5 temporally orthogonal ripple combinations (TORCs, references; 30 samples, 5 octaves, 0.75 sec duration, 0.35 sec inter-stimulus interval) (Klein et al. 2006) by licking a water spout (Figure 3.1A). Licks were detected by a piezoelectric sensor glued to the water spout. Licks occurring within the target window, 100ms after the presentation of the target and within 1.4 seconds of target onset, were rewarded with 1-3 drops of a 2:1 solution of water and Ensure. Licks occurring during the reference window, 1.1-4.5 seconds, during the presentation of TORC stimuli resulted in a 5-8 seconds timeout. The inter-trial interval was 2.5 ± 0.5 seconds. The number of TORCs per trial was distributed randomly with a flat hazard function to prevent behavioral timing

strategies. During the training phase, the TORC sample masking the target varied randomly in each trial to prevent animals from using TORCs' spectro-temporal features to identify targets. Target tone frequency (100 Hz-20000 Hz) was fixed within a behavioral block but varied across blocks. During training, target frequency was chosen at random to span the frequency range of a ferret. During electrophysiological recordings, target tone frequency was selected to best match the best frequency of the recording site. At the beginning of training, the tone was presented at +40 dB signal-to-noise ratio (SNR; ratio of peak-to-peak amplitude) relative to the TORCs. This difference was gradually reduced over the course of two-three weeks until the animal consistently performed above chance (three behavioral blocks with performance yielding to > 0.5 discrimination index, see below) at 0dB SNR.

For the tone-in-noise variation of the task, the SNR of the tone with respect to the overall level of the TORCs (fixed at 55 or 60dB SPL depending on the animal) varied between +5 and -20dB SNR, in 5dB steps. Each session included five target/noise SNRs. To manipulate task difficulty within each session, the probability of each of the five target/noise SNRs varied, yielding two difficulty conditions: a high SNR ("easy") condition in which 60% of the trials the target/noise SNR was either the highest or the second to the highest SNR within the target/noise SNR distribution; and a low SNR ("hard") condition in which the two lowest target/noise SNRs occurred in 60% of the trials. For example, for ferret B, the distribution after training was completed was kept between 0 and -20dB SNR, so that in the easy condition 0 and -5dB SNR targets would appear 60% of the time, -10dB SNR 20% of the time, and the remaining -15 and -20dB SNR would be presented another 20% of the trials. During electrophysiological tone-in-

noise experiments, the tone was embedded in a single TORC sample, which also occurred in the reference period. We confirmed that animals were not biased to respond to this TORC exemplar in the reference phase.

Sound presentation

Behavioral training and subsequent neurophysiological recording took place in a sound-attenuating chamber (Gretch-Ken) with a custom double-wall insert. Stimulus presentation and behavior were controlled by custom MATLAB software (code available at https://bitbucket.org/lbhb/baphy). Digital acoustic signals were transformed to analog (National Instruments), amplified (Crown), and delivered through two free-field speakers (Manger, 50-35 000 Hz flat gain) positioned ±30 degrees azimuth and 80 cm distant from the animal. Stimuli were presented either from the left or the right speaker, contralaterally to the recording site. Sound level was equalized and calibrated against a standard reference (Brüel & Kjær).

Pupil recording

During experiments, infrared video of one eye was collected for offline measurement of pupil size. Ambient light was maintained at a constant level to prevent light-evoked changes in size and maximize dynamic range. Recordings were collected using a CCD camera (Adafruit TTL Serial Camera 397) fitted with a lens (M12 Lenses PT-2514BMP 25.0 mm) whose focal length allowed placement of the camera 10 cm from the eye. To improve contrast, the imaged eye was illuminated by a bank of infrared LEDs. Ambient luminance was provided using a ring light (AmScope LED-144S). At the

start of each recording day, the intensity of the ring light was set to a level (~1500 lux measured at the recorded eye) chosen to give a maximum dynamic range of pupil sizes. Light intensity remained fixed across the recording session.

Neurophysiology

After animals demonstrated to perform consistently above chance (DI > 0.5 for five consecutive blocks), a small craniotomy was opened to access either primary auditory cortex (A1) (Schwartz & David 2018) or central and external nuclei of the inferior colliculus (IC) (Slee & David 2015a). Extracellular spontaneous and evoked neuronal activity was recorded in non-anesthetized ferrets either using a tetrode (Thomas Recording Inc.), or a linear 64-channel silicon probe (Shobe et al. 2015). The impedance of the tetrode was measured to be 1-2 MOhm, and the 64-channel probe was electroplated to reach a 0.7-MOhm impedance in each electrode. The tetrode or the probe were independently moved through the tissue via a motor system (Alpha-Omega).

Amplified (AM Systems) and digitized (National Instruments) electrophysiological signals were stored using the open-source data acquisition software MANTA (Englitz et al. 2013) or Open Ephys (Black et al. 2017). Recording sites were confirmed as being in A1 based on tonotopy and relatively reliable and simple response properties (Atiani et al. 2014; Shamma et al. 1993). Recording location in the IC were determined by tonotopic maps and basic tuning properties (Aitkin & Moore 1975; Aitkin et al. 1975; Moore et al. 1983; Slee & David 2015a). Neurons in the central nucleus of the IC (here referred to as ICC) receive input from the auditory brainstem, and have a characteristic short response latency, dorsal-ventral tonotopy, and narrow bandwidth

tuning. Conversely, regions around the central nucleus do not receive direct ascending input and present longer response latencies, considerably less sharp tuning, lack consistent tonotopic organization; these areas were grouped together as NCIC (Slee & David 2015a).

For tetrode recordings, upon unit isolation, a series of brief (100-ms duration, 200-400ms inter-stimulus intervals, 50 dB SPL) tones and/or narrowband noise bursts were used to determine the range of frequencies that evoked the strongest response, the best frequency (BF) of the unit(s). If a unit(s) in the site did not respond to the sound stimuli (that being an evoked increase or decrease in activity compared to spontaneous activity either during or right after sound presentation), the electrode was moved to a new recording depth with small, 5um incremental steps. For the 64-channel recordings, we lowered the entire depth of the probe (1 mm) such that it spanned the depth of auditory cortex or the inferior colliculus. Frequency tuning across the 64 channels varied. For the purposes of this study, we considered units isolated from the 64-channel as individual units.

Spike sorting

Putative spikes were sorted offline by band-pass filtering the raw trace (300–6000 Hz). Single units were extracted from the continuous signal by collecting all events ≥4 standard deviations from zero. To separate single units and stable multi units from the electrode signal, we used the Catamaran clustering program (kindly provided by D. Schwarz and L. Carney) (Schwarz et al. 2012) for tetrode recordings, and the software KiloSort (Allen et al. 2018; Pachitariu et al. 2016) for the 64-channel recordings. In both cases, units were defined based on visual inspection of traces and by having less than 1%

of inter-spike intervals and less than 2.5% inter-spike intervals shorter than 0.75ms for single units and multi units, respectively. Stability of single-unit isolation was verified by examining waveforms and interval histograms. If isolation was lost during a behavioral block, only activity during the stable period was analyzed.

Analysis

Task performance analysis

Behavioral performance was measured using signal detection theory (Signal Detection Theory for Everyman? 1993; Stanislaw & Todorov 1999). Hits or false alarms occurred when the animal licked the piezo waterspout upon presentation of the target or the reference stimuli, respectively. Misses or correct rejections (CR) occurred when the animal did not lick following presentation of the target or reference. Hit rate (HR) was calculated as the proportion of licks that occurred upon presentation of the target sound stimuli (P(licks|target)) during the target window (starting 100ms after target onset, 1.4-sec duration), according to the formula HR=hits/(hits+misses). False alarm rate (FAR) was calculated as the proportion of licks that occurred during presentation of a reference TORC sound or between TORCs (P(licks|reference)) during the reference window (1.1-4.5 seconds), FAR=FAs/(FAs+CRs). Sensitivity (*d*'), a measure of the animals' ability to discriminate between the signal (target) and the noise (non-target, both external and internal), was measured by taking the difference between the z-scored hit rate and the false alarm rate (*d*'=ZHR-ZFAR). When the animal cannot discriminate at all, *d*'=0. Animals were considered trained to the task and ready to undergo electrophysiological recordings when they performed consistently above chance for three consecutive session.

Pupil analysis

Pupil size was measured using custom MATLAB software (code available at https://bitbucket.org/lbhb/baphy). For each video, an intensity threshold was selected to capture pupil pixels and exclude the background image. During initial recordings, the threshold was selected manually. Because we observed that the first peak in the intensity histogram of the image generally corresponded to the pupil, during later recordings we automatically positioned the threshold at the first "valley" in the intensity histogram of each video frame. Each frame was smoothed by a Gaussian filter to remove shot noise before thresholding, then segmented by Moore boundary tracing (Gonzalez & Woods 2010). The segment with largest area was identified as pupil. We measured pupil size as the length of the minor axis of an ellipse fit to this region. We chose to measure the minor axis of the elliptical ferret pupil over the major axis, because the latter was sometimes interrupted by the reflection of the IR light that was used to visualize the pupil. To avoid identifying shadows or other dark regions of the image as pupil, the search for largest-area segment was restricted to a rectangular region of interest surrounding the pupil identified in the preceding frame (Nguyen & Stark 1993).

Blink artifacts were identified by extremely rapid, transient changes in pupil size. The numerical derivative of the pupil trace was taken and bins with derivatives more than 6 standard deviations from the mean were marked as blinks. A one-second window was removed from the pupil trace around each blink event, and the segment was replaced by a linear interpolation between valid points at its edges.

The framerate of the camera used in the experiments had an upper limit of 30 frames/second. To compensate for variability in the actual framerate, we recorded a

timestamp at the start and end of each trial, then interpolated measurements of pupil size as necessary to maintain a constant interval between measurement throughout each recording.

When comparing pupil and neural data, a 750-ms offset was applied to pupil trace to account for the lagged relationship between changes in pupil size and neural activity in auditory cortex (McGinley et al. 2015a).

Spectro-temporal receptive fields

Tuning properties of A1 and IC neurons were characterized by the spectro-temporal receptive field (STRF) estimated by reverse correlation between time-varying neuronal spike rate and the rippled noise used as reference sounds during behavior. For a stimulus with spectrogram, $s(x,t)$, and instantaneous firing rate r(t), the STRF, $h(x,u)$ is defined as the linear weight matrix that predicts the firing rate response as follows:

$$r(t) = \sum_{x=1}^{X} \sum_{u=1}^{U} h(x,u)\, s(x,t-u) \qquad (Eqn.\ 3.1),$$

where u is the time lag of the convolution kernel. Each weight of h indicates the gain applied to frequency channel x at time lag u to produce the predicted response. Positive values indicate components of the stimulus correlated with increased firing, and negative values indicate components correlated with decreased firing.

While we did not use the STRF to measure or quantify the state modulation, we used it to characterize units' responsiveness to auditory sounds and to determine tonotopy. We also use the STRF model to estimate if a unit after the post hoc sorting analysis was to considered on-BF, where its best frequency was near the frequency of the

target tone, or if it was off-BF, with a distance between unit's BF and target tone frequency of 0.5 octaves or more.

State models

We fit a generalized linear model in which time-varying state variables, pupil-indexed arousal, $p(t)$, and task engagement, $b(t)$, were used to re-weight each neuron's mean evoked response to each noise stimulus, $r_0(t)$, and spontaneous rate, s_0, to generate a prediction of the single-trial spike rate, $r_{full}(t)$, at each point in time. The model included multiplicative gain parameters (g) and DC offset parameters (d) to capture both types of modulation. We refer to this model as the *full model*:

$$r_{full}(t) = s_0\left(d_0 + d_p p(t) + d_b b(t)\right) + r_0(t)\left(g_0 + g_p p(t) + g_b b(t)\right) \qquad (Eqn.\ 3.2).$$

With additional constant terms g_0 and d_0, this model required a total of six free parameters. For comparison, we calculated a state-independent model here referred to as the *null model*, in which the state variable regressors were shuffled in time, effectively reducing the model to

$$r_{null}(t) = s_0 + r_0(t) \qquad (Eqn.\ 3.3).$$

Because shuffling removes any possible correlation between state and neural activity, gain and offset parameters are reduced to d_0=g_0=1 and d_p=d_b=g_p=g_b=0. In practice fitting to the shuffled data produces parameter values slightly different from zero, and controls for noise in the regression procedure.

We also considered two *partial models*, one to predict responses based on pupil size only, $r_p(t)$, and one to predict responses based on behavior only, $r_b(t)$, in which the other regressor was shuffled in time. Thus, one model accounted for pupil only

$$r_p(t) = s_0\left(d_0 + d_p p(t)\right) + r_0(t)\left(g_0 + g_p p(t)\right) \quad (Eqn.\ 3.4),$$

and the final model accounted for task engagement only:

$$r_b(t) = s_0\left(d_0 + d_b b(t)\right) + r_0(t)\left(g_0 + g_b b(t)\right) \quad (Eqn.\ 3.5).$$

These models tested the effects of a single state variable while ignoring the other.

By comparing performance of the full model to each partial model, we could determine the unique contribution of each state variable to the neuron's activity. We used a 20-fold cross validation procedure to evaluate model performance. The model was fit to 95% of the data and used to predict the remaining 5%. Fit and test data were taken from interleaved trials. This procedure was repeated 20 times with non-overlapping test sets, so that the final result was a prediction of the entire response. Model performance was then quantified by the fraction of variance explained, *i.e.*, the squared correlation coefficient, R^2, between the predicted and actual time-varying response. Variance uniquely explained by single state variables was calculated as the difference between R^2 for the full model and for the partial model in which the relevant variable was shuffled in time.

Modulation Index

To quantify the modulatory effects of task and arousal on the firing rate of A1 and IC units, we computed a modulation index (MI) (Otazu et al. 2009; Schwartz & David 2018). MI was defined as the difference between the mean response to the same stimuli between two conditions, α and β, normalized by their sum,

$$MI_{\alpha\beta} = \left(r_\alpha - r_\beta\right) \Big/ \left(r_\alpha + r_\beta\right) \quad (Eqn.\ 3.6).$$

MI could be calculated between different behavioral blocks or between state conditions. In the case of task engagement, MI was calculated between active and passive conditions, MI_{AP}. For arousal, data from an experiment were divided at the median pupil

diameter, and MI was computed between large pupil (pupil > median, high arousal) and small pupil (pupil < median, low arousal), MI_{LS}. To quantify differences between the first and the second passives, MI was computed between passive 1 and passive 2, MI_{P1P2}.

To quantify changes in firing rates due to unique contributions of arousal or task, we used MI to test how well the regression model could predict state-related changes in neural activity. The modulation between conditions α and β predicted by the full model, is denoted $MI_{\alpha\beta}\ full$, where α and β are either active/passive ($MI_{AP}\ full$) or large/small pupil ($MI_{LS}\ full$). Similarly, modulations between conditions α and β predicted by the pupil partial model or the behavior partial model, are denoted $MI_{\alpha\beta}\ pupil\ only$ and $MI_{\alpha\beta}\ task\ only$, respectively. The MI uniquely predicted by including task engagement as a regressor is

$$MI_{AP}\ task\ unique = MI_{AP}\ full - MI_{AP}\ pupil\ only \qquad (Eqn.\ 3.7),$$

that is, MI predicted by the full model minus the MI predicted by a model in which behavior condition, but not pupil, is shuffled. The net result is the MI predicted by task engagement above and beyond modulation predicted by changes in arousal alone. Similarly, MI uniquely predicted by including pupil size as a regressor is

$$MI_{LS}\ pupil\ unique = MI_{LS}\ full - MI_{LS}\ task\ only \qquad (Eqn.\ 3.8).$$

Significant effects of regressing out pupil-indexed arousal were quantified by comparing the signed-normalized differences between $MI_{AP}\ task\ unique$ and $MI_{AP}\ task\ only$ using a one-sided Wilcoxon signed-rank test with an α = 0.05, including zero-differences between the two (Pratt treatment). Differences of the same quantities across areas were quantified using rank-sum test. Sign normalization was achieved by

multiplying the difference between $MI_{AP}\ task\ unique$ and $MI_{AP}\ task\ only$ in each unit by the sign of their mean.

