Effects of attention and arousal on neural activity in auditory cortex

By

Leah Phillips Schwartz

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

June 19, 2020

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

LEAH PHILLIPS SCHWARTZ

has been approved on June 19, 2020

Advisor, Stephen V. David, Ph.D.

Member and Chair, Laurence O. Trussell, Ph.D.

Member, Gary L. Westbrook, M.D.

Member, Frederick J. Gallun, Ph.D.

External Member, Atheir Abbas, Ph.D.

Table of Contents

Acknowledgements

My mentor, Dr. Stephen David, generously allowed me to join his laboratory and was involved with all stages of this project. Without his deep curiosity about the auditory system, skill as an applied mathematician, programmer, and physiologist, and patience with my false starts and moments of doubt, the work in this dissertation could not have been started or completed.

I am also indebted to many members of the David Lab for training and emotional support. Daniela became a true friend during the years we overlapped as students. I'll never forget the energy she puts into understanding everything and everyone she encounters. Sean provided a model of scientific integrity. In addition to sharing their technical expertise, Brad, Luke, and Jean were each there for crucial conversations about how to keep going. Jesyin's friendship and understanding meant a lot to me. Ivar and Henry helped make the lab a lively and inviting place that I wanted to join. It's been exciting to see Charlie, Mateo, and Greg taking the lab's work in new directions.

Working as a technician in Dr. Susan Voglmaier's laboratory inspired my interest in neuroscience. I will always be thankful that she believed in me and supported my decision to pursue a PhD, in spite of my unconventional career path.

OHSU's Neuroscience Graduate Program gave me a home and an education whose quality I did not deserve. I am particularly indebted to Drs. Laurence Trussell, Rowland Taylor, Gary Westbrook, and Erick Gallun, for thoughtful questions, advice, and encouragement during this project as members of my dissertation committee.

Daisy and Ted Miller provided financial support during the first years of this project through their donations to the ARCS Foundation Oregon Chapter. Thank you for your generosity and your enthusiasm for scientific research.

To Dr. Adrian Larsen Sanchez, Dr. Jodi DeMunter, and the community I've found here, especially Cait, Janelle, and Rachel: It was the best thing that happened in graduate school. You are the reason it was possible, and why it was worth it.

To my extended family, my sisters, and most of all my parents: I doubt this is what you expected, but it is the book these years produced. Thank you for being there for me throughout graduate school, and so many times before. There is no way to say or measure how much you've given me.

Abstract

The auditory system uses sound to guide behavior. However, the aspects of the sensory environment that are relevant to an organism's behavioral goals change over time. How does the brain allow us to understand the words of one particular speaker in a conversation-filled room, or tune out ambient noise as we drift to sleep? In this dissertation, we explore the source of this flexibility by examining how attention and arousal affect neural activity in one region of the brain that processes sound.

We recorded extracellular, single-neuron spiking activity in primary auditory cortex (A1) of non-anesthetized ferrets trained on a selective attention task. Stimuli consisted of two streams of tones in noise that differed in frequency, location, and envelope. Ferrets were rewarded for responding to tones in one stream, but punished with a time-out for responding to tones or noise at a different frequency. When ferrets attended to sound that a neuron was tuned to, the neuron's response to the noise distractor at the frequency decreased, improving the discriminability of tones from noise in the attended stream. During the task, spontaneous and sound-evoked activity increased relative to neural activity during passive exposure to the same sounds, indicating that task engagement effects differed from attentional effects. Varying the difficulty of the task by changing the signal-to-noise ratio of tones in noise also affected spontaneous activity, but not the neural response to tones or noise.

Pupil size has been shown to track rapid changes in cognitive state and neural activity in multiple brain regions, including auditory cortex. To examine effects of arousal state on neural excitability, we therefore recorded responses to repeated natural

sounds and pupil size. In most neurons, spontaneous and sound-evoked activity in A1 increases when pupil was large. In addition, when pupil was large, the neural response to sound became more reliable across sound repetitions, and the activity of neurons became less synchronized. During some recordings, we observed an increase in saccades and constricted pupil that may have indicated sleep onset, which was sometimes accompanied by changes in neural activity distinct from those involved in changes in waking arousal. To examine effects of arousal state on neural selectivity, we recorded response to tone pips at varying frequencies and levels. Across the population of recorded neurons, there was a small decrease in acoustic threshold when pupil was large, but no change in best frequency or spectral bandwidth.

Taken together, these results suggest that pupil-linked arousal and attention have distinct effects on neural activity in A1. Pupil size tracks the gross level of activity evoked by auditory stimuli: representations of sound in A1 become more salient to the rest of the brain when pupil is large. Arousal does not produce changes in best frequency or spectral bandwidth. This contrasts with effects of attention on stimulus selectivity that we observed in A1, and that have been observed by others in previous work on learning and task engagement. Decreased neural variability may also support a more sensitive representation of sound in high-arousal states. The small decrease in acoustic threshold could indicate that the auditory system shifts to emphasize sound detection over sound discrimination in high-arousal states. Future experiments could contrast arousal and attention during behavior, examine whether arousal effects are dependent on learning, and test how effects on neural activity change at different levels of the auditory system.

Chapter 1: Introduction

The auditory system underlies human perception of speech, music, and environmental sounds. The auditory system transduces sound into nerve impulses, computes features such as pitch and location, and performs additional processing depending on what features of sound are relevant to behavior. In this dissertation, we contribute to knowledge of this additional processing by describing and analyzing how two aspects of behavioral state – attention and arousal – affect neural responses to sound in primary auditory cortex (A1).

Auditory processing begins in the inner ear, which converts variation in pressure waves carried by air or water into action potentials on the eighth cranial nerve. The inner ear also breaks down sound into its frequency components, such that information about different vibrational rates of sound is carried by different nerve fibers (Schnupp et al., 2011a). From the cochlea, the auditory pathway splits to travel through the several brainstem regions (cochlear nucleus, superior olivary complex, and nucleus of lateral leminiscus). Auditory processing performed in the brainstem includes computing sound localization cues that require comparison between input from both ears, and integrating sound with other sensory information (Oertel and Doupe, 2013). The auditory pathway then converges in the midbrain inferior colliculus (IC) before continuing to the medial geniculate body of the thalamus and on to auditory cortex (AC) (Fig. 1.1).

In various mammals, AC receives cortico-cortical inputs from visual cortex (Bizley and King, 2009), other auditory regions (Scheich et al., 2007), frontal cortex (Kaas and Hackett, 2000), and cholinergic input from nucleus basalis (Winer and Lee,

2007; Bajo et al., 2014), all of which may relay information about the behavioral state of the animal to AC. AC has been divided into areas based on connectivity, cytoarchitecture, and physiology (Kaas and Hackett, 2000; Hackett, 2011). These areas have been organized into hierarchies based on their physiology and laminar pattern of connections (e.g., whether connections from one area target layer IV of another area). A common motif observed across mammals is a primary core region, including A1 and other fields, projecting to secondary belt regions and (in some species) tertiary parabelt regions. Response latency and spectral bandwidth increase across this axis.

Behavioral modulation of sensory coding in auditory cortex

A sensory neuron's receptive field is the way in which sensory stimuli alter the neuron's activity. The term originally referred to the range of peripheral inputs that drove a neuron's firing, and was later extended to the properties of stimuli to which the neuron responded (Hartline, 1938). For example, the firing rate of neurons in auditory cortex depends on the frequency, location, and level of sounds in the environment, as well as the rate at which some sound features change over time (Schreiner and Winer, 2007). One representation of a receptive field is a tuning curve, which describes the neuron's mean rate of generating action potentials as a preselected feature of a stimulus varies. Other representations of receptive fields include multiple dimensions, and are estimated using methods that do not require varying only one feature of a stimulus at a time (deCharms et al., 1998; Chichilnisky, 2001).

It is often useful to consider single-neuron receptive fields as static entities determined by feed-forward processing. Understanding how neuron's response properties

change as one moves from the periphery to more central regions can lead to models of how neural circuits generate receptive fields. An example is the shift from the spot-like, center-surround receptive fields of the retina to the more complex receptive fields of primary visual cortex, including neurons that respond to lines oriented in a particular direction (Hubel and Wiesel, 1962). Within the auditory system, many studies of the emergence of tuning to cues relevant to sound location – which are not encoded directly in the cochlea – also follow this paradigm (Schnupp et al., 2011b).

Human auditory perception is modified by experiences such as early language exposure and musical training, and some form of receptive field plasticity may underlie these phenomena (Schnupp et al., 2011c). Shifts in the frequency tuning of A1 neurons have been demonstrated in animals that have learned to associate a specific frequency of tone with a shock or water reward: typically, the neuron's response to sounds at and near the frequency of the tone increases (Weinberger, 2004; Rutkowski and Weinberger, 2005). Similar tuning shifts can be induced when tones are paired with release of acetylcholine in cortex, which causes a shift in the balance of excitation and inhibition in intracortical inputs (Froemke et al., 2007). This type of receptive field plasticity is widespread in A1: at the population level, pairing acetylcholine release with a tone frequency increases the percentage of neurons tuned to the frequency (Kilgard and Merzenich, 1998). Although not as extensively studied, pairing tones with stimulation of the noradrenergic locus coeruleus can also cause shifts in frequency tuning in A1 (Martins and Froemke, 2015). In addition to shifts in frequency tuning, plasticity in cortical neurons' responses to other sound features, such as intensity and location, follow behavioral training (Schreiner and Polley, 2014).

In addition to training, responses of neurons in auditory cortex are modulated when subjects are engaged in auditory tasks (Fig. 1.2). In early experiments on this phenomena, ferrets were presented with noise bursts periodically interrupted by tones (Fritz et al., 2003). Licking during a tone was punished with a mild shock, while licking during noise was rewarded with water. Engaging in the task increased the neuronal response to sound at the frequency of the tone. Unlike classical conditioning, which produces changes in A1 tuning that can persist for days (Weinberger, 2004), these effects of task-related plasticity are dynamic: receptive fields often returned to baseline after behavior ended, and the same cell could be induced to change its receptive field in multiple ways by changing the stimulus. Rapid, reversible plasticity in the frequency tuning of A1 neurons has also been shown in tasks requiring motor activity after only one of two tone frequencies (Fritz et al., 2005), chords (Fritz et al., 2007b), and sequences of two tones (Yin et al., 2014). Subsequent experiments indicated that the sign of changes in A1 frequency selectivity depend on reward valence (David et al., 2012). When ferrets received water for licking during the tone, but were punished with a timeout for licking during noise, the neuronal response to the tone frequency decreased. This effect followed from the structure of the task. Suppressing the response to the tone increases the salience of the response to noise. Thus, in both cases, A1 changes to enhance the representation of sounds that require inhibiting motor activity to avoid punishment.

Evidence of task-related plasticity and learning in A1 challenges the picture of this region as an acoustic analyzer that extracts features for higher-level decision-making areas. In ferrets, behavioral modulation of neural activity during auditory tasks has been also been shown in IC (Slee and David, 2015) and secondary and tertiary regions of AC

(Atiani et al., 2014; Elgueda et al., 2019) using similar auditory tasks. Both IC and A1 show local changes in spectrotemporal tuning and gain during behavior. Secondary and tertiary regions of AC show greater modulation of sound-evoked responses during behavior than A1. Given that there are feedback connections between A1 and secondary and tertiary regions of ferret AC (Bizley et al., 2015) and between AC and IC (Bajo et al., 2006), it is not clear where in the ascending auditory pathway behavioral effects first appear.

Task-related plasticity in A1 enhances acoustic features carried by one type of stimulus (the foreground) to maximize its discriminability from others (the background). Depending on the reward structure of the task, this may be achieved by increasing or decreasing the response to the acoustic feature: either can enhance the contrast between the feature and others. In Chapter Two, we examine the issue of task-related plasticity in the context of a selective-attention task.

Auditory attention

Selective attention can be described operationally as the ability to respond to one class of stimulus while ignoring others. Humans use selective attention when we follow the content of one speaker's voice in a crowded room or monitor a visual scene for objects of one shape or color (Pashler, 1998). Selective attention is only one aspect of behavioral state that affects perception, and is dependent on other cognitive processes. For example, auditory attention requires separating the mixture of sound that reaches the ear into distinct sources (Shinn-Cunningham, 2008; McDermott, 2009). Sound source separation is itself a complicated problem for the brain to solve, given that natural sounds

are spread out over a wide range of frequencies. Since energy adds where the frequency content of sounds overlaps, energy from one sound can mask others.

In early experimental work on auditory selective attention (the cocktail-party problem), experimenters used headphones to present different speakers' voices to each ear, or a mixture of voices to both ears and gave listeners instructions to attend to one voice and ignore the other (Cherry, 1953). Listeners were generally unable to repeat back the content of the ignored message. Shortly after, it was demonstrated that listeners showed better comprehension of the ignored voice when it contained their name (Moray, 1959). These experiments and others suggested that although not all stimuli present at the auditory periphery are available to conscious awareness, some representations of ignored stimuli are preserved, even to the point of allowing some comprehension of high-level features such as words. In practice, however, it is often difficult to exclude the possibility that awareness of the ignored stimulus results from attentional lapses or the decision to direct attention to the unattended stimuli (Pashler, 1998).