Significantly modulated units were determined by comparing Pearson's r coefficients associated with the full model and with the difference between the full model and the task and pupil partial models using jackknifed t-test with an $\alpha = 0.05$ (Efron & Tibshirani 1986).

Data preprocessing and analysis were performed using custom MATLAB and Python scripts. Neural and pupil activity were binned at 20 samples/s before analysis. A Python library for the modeling portion of this analysis is available at https://github.com/LBHB/NEMS/.

Results

Changes in pupil size track task engagement

To study interactions between arousal and task engagement, we first trained four adult ferrets on a go/no-go auditory task. Animals reported the presence of a target tone and ignored broadband noise reference stimuli presented from a single speaker (Figure 3.1A, B). They were required to withhold licking a water spout during the presentation of the reference sounds, and were given liquid reward for licking during the target window (100-1500ms after target onset). Targets were pure tones, either presented alone (tone versus noise discrimination, Ferret L) or embedded in reference stimuli (tone-in-noise detection, Ferrets R, B, and T). After training, all animals performed consistently above chance (Figure 3.1C).

To track and record changes in pupil size, an infrared video camera was used to image the eye contralateral to the speaker emitting auditory stimuli (Figure 3.1A). Constant luminance was maintained throughout the experiment. Pupil was recorded when animals either passively listened to the task stimuli or actively performed the task. The distribution of pupil sizes differed between passive and active conditions (Figure 3.1D), such that mean pupil size was consistently larger during the active condition (Figure 3.1E). Within active behavioral blocks, pupil varied with task performance. While average values of pupil size were similar for false alarm and hit trials, values of pupil size during miss trials were lower, resembling values recorded during passive trials (Figure 3.1E, F). Furthermore, within trials, pupil dynamics also displayed a strong dependence on task condition and behavioral performance. During both hit and false alarm trials, pupil size increased rapidly following the trial onset. It remained mostly unchanged during passive trials and miss trials (Figure 3.1G-I). Thus, pupil size tracked slow changes in task-engagement (active versus passive blocks) as well as more rapid changes in trial-by-trial performance during the active condition.

Diversity of task engagement and arousal effects across A1 and IC neurons

We recorded single-unit (SU) and multiunit (MU) extracellular activity from the primary auditory cortex (A1, total of 129 units, 72 SU and 57 MU) and the inferior colliculus (IC, total of 66 units, 41 SU and 25 MU) of ferrets, while they switched between active engagement and passive listening. In both brain regions, activity was recorded from acute penetrations using either high impedance tetrode electrodes or 64-channel linear silicon arrays. In the IC, we recorded from central (ICC, total of 18 units,

14 SU and 4 MU) and non-central regions (NCIC, total of 48 units, 27 SU and 21 MU) (Slee & David 2015a). Results from the IC are here presented jointly between the two sub-regions, unless otherwise specified.

Because we were interested in studying how behavior state-dependent modulation of neural activity varied over time, we recorded activity in the same units over multiple passive and active blocks. Only well-isolated single units and stable multiunits were included in the analyses (see *Materials and Methods* for isolation criteria). To monitor changes in arousal, we recorded pupil diameter throughout each experiment. In order to analyze neural responses to acoustic stimulation under different behavioral conditions, the peri-stimulus time histogram (PSTH) was computed over responses to noise stimuli for each neuron. PSTHs were calculated for passive trials and active trials that were hits. On hit trials, animals did not lick during the reference period (TORCs noise stimuli). Thus this analysis minimized the possibility that lick-related motor signals confounded our results (Nelson et al. 2013; Schneider et al. 2014).

For many units the activity appeared to be modulated by changes in task engagement, as would be expected from previous behavioral studies (David et al. 2012; Fritz et al. 2003; Slee & David 2015a). During active behavior, responses to the noise stimuli either increased or decreased, and then returned to their baseline passive state during subsequent passive periods (example 1, Figure 3.2 A-C). For other units, however, no consistent change in activity was observed between active and passive conditions. In these cases, firing rate could change between behavioral blocks, but there was no significant change between responses for active versus passive conditions (example 2, Figure 3.2 E-G). Thus, despite our controls for sound acoustics and motor activity, the

firing rates of some units varied over time, but in a way that did not appear to be related to task engagement.

We wondered if changes in activity that could not be attributed to task engagement could instead be explained by fluctuations in pupil-indexed arousal. A simple way to investigate if arousal affects firing rates is to divide all trials (both hits and passive trials) into two groups based on the mean pupil size during each trial and compute a PSTH for all units in each respective group. One group was made for trials with pupil size smaller than the median, and another for trials with pupil size larger than the median. For both example units in Figure 3.2, this analysis showed that activity was enhanced when pupil was large (large versus small pupil PSTH, Figure 3.2D, H), indicating that pupil size was positively correlated with firing rate. Thus, we hypothesized that changes in pupil diameter accounted for fluctuations in firing rate when task engagement could not.

Because the active behavior state was correlated with large pupil (Figure 3.1E), we could not dissociate the effects of task engagement and arousal using just the raw PSTH. Therefore, to test our hypothesis that pupil accounted for changes in neural activity following task engagement, we fit a generalized linear model (full model, *Eqn. 3.2*, see *Materials and Methods*) in which the mean response to each presentation of the reference stimuli was modulated by task engagement and arousal (pupil size). Task engagement was modeled as a discrete regressor (active or passive) and pupil as a continuous regressor. This approach allowed us to measure the relative effects of arousal and task engagement on firing rate. Additionally, incorporating pupil as a continuous variable avoided arbitrarily dividing the data into large and small pupil groups.

Model performance was measured by the correlation coefficient (Pearson's r) between the predicted and actual spike rates averaged across all 20-fold cross-validation sets. To test for significant contributions of each regressor, we compared the performance of the full model, which included both pupil and task engagement variables, to the performance of a state-independent model, which only included the PSTH of the responses with pupil and task shuffled in time (null model, *Eqn. 3.3*, see *Materials and Methods*), and of two partial models, which only included the effects of a single state variable while shuffling the values of the other in time (pupil or task partial models, *Eqns. 3.4, 3.5*, respectively, see *Materials and Methods*).

We found that some units were driven more by task engagement and others by pupil-indexed arousal. Figure 3.2 shows the activity and pupil size dynamics of two A1 example units recorded simultaneously over the course of a 1-h recording. For units such as example 1 in Figure 3.2A-D, the modulation of firing rates in active and passive blocks was almost completely accounted for by the unique contribution of the task variable with little contribution of pupil-indexed arousal. Figure 3.2B shows model predictions plotted against the unit's firing activity averaged across trials in each behavioral block (each passive and active block were halved to show activity over time at higher time resolution). Responses predicted by the null model (Figure 3.2B, blue trace) did not capture the dynamics of this unit's modulation (Pearson's $r = 0.47$). However, when we added the task variable (task partial model; Figure 3.2B, purple trace), predictions significantly improved with respect to the null model (Pearson's $r = 0.54$; jackknifed t-test, $p < 0.05$). Interestingly, while we showed that pupil size positively related to firing rates (Figure 3.2D), adding pupil to the model (full model; Figure 3.2B,

black trace) did not improve prediction performance any further (Pearson's $r = 0.54$), suggesting that all variability in firing rate for this unit was accounted for by task engagement, and the apparent pupil effects in Figure 3.2D were in fact explained by task.

Conversely, for other units, such as the example in Figure E-H, the active/passive change in firing rate was significantly accounted for by arousal. A task partial model (Pearson's $r = 0.63$) did not lead to any significant improvement from the null model (Pearson's $r = 0.62$), suggesting that task was not a predictor of this unit's change in activity (Figure 3.2F). The pupil partial model, however, well captured this unit's modulation, leading to a significant improvement in prediction accuracy that did not differ from the full model (full model Pearson's $r = 0.73$; pupil partial model Pearson's $r = 0.72$; jackknifed t-test, $p < 0.05$). Thus, activity of example 1 and example 2 A1 units were predominantly modulated by task and arousal, respectively.

Next, we investigated how common task- and arousal-dominant modulations were across the population of A1 and IC units. Out of 129 units in A1 and 66 in IC, 51 (~40%) and 34 (~50%) were significantly modulated by one or both state variables, respectively (Figure 3.3A). Significance was tested by comparing the cross-validated variance explained (R^2) of the full model with the cross-validated R^2 estimated by the null model (jackknife t-test, $p < 0.05$; see *Materials and Methods*). Furthermore, by computing the difference between the cross-validated R^2 for the full model and for the two partial models separately, we were able to isolate the units with significant variance uniquely explained by either arousal or task alone (Figure 3.3A, B, purple and green, respectively). Two more significant groups were also found, one of units for which both task and arousal were uniquely contributing to the modulation (Figure 3.3A, B, black), and one for

which only the combined effect of task and pupil led to a significant state modulation (Figure 3.3A, B, dark grey). Interestingly, although the proportion of units with significant task modulation was comparable between A1 and IC, only about a third of significant state-modulation A1 units (16/51) showed unique arousal-dependent modulation compared to about two third for IC (21/34).

These results suggest that task- and arousal-dependent modulation of auditory processing is present in both the midbrain and cortex, and can be dissociated in about half of the sampled population of units that showed significant state modulation. However, state-dependent modulation in A1 was equally related to task or arousal, while in IC arousal signals make up a greater proportion of the state modulation.

Pupil-indexed arousal accounts for some apparent task engagement effects in both A1 and IC

So far, we have shown that a proportion of units in A1 and IC show state-dependent modulation of activity, with arousal being the dominant state variable in IC. Previous studies of task-related plasticity that have not measured pupil might have attributed the changes in firing rates across behavioral blocks to task engagement. Therefore, we asked to what extent arousal-related modulation could explain changes in activity between active and passive blocks that would be otherwise attributed to task alone.

To quantify the magnitude and sign of the modulation by each state variable, we compared the responses predicted by each model using a modulation index (MI), a metric that captured changes in both response gain and offset (*Eqn. 3.6*, see *Materials and*

Methods) (Otazu et al. 2009; Schwartz & David 2018). The modulation of activity captured by both pupil and task variables was computed from responses predicted by the full model ($MI_{AP}\ full$). The modulation captured by task alone was quantified by a modulation index computed from responses predicted by the partial task model, in which the variable pupil was shuffled in time ($MI_{AP}\ task\ only$, see *Materials and Methods*). The difference between $MI_{AP}\ full$ and $MI_{AP}\ task\ only$ yielded the residual and unique contribution of task engagement after accounting for pupil ($MI_{AP}\ task\ unique$, *Eqn. 3.7*).

To compare changes in the magnitude of the modulation index after accounting for arousal, we normalized the sign for each unit so that the mean of $MI_{AP}\ task\ only$ and $MI_{AP}\ task\ unique$ was positive. This normalization accounted for the bidirectionality of the modulation while avoiding bias that would result from normalizing by the sign for just a single condition. After sign normalization, accounting for arousal led to a significant decrease in the average modulation index across all units in both A1 and IC (Wilcoxon signed-rank test between values of $MI_{AP}\ task\ only$ and $MI_{AP}\ task\ unique$; A1: statistic = 1469.0, $p = 1.533\times10^{-10}$; IC: statistic = 785.0, $p = 0.041$; Figure 3.4A). Effectively, accounting for pupil led to a 40% and a 65% average reduction in the magnitude of MI in A1 and IC, respectively (A1: mean signed-normalized $MI_{AP}\ task\ only = 0.139$; mean signed-normalized $MI_{AP}\ task\ unique = 0.084$; IC: mean signed-normalized $MI_{AP}\ task\ only = 0.072$; mean signed-normalized $MI_{AP}\ task\ unique = 0.042$). The correction for pupil-related changes also led to an overall reduction in the number of units showing significant effects of task engagement.

Effects of removing pupil-related activity were not different between central and external regions of IC (rank-sum test, statistic = 0.115, $p = 0.900$). Moreover, there was no difference between groups of units for which the best frequency was near or far (> 2 octaves) from the target tone frequency (rank-sum test, A1: on-BF $n = 50$, off-BF $n = 79$, statistic = 0.121, $p = 0.904$; IC: on-BF $n = 26$, off-BF $n = 40$, statistic = -1.076, $p = 0.282$). Taken together these results suggest that arousal accounted for a significant portion of the change in activity between passive and active blocks in both A1 and IC, effect that was similar across regions of IC and independent on target frequency.

Task-related modulation is primarily positive in A1

Previous studies have reported either positive (Fritz et al. 2003, 2007b; Yin et al. 2014) or negative (Otazu et al. 2009; Slee & David 2015a) changes in spike rate associated with measurements of task-related modulation of auditory neurons' activity. Variability in task-dependent changes in auditory cortex has been explained by differences in task structure (David et al. 2012), behavioral effort (Atiani et al. 2009), and selective attention (Schwartz & David 2018). Given that we used a positive reinforcement structure similar to previous studies (David et al. 2012; Slee & David 2015a), we expected the average sign of the modulation to be about zero in A1 and negative in IC. However, we found that IC units were equally likely to be associated with negative or positive values of MI_{AP} *task only* (32/66 positive, 34/66 negative; median $MI_{AP}\ task\ only$ values = -0.002; Wilcoxon signed-rank test, statistic = 1054.0, $p = 0.742$). In contrast, A1 showed a significantly larger number of units with positive task-related modulation (85/129 positive, 44/129 negative; median $MI_{AP}\ task\ only$ values =

0.048; Wilcoxon signed-rank test, statistic = 2824.0, $p = 0.001$) (Figure 3.4B). Accounting for arousal made the positive modulation less pronounced and no longer significant in A1 (75/129 positive, 44/129 negative; median $MI_{AP}\ task\ unique$ values = 0.012; Wilcoxon signed-rank test, statistic = 3567.0, $p = 0.141$), while tipping it towards more negative values in IC (26/66 positive, 40/66 negative; median $MI_{AP}\ task\ unique$ values = -0.008; Wilcoxon signed-rank test, statistic = 823.0, $p = 0.071$).

Thus, values of task-related modulation were mostly positive in A1, meaning that during active engagement spontaneous and evoked activity tended to be enhanced with respect to passive listening. This positive effect of behavior was only partially reduced by accounting for arousal. Conversely, in IC significant state-modulate units were equally likely to be enhanced or suppressed during active engagement. Accounting for arousal did not significantly change this result.

Task- and arousal-related modulation may act via different neural pathways

In the examples in Figure 3.2, we showed that the effects of task engagement and arousal on firing rates are difficult to dissociate by simply looking at raw activity. The regression analysis helped us disambiguate unique contributions of these state variables to firing rate modulation in several units. Next, we wondered if arousal and task operated via functionally separate or common pathways. While identifying the specific circuits involved in this modulation was outside the scope of this work, we reasoned that if the two modulations were carried through the same pathway, the activity of one unit modulated by arousal would also be modulated by task. To test this prediction, we measured the correlation between $MI_{AP}\ task\ unique$ (see above) and

$MI_{LS}\ pupil\ unique$ (Figure 3.4), which quantified the unique contribution of pupil after regressing out task engagement effects (*Eqn. 3.8*, see *Materials and Methods*). We found no significant correlation between these quantities in either area, (A1: Pearson's r = -0.148, p = 0.100; IC: Pearson's r = -0.083, p = 0.510). The absence of a correlation suggests that separate neural circuits mediate the task and arousal state modulation in A1 and IC.