The human auditory system uses multiple acoustic features to separate and attend to sound sources. A common strategy to identify the cues used in the cocktail-party task is to manipulate one feature of recorded voices used in the task and test whether it makes the task more difficult (e.g., whether subjects are more likely to report content from the unattended voice) (Pashler, 1998). Introducing interaural level and timing differences, which simulate differences in the location of the sound sources, make the cocktail-party task easier, as does introducing differences in vocal pitch. Another strategy, adopted from studies of spatial attention in the visual system, is to present subjects with a cue that directs attention towards a particular acoustic feature at the start of the experiment or

trial. When humans are cued to focus attention at a particular location, the speed with which they detect sounds shows small improvements, as does the accuracy of frequency discriminations of sounds at the location (Spence and Driver, 1994). Cueing subjects about which frequency to attend to similarly improves reaction times, particularly for stimuli presented at fast repetition rates (Woods et al., 2001). Temporal coherence (whether energy at multiple frequencies changes over time following the same pattern) also aids in separating sound sources and directing attention (Shamma et al., 2011).

Behavioral studies suggest that auditory attention suppresses distractors that are far from the locus of attention. In probe-signal experiments (Greenberg and Larkin, 1968), listeners are asked to detect tones in a noise masker. While the majority of noise bursts that include tones use a tone at one frequency, some contain tones at higher or lower frequencies. Listeners' error rate increases for frequencies far from the most common tone, suggesting that they have focused their attention on a narrow band around its frequency. This decrease in sensitivity has been characterized as an auditory filter whose shape resembles cochlear filters (Dai et al., 1991; Hafter et al., 2007). When subjects are required to monitor a wider range of frequencies for the target, there is a small increase in the bandwidth of these attentional filters (Hafter et al., 2007).

In Chapter Two, we used a non-human animal model to explore the neural mechanisms of auditory selective attention. There have been relatively few studies of this issue at a single-neuron level, perhaps due to the difficulty of training animals to selectively attend in a laboratory setting, and those that have been conducted used coarse contrasts between sensory modalities or sound features (for review, see Osmanski and Wang, 2015). In addition, few studies have examined effects of attention on neural tuning

for stimulus features. As early as the 1950s, it was observed that some cortical neurons in non-anesthetized, non-restrained cats were driven by sound only when the cats "appeared to be paying attention" (Hubel et al., 1959). Subsequent single-unit studies have differed in the controls used to ensure that animals are attending, with some experiments (Benson and Hienz, 1978) using a combination of rewards for responding to attended stimuli and punishment for responding to the distractor and others (e.g., Beaton and Miller, 1975) only altering reward contingencies. In addition, some early single-unit studies compared attention to auditory and visual modalities (Hocherman et al., 1976). In spite of these differences, the studies generally observed increases in neural activity evoked by the attended auditory stimulus. Attention has also been shown to modulate the correlated activity of A1 neurons, which may affect the discriminability of population responses to attended and unattended stimuli (Averbeck et al., 2006; Downer et al., 2017).

Population measurements of neural activity in humans suggest that attention modulates tuning in AC to enhance task-relevant features. When humans are instructed to attend to either high-frequency or low-frequency tones, activity in voxels in primary auditory cortex whose tuning is closer to the attended frequency increases relative to voxels far from the attended frequency (Da Costa et al., 2013). Modulation of neural tuning occurs in AC during a realistic cocktail-party task (Mesgarani and Chang, 2012; O'Sullivan et al., 2019). Mesgarani and Chang (2012) recorded neural activity in humans implanted with electrodes on the cranial surface for assessment of epilepsy. They modeled the receptive fields of neural populations as subjects passively listened to voices, then used the model to reconstruct the neural representation of the stimulus as subjects attended to one of two simultaneous voices. Although the reconstructed stimulus

contained a mixture of the attended and unattended stimuli, it more closely matched the attended voice. Compared to nonprimary auditory cortex (superior temporal gyrus), recording sites in primary auditory cortex (Heschl's gyrus) show greater selectivity for one speakers' voice and less attentional modulation, suggesting a processing hierarchy (O'Sullivan et al., 2019). Similar results were obtained using magnetoencephalography (MEG) (Ding and Simon, 2012).

Studies of the effect of attention on intrinsic neural oscillations also suggest that auditory attention enhances the response to task-relevant features of sound, in this case their temporal dynamics. Monkeys that selectively attend to rhythmic sounds show an increase in the coherence of neural oscillations in AC at the rate of the attended stimulus (Schroeder and Lakatos, 2009; Lakatos et al., 2013). The increase in coherence selectively affects neurons that match the frequency of the attended stream. This suggests that neural oscillations that are in phase with attended sound align their phase, providing a boost in neural excitability just before the attended stimulus occurs. There is also an increase in the coherence of neural activity when humans attend to one speaker in a mixture of voices (Zion Golumbic et al., 2013). Because speech contains information at multiple timescales (Rosen, 1992; Smith et al., 2002) the particular mechanisms by which this increase in coherence change neural responses to the attended voice are more complex than in the case of the rhythmic stimuli used in primate experiments. Although these experiments do not explicitly study tuning, they are broadly consistent with results above, as both show an increase in information about attended sounds in AC.

In Chapter Two, we build on previous research by comparing neural activity in A1 across attention conditions. The work that demonstrated task-related plasticity in

ferret IC and AC, described in the previous section of this introduction, compared receptive fields in behaving and passively listening animals or across tasks that differed in reward structure. However, comparing neural responses during behavior and passive listening is not the same as comparing responses across behavioral conditions in which attention is directed to different features of sound. Some of the neural mechanisms involved in selective attention may be used to perform these tasks (Fritz et al., 2007a). In the visual system, comparing neural responses across attention conditions has shown a variety of effects including multiplicative gain increases (McAdams and Maunsell, 1999) as well as effects on spatial receptive field shape and feature selectivity (Womelsdorf et al., 2006; David et al., 2008). Chapter Two examines effects of attention on the discriminability of neural responses to the attended stimulus and distractors, gain, and spontaneous activity.

Pupil as an index of behavioral state

In Chapter Three of this dissertation, we examine correlations between pupil size and various response properties of neurons in primary auditory cortex. In humans and other mammals, pupil size is controlled by the autonomic nervous system (Loewenfeld and Lowenstein, 1999; Mathôt, 2018). Although pupillary reflexes are involved in regulating the intensity and focus of images projected on the retina, pupil size also varies under constant luminance and visual fixation. In humans, such non-luminance-mediated changes in pupil size, like other measurements of autonomic activity, have been proposed as an index of various aspects of behavioral state, including fatigue (Lowenstein et al., 1963; Lowenstein and Loewenfeld, 1964), emotion (Hess and Polt, 1960; Bradley et al.,

2008; Mauss and Robinson, 2009), and cognitive load (Kahneman, 1973). Within auditory neuroscience, pupil size has recently attracted attention for two reasons: as a measurement of listening effort applicable to audiology (Winn et al., 2015, 2018), and as an indirect index of cortical neuromodulation (McGinley et al., 2015a). Because these two lines of research inspired the experiments in Chapter Three, the next two sections of this introduction describe and critique them in detail.

The term "arousal" is used in some of this literature to indicate the state of the brain associated with a particular pupil size (McGinley et al., 2015b). Other operational definitions of arousal involve measurements of autonomic nervous activity besides pupil, the transition from sleep to wakefulness, and responses to emotionally engaging as opposed to neutral stimuli (Satpute et al., 2019). Although these definitions describe overlapping physiological processes rather than a single, simple state, they are conceptually united by their attempt to measure generalized changes in brain state and receptivity to sensory stimuli (McGinley et al., 2015b). By contrast, attention involves the ability to selectively focus on particular aspects of the environment, as demonstrated in the ability to respond to perceive particular stimuli more efficiently or accurately than others (Cherry, 1953; Pashler, 1998). Selective attention and arousal overlap in some of their physiological effects: for instance, both increase firing rates and desynchronize neural activity across populations (Harris and Thiele, 2011; McGinley et al., 2015b). However, as we will show below, pupil size is a measurement of arousal state that is sensitive to a wide range of behavioral manipulations besides changing the locus of attention. We therefore suggest that physiologists can use pupil size to distinguish neural

correlates of attention and task engagement from changes in behavioral state that do not depend on processing sensory information related to a particular task.

Pupil and listening effort

Numerous experiments have tracked pupil size as humans engage in controlled behaviors (Beatty, 1982a; Mathôt, 2018). The pupil dilations and constrictions shown in these experiments are relatively small compared to constrictions in response to changes in luminance or focusing on a nearby object, but they appear across a wide variety of tasks. For example, pupil dilations have been shown as humans perform mental arithmetic (Hess and Polt, 1964), recall items from short-term memory (Kahneman and Beatty, 1966), and detect near-threshold sensory signals (Hakerem and Sutton, 1966).

During the 1960s and 1970s, psychologists developed the theory that pupil dilation tracks cognitive load (Kahneman, 1973; Beatty, 1982a). According to this theory, transient pupil dilations occur when subjects' processing resources are allocated to a task, and the amplitude of the dilation depends on both the amount of resources devoted to it and the total resources available to the subject at a particular time. Support for the theory was found in both the onset of pupil dilations and the sensitivity of their amplitude to task difficulty. For example, when subjects were asked to monitor for appearance of a visual target while adding one to a number, their pupils dilated just after being presented with the number to transform (Kahneman et al., 1967). Subjects were most likely to miss the target during the same epoch, suggesting that multitasking had put strain on their cognitive resources. When participants performed the visual-detection task alone, they showed no performance deficit and less pupil dilation.

Given that auditory processing requires cognitive resources, it is not surprising that similar relationships between pupil dilation and listening tasks have been observed. In an experiment from the same era as the one just described (Kahneman and Beatty, 1967), subjects indicated whether two tones differed in frequency. The frequency of one tone was held constant while the other varied between trials. Subjects' error rate increased as the frequency separation between the tones decreased. Examining the mean pupil diameter across multiple trials revealed that subjects' pupils reliably dilated during each trial. The magnitude of dilation was also related to the frequency separation between the tone, and closely tracked changes in error rate. The authors suggested that both pupil size and error rate were influenced by the effort required to perform the task: the increase in error rate indicated that task became more difficult, and pupil dilation revealed the additional effort that subjects exerted. A systematic review of human studies that measured pupil during auditory tasks noted associations between pupil size and a wide range of task and individual factors (Fig. 1.3) that can be related to cognitive load, reflecting the broad definition of this concept (Zekveld et al., 2018).

Several recent studies have used pupillometry to measure listening effort (i.e., allocation of cognitive resources during challenging auditory tasks). These studies take it as a given that the amplitude of pupil dilation indexes cognitive load, and use comparisons across subjects or tasks to explore factors that affect listening effort. For example, Winn et al. examined the effect of spectral resolution of speech on pupil dilation (Winn et al., 2015). Cochlear implants (CIs) provide poor spectral resolution due to constraints on the number of independent electrical channels accommodated by the cochlea. Participants with normal hearing were asked to repeat speech in which a noise

vocoder (Shannon et al., 1995) was used to reduce the number of frequency bands in the signal, simulating the output of a CI. The task consistently evoked larger and faster pupil dilation as spectral resolution declined. Pupil diameter was not tightly locked to performance: although speech intelligibility decreased with spectral resolution across the population, individual intelligibility scores were not associated with pupil measures. In addition, the effect of spectral resolution on pupil diameter was observed even when analysis was restricted to correct trials. These results imply that subjects exerted greater effort to understand speech at low spectral resolution and that the degree of effort was independent of task performance.

A separate series of studies examined listening effort for speech in noise. Experimenters measured pupil dilation in listeners with normal hearing that repeated sentences masked at both ears by stationary (constant amplitude and frequency) noise (Zekveld et al., 2010). SNR was fixed using an adaptive procedure to obtain set levels of speech intelligibility. As expected, at lower SNRs and decreased speech intelligibility, pupil showed a small increase in task-evoked dilation, suggesting greater effort. A follow-up experiment (Koelewijn et al., 2012) used a similar design to contrast three maskers: stationary noise, noise modulated by the amplitude envelope of speech, and a competing voice. When SNR was adjusted to equalize the intelligibility level across all maskers, speech-modulated and stationary noise evoked identical pupil dilations in young subjects. Surprisingly, the speech masker evoked greater pupil dilation than either noise masker, suggesting that subjects were exerting different levels of effort to achieve the same intelligibility level in the two conditions. Similar results were obtained for a group of hearing-impaired subjects (Koelewijn et al., 2014). The result suggests that listeners

exerted greater effort to achieve the same level of speech intelligibility in the presence of interfering speech, compared to interfering noise.

What accounts for the observed difference in effort? One possibility is that the competing-voice masker obscured overlapping spectral energy in the speech target more effectively than the noise maskers. Speech and speech-modulated noise, unlike stationary noise, provide temporary release from such energetic masking during dips in amplitude. The ability to "listen in the dips" increases the intelligibility of speech in fluctuating maskers as compared to stationary noise at an identical SNR (Festen and Plomp, 1990). Consistent with this effect, Koelwijn, et al. (2012) observed a 4-8 dB improvement in speech reception threshold (SRT) for speech-modulated noise and competing-voice maskers relative to stationary noise (that is, the SNR for these maskers had to be set 4-8 dB lower than the SNR for stationary noise to achieve the same error rate). However, participants also showed a small (< 2 dB) improvement in SRT for the competing-voice masker over speech-modulated noise, indicating that speech was, if anything, a less effective energetic masker than either noise type. The authors thus argue that the competing voice's effect on pupil diameter stems from informational masking: interference between two signals that cannot be explained by overlapping spectral energy.

In particular, the competing-voice task may place demands on the subject's ability to track spectrotemporal features of the target and masker, an ability thought to underlie selective attention to competing sound sources (Shinn-Cunningham and Best, 2008). The target may share spectrotemporal features with the speech masker that it does not share with speech-modulated noise, increasing the effort of separating masker from target. In addition, the competing-voice task removes spatial cues known to aid speech

intelligibility in multi-talker situations (Cherry, 1953; Bronkhorst, 2000), which may increase listening effort.

Cocktail-party tasks also suggest pupil diameter reflects demands on auditory selective attention. In one experiment (Koelwijn, et al 2014b), listeners were played two voices at separate ears and told to repeat the words of one or both speakers. Task instructions thus determined whether subjects selectively attended to one voice or divided attention between the two. In a control condition meant to minimize demands on attention, subjects were exposed to only one voice. Dividing attention evoked larger pupil dilation than attending to one voice, which in turn evoked larger pupil dilation than the control. Adding a fluctuating noise masker to both voices increased error rate and preserved the relationship between pupil dilation in the three attention conditions. Varying the location, onset time, or speaker of the attended speech stream also increased pupil dilation, suggesting that maintaining spatial and spectrotemporal regularity across a longer timescale reduced effort (Koelwijn, et al 2015).

Pupil diameter can be non-invasively and relatively inexpensively measured in humans, and is therefore an appealing option as a clinical tool to measure listening effort. Audiologists could use the degree of pupil dilation during auditory tests to choose between CI or hearing-aid signal-processing strategies even when their effect on test performance has reached a ceiling (Winn et al., 2015). However, a wide variety of states are associated with pupil dilation, suggesting that interpretation of pupil size as a measure of listening effort is not straightforward. A number of studies have linked task-evoked pupil dilations to changes in cognitive variables more specific than cognitive load, including shifts in perception of ambiguous stimuli (i.e., pupil dilates at the moment

when one's interpretation of a visual illusion changes) (Einhauser, et al 2008), lapses in attention (Wierda, et al 2012), bias in sensory signal detection (de Gee, Knappen, & Donner 2013), confidence in perceptual judgments (Lempert, Chen, & Fleming 2015), and uncertainty about the underlying state of a dynamic environment (Nassar, et al 2012; Browning, et al 2015). It is not clear if these results fit within the cognitive-load framework (or its particular application to listening effort) or represent alternative theories about what pupil encodes.

Pupil dilation in response to sensory stimuli is part of a collection of movements and physiological reactions (e.g. head-turning, reduced heart rate) that orient subjects towards an event and prepare them to respond (Mathôt, 2018). In owls, pupil initially dilates in response to repeated sounds, habituates, and dilates again when sounds at a different frequency or location are introduced, suggesting that pupil encodes novelty (Bala and Takahashi, 2000). In humans, the amplitude of pupil dilation evoked by sounds is correlated with loudness rather than the degree of salience (Liao et al., 2016). Human pupils also dilate when sequences of repeated sounds become random, but not vice versa, suggesting that pupil encodes expectations (Zhao et al., 2019). These results are challenging to the listening-effort literature, since novel, loud, and unexpected sounds do not require a great deal of subjective effort to detect.

More comparison across studies could be useful for placing results from the listening-effort literature in context, particularly for experiments or applications where pupil size is observed outside controlled behavior. An analogy might be made to the problem of reverse inference in functional magnetic resonance imaging (fMRI) (Poldrack, 2006). Many fMRI experiments use internal controls to establish that a

particular brain region is activated by a particular cognitive process (forward inference). Sometimes, data established by forward inference in one experiment is used in a separate experiment to infer from activation of the brain region that a subject is engaged in the cognitive process (reverse inference). Since brain regions may be involved in more than one cognitive process at different times, reverse inference in this crude form is not valid. However, pooling data across studies allows one to estimate the region's selectivity for particular processes. Pupil, similarly, may be engaged by different aspects of tasks to different degrees. Data on the selectivity of pupil dilation for listening effort (as opposed to other processes) or the prior probability that particular auditory tasks will engage listening effort would be helpful for understanding pupil's utility as an indicator of listening effort.

Pupil and sub-states of wakefulness

A series of recent studies in mice document correlations between pupil size and single-neuron activity in various areas of the brain, including visual and auditory cortex (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015). Measurements of pupil size are correlated with changes in the number of action potentials spontaneously emitted by neurons or evoked by auditory and visual stimuli, shared variability between pairs of neurons, and the degree to which population activity is dominated by slow, correlated firing. In a jointly-authored review, the authors of the mouse papers suggest that pupil offers a physiological measurement of sub-states of waking or arousal, analogous to but distinct from the sub-states of sleep (McGinley et al., 2015b).

Pupil size is correlated with multiple other indicators of behavioral state, some of which also predict changes in neural activity in sensory cortex. In mice, pupil is consistently dilated when animals engage in activities involved in active exploration of the environment, including running, whisking, and licking a water reward (Lee and Margolis, 2016; Stringer et al., 2019). Given that rodents' whisking is coordinated with sniffing (Deschenes et al., 2012) – another aspect of exploratory behavior – it is likely that respiration rate is also correlated with pupil size. Like pupil, locomotion is associated with changes in the excitability of single neurons in visual and auditory cortex (Niell and Stryker, 2010; Schneider et al., 2014). In auditory cortex, pupil diameter predicts more neural variability than locomotion, perhaps because pupil size can vary when an animal is not moving (McGinley et al., 2015a). However, spontaneous facial movements predict more neural variability in visual cortex than pupil size (Stringer et al., 2019).

Unlike most human studies, this literature generally measures spontaneous changes in absolute pupil size rather than dilations evoked by particular task events. On its face, the reported pupillary movements resembles human literature on spontaneous changes in pupil size during fatigue rather than listening effort, since it involves changes in pupil size that occur slowly across a large dynamic range (Lowenstein et al., 1963; Lowenstein and Loewenfeld, 1964). One study that used air puffs to induce fast pupil dilation in mice found similar increase neural activity in visual cortex as when behavioral state was allowed to vary spontaneously (Vinck et al., 2015). It is possible that the dilations recorded in listening-effort experiments are accompanied by only a subset of the physiological processes associated with spontaneous pupil movements in mice.

Pupil as an index of neuromodulatory tone

Activity in neuromodulatory centers has been proposed as a mechanism underlying observed correlations between pupil size and the level of neural activity in sensory cortex (McGinley et al., 2015b). In particular, it has been suggested that neuromodulators are released in sensory cortex at the same time that they are released in other brain regions upstream of pupil control circuits. This hypothesis suggests that pupil serves as an indirect marker of neuromodulation that affects sensory cortex.

Most available evidence for the hypothesis focuses on the locus coeruleus (LC), a source of noradrenergic input to cortex (Larsen and Waters, 2018). During behavior, LC neurons switch between bursts of stimulus-locked action potentials and relatively slow, stimulus-independent firing when task engagement wanes, suggesting that LC is involved in the switch between optimizing behavior on a task and exploring the environment for new sources of reward (Aston-Jones and Cohen, 2005). Baseline pupil size is larger when humans are engaged in exploratory behavior, consistent with the hypothesis that LC drives baseline changes in pupil size (Jepma and Nieuwenhuis, 2011).

Physiology also indicates that LC plays a causal role in pupil dilation. In humans, pupil dilation is correlated with blood-oxygen level dependent (BOLD) fMRI signal in LC (Murphy et al., 2014; de Gee et al., 2017). In macaques, activity in single LC neurons precedes pupil dilation by several hundred milliseconds (Joshi et al., 2016). Activating LC, either by electrical stimulation (Joshi et al., 2016; Reimer et al., 2016) or optogenetics (Lovett-Barron et al., 2017) induces a transient pupil dilation with a latency of ~1 second. Brief, high-frequency (50 or 333 Hz) stimulation of LC induces a more rapid dilation than tonic (< 5 Hz) LC stimulation, consistent with the hypothesis that

stimulus-locked, phasic activity in LC drives the fast, rapidly-decaying pupil dilations associated with changing cognitive load (Liu et al., 2017).

LC acts on pupil through established circuits (Fig. 1.4). Pupil size is determined by the balance of activity in two smooth muscles of the iris: the sphincter and dilator pupillae (Mathôt, 2018). The dilator is innervated by the superior cervical ganglion (SCG), a component of the of the sympathetic nervous system, while the sphincter is innervated by the ciliary ganglion (CG), a component of the parasympathetic nervous system. CG receives input from cholinergic neurons of the Edinger-Westphal nucleus (EW). LC projects to both the constriction pathway (via EW) and the dilation pathway. In rats, injecting EW with a noradrenergic antagonist plus removal of the SCG renders the pupil insensitive to LC stimulation, suggesting that LC acts on EW to dilate pupil (Liu et al., 2017). Removal of the ipsilateral, but not contralateral, SCG reduces the degree of LC-induced pupil dilation, as well as the degree to which LC-induced pupil dilation is lateralized, suggesting that LC affects ipsilateral pupil size through both the constriction and dilation pathways, but contralateral pupil dilation only via the parasympathetic pathway (Liu et al., 2017).

Pupil indexes activity of multiple neuromodulators and brain regions that act on sensory cortex, not just noradrenaline. Activity in cholinergic terminals in mouse visual and auditory cortex is correlated with pupil dilation (Reimer et al., 2016; Larsen et al., 2018), as is activity in dopaminergic (substantia nigra and ventral tegmental area) and cholinergic (basal forebrain) regions in humans (de Gee et al., 2017). Slow changes in pupil size track acetylcholine release in sensory cortex, while faster changes in pupil size track noradrenaline release (Reimer et al., 2016). Microstimulation experiments in non-

human primates suggest that additional regions involved in central control of pupil size include the superior colliculus (Wang et al., 2012), frontal eye fields (Lehmann and Corneil, 2016), and prefrontal cortex (Ebitz and Moore, 2017).

If pupil does not offer an interpretable index of a single neuromodulator or activity in a particular circuit, what is its utility to physiologists? One possibility is that pupil can be used in particular experiments to gauge the type or degree of neural variability that are associated with uncontrolled fluctuations in internal state rather than behavioral factors such as task engagement, attention, or reward. For this purpose, the redundancy between the information in pupil size and other behavioral outputs such as running and respiration rate noted above would be an advantage, since it would suggest that pupil size reflects multiple types of uncontrolled variability. Such controls are particularly well-motivated for studies of effects of behavioral effects of the response properties of neurons in auditory cortex. As noted above, number of studies have used comparisons between passively-listening and behaving animals to demonstrate behavioral effects on receptive fields in A1. However, it is possible that animals are less alert during passive listening than they are when they behave. If so, this would introduce systematic differences between the state of the animal in the two conditions unrelated to the structure of the task, which could be indicated by changes in pupil size.

The experiments described in Chapter Three analyze variability in pupil size and neural responses to sound in passively-listening ferrets, the same species originally used to demonstrate effects of task engagement on frequency tuning in A1. These experiments explore the effects of changes in pupil state on overall excitability, frequency selectivity, and sound level response threshold. These experiments therefore contribute towards the

general goal of characterizing how multiple changes in behavioral state affect the representation of sound in auditory cortex.

Figures

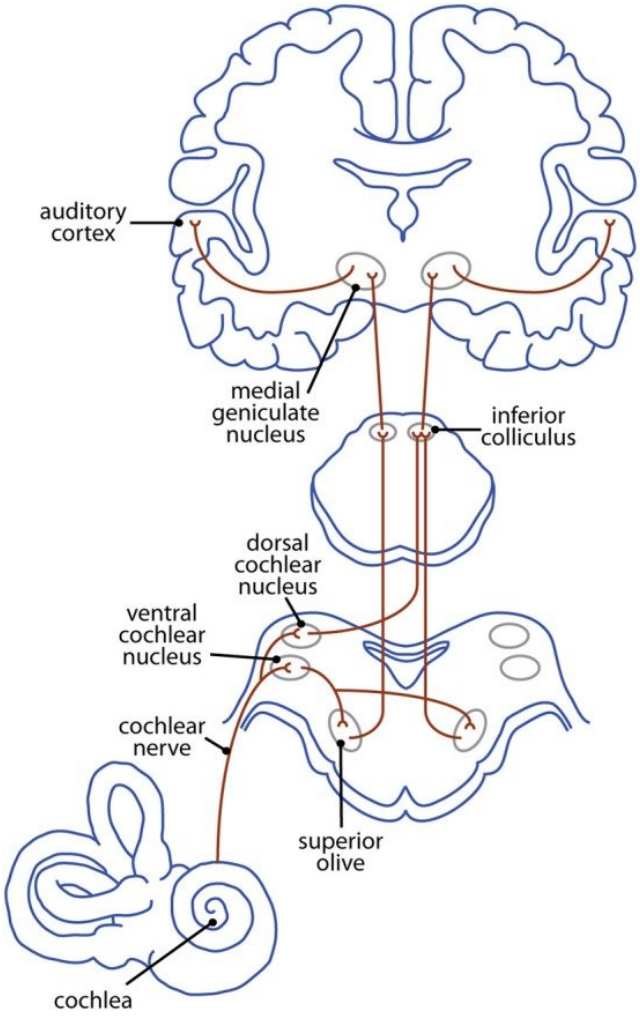

Figure 1.1: Ascending auditory system. Simplified schematic of ascending auditory system through auditory cortex. Reproduced from (Butler and Lomber, 2013).

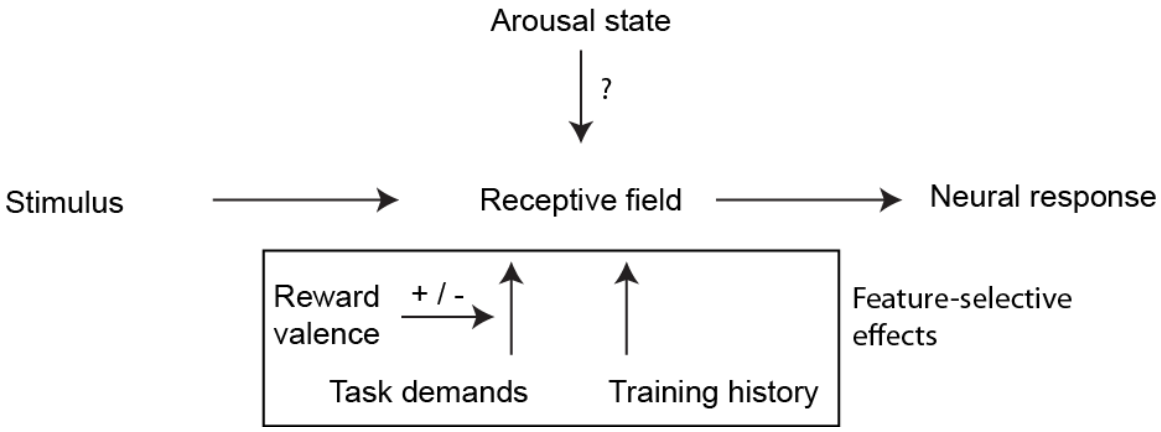

Figure 1.2: Task-related plasticity in receptive fields. Schematic of factors affecting receptive field shape during behavior.