Relationship between state modulation and auditory responsiveness

While all units were recorded from areas functionally and anatomically characterized as auditory, it is well-known that within the same sensory areas and circuits different subpopulations of neurons serve distinct functions. Within A1 and IC, neurons vary substantially in the degree of auditory responsiveness (Atiani et al. 2014; Gruters & Groh 2012). We asked whether the magnitude of the state-dependent modulation in each recorded unit was related to its auditory responsiveness. The variance explained by the null model was used as a proxy for auditory responsiveness, since it described how accurately the PSTH response to the noise stimuli predicted activity on a single trial. The null model fit the average activity during the reference portion of the stimulus across the whole length of the recording. Therefore, small values of null model R^2 were associated with units whose activity during the reference period was not reliable across the experiment. In IC but not in A1, units that responded less reliably to sound were also those whose modulation was explained best by changes in state (A1: Pearson's r = -0.084, p = 0.343; IC: Pearson's r = -0.344, p = 0.005; Figure 3.6). These results also held true when the signal-to-noise ratio of the spectro-temporal receptive field (STRF),

measured during passive listening, was instead used as a measure of auditory responsiveness (A1: Pearson's r = -0.093, p = 0.297; IC: Pearson's r = -0.260, p = 0.035; data not shown) (Klein et al. 2000). Thus, while state-dependent modulation in A1 did not depend on units' sensory responsiveness, state-dependent effects in IC were more common in units with a weaker sensory component.

Are "persistent" state effects explained by arousal?

Previous studies in both A1 (Fritz et al. 2003) and even more so in IC (Slee & David 2015a) have reported examples of task-related modulation persisting in passive blocks following active engagement (referred to as "post-passive" blocks). These effects have been interpreted as persistent task-related plasticity, but they were highly variable and difficult to attribute to any specific feature of the behavior. The examples in Figure 3.2 suggest that some persistent changes following task engagement may be explained by fluctuations in arousal. After the first active block, only for example 1 firing rates returned to values comparable to the first passive (Figure 3.2A, B). For the example unit 2, firing rates continued to decrease as the experiment progressed (Figure 3.2 E, F). For this example, we showed that the modulation was tracked and largely accounted for by changes in pupil size rather than being a consequence of the animal switching behavioral context (Figure 3.2F). However, if we were not measuring pupil, we may have also concluded that activity was suppressed due to persistent effects of task engagement.

To measure the presence of persistent task-related modulation – and the extent to which arousal accounts for it – we compared the activity between the first and second passive blocks. We performed the same regression analysis as above, but only considered

data from passive blocks before (P1) and after (P2) a behavioral block, treating the passive block (P1 or P2) as a state variable. Modulation index was computed either for responses predicted by a partial model, based only on passive block position ($MI_{P1P2}\ block\ only$), or for the unique contribution of passive block position. The latter, unique component ($MI_{P1P2}\ block\ unique$) was computed from the difference in MI between a full model and the partial model accounting only for pupil-related effects (Figure 3.7A). To assess the effect of pupil on the magnitude of MI between passive conditions, we used the same strategy as for the active/passive analysis above. We normalized the sign of MI for each unit so that the mean of $MI_{P1P2}\ block\ only$ and $MI_{P1P2}\ block\ unique$ was positive. In both areas, the magnitude of firing rate modulation between passive blocks was significantly reduced after accounting for effects of pupil (A1: statistic = 1504.0, $p = 1.673\times10^{-5}$; IC: statistic = 309.0, $p = 1.376\times10^{-5}$; Wilcoxon signed-rank test). Effectively, accounting for arousal led to a 20% and a 48% reduction in activity modulation in A1 and PEG, respectively (A1: mean signed-normalized $MI_{P1P2}\ block\ only$ = 0.140; mean signed-normalized $MI_{P1P2}\ block\ unique$ = 0.100; IC: mean signed-normalized $MI_{P1P2}\ block\ only$ = 0.100; mean signed-normalized $MI_{P1P2}\ block\ unique$ = 0.052).

To evaluate the time course of arousal-dependent modulation of firing rate across the length of the experiment, we considered a subset of A1 and IC units for which recordings were stable across five consecutive passive/active behavioral blocks (45 units in A1; 31 units in IC, Fig. 3.5B). The sign of MI for each unit was normalized so that responses were larger during active conditions, and the average modulation index across the three passive blocks was 0. In both areas, after accounting for the contribution of

arousal, the unique contribution of task to firing rate modulation was reduced (Figure 3.7B). In IC, the average persistent change in firing rate was completely accounted for by arousal. Taken together, these results suggest that some post-behavior effects previously reported as persistent plasticity induced by behavior can in fact be explained by fluctuations in the state of arousal.

Discussion

The primary goal of this study was to explore how arousal contributed to task-related changes in neural activity in the primary auditory cortex (A1) and the auditory midbrain (IC). Several previous studies have shown that transitioning from passive listening to active behavior leads to sizable changes in neural activity in A1 (Fritz et al. 2003; Niwa et al. 2012; Otazu et al. 2009) and IC (Ryan & Miller 1977; Slee & David 2015a). These changes can be specific to properties of the task stimuli (Fritz et al. 2003; Jaramillo & Zador 2011; Lee & Middlebrooks 2011) or depend on the structural elements of the task, such as reward contingencies (David et al. 2012), task difficulty (Atiani et al. 2009) or selective attention (Downer et al. 2017; Hocherman et al. 1976; Schwartz & David 2018). While it is clear that sensory and behavioral context shapes auditory representation, a coherent theory of how multiple state variables interact to influence sensory coding has yet to be formed.

Changes in firing rates associated with different behavioral contexts are likely to reflect multiple aspects of internal state (Cohen & Maunsell 2009; McGinley et al. 2015b; Niell & Stryker 2010; Polack et al. 2013). Factors like arousal, locomotion, and whisking have been shown to change smoothly in non-anesthetized animals over the course of an

experiment, tracking fluctuations of sensory neurons' membrane potentials and behavioral performance (Cohen & Maunsell 2010; Goris et al. 2014; McGinley et al. 2015a; Niell & Stryker 2010; Poulet & Petersen 2008; Reimer et al. 2014; Vinck et al. 2015).

One major obstacle to understanding the relationship between state variables and firing rate is that they often covary. To overcome this obstacle and dissociate the effects of arousal and task engagement on cortical and midbrain activity, we used a step-wise linear regression approach in which arousal, indexed by pupil size, and task were counted as independent regressors. In nearly half of recorded A1 and IC units, arousal and task engagement contributed substantially to the modulation of spontaneous and evoked activity between blocks of active engagement and passive listening. Furthermore, our data suggest that arousal and other effects of task engagement operate via distinct feedback circuits. Thus, task engagement and arousal are effectively two state variables that can be dissociated in most state-modulated neurons.

The main consequence of our findings is that changes in firing rates previously attributed to task engagement alone, might have been the results of fluctuations in internal state, specifically in arousal, as indexed by pupil size. However, arousal did not account for all the activity modulation between active and passive blocks, suggesting that other factors might play a role in shaping sound processing. Our approach provides a general method for dissociating the influence of continuous internal state variables on sensory representation and their relationship with task-related discrete variables often imposed by the experimenter. The same method can be adopted to explore the contribution of other state variables (*e.g.*, task difficulty, behavioral performance, reward

size, and selective attention) measured in normal conditions as well as in perturbation experiments to help identify the circuits through which such variables operate.

Comparison with previous studies

Previous studies that used similar paradigms to the current experiment to measure the effects of behavior on auditory representation, described local changes in frequency tuning measured in the spectro-temporal receptive fields (STRF) (David et al. 2012; Fritz et al. 2003; Slee & David 2015a). Fritz *et al*., showed that in A1 neural activity near the frequency of the target tone in response to reference stimuli was enhanced when ferrets were engaged in an avoidance auditory task with respect to passive listening (Fritz et al. 2003). The sign of the local modulation of the evoked activity was later found to be contingent to the structure of the task (David et al. 2012). Conversely to the avoidance task, positive reinforcement version of the task (similar to the one used in the current study) led to a suppression of evoked responses instead of enhancement at the target frequency if placed near the best frequency of the unit. Similarly, in the IC, local changes in the STRF were also found to be suppressive during behavior (Slee & David 2015a). However, while for A1 the effects of engaging in a task on overall excitability were equally likely to be suppressive or enhancing, in IC global, untuned changes were predominantly suppressive (Slee & David 2015a). In fact, in several IC units, the overall modulation could be described by global suppression of firing rates across frequencies rather than a shift in turning (David 2018; Slee & David 2015a).

Here, we found a different pattern of modulation. Activity in A1 tended to be more enhanced during active behavior compared to passive blocks, even after accounting

for arousal. In IC activity was equally likely to be enhanced or suppressed. This discrepancy with previous findings could reflect differences in task stimuli and/or training. The previous studies used exclusively pure-tone targets, while this study used tones embedded in a reference noise. Reward associations can also impact sensory responses (Baruni et al. 2015). The fact that the noise stimuli were paired with positive reward could have impacted changes imposed on responses to the noise when it was presented as a reference sound. Moreover, the size of the reward itself has been shown to influence firing rate in response to sensory stimulation (Metzger et al. 2006). Further studies in which reward associations are targeted systematically and accounted for in the encoding model may clarify the differences observed across these studies.

In the current study, significant state-dependent A1 units were equally likely to be modulated by task, by arousal, or by both. In IC, however, we found that a larger portion of state-dependent changes could be explained by arousal. Furthermore, in IC but not in A1, units with less reliable auditory responsiveness were those with a stronger state-modulation component. This result is perhaps not surprising given that the majority of IC units in our sample were recorded from regions of the IC marked as non-central (NCIC). These are dorsal and lateral regions of IC that receive input from brain neuromodulatory circuits such as the pedunculopontine and latero-dorsal tegmental nuclei (Motts & Schofield 2009), as well as multisensory information from somatosensory, visual, and oculomotor centers (Gruters & Groh 2012). When we compared the magnitude of the state-modulation between the two regions, we did not find any statistically significant difference. This result is in line with previous findings by Slee and David (2015), who

also found no difference in the local and global task-related modulation of evoked responses. Further experiments, however, which include a larger sample of units across the different areas would be needed to confirm this finding in the context of arousal-related modulation.

Possible circuits mediating task and arousal-related modulation in A1 and IC

Probing the neural circuits underling task and arousal-related modulation was outside the scope of this study. However, by measuring the relationship between the unique contributions of each state variable, we were able to conclude that the effects of task and arousal are dissociated and may be carried via different pathways.

Several previous studies have shown the role of neuromodulators in inducing short-term changes in activity and sensory tuning in auditory cortical regions (Kuchibhotla et al. 2017; Reimer et al. 2016) and midbrain (Gittelman et al. 2013; Habbicht & Vater 1996; Hurley & Pollak 2005). Cholinergic fibers projecting to the auditory cortex from the nucleus basalis play a key role in rapid switching between passive listening and active engagement by modulating the activity of different populations of cortical inhibitory interneurons (Kuchibhotla et al. 2017). Furthermore, the activity of cholinergic and noradrenergic terminals in auditory cortex was found to be elevated during pupil dilation and reduced during pupil constriction (Reimer et al. 2016). Given that we find that modulation of task and pupil often covary, it is reasonable to hypothesize that cholinergic and noradrenergic pathways enable these aspects of cognition.

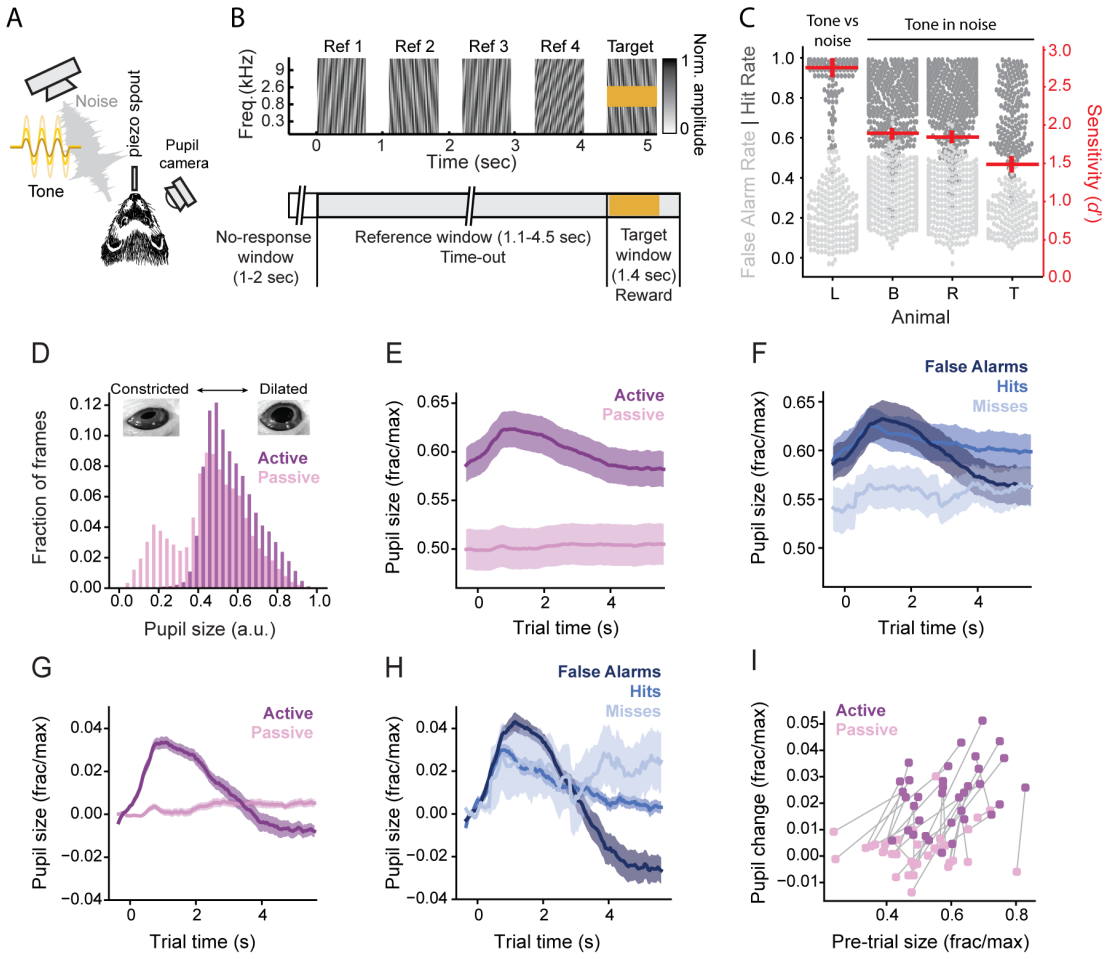

Figure 3.1: Pupil size correlates with task engagement.

A. Schematic of head-fixed behavioral setup, including a free-field speaker, a piezo spout to register licks and deliver the reward, and an infrared video camera for pupillometry**. B.** Schematic of the task structure. False alarms before the target window were punished with a time out and hits were rewarded with water. **C.** Swarm plot showing behavioral performance (hit rate or false alarm rate) for each animal. Each plot is a behavioral session. Red lines indicate mean and standard error of the mean (s.e.m.) of the sensitivity (*d*') for each animal. All animals performed above chance (*d*' = 0). **D.** Histogram of pupil size during passive and active behavior averaged across all animals and behavior sessions. **E.** Normalized mean pupil size across all trials and animals. Shading indicates s.e.m. **F.** Average pupil size during active behavior, grouped according to performance. Axis as in E. **G.** Trial time-course of average pupil size normalized to pre-trial size (averaged across 0.35-sec pre-stimulus window) for active and passive behavioral blocks.

H. Trial time-course of average pupil size as in G but during active behavioral blocks and grouped according to performance. **I.** Pre-trial pupil size plotted against pupil change per trial (averaged across 3-sec window during stimulus presentation). Values were normalized to peak pupil size in each session. Each dot represents a behavioral block, and passive and active blocks within the experiment are connected by a line.

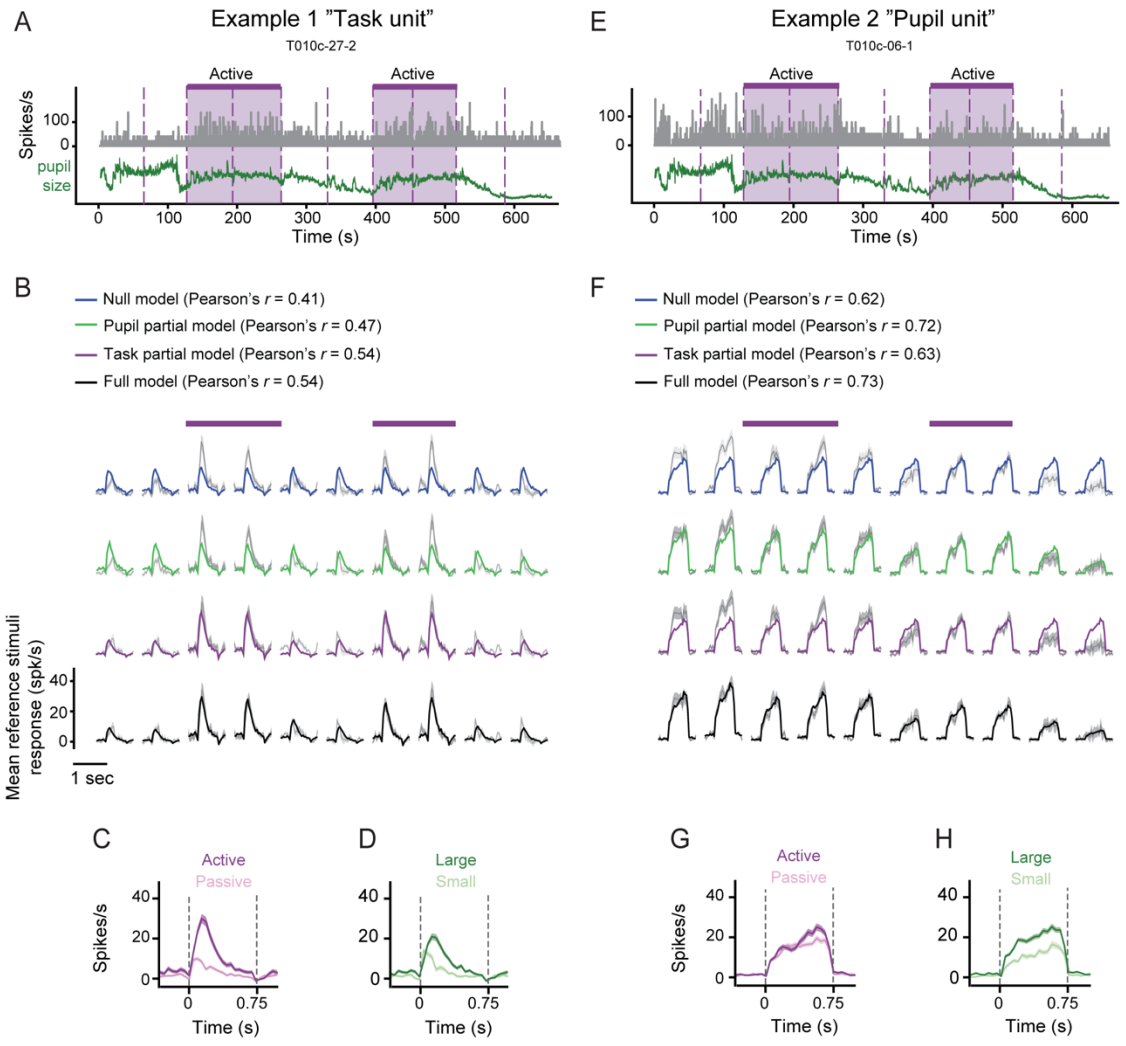

Figure 3.2: Example A1 units recorded simultaneously

A. Firing rate (grey) during the reference period and time-matched pupil size (green) are plotted for the duration of one example experiment (total time of the recording ~1h, but intra-trial time and target responses are not shown). Purple shading highlights active blocks. Dashed lines delineate halved recording behavioral blocks, times that were used to compute partial behavioral blocks' PSTHs in B. **B.** Partial PSTHs (grey) generated by averaging responses across trials in each halved behavioral block as shown in A. Shading represents standard error of the mean (s.e.m.). Solid colored lines are the responses predicted by the null (blue), pupil partial (green), task partial (purple), or full (black) models. Prediction accuracy is quantified by the correlation coefficient (Pearson's *r*). Solid purple lines above PSTHs indicate active blocks as in A. **C.** PSTHs of example 1

unit's activity averaged across all passive and active blocks. **D.** PSTHs of example 1 unit's activity when pupil size was larger or smaller than its median across the experiment. **E-H.** Same as in A-D for example Neuron 2. In C-E and H-J, dashed grey lines indicate beginning and end of reference noise stimulus.

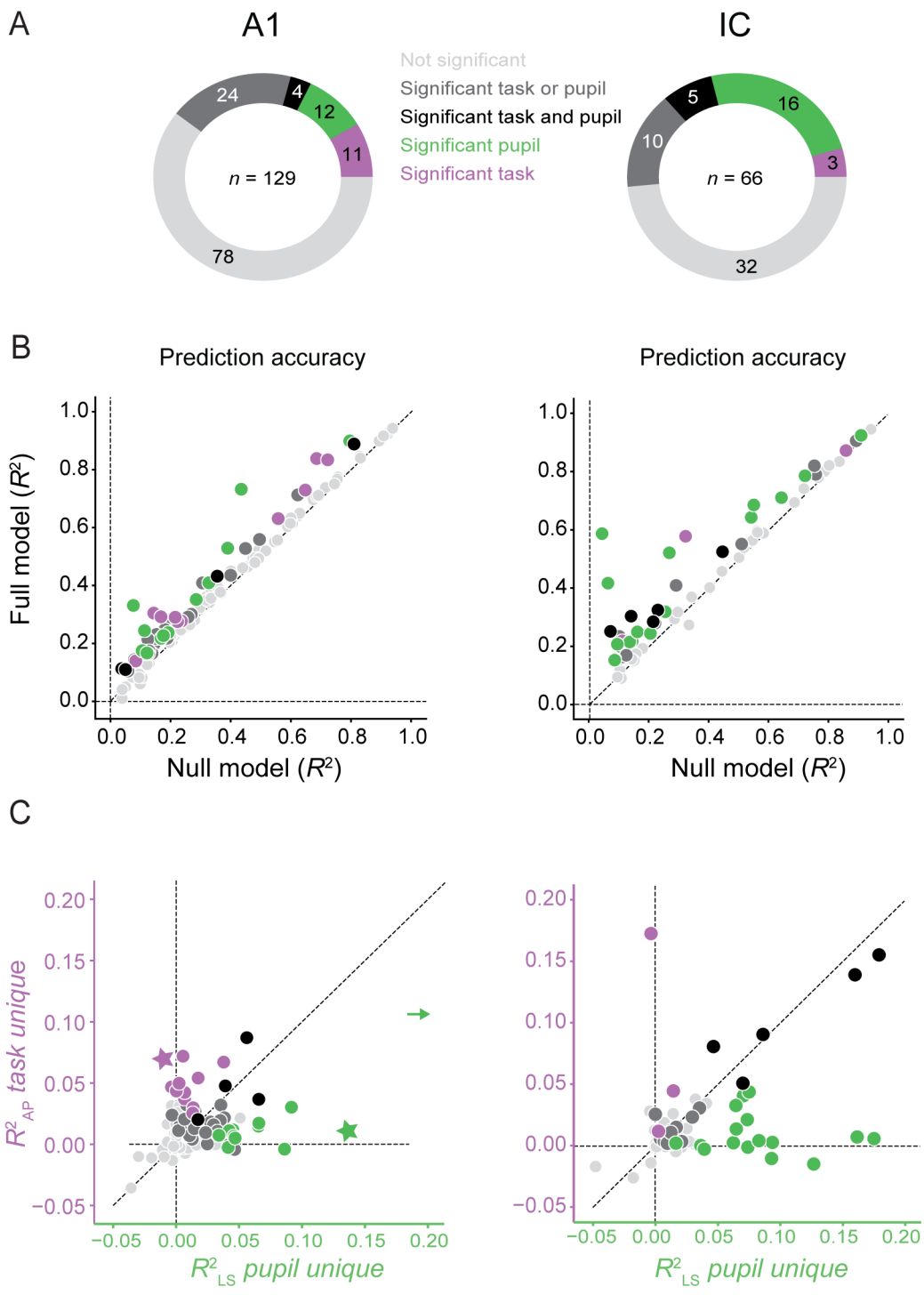
A
A1
IC
Not significant
Significant task or pupil
Significant task and pupil
Significant pupil
Significant task
24
4
12
11
n = 129
78
5
16
10
3
n = 66
32
B
Prediction accuracy
Prediction accuracy
Full model (R²)
Null model (R²)
Null model (R²)
C
R² AP task unique
R² LS pupil unique
R² LS pupil unique

Figure 3.3: Task and pupil-related modulation of firing rates in A1 and IC

A. Number of units significantly modulated by task, arousal, or both in A1 (left) and IC (right). The total number of recorded units is reported in the center of the doughnut plots. Significance was determined by cross-validated jackknifed *t*-test, at a cut-off $p < 0.05$. Purple and green symbols correspond to units with significant unique modulation of task or pupil, respectively. Black symbols correspond to units for which both unique task and pupil contributions are significant. Dark grey symbols correspond to units for which the contribution of task and pupil to model prediction is significant only when both variables are present in the model. Light gray indicates units for which there is no significant improvement in prediction accuracy between full and null models. **B.** Variance explained (R^2) in single-trial activity by full model versus null model. Each symbol represents a unit in AC (left) or IC (right; circles for NCIC and triangles for ICC units). Colors as in A. **C.** Active/passive unique variance explained by pupil size ($R^2_{AP}\ pupil\ unique$, x-axis) plotted against unique variance explained by task engagement ($R^2_{AP}\ task\ unique$, y-axis). Star symbols correspond to examples in Figure 3.2A-D (purple) and E-H (green). Arrow indicates unit with significant $R^2_{AP}\ task\ unique$ value of 0.23.

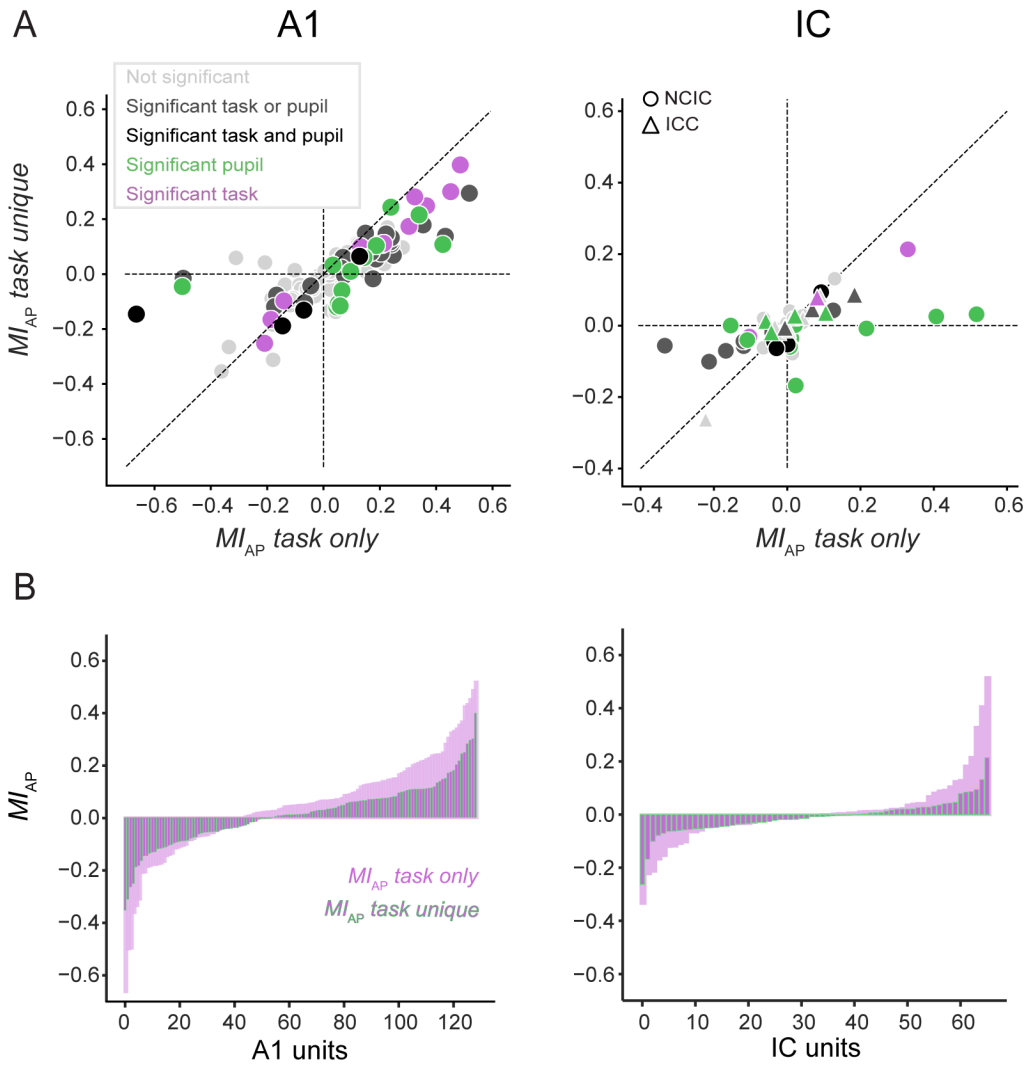

Figure 3.4: Arousal accounts for some firing rate modulation between active and passive behavior

A. Modulation index computed with responses predicted by the task partial model in which pupil is shuffled (MI_{AP} *task only,* x axis), plotted against task unique modulation in which pupil size modulation is regressed out (MI_{AP} *task unique*, y axis) for A1 (left) and IC (right; circles for NCIC and triangles for ICC units). Purple and green symbols correspond to units with significant unique contributions of task or pupil, respectively. Black symbols correspond to units for which both unique task and pupil contributions are significant independently. Dark grey symbols correspond to units for which the

contribution of task and pupil to model prediction is significant only when both of them are present in the model. Light gray indicates units for which there is no significant improvement in prediction accuracy between full state-dependent and state-independent models. Significance tested by cross-validated jackknifed *t*-test, $p < 0.05$. To quantify the effects of accounting for pupil in our model, we run a Wilcoxon signed-rank test on signed-normalized values of *MI*. A1: $p = 2.816 \times 10^{-11}$; IC $p = 0.003$. **B.** $MI_{AP}\ task\ unique$ and $MI_{AP}\ task\ only$ sorted according to their magnitude for each unit in A1 and IC. Accounting for pupil-indexed arousal reduces activity modulation between active and passive blocks by 40% and 65% in A1 and IC, respectively.