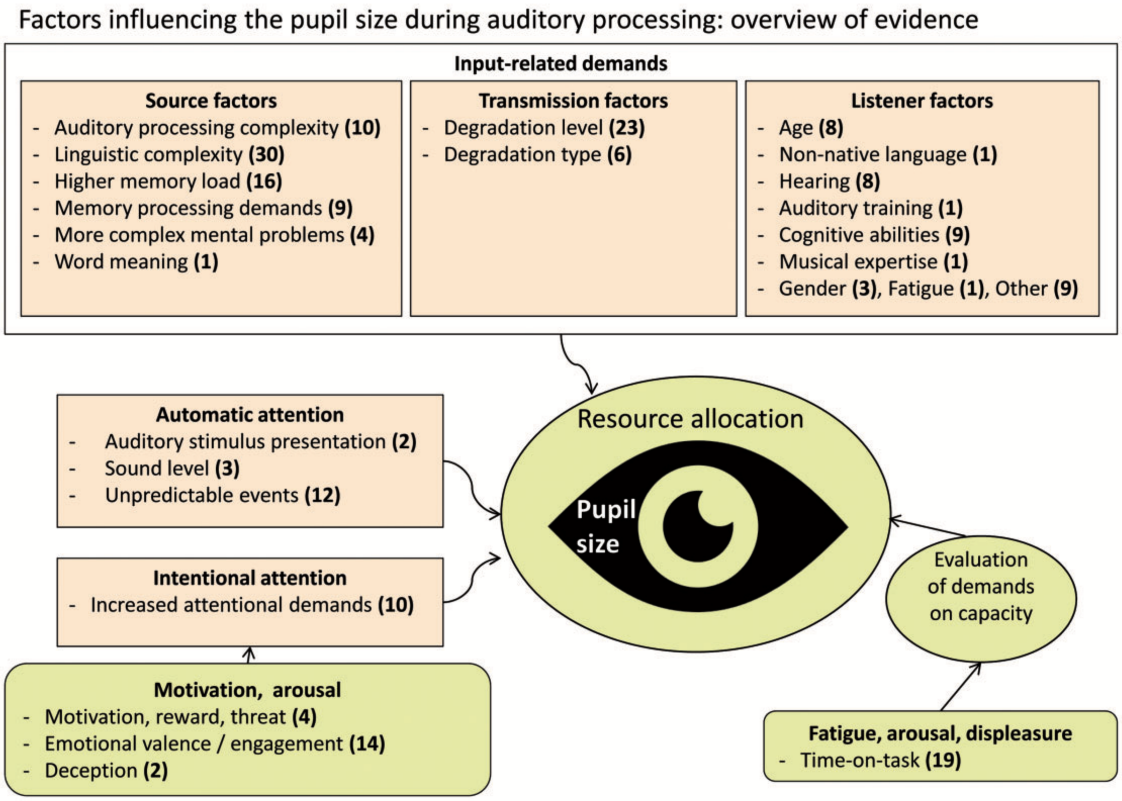

Figure 1.3: Factors influencing pupil size during auditory tasks. Results of a systematic review of individual and task-related factors linked to pupil size during auditory processing in humans. The number of publications supporting each factor is indicated in parentheses. Reproduced from (Zekveld et al., 2018).

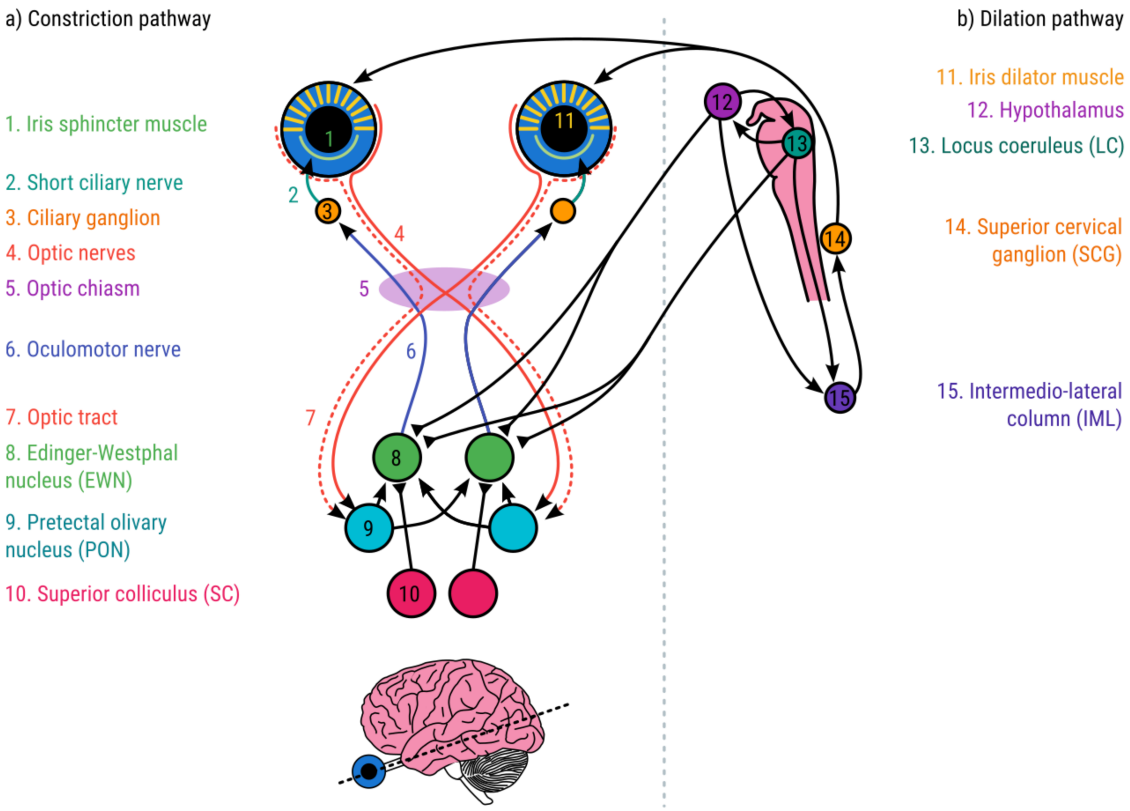

Figure 1.4: Neural circuits controlling pupil size. Schematic of neural circuits involved in reflexive and cognitive modulation of pupil size. Reproduced from (Mathôt, 2018).

This chapter was adapted from Schwartz et al. (2020) which was published in Journal of Neurophysiology 123(1):191-208. doi: 10.1152/jn.00595.2019.

Chapter 2

Focal suppression of distractor sounds by selective attention in auditory cortex

Leah P. Schwartz[1], Stephen V. David[2]

[1]Neuroscience Graduate Program, Oregon Health and Science University

[2]Oregon Hearing Research Center, Oregon Health and Science University

Acknowledgements: This work was supported by grants from the National Institutes of Health (R01 DC014950, F31 DC016204) and a fellowship from the ARCS Foundation Oregon Chapter. The authors would like to thank Henry Cooney, Brian Jones, and Daniela Saderi for assistance with behavioral training and neurophysiological recording and Sean Slee for comments on task design and data analysis. The authors declare no competing financial interests.

Abstract

Auditory selective attention is required for parsing crowded acoustic environments, but cortical systems mediating the influence of behavioral state on auditory perception are not well characterized. Previous neurophysiological studies suggest that attention produces a general enhancement of neural responses to important target sounds versus irrelevant distractors. However, behavioral studies suggest that in the presence of masking noise, attention provides a focal suppression of distractors that compete with targets. Here, we compared effects of attention on cortical responses to masking versus non-masking distractors, controlling for effects of listening effort and general task engagement. We recorded single-unit activity from primary auditory cortex (A1) of ferrets during behavior and found that selective attention decreased responses to distractors masking targets in the same spectral band, compared to spectrally distinct distractors. This suppression enhanced neural target detection thresholds, suggesting that limited attention resources serve to focally suppress responses to distractors that interfere with target detection. Changing effort by manipulating target salience consistently modulated spontaneous but not evoked activity. Task engagement and changing effort tended to affect the same neurons, while attention affected an independent population, suggesting that distinct feedback circuits mediate effects of attention and effort in A1.

Introduction

Humans and other animals are able to focus attention on one of multiple competing sounds in order to resolve details in behaviorally important signals (Cherry, 1953; Dai et al., 1991; Shinn-Cunningham and Best, 2008). Studies in humans have found that when attention is directed to one of two competing auditory streams, local field potential (LFP), MEG, and/or fMRI BOLD responses to task-relevant features are enhanced, and responses to distractor stimuli are generally suppressed (Ding and Simon, 2012; Mesgarani and Chang, 2012; Da Costa et al., 2013). However, little is known about the effect of attention when distractors compete in the same spectral band as target sounds, an important problem for hearing in natural noisy environments (Shinn-Cunningham and Best, 2008). Human behavioral studies suggest auditory attention does not uniformly suppress all distractor sounds, but instead preferentially suppresses distractors near the locus of attention (Greenberg and Larkin, 1968; Kidd et al., 2005). A small number of studies in behaving animals have found that attention improves coding of task-relevant versus irrelevant features at the population level, observed through changes in multiunit and LFP synchrony (Lakatos et al., 2013) and inter-neuronal correlations (Downer et al., 2017). As with the human studies, this work did not distinguish between features near and far from the locus of attention, and it is not known if the same mechanism operates in both cases. Sounds of similar frequency are encoded by topographically interspersed neurons in cortex (Bizley et al., 2013), and analysis at the level of single neuron responses is important for understanding their representation.

Several studies have identified changes in single-neuron activity in primary auditory cortex (A1) during behavior. Neurons can enhance or suppress responses to task-relevant spectral or spatial sound features. However, most of this work has relied on comparisons between passive listening and behavioral engagement (Fritz et al., 2003; Otazu et al., 2009; Lee and Middlebrooks, 2010; David et al., 2012; Atiani et al., 2014) or between behaviors with different structure (Fritz et al., 2005; David et al., 2012; Rodgers and DeWeese, 2014). It is difficult to attribute changes in neural activity to selective attention because other aspects of internal state also change between conditions, including arousal, effort, rules of behavior, and associated rewards. Moreover, the specific acoustic task has differed between studies, ranging among tone detection, modulation detection, tone discrimination, and tone-in-noise detection. Changing task sound features could recruit different feedback systems or require different auditory areas with specialized coding properties (Tian et al., 2001; Bizley et al., 2005), making a comparison between these studies difficult.

To link specific aspects of behavioral state to changes in auditory coding, we developed an approach that isolates the effects of selective attention to sound frequency from general task engagement (Fritz et al., 2003; Otazu et al., 2009) and behavioral effort (Atiani et al. 2009). We trained animals to perform a tone-in-noise detection task in which they heard the same sound sequences but switched attention between target tones at different frequencies. To compare effects of attention and listening effort, we manipulated target salience while requiring detection of the same target tone (Atiani et al., 2009). We recorded single unit activity in A1 during behavior. During manipulation of selective attention, we observed suppression specifically of responses to distractors

near the target frequency rather than a generalized suppression of all distractors. When task difficulty was varied, we saw changes in tonic spike rate, rather than sound-evoked activity.

The noise stimuli developed for these behaviors contain natural temporal dynamics and were designed to permit characterization of the sensory filter properties of neurons reflected in their sound-evoked activity (David and Shamma, 2013). We used these spectro-temporal receptive field models to compare the encoding properties of A1 neurons between behavior conditions. This analysis revealed changes in neuronal filter properties consistent with changes in the average spontaneous and sound-evoked activity.

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.

Animal preparation

Young adult male ferrets were obtained from an animal supplier (Marshall Farms, New York). A sterile surgery was then performed under isoflurane anesthesia to mount a post for subsequent head fixation and to expose a small portion of the skull for access to auditory cortex. The head post was surrounded by dental acrylic or Charisma composite, which bonded to the skull and to a set of stainless steel screws embedded in the skull. Following surgery, animals were treated with prophylactic antibiotics and analgesics under the supervision of University veterinary staff. The wound was cleaned and bandaged during a recovery period. After recovery (approximately 2 weeks), animals were habituated to a head-fixed posture for about two weeks.

Auditory selective attention task

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 software (Matlab). 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. Sound level was equalized and calibrated against a standard reference (Brüel & Kjær).

Three ferrets were trained to perform an auditory selective attention task modeled on studies in the visual system (Moran and Desimone, 1985; McAdams and Maunsell, 1999; David et al., 2008), in which they were rewarded for responding to target tones masked by one of two simultaneous, continuous noise streams and for ignoring catch tones masked by the other stream (Fig. 2.1). The task used a go/no-go paradigm, in which animals were required to refrain from licking a water spout during the noise until they heard the target tone (0.5 s duration, 0.1 s ramp) at a time randomly chosen from a set of delays (1, 1.5, 2, … or 5 s) after noise onset. To prevent timing strategies, the target time was distributed randomly with a flat hazard function (Heffner and Heffner, 1995). Target times varied across presentations of the same noise distractors so that animals could not use features in the noise to predict target onset.