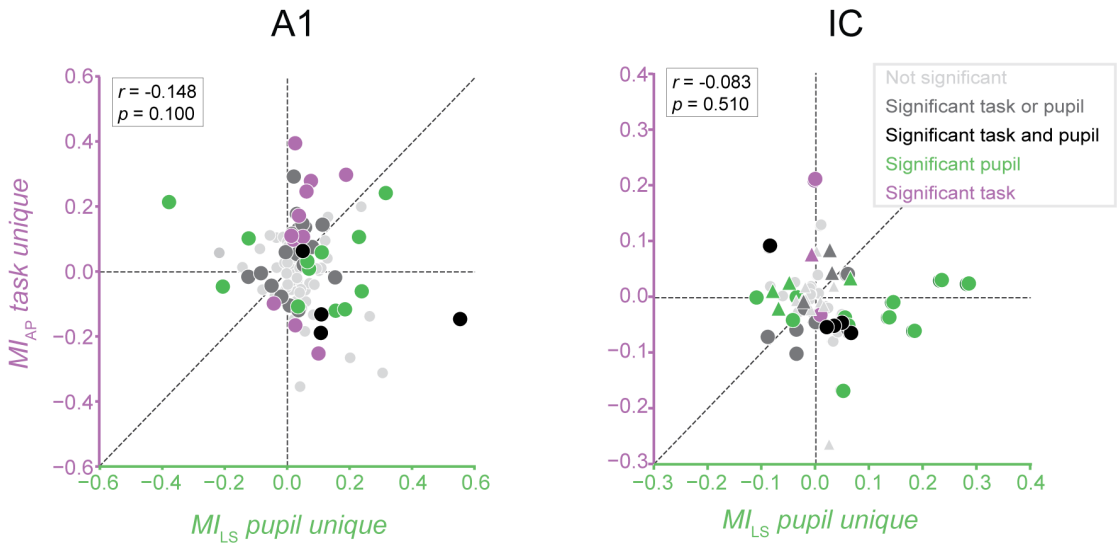

Figure 3.5: Unique task and arousal modulations are uncorrelated

Unique contribution of task (MI_{AP} $task\ unique$) plotted against unique contribution of pupil size (MI_{LS} $pupil\ unique$) to activity modulation for A1 (left) and IC (right). The two quantities were not significantly correlated as quantified by cross-correlation. A1: Pearson's r = -0.148, p = 0.100; IC Pearson's r = -0.083, p = 0.510. Colors as in Figure 3.4.

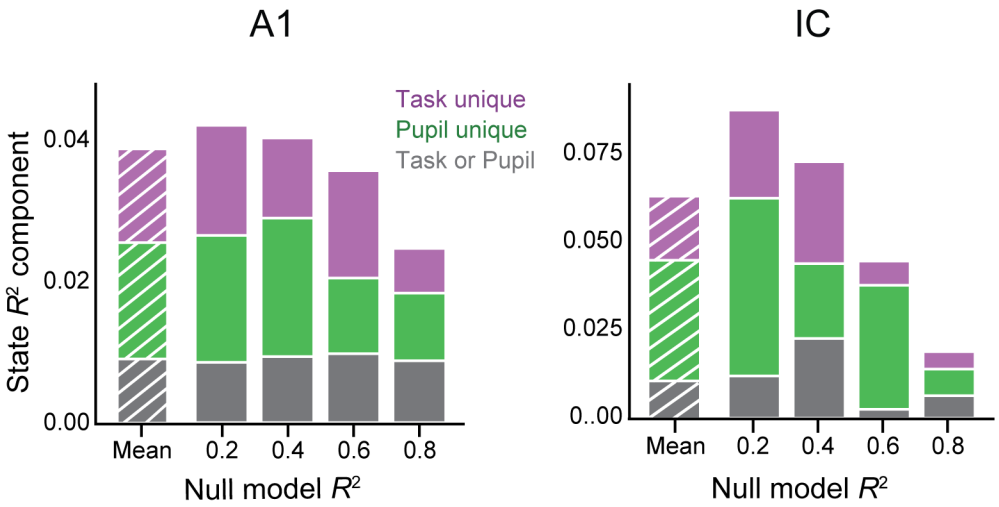

Figure 3.6: Task and arousal-related modulation in A1 and IC

Mean variance explained (striped bar) for significant units with unique contribution of task (purple), pupil (green), or ambiguous (grey) for A1 (left) and IC (right). Solid bars represent unique contribution of each state variable as a function of binned null model R^2, a measure of auditory responsiveness. Colors as in Figure 3.4.

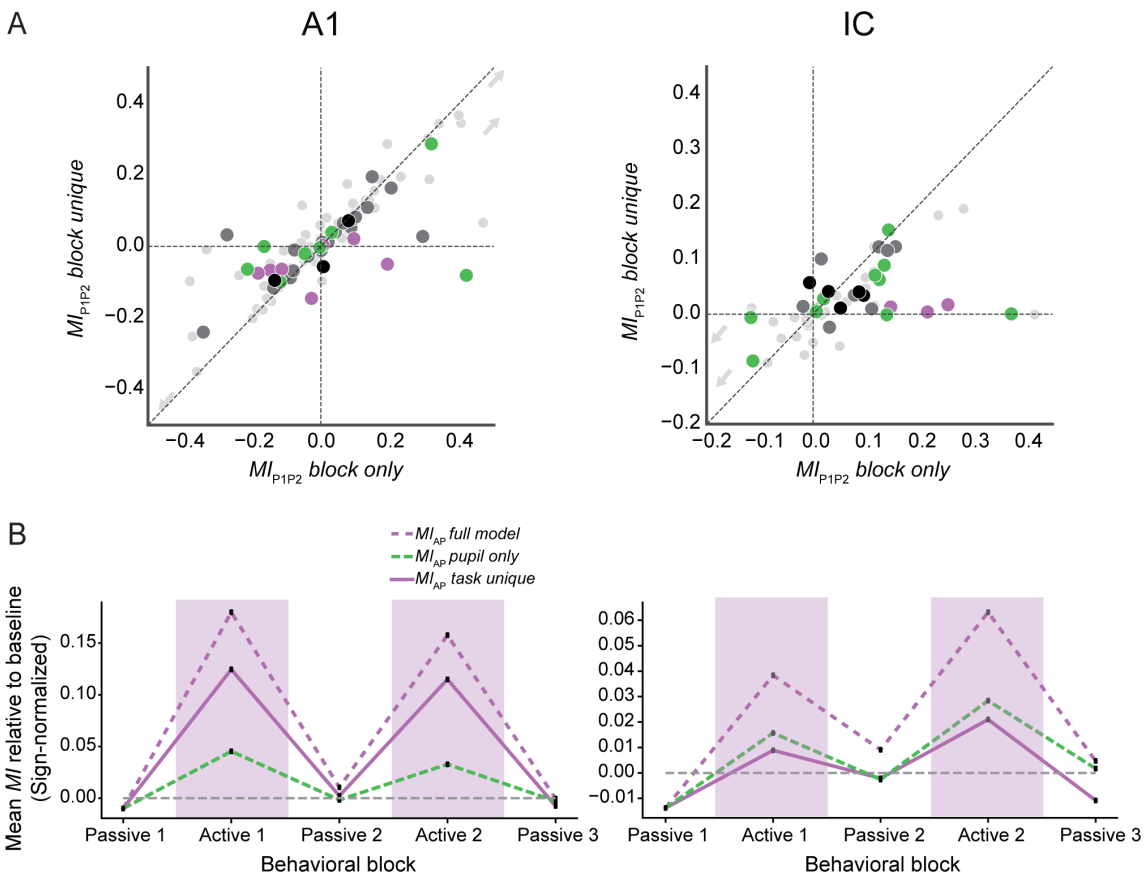

Figure 3.7: Persistence of task-related modulation

A. Modulation index computed with responses predicted by the passive block partial model in which pupil is shuffled (MI_{P1P2} *block only,* x axis), plotted against passive block unique modulation in which pupil size modulation is regressed out (MI_{P1P2} *block unique*, y axis) for A1 (left) and IC. Colors as in Figure 3.4. **B.** Mean *MI* of each behavioral block normalized to the mean value for the three passive conditions plotted for a subset of units in A1 (n = 45; left) and IC (n = 31; right) for which recordings were stable across five consecutive behavioral blocks. Mean *MI* is sign-normalized to have mean active/passive or large/small response > 0 across blocks to account for activity enhancement versus suppression. 32/45 units in A1 and 25/31 units in IC had significant state modulation (jackknifed *t*-test).

Chapter 4: Conclusions

It is well-established that neurons tasked with processing sensory information are not mere analyzers of the individual physical components that make up a stimulus. Hearing, is an active process that requires integration between physical sound properties and information about the context in which the sound is presented. This thesis explored two aspects of the role of sensory and behavioral context in modulating sound representation.

We used the ferret as animal model and recorded extracellular single- and multi-unit activity across three regions of the auditory brain: the inferior colliculus (IC), primary auditory cortex (A1), and secondary auditory cortex (PEG). In the first study (Chapter 2), we investigated the neural signature of streaming sequences of complex sound based on repetition in A1 and PEG. Our results shed new light into the possible mechanisms through which the brain uses repetition as a streaming cue. The second study (Chapter 3) adds new evidence to the current understanding of how behavior shapes sound representation by modulating neural activity in IC and A1. Our results indicate that much of the activity modulation classically attributed to task engagement was in fact uniquely explained by changes in the internal state of the animal linked to arousal.

Neural signature of streaming repeated-embedded noise in A1 and PEG: limitations of the study and future directions

Frequency separation and presentation rate are key perceptual cues for segregating alternating sequences of pure tones (Bregman 1978b; Bregman et al. 2000; Oberfeld 2014; van Noorden 1975), and lead to spatially segregated neural responses in

the cochlea (Cusack & Roberts 2000) and in the auditory cortex (Bee & Klump 2004; Bidet-Caulet & Bertrand 2009; Fishman et al. 2001; Micheyl et al. 2007a). In a more naturalistic scenario, however, the challenge is to separate complex sounds that occur simultaneously (*e.g*., in a cocktail party (McDermott 2009)). In this scenario, listeners need to direct their attention to a chosen target (*e.g.*, a friend's voice) while selectively ignoring multiple sounds in the background (*e.g*., music playing), often in conditions when these distractors are acoustically more salient (Bee & Micheyl 2008; Cherry 1953). Psychophysical and physiological studies conducted in humans that used simultaneous, speech multi-talker paradigms, demonstrated that even if the frequency spectra overlap, one stream of speech can be perceptually segregated from another as long as they differ on any other perceptual dimension such as pitch, timbre, or spatial location (Carlyon 2004; Ding & Simon 2012; Mesgarani & Chang 2012; Micheyl et al. 2007a).

Another cue for stream formation is temporal patterning in the ongoing progress of overlapping sound sequences (Bendixen et al. 2010; Szalárdy et al. 2014). Sounds emanating from the same source likely share statistical regularities, and often repeat. McDermott *et al*. (2011) used a set of artificially-generated stimuli that lacked the major grouping cues for auditory streaming but retained the spectro-temporal structure of natural sounds to show that simple repetition is sufficient to recover sound sources from mixtures in humans (McDermott et al. 2011). Listeners were asked to identify a probed sound from a mixture of sequences of overlapping sounds spectro-temporally diverse. Only if the probed sound was repeated in the sequence in the presence of a different "distractor" sound, participants were able to successfully report the occurrence of that sound in the mixture. This finding suggested that sound sources can emerge as

foreground and be perceptually separated if they occur repeatedly, and if are mixed within background sounds.

Where does this phenomenon emerge in the brain? What are the neuronal correlates of the perceptual "pop-out" of sound repetition? In Chapter 2, we examined the neural representation of sound sequences similar to those used by McDermott *et al*. (2011). To our knowledge, our study is the first attempt to understand the underlying neural signature of repetition-based streaming of complex, temporally and spectrally-overlapping sound sequences.

Behavioral paradigm and perceptual streaming analysis

The behavioral paradigm we used was adapted from McDermott *et al*. (2011). We used the same stimuli, artificially generated sounds with similar spectro-temporal properties of natural sound. The stimuli had high point-to-point correlation in frequency content and time, but lacked second order statistics that could function as known cues for streaming (*e.g.*, harmonic structures (Moore et al. 1986) and common onset (Darwin & Carlyon 1995)). Ferrets had to report the presence of the repetition of one of twenty noise samples by licking a water spout. A potential pitfall of this behavioral paradigm, is that it did not directly test for the perceptual ability of the animals to stream sounds based on repetition. However, it does indicate that ferrets were able to detect the repetition embedded in a mixture.

Training ferrets to report the identity of a repeating stimulus would likely be too difficult. Therefore, to better understand the processes underlying the perceptual categorization of repeating sounds, one could perform a classification image (CIm)

analysis using our behavioral data. A CIm is a correlation technique first developed for the analysis of visual stimuli (Ahumada & Lovell 1971; Murray 2011), and recently adapted to study the identification of functional fine acoustic cues in speech perception (Varnet et al. 2013). This method applied to our behavioral data would allow us to directly estimate the portions of the sound sequence that modulate the behavior of the ferret perceptual system. Therefore, the results of this analysis could confirm whether ferrets are indeed a good model for streaming of repeating sounds by showing that their behavior conforms to the theoretical ideal strategy for detecting the repeated noise. In addition, CIm analysis could potentially explain the high false alarm rate we measured in both animals. The high false alarm rate could be related to the fact that random sequential samples in the reference phase may contain repeated spectro-temporal patterns by chance, inducing animals to perceive a target. If this were the case, false alarm responses would be more likely to occur after two different samples occurring in sequence with a highly correlated spectrogram. In summary, CIm analyses could inform about strategies used by ferrets in to perform the task, increasing our confidence in concluding that the ferret is indeed a good anima model for streaming repeated sounds. Furthermore, it could suggest modifications of the task to improve behavioral performance.

Streaming analysis

By analyzing the raw spike rate evoked by the presentation of the continuous and overlapping sequences of noise samples, we showed that responses to a given target noise were more suppressed when the noise sample was presented in the repeating phase compared to the random phase. This result was not surprising in light of what is known

about the wide-spread phenomenon of stimulus adaptation in the auditory brain (Pérez-González & Malmierca 2014). However, to test the hypothesis that the two streams – one with a repeating sound (referred to as the foreground stream) and another with a non-repeating mixture of random samples (referred to as the background stream) – were represented differently in the auditory cortex, we had to model the responses to the two streams independently. We constructed a Bayesian model with a normal prior and fit a stream-specific gain to each predicted response. By doing so, we were able to show that, even though in most units both responses to each stream were suppressed by the repetition, the response to the foreground stream was less suppressed compared to the background stream. This effect was stronger in PEG than A1 and for units that were tuned to the target. We thus concluded that this selective foreground enhancement is the signature of streaming that emerges A1 and is refined in PEG.

The challenge of identifying a signature of streaming of simultaneously occurring sounds has been previously addressed in humans with a similar approach (Ding & Simon 2012). These authors asked human subjects to listen to one of two competing speakers, either of different or the same sex, and recorded brain activity via magnetoencephalography (MEG). To investigate the timing and spatial information of the neural encoding process, they fit a separate spectro-temporal receptive field (STRF, or more precisely a "TRF", given that MEG data could not resolve spectral tuning) model for each of the two simultaneously presented speech streams. Neural activity selectively synchronized with the speech of the single speaker to which the attention of the listener was directed. Furthermore, the latency and source location of the two components

suggested a hierarchy of auditory processing in which the representation of the attended object emerges from core to posterior auditory cortex (Ding & Simon 2012).

Another approach used to investigate the neural signature of streaming is given by a decoding analysis with stimulus reconstruction (Ding & Simon 2012; Mesgarani & Chang 2012; Mesgarani et al. 2009). A decoding model describes the same relationship between stimulus and response that an encoding model such as the STRF describes, just backwards. It typically uses the neural output to reconstruct the input sound stimulus. If the reconstruction of the envelope has a higher correlation to the envelop of the attended stream rather than the combined stream making up the physical stimulus, it suggests that the stimulus mixture is neutrally segregated. In their human study using MEG, Ding and Simon (2012) found that this was indeed the case. Similar results were also obtained by Mesgarani and Chang (2012) using data collected via electrocorticography (ECoG) (Mesgarani & Chang 2012).