Noise streams were constructed from narrowband noise (0.25-0.5 octave, 65 dB peak SPL) modulated by the envelope of one of 30 distinct ferret vocalizations from a library of kit distress calls and adult play and aggression calls (David and Shamma, 2013). The envelope fluctuated between 0 and 65 dB SPL, and its modulation power spectrum was low-pass with 30 dB attenuation at 10 Hz, typical of mammalian vocalizations (Singh and Theunissen, 2003). Thus the spectral properties of the noise streams were simple and sparse, while the temporal properties matched those of ethological natural sounds. To maximize perceptual separability (Shamma et al., 2011), the streams were generated using different vocalization envelopes, centered at different

frequencies (0.9-4.3 octave apart) and presented from different spatial locations (±30 degrees azimuth).

In a single block of behavioral trials, the target tone matched the center frequency and spatial position of one noise stream. It was switched to match the other stream in a subsequent block. The majority of trials (80-92%) contained a *cue tone* target with relatively high signal-to-noise ratio (SNR, 5 to -5 dB peak-to-peak, relative to reference noise). For all tones, SNR was calculated locally in period of the noise that overlapped the tone. This definition of SNR produced relatively stable performance at a given SNR. The remainder of trials contained *probe tone* targets with low SNR (-7 to -12 dB), requiring focused attention on the target stream. The exact target SNR was adjusted for each animal and target frequency so that the cue tone was super-threshold (>90% hit rate) and the probe tone was closer to threshold (70% hit rate). The large number of high-SNR cue targets provided a cue for attention and were important to maintain motivation, which flagged if animals were subjected to a large number of low-SNR probe targets. A random subset of trials (8-15%) included a *catch tone* in the non-target stream, occurring before the target and identical to the probe tone in the opposite trial block (-7 to -12 dB SNR). To avoid perceptual grouping of the target and catch tone, the interval between the two tones was jittered on each trial that contained both a target and catch. Responses to the cue and probe tones were rewarded with water (response window 0.1-1 s following tone onset). Responses to the catch tone or to the reference before the target resulted in no reward and were punished with a brief timeout (5-10 s) before the next trial. A preferential response to the probe versus catch tone indicated selective attention to the target stream (Moran and Desimone, 1985; McAdams and Maunsell, 1999; David et al.,

2008). Trial blocks began by requiring correct behavior on five trials with only the cue target to direct attention to a single stream. Cue trials were discarded from subsequent analysis.

Behavioral performance was quantified by hit rate (correct responses to targets vs. misses), false alarm rate (incorrect responses prior to the target), and a discrimination index (DI) that measured the area under the receiver operating characteristic (ROC) curve for hits and false alarms (Fig. 2.1C-D, (Yin et al., 2010; David et al., 2012)). To compute DI, each time when a target could occur was identified in each trial. The first lick during each trial was treated as a hit (response following target onset) or false alarm (response to noise at a time when a target could occur), depending on whether it fell in a bin before or after tone onset. Trials with licks prior to any possible target window were punished as false alarms but classified as invalid and excluded from behavioral analysis. The probability of a hit was computed as a function of the latency after tone onset, and the probability of a false alarm was computed for latency relative to times when targets could occur. These probabilities were then used to construct the ROC curve. A DI of 1.0 reflected perfect discriminability and 0.5 reflected chance performance.

Behavioral statistics were computed separately for the three tones (cue, probe, and catch). Target responses at the last possible time (5 sec) were discarded, as catch responses and false alarms never occur at that latency. Only sessions with above-chance performance on cue tones (DI>0.5, p>0.05, jackknife t-test) were included in analysis of neurophysiology data. During about 30% of behavior blocks, animals' behavioral state lapsed at some point (indicated by missing 5-10 cue targets in a row, usually later in the day), and the experimenter provided a reward manually following a target tone to re-

engage behavior. Typically, a single reminder trial was adequate. If animals failed to re-engage after multiple reminders (up to 10), the behavioral block was ended. These reminder trials were excluded from the analysis of neurophysiology data. Truncating all data acquired after reminder trials to control for long-term effects of the reminder increased noise in the analysis of neurophysiology data but did not significantly change any of the results observed across the neural population. Selective attention to the target stream was confirmed by larger DI for the probe tone than the catch tone across two behavioral blocks with reversed probe and catch tones.

As expected, animals were better able to report the cue tone, but also responded preferentially to the probe tone versus the catch tone (Fig. 2.1E). The combination of spatial and spectral streaming cues maximized behavioral attention effects. Animals were also able to perform tasks with only spectral separation between streams, but behavioral effects were weaker (data not shown). If only spatial cues were used, the two streams were fused, producing a strong percept of a single noise stream moving in space. This effect produced a spatial release from masking, increasing the salience of probe and catch tones and eliminating the need for selective attention. Animals were not tested on a spatial-only task because it appeared that attention would not be required for probe detection.

Variable effort task

Two animals were trained on a variant of the tone-in-noise task, in which the frequency and location of the target tone was fixed, but target SNR was varied between blocks. For this task either one or two noise streams were used (defined as above), but the

target was always masked by the same stream. In a high-SNR condition, the target tone level was well above threshold on 80-90% of trials (10 to -5 dB SNR, measured by peak-to-peak amplitude). For a small number of trials (10-20%), a low-SNR probe target (0 to -15 dB SNR, 10 dB below high SNR) was used to measure behavioral sensitivity to a near-threshold target. Detection threshold varied according to the frequency of the target tone, and probe targets were chosen to be 10 dB above threshold. In the low-SNR condition, the majority of trials used low-SNR targets, and the remainder used high-SNR targets. While the exact level of high- and low-SNR targets varied across days, they were always fixed between blocks on a given day, so that performance could be compared between identical stimuli.

Behavioral performance was assessed using DI, as for the selective attention task. A change in effort was indicated by comparing DI for the low-SNR target between low-SNR and high-SNR conditions. Greater effort in the low-SNR condition was indicated by higher DI, reflecting adjustment of behavioral strategy to detect the difficult, low-SNR targets more reliably.

Neurophysiological recording

After animals demonstrated reliable selective attention behavior (DI>0.5 for at least three successive sessions), we opened a small craniotomy over primary auditory cortex (A1). Extracellular neurophysiological activity was recorded using 1-4 independently positioned tungsten microelectrodes (FHC). Amplified (AM Systems) and digitized (National Instruments) signals were stored using open-source data acquisition software (Englitz et al., 2013). Recording sites were confirmed as being in A1 based on

tonotopy and relatively reliable and simple response properties (Shamma et al., 1993; Atiani et al., 2014). Single units were sorted offline by bandpass filtering the raw trace (300-6000 Hz) and then applying PCA-based clustering algorithm to spike-threshold events (David et al., 2009).

Upon unit isolation, a series of brief (100-ms duration, 100-ms interstimulus interval, 65 dB SPL) quarter-octave noise bursts was used to determine the range of frequencies that evoked a response and the best frequency (BF) that drove the strongest response. If a neuron did not respond to the noise bursts, the electrode was moved to a new recording depth. For the selective attention task, one noise stream was centered at the BF and at the preferred spatial location (usually contralateral). The second stream was positioned two octaves above or below BF, usually outside of the tuning curve. Occasionally, neurons responded to noise bursts across the entire range of frequencies measured, and a band with a very weak response (< 1/2 BF response) was used for the second stream. Thus, task conditions alternated between *attend RF* (target at BF and preferred spatial location) and *attend away* (target in the non-preferred stream). For the variable effort task, the noise configuration was the same, but the target was always centered over neuronal BF.

The order of behavior conditions (attend RF and attend away blocks for selective attention, low- and high-SNR blocks for variable effort) was randomized between experiments to offset possible bias from decreased motivation during later blocks (note gradual decrease in DI across blocks in Fig. 2.1D). We recorded neural activity during both behavior conditions and during passive presentation of task stimuli pre- and post-behavior. Thirty identical streams, with frozen noise carriers were played in all behavior

conditions. Of the 30 noise streams, 29 were repeated 1-2 times in each behavior block. One stream was presented 3-10 times, permitting a more reliable estimate of the PSTH response during neurophysiological recordings. Noise stimuli presented on incorrect (miss or false alarm) trials were repeated on a later trial, and a repetition was complete only when all the noise stimuli were presented on correct (hit) trials. Data from the repeated stimuli were used as the validation set to evaluate encoding model prediction accuracy (see Spectro-temporal receptive field analysis, below).

Evoked activity analysis

Peri-stimulus time histograms (PSTHs) of spiking activity were measured in each behavioral condition, aligned to reference and target stimulus onsets (Fig. 2.2). Because target and catch tones were embedded in the reference sound at random times, reference responses were truncated at the time of tone onset. We compared the PSTH during three epochs (Fig. 2.3): spontaneous (0-500 ms before reference onset), reference-evoked (0-2000 ms after reference onset, minus spontaneous) and target-evoked (0-400 ms after target onset minus 0-500 ms before target onset). A longer target window did not affect changes measured in the target response, but the short target window minimized potential confounds from motor signals associated with licking, which typically had latency > 400 ms for probe tones.

To measure changes in mean spontaneous and evoked activity for single neurons, we measured a *behavior modulation index*, the fraction change (Otazu et al., 2009) in spontaneous activity and evoked responses between behavior conditions:

$$d = \frac{\bar{r}_{\mathrm{A}} - \bar{r}_{\mathrm{B}}}{\bar{r}_{\mathrm{A}} + \bar{r}_{\mathrm{B}}} \qquad (2.1)$$

The subscripts *A* vs. *B* refer to experimentally controlled behavioral conditions (e.g., attend RF vs. attend away, active vs. passive, low SNR vs. high SNR). Significant differences between behavior conditions were assessed by a jackknifed *t*-test, and significant average changes across a neural population were assessed by a Wilcoxon sign test. We compared results of the sign test for population data to a jackknifed *t*-test, and found similar results.

To measure changes in baseline rate and response gain, we modeled the time-varying response to a stimulus in behavior condition *A*, $r_A(t)$, as the response in condition B, $r_B(t)$, scaled by a constant gain, g, plus a constant offset, *d* (Slee and David 2015),

$$r_A(t) = g\left[r_B(t) - r_{B,0}\right] + r_{B,0} + d \qquad (2.2)$$

The gain term was applied after subtracting the spontaneous rate, $r_{B,0}$, so that it impacted only sound-evoked activity. We used least-square linear regression to determine the optimal values of *d* and *g* that minimized mean squared error over time:

$$d, g = argmin \sum_t \left[r_A(t) - g\left[r_B(t) - r_{B,0}\right] - r_{B,0} - d\right]^2 \qquad (2.3)$$

Identical noise stimuli were presented in each behavioral condition, but variability in performance did not always permit presentation of a complete stimulus set in all behavioral conditions. To control for any possible difference in sound-evoked activity, these analyses were always applied to the subset of data with identical stimuli in both behavior conditions for a given experiment. For comparison of average PSTH (Figs. 2.2C, 2.6C), responses were normalized by mean noise-evoked activity in the passive condition before averaging.

We used signal detection analysis to measure neural discriminability (*d*') of task-relevant sounds, based on single-trial responses to tone and reference noise stimuli (Niwa et al., 2012),

$$d' = \frac{\bar{r}_T}{\sigma_T} - \frac{\bar{r}_N}{\sigma_N} \tag{2.4}$$

where $\bar{r}_T$ and $\bar{r}_N$ were the average response to tone and noise stimuli, respectively, and σ_T and σ_N were the standard deviation of responses across trials. For the selective attention data, *d*' was computed for the probe tone in the attend RF condition and compared to *d*' for the catch tone in the attend away condition. For the variable SNR data, *d*' was compared between the low-SNR target in the low-effort (high-SNR) and high-effort (low-SNR) blocks. Neural responses to probe and catch tones were computed as the mean spike rate during 0-400 ms following tone onset. This relatively short window avoided possible motor artifacts from licking during behavior. Reponses to noise were computed during 400-ms windows prior to probe or catch, during which time a tone could occur on a different trial.

Only data from correct trials were analyzed, although no significant differences were observed for data from incorrect trials. The majority of incorrect trials were false alarms with relatively short duration. Because stimuli were terminated immediately after a false alarm, their relative contribution to the overall data set size was limited. Probe and catch tones were relatively rare during selective attention behavior (8-20% of trials). This limitation, combined with their low SNR, made it difficult to measure reliable probe responses in some neurons. Only 43/54 neurons with at least 5 presentations of probe and target tones during behavior were included in the *d*' analysis. In addition, during early

experiments, probe tones were not presented during passive listening (7/43 neurons), and these were also excluded from the target BMI analysis (Fig. 2.3C).

Spectro-temporal receptive field analysis

Vocalization-modulated noise was designed so that the random fluctuations in the two spectral channels (noise stream envelopes) could be used to measure spectro-temporal encoding properties. The *linear spectro-temporal receptive field* (*STRF*) is a current standard model for early stages of auditory processing (Aertsen and Johannesma, 1981; Theunissen et al., 2001; Machens et al., 2004; David et al., 2009; Calabrese et al., 2011). The linear STRF is an implementation of the generalized linear model (GLM) and describes time-varying neural spike rate, $r(t)$, as a weighted sum of the preceding stimulus spectrogram, $s(x,t)$, plus a baseline spike rate, b,

$$r(t) = \sum_{x,u} h(x,u)s(x,t-u) + b \tag{2.5}$$

The time lag of temporal integration, u, ranged from 0 to 150 ms. In typical STRF analysis, the stimulus is broadband and variable across multiple spectral or spatial channels, x. Here, the stimulus is composed of just two time-varying channels, and a spectrally simplified version of the STRF can be constructed in which x=1…2 spans just these two channels. The encoding model for a single spectral band is referred to simply as a temporal receptive field (TRF, (Ding and Simon, 2012; David and Shamma, 2013)), but because the current study included two bands, we continue to refer to these models STRFs. Analytically, this simplified STRF can be estimated using the same methods as for standard STRFs. For the current study, we used coordinate descent, which has proven

effective for natural stimuli (David et al., 2007; Thorson et al., 2015). Spike rate data and stimulus spectrograms were binned at 10 ms before STRF analysis.

The ability of the STRF to describe a neuron's function was assessed by measuring the accuracy with which it predicted time varying activity in a held-out validation dataset that was not used for model estimation. The *prediction correlation* was computed as the correlation coefficient (Pearson's R) between the predicted and actual average response. A prediction correlation of R=1 indicated perfect prediction accuracy, and a value of R=0 indicated chance performance.

To measure effects of behavioral state on spectro-temporal coding, we estimated *behavior-dependent STRFs*, by estimating a separate STRF for data from each behavioral condition (attend BF vs. attend away or high-SNR vs. low-SNR). Thus two sets of model parameters were estimated for each neuron, e.g., in the case of selective attention data, filters $h_{RF}(x,u)$ and $h_{away}(x,u)$, and baseline rates b_{RF} and b_{away}. Predication accuracy was assessed using a validation set drawn from both behavior conditions, using each STRF to predict activity in their respective behavioral state. Significant behavioral effects were indicated by improved prediction correlation for behavior-dependent STRFs over a *behavior-independent STRF* estimated using data collapsed across behavior conditions. Behavior-dependent changes in tuning were measured by comparing STRF parameters, $h(x,u)$ and b, directly between conditions.

Competing behavior-dependent and –independent models were fit and tested using the same estimation and validation data sets. We used coordinate descent for fitting, which has previously been demonstrated as useful for fitting nonlinear encoding models using natural and naturalistic stimuli (David et al., 2007; Willmore et al., 2010;

Thorson et al., 2015). Significant differences in prediction accuracy across the neural population were determined by a Wilcoxon sign test.

Results

Ferrets can selectively attend among competing auditory streams

We trained three ferrets to perform an auditory selective attention task requiring detection of a tone in one of two simultaneous noise streams. Task stimuli were composed of two simultaneous tone-in-noise streams (Fig. 2.1A-B). During a single block of trials, animals were rewarded for responding to tones masked by one stream (the *target stream*) and punished for responding to the other (the *non-target*). Reward contingencies were reversed between blocks. The task therefore allowed comparison of neural responses to identical sensory stimuli while animals reported targets occurring in only one of the two streams (Moran and Desimone, 1985).

Spectral, spatial, and temporal cues were used to maximize perceptual separability of the streams. Both streams consisted of tones embedded in narrowband noise (0.25-0.5 octave) modulated by the temporal envelope of natural vocalizations (David and Shamma, 2013). To facilitate perceptual segregation, the streams differed in center frequency (0.9-4.3 octave separation), location (±30 degrees azimuth) and temporal envelope dynamics (one of 30 envelopes from a vocalization library, chosen randomly on each trial).

The task employed a go/no-go paradigm (David et al., 2012; Slee and David, 2015). Ferrets initiated each trial by refraining from licking a water spout for a random period (1-3 s). They were then simultaneously presented with the target and non-target noise streams (Fig. 2.1B). On each trial, a target tone at the center frequency and spatial location of the target stream was presented at a random time after noise onset (1-4 s). The majority (80-92%) of trials contained a high signal-to-noise ratio (SNR, 5 to -4 dB) *cue*

tone, while the remaining trials contained a low-SNR *probe tone* (-7 to -12 dB). On a random subset of trials (8-15%), the target was preceded by a low-SNR *catch tone* at the frequency and location of the non-target stream, with SNR matched to the probe tone. The specific SNR was manipulated between experiments to produce near-threshold behavior, but probe and catch tones were presented at identical SNRs within a single experiment. Licks that occurred before target tone onset (including catch tone periods) resulted in termination of the trial and punishment with a brief timeout (5-10 s). Licks that promptly followed target tone onset were rewarded with water (0.1-1 s following target onset).

A behavioral session consisted of two blocks, with the attended stream switching between blocks. The frequency, location, and level of the noise bands and the probe and catch tones remained the same across blocks. Only the reward contingencies, indicated by the frequency of the cue tone, were reversed: the target stream became the non-target, and the behavioral meaning of the probe and catch tones was reversed. The task therefore allowed comparison between responses to identical noise and tone stimuli under different internal states, a key requirement of a selective attention task (Moran and Desimone, 1985; McAdams and Maunsell, 1999).

We verified selective attention by comparing behavioral responses to the probe and catch tones. In humans, knowledge of the location or frequency of an attended sound affects the speed and accuracy with which the sound is detected relative to distractors (Greenberg and Larkin, 1968; Scharf et al., 1987; Dai et al., 1991; Spence and Driver, 1994; Woods et al., 2001; Kidd et al., 2005). We expected a similar improvement in behavioral discriminability for the target over catch tone. To test this prediction, we

calculated a discrimination index (DI), which quantified the area under the receiver-operating characteristic (ROC) curve for discrimination of each tone class from the narrowband noise background (Fig. 2.1C-D, (Yin et al., 2010; David et al., 2012)). A DI of 1.0 reflected perfect discriminability and 0.5 reflected chance performance. As expected, animals were able to report the cue tone more reliably than the low-SNR tones, but they also responded preferentially to the probe tone versus the catch tone. Mean performance of all three ferrets showed significantly greater DI for the probe tone versus catch tone and no difference between attention to left and right streams (Fig. 2.1E).

The order of behavioral conditions was randomized across days, but we considered the possibility that increased satiety might lead to a decrease in DI over the course of behavior during a single day. We did observe a decrease in overall DI between the first and second behavioral block for one animal (mean DI, animal 1: 86.4 vs. 82.7**, animal 2: 78.0 vs. 78.3, animal 3: 76.1 vs. 74.8, **$p<0.01$, jackknife *t*-test, $p>0.05$), but the random ordering of behavioral blocks controlled for this trend. We also considered the possibility of changes in DI over the course of a single behavioral block. One animal showed a trend toward decreased DI in the second half of each block (animal 1: 86.7 vs. 83.2), while two showed a trend toward increased DI (animal 2: 77.1 vs. 80.6, animal 3: 74.6 vs. 76.7). However, none of these within-block changes was significant (jackknife *t*-test, $p>0.05$), and performance was broadly stable over time.

A1 single-unit responses are selectivity suppressed for the attended distractor stream

We recorded single-unit activity in primary auditory cortex (A1) of the three trained ferrets during selective attention behavior to determine how neural activity changed as attention was switched between noise streams. For each unit, one noise stream was centered over its best frequency (BF) and was presented from a location contralateral to the recording site. The other stream was presented in a frequency band outside the neuron's frequency tuning curve (0.9-4.3 octaves from BF) and ipsilateral to the recording site. Thus, the task alternated between an *attend receptive field (RF)* condition (target stream in the RF) and *attend away* condition (target stream outside the RF). We recorded activity during both behavior conditions and during passive presentation of task stimuli (two behavior conditions: n=54, at least one behavior condition: n=94 neurons). The order of attend RF and attend away blocks was randomized across experiments to avoid bias from changes in overall motivation over the course of the experiment.

Changes in behavioral state can influence spontaneous spike rate and/or the sound evoked activity of single-units (Ryan et al., 1984; Rodgers and DeWeese, 2014). We measured the effects of selective attention by comparing spontaneous and noise-evoked activity (after subtracting spontaneous rate) between attend RF and attend away conditions. We measured the effects of task engagement by comparing activity between behaving and passive conditions. Because identical noise stimuli were played in all behavioral conditions, this controlled for any difference in sound-evoked responses. We

measured noise-evoked responses only during the periods prior to the occurrence of target and catch tones.

When animals attended to a noise stream in a neuron's RF, the response of many neurons to the noise stimuli decreased (Fig. 2.2A-B). Of the 54 neurons with data from both attend-RF and attend-away conditions, 16 showed significant changes in baseline rate and 26 showed significant changes in noise-evoked response ($p<0.05$, jackknife *t*-test). The average peristimulus time histogram (PSTH) response computed from the activity of neurons that underwent a change in either spontaneous or evoked activity followed a pattern consistent with the examples, showing no consistent change in spontaneous rate but a decreased evoked response (Fig. 2.2C). The change in evoked activity was roughly constant over time, occurring with about the same latency as the sensory response itself.

We quantified behavior-dependent changes in neural activity using a *behavior modulation index* (BMI, Eq. 2.1), computed as the ratio of the difference in neural activity between conditions to the sum of activity across conditions (Otazu et al., 2009). For selective attention comparisons, BMI was calculated as the difference between attend RF and attend away. Thus, BMI greater than zero indicated greater neural activity in the attend RF condition, and negative BMI indicated greater activity in the attend away condition. A value of 0.5 or -0.5 indicated complete suppression of responses in the attend away or attend RF condition, respectively. For the 16 neurons showing significant changes in spontaneous activity, median BMI was not significantly different from zero ($p>0.5$, sign test, Fig. 2.3A, D, shaded bars). However, for the 26 neurons showing significant changes in noise-evoked response, BMI was significantly decreased in the

attend RF condition (-0.12, p=0.005, approximately 20% suppression, Fig. 2.3B, E). A decrease in the noise-evoked response was also observed across the entire selective attention dataset (median BMI -0.06, p<0.001, n=54). Because the stimuli were designed so that only the RF stream evoked a neural response, this change reflected suppression of responses to the noise stream that masked the attended target. In addition to measuring changes in the mean evoked response, we also compared changes in the gain of noise evoked responses (Slee and David, 2015), which showed a similar suppression in the attend RF condition (median gain change -14%, p<0.001, Fig. 2.4).

There was no systematic effect of selective attention on responses to the probe tone (target during attend RF vs. catch during attend away), indicating that suppression was selective for the masking noise in the attended stream (Fig. 2.2D, 2.3C). Across the entire set of recordings with a sufficient number of probe and catch tones to measure responses in both conditions (n=36 neurons with at least 5 presentations of probe and catch tones in passive and both active conditions), the tone response changed significantly between attention conditions in 12 neurons, but median BMI was not significantly different from zero (p>0.5, Fig. 2.3F).

These attention-related effects were distinct from changes in neural activity related to task engagement. When compared to passive listening, the median spontaneous spike rate and noise evoked response both increased significantly (Fig. 2.3G-I, spontaneous: median BMI 0.17, p=0.002; noise-evoked: median 0.11, p=0.005; probe-evoked: median 0.14, p=0.09; sign test). Because the magnitude of spontaneous rate changes was the same for both attention conditions, there was no difference between attend RF and attend away conditions (Fig. 2.5A). Thus, the suppression of noise-evoked

activity between selective attention conditions contrasted with a general increase in spontaneous activity and excitability during task engagement.

The effects of selective attention and task engagement varied substantially across neurons. To better understand the interplay of these effects across the population, we compared the BMI for the noise-evoked response between behavior (averaged across attention conditions) and passive listening across neurons to the BMI between attention conditions (Fig. 2.5B). We found no correlation between these effects (r=0.031, p>0.5, jackknife t-test), consistent with a system in which effects of selective attention and task engagement operate on different subsets of A1 neurons. We also considered the possibility that the magnitude of behavior effects might depend on tuning bandwidth or the recording depth in cortex. However, no significant relationship was observed (data not shown).

Distractor suppression enhances neural discriminability of target versus distractors

We hypothesized that if responses to the distractor noise were suppressed relative to the probe tone when attention was directed to the neuronal BF, then the neural response to the noise versus tone-plus-noise stimuli should be more discriminable when attention was directed to the RF (Fig. 2.5C). To assess discriminability of these responses, we computed a neurometric *d*' for discrimination of the RF (probe/catch) tone from distractor noise, based on single-trial neural responses during attend RF vs. away conditions (Fig. 2.5D-E). Previous work has used this approach to measure improvements in neural discriminability following engagement in a temporal modulation detection task (Niwa et al., 2012). Across all behavioral sessions, neural *d*' increased slightly between

the two attention conditions (median change 0.03, p=0.02, sign test). However, the selective attention task was difficult for animals to perform during neurophysiological recordings, and behavioral performance varied between experiments. If we considered only sessions in which animals showed a significantly greater DI for the probe target than for the catch tone (p<0.05, jackknife t-test), this subset of neurons showed a greater increase in *d'* in the attend RF condition (median change 0.1, p=0.02, sign test, Fig. 2.5E). This larger change in *d*' suggests that selective attention improves neural discriminability in A1 during the tone-in-noise detection task, and the degree of improvement depends on the animal's performance.

A1 spontaneous activity, but not gain, is modulated by changes in task difficulty

Previous work has suggested that behavioral effort, driven by changing task difficulty, can also modulate neuronal activity in A1 (Atiani et al., 2009) and other sensory cortical areas (Chen et al., 2008). To assess the effects of changing effort in the current tone-in-noise context, we recorded A1 single-unit activity during a variant of the tone-detection task. In this case, the target stream was fixed, but the task varied from easy to hard by changing the SNR of the target tone relative to the noise stream from high (10 to -5 dB) to low (0 to -15 dB) between behavioral blocks (Fig. 2.6A). To probe differences in effort associated with changes in difficulty, the high-SNR condition included a small number of low-SNR targets (10-20%, (Spitzer et al., 1988)). We could then assess listening effort by comparing performance for identical low-SNR targets between the two conditions. In both animals, DI was consistently higher when the low-

SNR target was more likely to occur, consistent with greater effort during blocks when target detection was more difficult (Fig. 2.6B, (Spitzer et al., 1988)).