In our study, a decoding analysis could complement the encoding approach we used, potentially revealing how the relative enhancement of the repeating stream allows for the separation of the two streams. Specifically, I predict that for units with positive foreground enhancement the stimulus reconstruction would be more successful for the foreground stream compared to the background stream, matching perception. However, in order to avoid bias towards the foreground stream, an experimental design in which the background stimulus is always composed by the same set of stimuli would be necessary.

Role of attention in streaming of repeated-embedded noise

The human studies mentioned above showed that directing listeners to attend to one of two streams presented simultaneously cued subjects to listen to one or the other stream of speech (Ding & Simon 2012; Mesgarani & Chang 2012). These studies argued that streaming a complex sound such as streams of speech requires – or at least is facilitated by – directed attention. However, in a follow-up study to McDermott *et al*. (2011), Masutomi and colleagues found that recovering the identity of noise stimuli using repetition as the only cue was robust to inattention (Masutomi et al. 2015). The authors were able to show that human listeners continued to perform the task at criterion even when distracted by a visual decoy task.

In the current study, we did not record neural activity during active engagement. We only recorded neural activity from trained or task naïve ferrets passively listening to the task stimuli. However, it is possible that active task engagement as well as changes in arousal could alter the representation of the two streams, such that the foreground stream is even more enhanced relative to the background stream. To test this hypothesis, future studies will have to include both passive and active presentation of the same stimuli, as well as the recording of pupil fluctuations to track arousal.

Task and arousal-related modulation of sound representation in A1 and IC: limitations of the study and future directions

Task engagement and arousal are two state variables that have been independently associated with changes in sensory representation at the level of single and multiunit activity (Fritz et al. 2003; McGinley et al. 2015a; Niwa et al. 2012; Otazu et al. 2009;

Reimer et al. 2016; Vinck et al. 2015). Previous studies using behavioral paradigms similar to the current study, have reported changes in excitability and frequency tuning in A1 and IC units when responses to reference stimuli were compared between active and passive blocks (Atiani et al. 2009; David et al. 2012; Fritz et al. 2003; Slee & David 2015a). These changes were attributed to the rapid adaptation of auditory neurons to the current behavioral demand, in order to optimize the sensory representation of relevant task stimuli. For some units these changes were reported to persist for several minutes after the active behavior, while for others, activity and tuning changes were observed to rapidly regress back to baseline (passive condition) (Fritz et al. 2003; Slee & David 2015a). One interpretation for such variability might be that the source of modulation is a non-auditory input carrying information about another changing state variable – related or unrelated to task engagement, such as arousal.

Given that we observed a distinct variability in pupil size associated with the ferret's engagement with the task, we wondered if this continuous variable, often associated with changes in arousal (McGinley et al. 2015a), could explain ongoing changes in auditory neurons' activity that are not explained by task engagement alone. To separate the relative contributions of task engagement and arousal indexed by pupil diameter, we used a step-wise approach to fit a linear model with task (active/passive) and arousal (pupil size) as discrete and continuous variables, respectively. In nearly half of the recorded units in A1 and IC, activity was modulated by task, arousal, or both. Surprisingly, in IC the majority of the modulation of firing rate between active and passive blocks was explained by changes in pupil-indexed arousal, and was more pronounced in units with weaker auditory responses. In light of these results, we

conclude that previous findings that assigned global changes in evoked activity to task engagement alone may in fact be due to changes in arousal.

Behavior and pupil recordings

Similar to previous studies (Joshi et al. 2016; McGinley et al. 2015a), we found that on average pupil size was significantly larger during active engagement with respect to passive listening to the task stimuli. During active trials pupil size increased steadily prior to the target and the final choice with concomitant motor response. De Gee *et al*. showed that pupil dilation during decision formation was bigger before 'yes' than 'no' choices, and the magnitude of this effect was reflected in the individual criterion such that it was strongest in conservative subjects, those who chose yes against their bias (de Gee et al. 2014).

We found that pupil size increased during trials that resulted in hits or false alarm responses compared to passive listening and miss choices. A more detailed analysis in which every TORC stimulus is considered as a yes/no choice in relationship with pupil size is needed to better understand how pupil dilation relates to trial-by-trial decision making. For example, to investigate the extent to which pre-trial pupil baseline and time-course of pupil change could predict upcoming behavior, we could train a simple linear decoder and measure its accuracy in correctly classifying trial decision's outcome (Schriver et al. 2018).

Tone-in-noise and tone-versus-noise behavior

The behavioral paradigm used in this study was similar to the one used in other studies conducted in our laboratory (David et al. 2012; Slee & David 2015a). However, to test for effects of task engagement in conditions closer to psychophysical thresholds, we embedded the target tone in one of 30 reference TORC stimuli in our collection (Atiani et al. 2009). Overall block difficulty was controlled by changing the signal-to-noise ratio (SNR) of the target/TORC sound intensity. Three ferrets were trained to the tone-in-noise variation of the task as well as to the original tone-versus-noise variant (David et al. 2012; Slee & David 2015a). One ferret was trained only to the tone-versus-noise variant of the task. Not surprisingly, animal performance was lower for animals trained to the tone-in-noise task compared to performance for the animal trained to tone-versus-noise alone. In order to further understand the relationship between performance and task difficulty, a more in-depth analysis across subjects' performance in each SNR category is needed.

One study conducted in ferret auditory cortex (AC) showed that task difficulty modulated neural activity (Atiani et al. 2009). Stronger task-related modulations correlated with greater difficulty and better behavioral performance. Specifically, the authors found an overall decrease in gain under active performance accompanied by a selective enhancement at the target frequency when it matched the best frequency of the unit. These results suggest a model in which the brain adapts to a noisy background in order to increase the sensitivity to the target tone enhanced as a foreground. Furthermore, Atiani and collaborators' results indicate that as perceptual demand increases, greater

effort is required to perform the task, and the increased effort is reflected in the cortical activity (Atiani et al. 2009).

In our analysis we did not measure tuning changes in response to task engagement and arousal fluctuations. However, given that we used the same reference stimuli used in Atiani *et al*., an STRF analysis could further characterize the contribution of arousal in shaping the tuning of sound-evoked activity. Our analysis revealed a predominantly positive global modulation of activity during active engagement in A1, persisting even after accounting for the contribution of arousal. Given that the task was a positive reinforcement approach task previously shown to yield opposite results to the negative reinforcement avoidance task (David et al. 2012), I would expect local changes in STRF tuning near the frequency of the target to be negative, especially during blocks of low SNR. This would maximize the difference between target and reference stimuli, increasing the perceptual sensitivity for target detection.

Population code and noise correlation analysis

For some of our recordings we used a 64-channel array acutely positioned through the depth of A1 or IC. However, for the purposes of this study we considered single unit and stable multiunit recorded with the array independently. Simultaneous recordings of multiple neurons allow for population analyses that might reveal further insight into the relationship between state variables and sensory representation, and provide information about how the local circuit is engaged.

Several studies have shown that representations of sensory stimuli are more faithful when neural cortical activity is desynchronized due to lower pairwise noise

correlations (Averbeck et al. 2006; Bagur et al. 2018; Cohen & Maunsell 2009; Downer et al. 2017; Goard & Dan 2009; Jacobs et al. 2018; Marguet & Harris 2011; McGinley et al. 2015a; Reimer et al. 2014; Vinck et al. 2015; Zagha et al. 2013). Noise correlations quantify the correlation between two neurons' responses to a given stimulus.

A common interpretation of the meaning behind neural desynchronized state is that it improves the signal-to-noise ratio of the neural code by reducing correlated fluctuations in neural activity, thereby leading to more accurate decisions (Cohen & Maunsell 2009; Mitchell et al. 2009). In A1, average noise correlation decreases when animals perform a task compared to passive listening (Downer et al. 2017; Jacobs et al. 2018; McGinley et al. 2015a). In a recent preprint, Jacobs *et al*., 2018 found that the desynchronization of activity appears to be related to active engagement rather than behavioral accuracy, such that it is more pronounced during hits and false alarms, but not during miss trials. Furthermore, while neural desynchronization was found to be widespread across cortical regions, sensory and non-sensory, arousal indexed by pupil did not fully account for it (Jacobs et al. 2018).

Future work involving simultaneous recordings across hundreds of neurons and measurements of noise correlations across neural cortical and subcortical circuits will be needed to investigate the physiological processes underlying the relationship between arousal, task engagement, and sensory encoding. Furthermore, designing experiments that can control for and measure different state variables known to affect sensory representation, such as locomotion (McGinley et al. 2015a; Reimer et al. 2014; Vinck et al. 2015), behavioral performance (Atiani et al. 2009; de Gee et al. 2014), and selective

attention, will be needed to achieve a more complete picture of how sound is represented in a brain that needs to adapt to the current contextual demand.

References

Aertsen AM, Johannesma PI. 1981. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol. Cybern.* 42(2):133–43

Agus TR, Thorpe SJ, Pressnitzer D. 2010. Rapid formation of robust auditory memories: insights from noise. *Neuron*. 66(4):610–18

Ahumada A, Lovell J. 1971. Stimulus Features in Signal Detection. *J. Acoust. Soc. Am.* 49(6B):1751

Aitkin LM, Moore DR. 1975. Inferior colliculus. II. Development of tuning characteristics and tonotopic organization in central nucleus of the neonatal cat. *J. Neurophysiol.* 38(5):1208–16

Aitkin LM, Phillips SC. 1984. The interconnections of the inferior colliculi through their commissure. *J. Comp. Neurol.* 228(2):210–16

Aitkin LM, Webster WR, Veale JL, Crosby DC. 1975. Inferior colliculus. I. Comparison of response properties of neurons in central, pericentral, and external nuclei of adult cat. *J. Neurophysiol.* 38(5):1196–1207

Akeroyd MA, Carlyon RP, Deeks JM. 2005. Can dichotic pitches form two streams? *J. Acoust. Soc. Am.* 118(2):977–81

Alain C, Arnott SR. 2000. Selectively attending to auditory objects. *Front. Biosci.* 5:D202-12

Alain C, Woods DL. 1994. Signal clustering modulates auditory cortical activity in humans. *Percept. Psychophys.* 56(5):501–16

Allen M, Chowdhury T, Wegener M, Moghaddam B. 2018. Efficient sorting of single-unit activity from midbrain cells using KiloSort is as accurate as manual sorting. *bioRxiv*. 303479

Anderse RA, Roth GL, Aitkin LM, Merzenich MM. 1980. The efferent projections of the central nucleus and the pericentral nucleus of the inferior collculus in the cat. *J. Comp. Neurol.* 194(3):649–62

Anderson LA, Christianson GB, Linden JF. 2009. Stimulus-specific adaptation occurs in the auditory thalamus. *J. Neurosci.* 29(22):7359–63

Andreou L-V, Kashino M, Chait M. 2011. The role of temporal regularity in auditory

segregation. *Hear. Res.* 280(1–2):228–35

Anstis SM, Saida S. 1985. Adaptation to auditory streaming of frequency-modulated tones. *J. Exp. Psychol. Hum. Percept. Perform.* 11(3):257–71

Antunes FM, Nelken I, Covey E, Malmierca MS. 2010. Stimulus-Specific Adaptation in the Auditory Thalamus of the Anesthetized Rat. *PLoS One*. 5(11):e14071

Appen RE. 1994. The Pupil: Anatomy, Physiology, and Clinical Applications, Volumes 1 and 2. *Arch. Ophthalmol.* 112(12):1517

Atiani S, David S V, Elgueda D, Locastro M, Radtke-Schuller S, et al. 2014. Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron*. 82(2):486–99

Atiani S, Elhilali M, David S V, Fritz JB, Shamma SA. 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*. 61(3):467–80

Averbeck BB, Latham PE, Pouget A. 2006. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7(5):358–66

Ayala YA, Malmierca MS. 2013. Stimulus-specific adaptation and deviance detection in the inferior colliculus. *Front. Neural Circuits*. 6:89

Bagur S, Averseng M, Elgueda D, David S, Fritz J, et al. 2018. Go/No-Go task engagement enhances population representation of target stimuli in primary auditory cortex. *Nat. Commun.* 9(1):2529

Bajo VM, Nodal FR, Bizley JK, Moore DR, King AJ. 2007. The ferret auditory cortex: descending projections to the inferior colliculus. *Cereb. Cortex*. 17(2):475–91

Bajo VM, Nodal FR, Moore DR, King AJ. 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13(2):253–60

Bakin JS, Lepan B, Weinberger NM. 1992. Sensitization induced receptive field plasticity in the auditory cortex is independent of CS-modality. *Brain Res.* 577(2):226–35

Bakin JS, Weinberger NM. 1990. Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Res.* 536(1–2):271–86

Bandyopadhyay S, Shamma SA, Kanold PO. 2010. Dichotomy of functional organization in the mouse auditory cortex. *Nat. Neurosci.* 13(3):361–68

Bar-Yosef O, Nelken I. 2007. The effects of background noise on the neural responses to

natural sounds in cat primary auditory cortex. *Front. Comput. Neurosci.* 1:3

Bar-Yosef O, Rotman Y, Nelken I. 2002. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J. Neurosci.* 22(19):8619–32

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M. 2016. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc. Natl. Acad. Sci.* 113(5):E616–25

Baruni JK, Lau B, Salzman CD. 2015. Reward expectation differentially modulates attentional behavior and activity in visual area V4. *Nat. Neurosci.* 18(11):1656–63

Beaton R, Miller JM. 1975. Single cell activity in the auditory cortex of the unanesthetized, behaving monkey: correlation with stimulus controlled behavior. *Brain Res.* 100(3):543–62

Beatty J. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91(2):276–92

Beauvois MW, Meddis R. 1996. Computer simulation of auditory stream segregation in alternating-tone sequences. *J. Acoust. Soc. Am.* 99(4 Pt 1):2270–80

Bee MA, Klump GM. 2004. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J. Neurophysiol.* 92(2):1088–1104

Bee MA, Klump GM. 2005. Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain. Behav. Evol.* 66(3):197–214

Bee MA, Micheyl C. 2008. The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *J. Comp. Psychol.* 122(3):235–51

Bendixen A, Denham SL, Gyimesi K, Winkler I. 2010. Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* 128(6):3658–66

Bendixen A, SanMiguel I, Schröger E. 2012. Early electrophysiological indicators for predictive processing in audition: a review. *Int. J. Psychophysiol.* 83(2):120–31

Best V, Ozmeral E, Gallun FJ, Sen K, Shinn-Cunningham BG. 2005. Spatial unmasking of birdsong in human listeners: energetic and informational factors. *J. Acoust. Soc. Am.* 118(6):3766–73

Bidet-Caulet A, Bertrand O. 2009. Neurophysiological mechanisms involved in auditory

perceptual organization. *Front. Neurosci.* 3(2):182–91

Bizley JK, Bajo VM, Nodal FR, King AJ. 2015. Cortico-cortical connectivity within ferret auditory cortex. *J. Comp. Neurol.*

Bizley JK, Cohen YE. 2013. The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14(10):693–707

Bizley JK, Nodal FR, Nelken I, King AJ. 2005. Functional organization of ferret auditory cortex. *Cereb. Cortex*. 15(10):1637–53

Bizley JK, Walker KMM, Nodal FR, King AJ, Schnupp JWH. 2013. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr. Biol.* 23(7):620–25

Black C, Voigts J, Agrawal U, Ladow M, Santoyo J, et al. 2017. Open Ephys electroencephalography (Open Ephys + EEG): a modular, low-cost, open-source solution to human neural recording. *J. Neural Eng.* 14(3):035002

Bregman AS. 1978a. Auditory streaming is cumulative. *J. Exp. Psychol. Hum. Percept. Perform.* 4(3):380–87

Bregman AS. 1978b. Auditory streaming: competition among alternative organizations. *Percept. Psychophys.* 23(5):391–98

Bregman AS. 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA. MIT Press ed.