We recorded from 88 neurons from two ferrets in both effort conditions and from 122 neurons in at least one effort condition. Mean PSTHs showed an increase in spontaneous activity during low- versus high-SNR blocks but no change in the average response evoked by noise or targets (Fig. 2.6C-D). The spontaneous rate of 47/88 units was significantly modulated by task difficulty ($p<0.05$, jackknife t-test). Among these cells, the spontaneous rate showed a significant decrease during low-SNR blocks (median BMI -0.16, $p=0.008$, sign test, Fig. 2.7A). Many neurons also showed an effect of task difficulty on the evoked responses to noise ($n=51$) or tones ($n=35$). However, there was no consistent average change in either sound-evoked response (Fig. 2.7B-C, $p=0.4$ and 0.09, respectively, for noise and tone responses). As in the case of selective attention, we observed increases in spontaneous activity and sound-evoked responses during behavior relative to passive listening (Fig. 2.7D-F). Thus, task engagement effects were similar to those of the selective attention task, but unlike selective attention, changing task difficulty impacted average spontaneous rate rather than the noise-evoked response (Fig. 2.7G).

As in the case of selective attention, effects of task difficulty varied substantially across neurons. To better understand the interplay of task engagement and effort across the neural population, we compared the change in noise-evoked activity during task engagement (active versus passive) to the change between difficulty conditions (Fig. 2.7H). In this case, we observed a significant negative correlation in BMI ($r=-0.45$, $p=0.004$). Neurons whose response increased during task engagement tended to be

suppressed in the more difficult, low-SNR condition. Conversely, neurons whose response decreased during engagement tended to produce an enhanced response in the low-SNR condition. Thus, although no consistent effects of effort on average evoked activity were observed, neurons affected by changes in difficulty were the same as those affected by task engagement.

The pattern of changes in neural activity associated with changes in task difficulty did not suggest an obvious impact on neural discriminability. For a complete comparison with the selective attention data, however, we also tested whether engaging in the more difficult, low-SNR task increased neural discriminability of target versus noise (Fig. 2.7I). We used the same measure of *d*' as for the selective attention data. Across the entire set of 88 neurons in the variable SNR dataset, we observed a small increase in *d*' (median change 0.04, p=0.05, sign test). When we considered only the subset of neurons for which DI for the low-SNR tone significantly improved in the low-SNR versus high-SNR condition, we observed a trend toward a greater increase in *d*', but this change was not significant (median change 0.07, p=0.06). Thus, the change in effort may be accompanied by increased neural discriminability between target and noise, but any effects were weaker than in the selective attention data.

Selective attention and task-difficulty effects are reflected in neuronal filter models

To investigate the effects of changing behavioral state on sensory coding in more detail, we computed spectro-temporal receptive field (STRF) models for each single unit, fit using activity evoked by the distractor noise under the different behavioral conditions (Fig. 2.8A). The linear STRF typically describes spectro-temporal tuning as the weighted

sum over several channels of a broadband stimulus spectrogram (see Eq. 2.5 and (Aertsen and Johannesma, 1981; Thorson et al., 2015)). For vocalization modulated noise, we employed a spectrally simplified version of the STRF that summed activity over just the two spectral channels that comprised the noise stimulus. Because one noise band was positioned outside of the neuron's receptive field, the STRF typically showed tuning to only one of the two stimulus channels (see examples in Fig. 2.8A).

To investigate the interaction between attention and spectro-temporal tuning, we calculated separate STRFs for data from each behavior condition. Model performance was assessed by measuring the accuracy (correlation coefficient, Pearson's R) with which the STRF predicted the neuron's time-varying spike rate. Effects of behavioral state were identified by comparing the predictive power of these behavior-dependent STRFs to performance of behavior-independent STRFs, for which a single model was fit across all behavior conditions. A neuron was labeled as showing a significant effect of attention if the behavior-dependent STRF predicted neural responses significantly better than the behavior-independent STRF ($p<0.05$, permutation test). Prediction accuracy was measured using a held-out validation dataset, so that any difference in prediction accuracy could not reflect overfitting to noise (Wu et al., 2006; David et al., 2009). Identical noise sounds were presented during passive listening and different task conditions. Thus, each STRF for a single neuron was estimated using identical stimuli, avoiding potential stimulus-related bias between estimates (Wu et al., 2006; David et al., 2007).

For the selective attention data, the behavior-dependent STRF performed significantly better in 26/54 neurons, and the mean prediction correlation was

significantly greater for the behavior-dependent model (mean R=0.31 vs.0.25, n=54, p<0.0001, sign test, Fig. 2.8B). We then compared the fit parameters of STRFs for neurons showing a significant improvement for the behavior-dependent model. For these neurons, peak gain of the STRF's temporal filter was lower for the attend RF versus attend away condition (Fig. 2.8C-D, p=0.007, permutation test). Baseline firing rate, which would reflect a change in spontaneous rate, did not change between attention conditions (p=0.3). Thus, we observed changes in STRFs that were consistent with changes in mean spontaneous and evoked firing rates (Fig. 2.3E). Moreover, there was no consistent change in the shape of the temporal response (Fig. 2.8C), indicating that selective attention primarily affected the magnitude of A1 noise-evoked responses, but not temporal tuning.

For the variable SNR task, we also observed improved prediction accuracy for the behavior-dependent model in a subset of neurons (51/88), and an overall improvement in mean prediction accuracy (mean R=0.34 vs. 0.29, n=88, p<0.0001, permutation test). For neurons with significantly better behavior-dependent STRFs, baseline firing rate increased during the less difficult, high-SNR task (p=0.02, permutation test), and there was no change in the STRF's temporal filter (p>0.5, Fig. 2.8E-F). Thus, as in the case of the selective attention data, the behavior-dependent changes observed in mean spike rate were reciprocated by changes in the STRF baseline, again with no systematic change in temporal filter properties.

Discussion

This study demonstrates that auditory selective attention suppresses the responses of neurons in primary auditory cortex (A1) to distractor sounds that compete with an attended target. In nearly 50% of A1 neurons, responses to distractor noise centered at the frequency of a target tone were suppressed relative to distractors in a different frequency band. Responses to the target tone were not suppressed, leading to an improvement in its neural detection threshold. Selective attention effects differed from those of task engagement and changing task difficulty, which produced more systematic changes in spontaneous rather than evoked activity and tended to affect a different population of A1 neurons.

A primary goal of this study was to isolate mechanisms producing task-related plasticity previously reported in A1 when animals engaged in auditory behavior (Fritz et al., 2003; Lee and Middlebrooks, 2010; David et al., 2012; Niwa et al., 2012). The current data demonstrate that task-related effects can in fact be broken down into components that reflect task engagement, selective attention and effort. These behavior-related changes have a similar total magnitude to those reported following task engagement in A1 (Fritz et al., 2003; Niwa et al., 2012). The same approach can be applied to measure the composition of behavior-related changes in more central belt and parabelt areas, where overall behavior effects are generally larger (Niwa et al., 2013; Atiani et al., 2014; Tsunada et al., 2016).

Mechanisms of auditory selective attention

Distractor suppression may be a general strategy used by the auditory system to perform a variety of tasks, including those that require selective attention. At face value, these results are inconsistent with human studies of attention during streaming of simultaneously presented sounds. When human subjects attend to one of two speech or non-speech streams, the neural representation of the attended stream is enhanced over the non-attended stream (Ding and Simon, 2012; Mesgarani and Chang, 2012; Da Costa et al., 2013). Local field potential and fMRI recordings in non-human primates show similar enhancement for selective attention between two auditory streams (Lakatos et al., 2013) or auditory versus visual streams (Rinne et al., 2017).

Several factors could explain this difference from the suppression observed in the current study. First, previous studies used field recordings that sum the activity of large neural populations, and the contribution of single neuron activity to these signals is not well understood. For example, the magnitude of evoked field potentials can be influenced the synchrony of local neural populations, independent of their firing rate (Telenczuk et al., 2010). Thus, a change in the amplitude of one signal does not necessitate a change of the same sign in the other. Second, several of the human studies focused on non-primary belt and parabelt areas of auditory cortex that may not undergo the same changes as A1. Finally, and perhaps most importantly, these differences may be explained by task structure. In the tone-in-noise task used in the current study, the noise does not provide useful information to the subject performing the task. A strategy of suppression thus may be helpful for enhancing contrast with the masked target tone (Durlach, 1963; Dai et al., 1991; de Cheveigné, 1993). Streaming speech, in contrast, requires a much different

listening strategy in which the semantic content of the signal must be encoded rather than suppressed. An alternative task in which animals must detect specific features in the noise stream (e.g., modulation patterns) may reveal enhanced responses similar to those observed in the speech studies. Neural populations encoding targets and spectrally similar maskers are difficult to isolate in large-scale field recordings, which may explain why the effects of attention on masker stimuli have not been characterized previously.

The distractor suppression we observed may be a neural correlate of a phenomenon observed in psychophysical studies of tone-in-noise detection. Humans can attend to a narrow frequency band surrounding a tone target: when listeners are cued to expect a tone at a given frequency, their ability to detect tone targets more than one critical bandwidth from the expected frequency dramatically decreases, effectively attenuating off-target sounds by 7 dB (Greenberg and Larkin, 1968; Scharf et al., 1987; Dai et al., 1991). More generally, the hypothesis that the auditory system suppresses neural responses to predictable distractors in order to amplify target signals is widespread in studies of psychoacoustics (Durlach, 1963; de Cheveigné, 1993; Shinn-Cunningham and Best, 2008). We observed suppression of responses to a ½-octave noise distractor during a tone detection. This suppression may reflect sharpening of frequency tuning curves when attention was shifted into the neuron's RF. If so, then the noise suppression would increase the neural representation of the tone target at the expense of sounds at nearby frequencies.

Some aspects of our experiment suggest caution before adopting this interpretation. First, our acoustic stimuli sampled frequency tuning very sparsely and did not permit direct measurement of changes in frequency tuning bandwidth across attention

conditions. Second, it is not certain that animals used frequency as the dominant cue to direct their attention, since our stimuli also provided spatial and envelope cues to distinguish the attended and non-attended stream (Nelson and Carney, 2007). Specific interactions between attention and spectral coding can therefore only be resolved by future studies that require attention exclusively to spectral cues and probe a larger range of stimulus frequencies.

Separability of engagement and attention effects

Previous studies of tone-detection behavior have identified changes in the selectivity of A1 neurons specific to the frequency of a target tone (Fritz et al., 2003; Atiani et al., 2009; David et al., 2012; Kuchibhotla et al., 2017), but it has remained unclear how much these changes in neural tuning reflect selective attention to the target frequency versus more global processes of task engagement. Here we have isolated these effects and shown that selective attention produces frequency-specific suppression of responses evoked by distractors. Task engagement also produced changes that were independent of the locus of attention. Engagement was equally likely to produce enhancement or suppression of sound-evoked activity. This overall stability of average evoked activity suggests that changes in cortical network activity are tempered by a homeostatic mechanism that maintains stability in the level of spiking activity across auditory cortex. The selective suppression of noise responses at the locus of attention, thus, may be accompanied by enhancement at non-attended location (Fritz et al., 2005). Changing behavioral effort also influenced noise-evoked responses in A1. However, effects were not consistent, and a change in effort was equally likely to increase or

decrease responses. Instead, greater behavioral effort lead to a decrease in spontaneous spike rate. The distinct effects of attention and effort on evoked versus spontaneous activity suggest that different modulatory circuits mediate these changes. Task engagement and adjusting effort could reflect large-scale changes in brain state that do not depend on the acoustic features of task stimuli. The influence of these global state variables may arise from circuits that mediate effects of arousal (Issa and Wang, 2008; McGinley et al., 2015b). In contrast to changes in global state, selective attention requires differential processing of acoustic features, and its effects cannot be uniform across the auditory system. Consistent with a system containing distinct global versus local modulatory top-down circuits, the neuronal populations affected by selective attention and task engagement are not correlated.

While we did observe a correlation in the magnitude of effects of task engagement and effort, there was substantial additional variability of behavioral effects across neurons that could not be explained by tuning properties or recording depth. Future studies that identify the location of neurons the cortical circuit more precisely, either by genetic labels (Natan et al., 2015; Kuchibhotla et al., 2017) or network connectivity (Schneider et al., 2014) may explain more of this variability. These approaches may also be used to confirm whether engagement and effort effects derive from the same source.

Effects of selective attention and effort could emerge at different stages of the auditory network. Multiple populations of inhibitory interneurons have been implicated in behavioral state modulation in A1, and signals reflecting different aspects of behavioral state could arrive through distinct inhibitory subpopulations (Pi et al., 2013;

Kuchibhotla et al., 2017). These signals could also arrive in different auditory brain areas. Engaging in a tone detection task changes activity in the inferior colliculus (IC), an area upstream from A1 (Slee and David, 2015), but it is not known if selective attention modulates IC activity. Studies comparing the same tasks across brain areas are critical for determining where behavior-mediated effects emerge in the auditory processing network. More generally, these results indicate that multiple aspects of task structure can influence activity in sensory cortical areas. Control and monitoring of behavioral state (arousal, reward, motor contingencies, attention) is required to assess the effects of a desired behavioral manipulation (David et al., 2012; Baruni et al., 2015; Luo and Maunsell, 2015; McGinley et al., 2015a).