Bregman AS, Ahad PA, Crum PA, O'Reilly J. 2000. Effects of time intervals and tone durations on auditory stream segregation. *Percept. Psychophys.* 62(3):626–36

Briley PM, Krumbholz K. 2013. The specificity of stimulus-specific adaptation in human auditory cortex increases with repeated exposure to the adapting stimulus. *J. Neurophysiol.* 110(12):2679–88

Brody CD, Romo R, Kepecs A. 2003. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr. Opin. Neurobiol.* 13(2):204–11

Brosch M, Schreiner CE. 1997. Time course of forward masking tuning curves in cat primary auditory cortex. *J. Neurophysiol.* 77(2):923–43

Byrd RH, Lu P, Nocedal J, Zhu C. 1995. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* 16(5):1190–1208

Calford MB, Semple MN. 1995. Monaural inhibition in cat auditory cortex. *J. Neurophysiol.* 73(5):1876–91

Campbell AW. 1905. *Histological Studies on the Localisation of Cerebral Function - Alfred Walter Campbell - Google Books*

Carlyon RP. 2004. How the brain separates sounds. *Trends Cogn. Sci.* 8(10):465–71

Chen C, Cheng M, Ito T, Song S. 2018. Neuronal Organization in the Inferior Colliculus Revisited with Cell-Type-Dependent Monosynaptic Tracing. *J. Neurosci.* 38(13):3318–32

Cherry EC. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *J. Acoust. Soc. Am.* 25(5):975–79

Cheveigné A de, Kawahara H, Tsuzaki M, Aikawa K. 1998. Concurrent vowel identification. I. Effects of relative amplitude and F0 difference. *J. Acoust. Soc. Am.* 101(5):2839

Christiansen SK, Jepsen MLL, Dau T. 2014. Effects of tonotopicity, adaptation, modulation tuning, and temporal coherence in "primitive" auditory stream segregation. *J. Acoust. Soc. Am.* 135(1):323–33

Cohen MR, Maunsell JHR. 2009. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 12(12):1594–1600

Cohen MR, Maunsell JHR. 2010. A Neuronal Population Measure of Attention Predicts Behavioral Performance on Individual Trials. *J. Neurosci.* 30(45):15241–53

Coleman JR, Clerici WJ. 1987. Sources of projections to subdivisions of the inferior colliculus in the rat. *J. Comp. Neurol.* 262(2):215–26

Cusack R, Deeks J, Aikman G, Carlyon RP. 2004. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J. Exp. Psychol. Hum. Percept. Perform.* 30(4):643–56

Cusack R, Roberts B. 2000. Effects of differences in timbre on sequential grouping. *Percept. Psychophys.* 62(5):1112–20

Dai HP, Scharf B, Buus S. 1991. Effective attenuation of signals in noise under focused attention. *J. Acoust. Soc. Am.* 89(6):2837–42

Daikhin L, Ahissar M. 2012. Responses to deviants are modulated by subthreshold variability of the standard. *Psychophysiology*. 49(1):31–42

Darwin C, Carlyon RP. 1995. *Auditory Grouping: The Handbook of Perception and Cognition*. Academic, New York: ed Moore BCJ. Vol 6 ed.

Darwin CJ. 1997. Auditory grouping. *Trends Cogn. Sci.* 1(9):327–33

Darwin CJ, Ciocca V. 1992. Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component. *J. Acoust. Soc. Am.* 91(6):3381–90

David S V. 2018. Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding. *Hear. Res.* 360:107–23

David S V, Fritz JB, Shamma S a. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 109(6):2144–49

David S V, Mesgarani N, Fritz JB, Shamma SA. 2009. Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *J. Neurosci.* 29(11):3374–86

de Gee JW, Knapen T, Donner TH. 2014. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc. Natl. Acad. Sci. U. S. A.* 111(5):E618-25

deCharms RC, Blake DT, Merzenich MM. 1998. Optimizing sound features for cortical neurons. *Science*. 280(5368):1439–43

Depireux DA, Simon JZ, Klein DJ, Shamma SA. 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J.* .... 85(3):1220–34

Ding N, Simon JZ. 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U. S. A.* 109(29):11854–59

Divenyi P. 2004. *Speech Separation by Humans and Machines*, Vol. 2004. Springer

Downer JD, Niwa M, Sutter ML. 2015. Task engagement selectively modulates neural correlations in primary auditory cortex. *J. Neurosci.* 35(19):7565–74

Downer JD, Rapone B, Verhein J, O'Connor KN, Sutter ML. 2017. Feature-Selective Attention Adaptively Shifts Noise Correlations in Primary Auditory Cortex. *J. Neurosci.* 37(21):5378–92

Duque D, Malmierca MS. 2015. Stimulus-specific adaptation in the inferior colliculus of the mouse: anesthesia and spontaneous activity effects. *Brain Struct. Funct.* 220(6):3385–98

Ecker AS, Tolias AS. 2014. Is there signal in the noise? *Nat. Neurosci.* 17(6):750–51

Efron B, Tibshirani R. 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Stat. Sci.* 1(1):54–75

Ehret G, Schreiner CE. 1997. Frequency resolution and spectral integration (critical band analysis) in single units of the cat primary auditory cortex. *J. Comp. Physiol. A.* 181(6):635–50

Elgueda D, Duque D, Radtke-Schuller S, Yin P, David S V, et al. 2019. State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. *Nat. Neurosci.*

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma S a. 2009a. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron*. 61(2):317–29

Elhilali M, Xiang J, Shamma SA, Simon JZ. 2009b. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol.* 7(6):e1000129

Eliades SJ, Wang X. 2003. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* 89(4):2194–2207

Eliades SJ, Wang X. 2005. Dynamics of Auditory–Vocal Interaction in Monkey Auditory Cortex. *Cereb. Cortex*. 15(10):1510–23

Eliades SJ, Wang X. 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. 453(7198):1102–6

Englitz B, David SV V, Sorenson MDD, Shamma SAA. 2013. MANTA − An open-source, high density electrophysiology recording suite for MATLAB. *Front. Neural Circuits*. 7:69

Farley BJ, Quirk MC, Doherty JJ, Christian EP. 2010. Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. *J. Neurosci.* 30(49):16475–84

Fay R. 2008. Sound source perception and stream segregation in nonhuman vertebrate animals. In *Auditory Perception of Sound Sources.*, eds. W Yost, R Fay, A Popper, pp. 307–23. Springer; New York

Fishman YI, Arezzo JC, Steinschneider M. 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J.*

Acoust. Soc. Am. 116(3):1656–70

Fishman YI, Reser DH, Arezzo JC, Steinschneider M. 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear. Res.* 151(1–2):167–87

Fritz J, Elhilali M, Shamma S. 2005a. Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206(1–2):159–76

Fritz J, Shamma S, Elhilali M, Klein D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6(11):1216–23

Fritz JB, David S V, Radtke-Schuller S, Yin P, Shamma SA. 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat. Neurosci.* 13(8):1011–19

Fritz JB, Elhilali M, David S V., Shamma SA. 2007a. Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear. Res.* 229(1–2):186–203

Fritz JB, Elhilali M, Shamma SA. 2005b. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J. Neurosci.* 25(33):7623–35

Fritz JB, Elhilali M, Shamma SA. 2007b. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* 98(4):2337–46

Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, et al. 2014. A cortical circuit for gain control by behavioral state. *Cell*. 156(6):1139–52

Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. 2010. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn. Affect. Behav. Neurosci.* 10(2):252–69

Gittelman JX, Perkel DJ, Portfors C V. 2013. Dopamine modulates auditory responses in the inferior colliculus in a heterogeneous manner. *J. Assoc. Res. Otolaryngol.* 14(5):719–29

Goard M, Dan Y. 2009. Basal forebrain activation enhances cortical coding of natural scenes. *Nat. Neurosci.* 12(11):1444–49

Gonzalez RC, Woods RE (Richard E. 2010. *Digital Image Processing*. Pearson/Prentice

Hall

Goris RLT, Movshon JA, Simoncelli EP. 2014. Partitioning neuronal variability. *Nat. Neurosci.* 17(6):858–65

Granholm E, Steinhauer SR. 2004. Pupillometric measures of cognitive and emotional processes. *Int. J. Psychophysiol.* 52(1):1–6

Griffiths TD, Warren JD. 2004. What is an auditory object? *Nat. Rev. Neurosci.* 5(11):887–92

Grill-Spector K, Henson R, Martin A. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* 10(1):14–23

Gruters KG, Groh JM. 2012. Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Front. Neural Circuits*. 6(December):96

Gutschalk A, Dykstra AR. 2014. Functional imaging of auditory scene analysis. *Hear. Res.* 307:98–110

Habbicht H, Vater M. 1996. A microiontophoretic study of acetylcholine effects in the inferior colliculus of horseshoe bats: implications for a modulatory role. *Brain Res.* 724(2):169–79

Hackett TA. 2011. Information flow in the auditory cortical network. *Hear. Res.* 271(1–2):133–46

Hackett TA, Stepniewska I, Kaas JH. 1998a. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 394(4):475–95

Hackett TA, Stepniewska I, Kaas JH. 1998b. Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.* 400(2):271–86

Hartmann WM, Johnson D. 1991. Stream Segregation and Peripheral Channeling. *Music Percept. An Interdiscip. J.* 9(2):155–83

Heil P, Rajan R, Irvine DR. 1994. Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hear. Res.* 76(1–2):188–202

Hocherman S, Benson DA, Goldstein MH, Heffner HE, Hienz RD. 1976. Evoked unit activity in auditory cortex of monkeys performing a selective attention task. *Brain Res.* 117(1):51–68

Hoffman MD, Gelman A. 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths

in Hamiltonian Monte Carlo

Huang CM, Fex J. 1986. Tonotopic organization in the inferior colliculus of the rat demonstrated with the 2-deoxyglucose method. *Exp. brain Res.* 61(3):506–12

Huffman RF, Henson OW. The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Res. Brain Res. Rev.* 15(3):295–323

Hurley LM, Pollak GD. 2005. Serotonin shifts first-spike latencies of inferior colliculus neurons. *J. Neurosci.* 25(34):7876–86

Jacobs EAK, Steinmetz NA, Carandini M, Harris KD. 2018. Cortical state fluctuations during sensory decision making. *bioRxiv*. 348193

Jaramillo S, Zador AM. 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* 14(2):246–51

Jepma M, Nieuwenhuis S. 2011. Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *J. Cogn. Neurosci.* 23(7):1587–96

Joshi S, Li Y, Kalwani RM, Gold JI. 2016. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*. 89(1):221–34

Kaas JH, Hackett TA. 2000. Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci.* 97(22):11793–99

Kahneman D, Beatty J. 1966. Pupil diameter and load on memory. *Science*. 154(3756):1583–85

Kanwal JS, Medvedev A V, Micheyl C. 2003. Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment. *Network*. 14(3):413–35

Kato HKKKK, Chu MWWWW, Isaacson JSSSS, Komiyama T. 2012. Dynamic Sensory Representations in the Olfactory Bulb: Modulation by Wakefulness and Experience. *Neuron*. 76(5):962–75

Khouri L, Nelken I. 2015. Detecting the unexpected. *Curr. Opin. Neurobiol.* 35:142–47

Klein DJ, Depireux DA, Simon JZ, Shamma SA. 2000. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J. Comput. Neurosci.* 9(1):85–111

Klein DJ, Simon JZ, Depireux DA, Shamma SA. 2006. Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J. Comput. Neurosci.* 20(2):111–36

Klepper A, Herbert H. 1991. Distribution and origin of noradrenergic and serotonergic fibers in the cochlear nucleus and inferior colliculus of the rat. *Brain Res.* 557(1–2):190–201

Knudsen DP, Gentner TQ. 2013. Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. *J. Neurophysiol.* 109(7):1690–1703

Kuchibhotla K V, Gill J V, Lindsay GW, Papadoyannis ES, Field RE, et al. 2017. Parallel processing by cortical inhibition enables context-dependent behavior. *Nat. Neurosci.* 20(1):62–71

Kuwabara N, Zook JM. 2000. Geniculo-collicular descending projections in the gerbil. *Brain Res.* 878(1–2):79–87

Lee C-C, Middlebrooks JC. 2011. Auditory cortex spatial sensitivity sharpens during task performance. *Nat. Neurosci.* 14(1):108–14

Lee CC. 2015. Exploring functions for the non-lemniscal auditory thalamus. *Front. Neural Circuits*. 9:69

Luck SJ, Chelazzi L, Hillyard SA, Desimone R. 2013. Neural Mechanisms of Spatial Selective Attention in Areas V1 , V2 , and V4 of Macaque Visual Cortex Neural Mechanisms of Spatial Selective Attention in Areas V1 , V2 , and V4 of Macaque Visual Cortex. . 24–42

Luethke LE, Krubitzer LA, Kaas JH. 1988. Cortical connections of electrophysiologically and architectonically defined subdivisions of auditory cortex in squirrels. *J. Comp. Neurol.* 268(2):181–203

Ma L, Micheyl C, Yin P, Oxenham AJ, Shamma SA. 2010. Behavioral measures of auditory streaming in ferrets (Mustela putorius). *J. Comp. Psychol.* 124(3):317–30

Malmierca MS, Anderson LA, Antunes FM. 2015. The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding. *Front. Syst. Neurosci.* 9:19

Malmierca MS, Cristaudo S, Pérez-González D, Covey E, Perez-Gonzalez D, Covey E.

2009. Stimulus-Specific Adaptation in the Inferior Colliculus of the Anesthetized Rat. *J. Neurosci.* 29(17):5483–93

Malmierca MS, Seip KL, Osen KK. 1995. Morphological classification and identification of neurons in the inferior colliculus: a multivariate analysis. *Anat. Embryol. (Berl).* 191(4):343–50

Marguet SL, Harris KD. 2011. State-Dependent Representation of Amplitude-Modulated Noise Stimuli in Rat Auditory Cortex. *J. Neurosci.* 31(17):6414–20

Marrone N, Mason CR, Kidd G. 2008. Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends Amplif.* 12(4):300–315

Masterton RB, Granger EM, Glendenning KK. 1992. Psychoacoustical contribution of each lateral lemniscus. *Hear. Res.* 63(1–2):57–70

Masutomi K, Barascud N, Kashino M, McDermott JH, Chait M. 2015. Sound Segregation via Embedded Repetition Is Robust to Inattention. *J. Exp. Psychol. Hum. Percept. Perform.*

May PJC, Tiitinen H. 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*. 47(1):66–122

McCabe SL, Denham MJ. 1997. A model of auditory streaming. *J. Acoust. Soc. Am.* 101(3):1611–21

McDermott JH. 2009. The cocktail party problem. *Curr. Biol.* 19(22):R1024-7

McDermott JH, Wrobleski D, Oxenham AJ. 2011. Recovering sound sources from embedded repetition. *Proc. Natl. Acad. Sci. U. S. A.* 108(3):1188–93

McGinley MJ, David S V, McCormick DA. 2015a. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*. 87(1):179–92

McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, et al. 2015b. Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron*. 87(6):1143–61

Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*. 485(7397):233–36

Mesgarani N, David S V, Fritz JB, Shamma SA. 2009. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J. Neurophysiol.* 102(6):3329–39

Mesgarani N, Fritz J, Shamma S. 2010. A computational model of rapid task-related

plasticity of auditory cortical receptive fields. *J. Comput. Neurosci.* 28(1):19–27

Metzger RR, Greene NT, Porter KK, Groh JM. 2006. Effects of reward and behavioral context on neural activity in the primate inferior colliculus. *J. Neurosci.* 26(28):7468–76

Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, et al. 2007a. The role of auditory cortex in the formation of auditory streams. *Hear. Res.* 229(1–2):116–31

Micheyl C, Hanson C, Demany L, Shamma S, Oxenham AJ. 2013a. Auditory stream segregation for alternating and synchronous tones. *J. Exp. Psychol. Hum. Percept. Perform.* 39(6):1568–80

Micheyl C, Kreft H, Shamma S, Oxenham AJ. 2013b. Temporal coherence versus harmonicity in auditory stream formation. *J. Acoust. Soc. Am.* 133(3):EL188-94

Micheyl C, Shamma SA, Oxenham AJ. 2007b. Hearing Out Repeating Elements in Randomly Varying Multitone Sequences: A Case of Streaming? BT - Hearing – From Sensory Processing to Perception. , eds. B Kollmeier, G Klump, V Hohmann, U Langemann, M Mauermann, et al., pp. 267–74. Berlin, Heidelberg: Springer Berlin Heidelberg

Micheyl C, Tian B, Carlyon RP, Rauschecker JP. 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*. 48(1):139–48

Mitchell JF, Sundberg K a, Reynolds JH. 2009. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron*. 63(6):879–88

Moerel M, De Martino F, Formisano E. 2014. An anatomical and functional topography of human auditory cortical areas. *Front. Neurosci.* 8:225

Moore BC, Glasberg BR, Peters RW. 1986. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* 80(2):479–83

Moore BCJ, Gockel HE. 2012. Properties of auditory stream formation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367(1591):919–31

Moore DR. 1988. Auditory brainstem of the ferret: Sources of projections to the inferior colliculus. *J. Comp. Neurol.* 269(3):342–54

Moore DR, Semple MN, Addison PD. 1983. Some acoustic properties of neurones in the ferret inferior colliculus. *Brain Res.* 269(1):69–82

Morest DK, Oliver DL. 1984. The neuronal architecture of the inferior colliculus in the cat:

defining the functional anatomy of the auditory midbrain. *J. Comp. Neurol.* 222(2):209–36

Motts SD, Schofield BR. 2009. Sources of cholinergic input to the inferior colliculus. *Neuroscience*. 160(1):103–14

Murray RF. 2011. Classification images: A review. *J. Vis.* 11(5):2–2

Näätänen R. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*. 38(1):1–21

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, et al. 2015. Complementary control of sensory adaptation by two types of cortical interneurons. *Elife*. 4:

Natan RG, Carruthers IM, Mwilambwe-Tshilobo L, Geffen MN. 2017. Gain Control in the Auditory Cortex Evoked by Changing Temporal Correlation of Sounds. *Cereb. Cortex*. 27(3):2385–2402

Nelken I. 2014. Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models. *Biol. Cybern.* 108(5):655–63

Nelken I, Ulanovsky N. 2007. Mismatch Negativity and Stimulus-Specific Adaptation in Animal Models. *J. Psychophysiol.* 21(3–4):214–23

Nelken I, Yaron A, Polterovich A, Hershenhoren I. 2013. Stimulus-Specific Adaptation Beyond Pure Tones. , pp. 411–18

Nelson A, Schneider DM, Takatoh J, Sakurai K, Wang F, Mooney R. 2013. A circuit for motor cortical modulation of auditory cortical activity. *J. Neurosci.* 33(36):14342–53

Nguyen AH, Stark LW. 1993. Model control of image processing: pupillometry. *Comput. Med. Imaging Graph.* 17(1):21–33

Niell CM, Stryker MP. 2010. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron*. 65(4):472–79

Niwa M, Johnson JS, O'Connor KN, Sutter ML. 2012. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J. Neurosci.* 32(27):9323–34

Niwa M, Johnson JS, O'Connor KN, Sutter ML. 2013. Differences between primary auditory cortex and auditory belt related to encoding and choice for AM sounds. *J.*

Neurosci. 33(19):8378–95

O'Sullivan JA, Shamma SA, Lalor EC. 2015. Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. *J. Neurosci.* 35(18):7256–63

Oberfeld D. 2014. An objective measure of auditory stream segregation based on molecular psychophysics. *Atten. Percept. Psychophys.* 76(3):829–51

Oliver DL. 1987. Projections to the inferior colliculus from the anteroventral cochlear nucleus in the cat: possible substrates for binaural interaction. *J. Comp. Neurol.* 264(1):24–46

Oliver DL. 2005. Neuronal Organization in the Inferior Colliculus. In *The Inferior Colliculus*, pp. 69–114. New York: Springer-Verlag

Oliver DL, Kuwada S, Yin TC, Haberly LB, Henkel CK. 1991. Dendritic and axonal morphology of HRP-injected neurons in the inferior colliculus of the cat. *J. Comp. Neurol.* 303(1):75–100

Osmanski MS, Wang X. 2015. Behavioral Dependence of Auditory Cortical Responses. *Brain Topogr.* 28(3):365–78

Otazu GH, Tai L-H, Yang Y, Zador AM. 2009. Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* 12(5):646–54

Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD. 2016. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*. 061481

Pérez-González D, Malmierca MS. 2014. Adaptation in the auditory system: an overview. *Front. Integr. Neurosci.* 8:19

Petreanu L, Gutnisky DA, Huber D, Xu N, O'Connor DH, et al. 2012. Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature*. 489(7415):299–303

Polack P-O, Friedman J, Golshani P. 2013. Cellular mechanisms of brain state–dependent gain modulation in visual cortex. *Nat. Neurosci.* 16(9):1331–39

Poulet JFA, Petersen CCH. 2008. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature*. 454(7206):881–85

Pressnitzer D, Hupé J-M. 2006. Temporal Dynamics of Auditory and Visual Bistability

Reveal Common Principles of Perceptual Organization. *Curr. Biol.* 16(13):1351–57

Rauschecker JP. 1998. Parallel Processing in the Auditory Cortex of Primates. *Audiol. Neurotol.* 3(2–3):86–103

Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci.* 97(22):11800–806

Read HL, Winer JA, Schreiner CE. 2002. Functional architecture of auditory cortex. *Curr. Opin. Neurobiol.* 12(4):433–40

Reimer J, Froudarakis E, Cadwell CRR, Yatsenko D, Denfield GHH, Tolias ASS. 2014. Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. *Neuron*. 84(2):355–62

Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, et al. 2016. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* 7(1):13289

Reynolds JH, Chelazzi L. 2004. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* 27:611–47

Ringach DL. 2009. Spontaneous and driven cortical activity: implications for computation. *Curr. Opin. Neurobiol.* 19(4):439–44

Roberts B, Brunstrom JM. 1998. Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *J. Acoust. Soc. Am.* 104(4):2326–38

Roberts B, Glasberg BR, Moore BCJ. 2002. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am.* 112(5 Pt 1):2074–85

Rockel AJ, Jones EG. 1973a. The neuronal organization of the inferior colliculus of the adult cat. I. The central nucleus. *J. Comp. Neurol.* 147(1):11–59

Rockel AJ, Jones EG. 1973b. The neuronal organization of the inferior colliculus of the adult cat. II. The pericentral nucleus. *J. Comp. Neurol.* 149(3):301–33

Rodgers CC, DeWeese MR. 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

Rubin J, Ulanovsky N, Nelken I, Tishby N. 2016. The Representation of Prediction Error in Auditory Cortex. *PLoS Comput. Biol.* 12(8):e1005058

Ryan A, Miller J. 1977. Effects of behavioral performance on single-unit firing patterns in inferior colliculus of the rhesus monkey. *J. Neurophysiol.* 40(4):943–56

Sachidhanandam S, Sreenivasan V, Kyriakatos A, Kremer Y, Petersen CCH. 2013. Membrane potential correlates of sensory perception in mouse barrel cortex. *Nat. Neurosci.* 16(11):1671–77

Saint Marie RL, Shneiderman A, Stanforth DA. 1997. Patterns of gamma-aminobutyric acid and glycine immunoreactivities reflect structural and functional differences of the cat lateral lemniscal nuclei. *J. Comp. Neurol.* 389(2):264–76

Saldaña E, Merchán MA. 1992. Intrinsic and commissural connections of the rat inferior colliculus. *J. Comp. Neurol.* 319(3):417–37

Schneider DM, Nelson A, Mooney R. 2014. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*. 513(7517):189–94

Schnupp JWH. 2008. Auditory Neuroscience: Sound Segregation in the Brainstem? *Curr. Biol.* 18(16):R705–6

Schnupp JWH, Nelken I, King AJ. 2012. *Auditory Neuroscience: Making Sense of Sound*. MIT Press

Schreiner CE, Mendelson JR. 1990. Functional topography of cat primary auditory cortex: distribution of integrated excitation. *J. Neurophysiol.* 64(5):1442–59

Schriver BJ, Bagdasarov S, Wang Q. 2018. Pupil-linked arousal modulates behavior in rats performing a whisker deflection direction discrimination task. *J. Neurophysiol.* 120(4):1655–70

Schwartz Z, David S V. 2015. Isolating Distinct Top-Down Control Circuits for Auditory Cortex. *Midwinter Meet. Assoc. Res. Otolaryngol.*

Schwartz ZP, David S V. 2018. Focal Suppression of Distractor Sounds by Selective Attention in Auditory Cortex. *Cereb. Cortex*. 28(1):323–39

Schwarz DM, Zilany MSAA, Skevington M, Huang NJ, Flynn BC, Carney LH. 2012. Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. *J. Neurosci. Methods*. 206(2):120–31

Semple MN, Aitkin LM. 1979. Representation of sound frequency and laterality by units in

central nucleus of cat inferior colliculus. *J. Neurophysiol.* 42(6):1626–39

Shamma S a, Elhilali M, Micheyl C. 2011. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34(3):114–23

Shamma SA, Fleshman JW, Wiser PR, Versnel H. 1993. Organization of response areas in ferret primary auditory cortex. *J. Neurophysiol.* 69(2):367–83

Shimaoka D, Harris KD, Carandini M. 2018. Effects of Arousal on Mouse Sensory Cortex Depend on Modality. *Cell Rep.* 22(12):3160–67

Shinn-Cunningham BG, Best V. 2008. Selective attention in normal and impaired hearing. *Trends Amplif.* 12(4):283–99

Shobe JL, Claar LD, Parhami S, Bakhurin KI, Masmanidis SC. 2015. Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *J. Neurophysiol.* 114(3):2043–52

Shuler MG, Bear MF. 2006. Reward timing in the primary visual cortex. *Science*. 311(5767):1606–9

Signal Detection Theory for Everyman? 1993. *J. Math. Psychol.* 37:

Singh PG, Bregman AS. 1997. The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *J. Acoust. Soc. Am.* 102(4):1943–52

Slee SJ, David S V. 2015a. Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in the Auditory Midbrain. *J. Neurosci.* 35(38):13090–102

Slee SJ, David S V. 2015b. Rapid task-related plasticity of spectro-temporal receptive fields in the auditory midbrain. *J. Neurosci.* 35(38):13090–102

Slee SJ, Young ED. 2013. Linear processing of interaural level difference underlies spatial tuning in the nucleus of the brachium of the inferior colliculus. *J. Neurosci.* 33(9):3891–3904

Snyder JS, Alain C, Picton TW. 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn. Neurosci.* 18(1):1–13

Snyder JS, Gregg MK, Weintraub DM, Alain C. 2012. Attention, awareness, and the perception of auditory scenes. *Front. Psychol.* 3:15

Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behav. Res. Methods. Instrum. Comput.* 31(1):137–49

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. 2018.

Spontaneous behaviors drive multidimensional, brain-wide population activity. *bioRxiv*. 306019

Sundberg K a, Mitchell JF, Reynolds JH. 2009. Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron*. 61(6):952–63

Sussman E, Steinschneider M. 2006. Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Res.* 1075(1):165–74

Syka J, Popelár J, Kvasnák E, Astl J. 2000. Response properties of neurons in the central nucleus and external and dorsal cortices of the inferior colliculus in guinea pig. *Exp. brain Res.* 133(2):254–66

Szalárdy O, Bendixen A, Böhm TM, Davies LA, Denham SL, Winkler I. 2014. The effects of rhythm and melody on auditory stream segregation. *J. Acoust. Soc. Am.* 135(3):1392–1405

Taaseh N, Yaron A, Nelken I. 2011. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PLoS One*. 6(8):e23369

Teki S, Barascud N, Picard S, Payne C, Griffiths TD, Chait M. 2016. Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. *Cereb. Cortex*

Teki S, Chait M, Kumar S, Shamma S, Griffiths TD. 2013. Segregation of complex acoustic scenes based on temporal coherence. *Elife*. 2:e00699

Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD. 2011. Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci.* 31(1):164–71

Thorson IL, Liénard J, David S V. 2015. The Essential Complexity of Auditory Receptive Fields. *PLoS Comput. Biol.* 11(12):e1004628

Tsunada J, Liu ASK, Gold JI, Cohen YE. 2016. Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat. Neurosci.* 19(1):135–42

Ulanovsky N, Las L, Farkas D, Nelken I. 2004. Multiple time scales of adaptation in auditory cortex neurons. *J. Neurosci.* 24(46):10440–53

Ulanovsky N, Las L, Nelken I. 2003. Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6(4):391–98

van Noorden L. 1975. *Temporal Coherence in the Perception of Tone Sequences*. Technical University Eindhoven

Varnet L, Knoblauch K, Meunier F, Hoen M. 2013. Using auditory classification images for

the identification of fine acoustic cues used in speech perception. *Front. Hum. Neurosci.* 7:865

Vinck M, Batista-Brito R, Knoblich U, Cardin JAA. 2015. Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*. 86(3):740–54

Vinje WE, Gallant JL. 2000. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*. 287(5456):1273–76

Wang D, Brown GJ. 2006. *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*. Wiley

Warren RM. 1970. Perceptual restoration of missing speech sounds. *Science*. 167(917):392–93

Weinberger NM. 1997. Learning-Induced Receptive Field Plasticity in the Primary Auditory Cortex. *Semin. Neurosci.* 9(1–2):59–67

Weinberger NM, Javid R, Lepan B. 1993. Long-term retention of learning-induced receptive-field plasticity in the auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 90(6):2394–98

Wierda SM, van Rijn H, Taatgen NA, Martens S. 2012. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proc. Natl. Acad. Sci. U. S. A.* 109(22):8456–60

Wimmer RD, Schmitt LI, Davidson TJ, Nakajima M, Deisseroth K, Halassa MM. 2015. Thalamic control of sensory selection in divided attention. *Nature*. 526(7575):705–9

Winer E. 1992. The Functional Architecture of the Medial Geniculate Body and the Primary Auditory Cortex. In *Springer Handbook of Auditory Research, Volume 1*

Winer J a, Larue DT, Diehl JJ, Hefti BJ. 1998. Auditory cortical projections to the cat inferior colliculus. *J. Comp. Neurol.* 400(2):147–74

Winer JA. 2005. Three Systems of Descending Projections to the Inferior Colliculus. In *The Inferior Colliculus*, pp. 231–47. New York: Springer-Verlag

Winer JA, Lee CC. 2007. The distributed auditory cortex. *Hear. Res.* 229(1–2):3–13

Winer JA, Schreiner CE. 2005. The Central Auditory System: A Functional Analysis. In *The Inferior Colliculus*, eds. JA Winer, CE Schreiner, pp. 1–68. New York: Springer-Verlag

Winkler I, Denham SL, Nelken I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* 13(12):532–40

Winn MB, Edwards JR, Litovsky RY. 2015. The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear Hear.* 36(4):e153-65

Yarden TS, Nelken I. 2017. Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. *PLOS Comput. Biol.* 13(3):e1005437

Yin P, Fritz JB, Shamma SA. 2010. Do ferrets perceive relative pitch? *J. Acoust. Soc. Am.* 127(3):1673–80

Yin P, Fritz JB, Shamma SA. 2014. Rapid spectrotemporal plasticity in primary auditory cortex during behavior. *J. Neurosci.* 34(12):4396–4408

Zagha E, Casale AE, Sachdev RNS, McGinley MJ, McCormick DA. 2013. Motor Cortex Feedback Influences Sensory Processing by Modulating Network State. *Neuron*. 79(3):567–78