Impact of behavioral state on neural coding

The changes in sound evoked activity associated with selective attention support enhanced neural discriminability of target tone versus distractor noise in A1. The absence of significant suppression in the target response alone does not imply a selective suppression of distractor responses. However, the increase in neural discriminability when attention is directed into the receptive field does indicate that any suppression is stronger for the distractor, increasing the difference in neural response between the two sound categories. Improvements in neural discriminability with similar magnitude have been observed previously in A1 following engagement in auditory detection tasks (Ryan et al., 1984; David et al., 2012; Niwa et al., 2012). Some early single-unit studies in monkey auditory cortex also support the view that enhanced responses occur selectively for stimuli that carry task-relevant information. In tasks that required discrimination

between spectral (Beaton and Miller, 1975) or spatial sound features (Benson and Hienz, 1978), neural responses were enhanced to the sound requiring a behavioral response. Moreover, activity in auditory cortex can explicitly encode behavioral choice (Bizley et al., 2005; Niwa et al., 2012; Tsunada et al., 2016). Thus, across several studies, a model has emerged in which neural discriminability of task-relevant stimulus features increases, and that emergent representation feeds directly into behavioral decisions. The parallel between neural signaling and behavioral output encourages a straightforward conclusion that changes in behavioral state serve primarily to enhance coding of behaviorally relevant categories. While a change in *d*' in the range 0.05-0.1 is not extremely large for a single neuron, this value represents the average increase per A1 neuron. Effects of this size can be substantial when compounded across an entire neural population (Shadlen and Newsome, 1998).

However, in the current study, the benefit of changes in neural coding is not always so clear. The shift in spontaneous rate associated with listening effort did not produce significant enhancement in neural discriminability. Changes in spontaneous activity that do not enhance discriminability have been reported previously, following switches between tasks that vary in difficulty (Rodgers and DeWeese, 2014). Studies involving switching targets between auditory and other sensory modalities have also reported mixed results, sometimes finding enhanced coding of the attended modality (O'Connell et al., 2014) and sometimes not (Hocherman et al., 1976; Otazu et al., 2009). The cortex contains a rich diversity of circuits for learning new behavioral associations and adapting to new contexts (Fritz et al., 2010; David et al., 2012; Jaramillo et al., 2014). An architecture that supports flexibility and multiplexing of behaviors likely

imposes additional constraints on behavior-related changes beyond enhanced sensory discriminability.

The analysis of behavior-dependent STRFs revealed task-related effects consistent with the changes in PSTH response gain between selective attention conditions and changes in spontaneous between variable SNR conditions. The STRF analysis also allowed us to identify any possible changes in temporal filter properties between behavior conditions. However, temporal response properties were largely stable. Thus, while we do observe changes in spectral tuning, top-down behavioral signals do not affect temporal tuning in the current task. It should be noted that the current task did not require attention to specific temporal features, and a task requiring discrimination of temporal features, such as modulation detection or discrimination, might have a different effect (Fritz et al., 2010; Niwa et al., 2012).

Comparison to studies of visual selective attention

Noise suppression may be viewed as a mechanism to bias competition between neural representations of the tone target and masking noise. Analogous effects have been observed in the primate visual system (Spitzer et al., 1988; Desimone and Duncan, 1995; Connor et al., 1996; Reynolds et al., 1999). When macaques attend to one of two stimuli in a the receptive field of a visual cortical neuron, the neural response shifts to resemble the response to the attended stimulus presented in isolation (Connor et al., 1996; Reynolds et al., 1999). Thus, when attention is directed within a visual receptive field, the spatial receptive field effectively shrinks, and responses to stimuli outside the locus of attention are attenuated. Similar effects are observed across visual cortex, growing

progressively greater in magnitude across areas V1, V2 and V4 (Motter, 1993; Luck et al., 1997). This narrowing of tuning has been modeled as enhanced surround inhibition (Sundberg et al., 2009). In the auditory system, stimulus bandwidth can be viewed as a dimension in sensory space (bandpass noise versus very narrowband tones), analogous to retinotopic space (Schreiner and Winer, 2007). By the logic of the current task, attention within the A1 receptive field is directed to narrowband versus broadband stimuli. A narrowing of spectral tuning bandwidth that would produce the distractor suppression reported here may be analogous to the shrinking of visual spatial receptive fields around the locus of spatial attention. The average BMI of -0.12 for noise responses in A1 falls between the magnitude of spatial attention effects in areas V2 and V4 measured using a similar statistic (Luck et al., 1997).

Simultaneous population recordings from V4 during selective attention behavior have revealed a decrease in noise correlations between pairs of neurons that encode stimuli at the locus of attention (Cohen and Maunsell, 2009), and similar effects were recently reported in A1 (Downer et al., 2017). Because the current data were collected serially from single neurons, it was not possible to measure inter-neuronal correlations. The effect of auditory selective attention on neural population activity in this sensory context remains an open question for future studies.

Figures

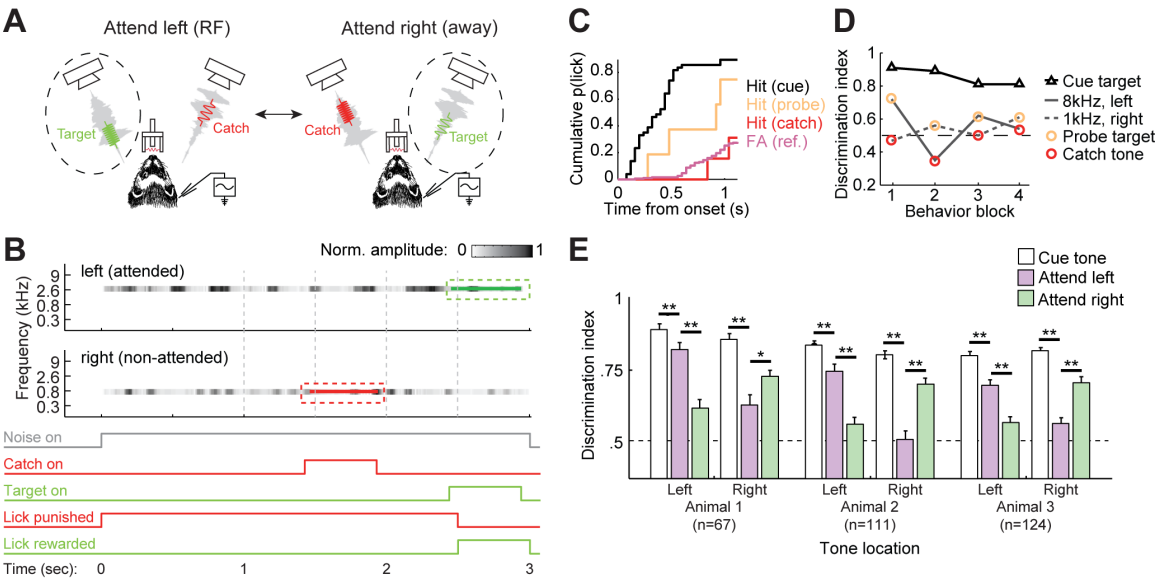

Figure 2.1: Selective attention behavior in ferrets. **A.** Configuration of selective attention behavior. Head-fixed ferrets were presented with simultaneous, continuous noise streams from two locations 30 degrees left and right of midline. A target tone was presented at the center frequency and location of the attended noise stream 1-5 s after noise onset. Animals were rewarded for licking following target onset. Responses prior to target onset were punished with a short timeout. **B.** Top shows spectrogram representation of stimuli during a single selective attention trial. Two narrowband (0.25 octave) noise streams with different center frequency were presented from different spatial locations (3000 Hz, left; 700 Hz, right). Noise level fluctuated from quiet (white) to loud (black) with dynamics drawn from natural vocalizations. The target tone (purple) appeared at a random time, centered in the attended stream (in this case the high frequency stream). To control for selective attention, a catch tone (green) appeared in the non-attended stream on a minority (8-15%) of trials. Bottom shows the time course of noise, catch and target

stimuli, as well as the time course of false alarm and hit windows during the same trial. **C.** Cumulative probability of hit or false alarm following onset of each task stimulus during one behavioral block. False alarms were measured during windows when a target could occur (vertical lines in B). **D.** DI was computed as the area under the receiver operating characteristic (ROC), calculated from lick probability curves in D. for each target and catch tone during four sequential behavioral blocks. Between blocks, the low-SNR probe target alternated between 8 kHz/contra and 1 kHz/ipsi, and was reversed with the catch tone. DI was higher for the probe than the catch tone in all blocks, consistent with shifts in selective attention. Trial block order (attend RF versus attend away) was randomized during neurophysiology experiments to prevent any bias from changes in motivation, which could produce a gradual decrease in DI across blocks. **E.** Bar plot shows average discrimination index (DI) for each tone category (cue, probe, catch), broken down by animal and attended location, indicating the accuracy with which animals reported the presence of a tone versus the noise. DI was consistently highest for a tone with high signal-to-noise ratio (SNR, -4 to 5 dB peak relative to noise) used to cue animals to the attended location. DI was lower for low-SNR probe targets (-7 to -12 dB), but consistently greater than for low-SNR catch tones, indicating that animals consistently allocated attention to the target stream (**$p<0.01$, sign test).

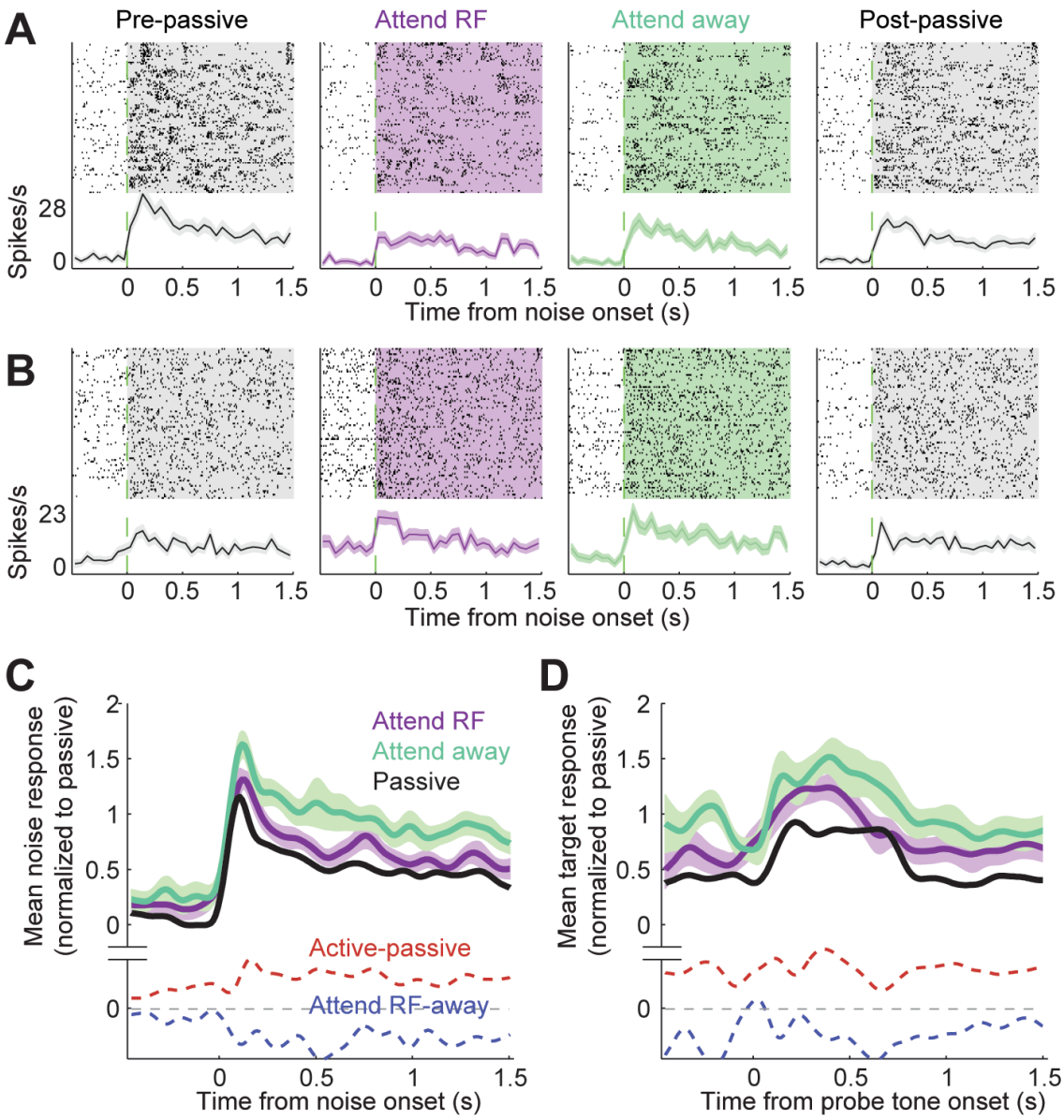

Figure 2.2: Sound-evoked activity during selective attention behavior. **A.** Comparison of spike raster plots and average PSTH responses to reference noise for one neuron, aligned to trial onset in different behavior conditions. Evoked activity was suppressed when the animal engaged in the task (passive versus attend RF or attend away). When attention was directed toward the noise stream that falls in the neuron's receptive field (attend RF) responses were further suppressed, relative to attention directed to the stream out of the

RF (attend away). Behavioral modulation index (BMI, Eq. 2.1) for attend RF versus away is -0.17. **B.** Responses of a second neuron are enhanced during task engagement, but again show relative suppression in the attend RF versus attend away condition (BMI -0.24). **C.** Average PSTH response to distracter noise across significantly-modulated units (34/54 units) in the different behavior conditions ($p<0.05$ spontaneous or noise response change, Bonferroni corrected jackknife *t*-test). Dashed lines show the average PSTH difference between active and passive conditions (red) and between attend RF and attend away (blue). Spontaneous rate increases during task engagement; reference-evoked activity is suppressed for attend RF versus attend away conditions. **D.** Average PSTH response aligned to probe target onset in the three conditions reflects pre-target differences in noise-evoked activity, but the magnitude of the target-evoked response does not change.

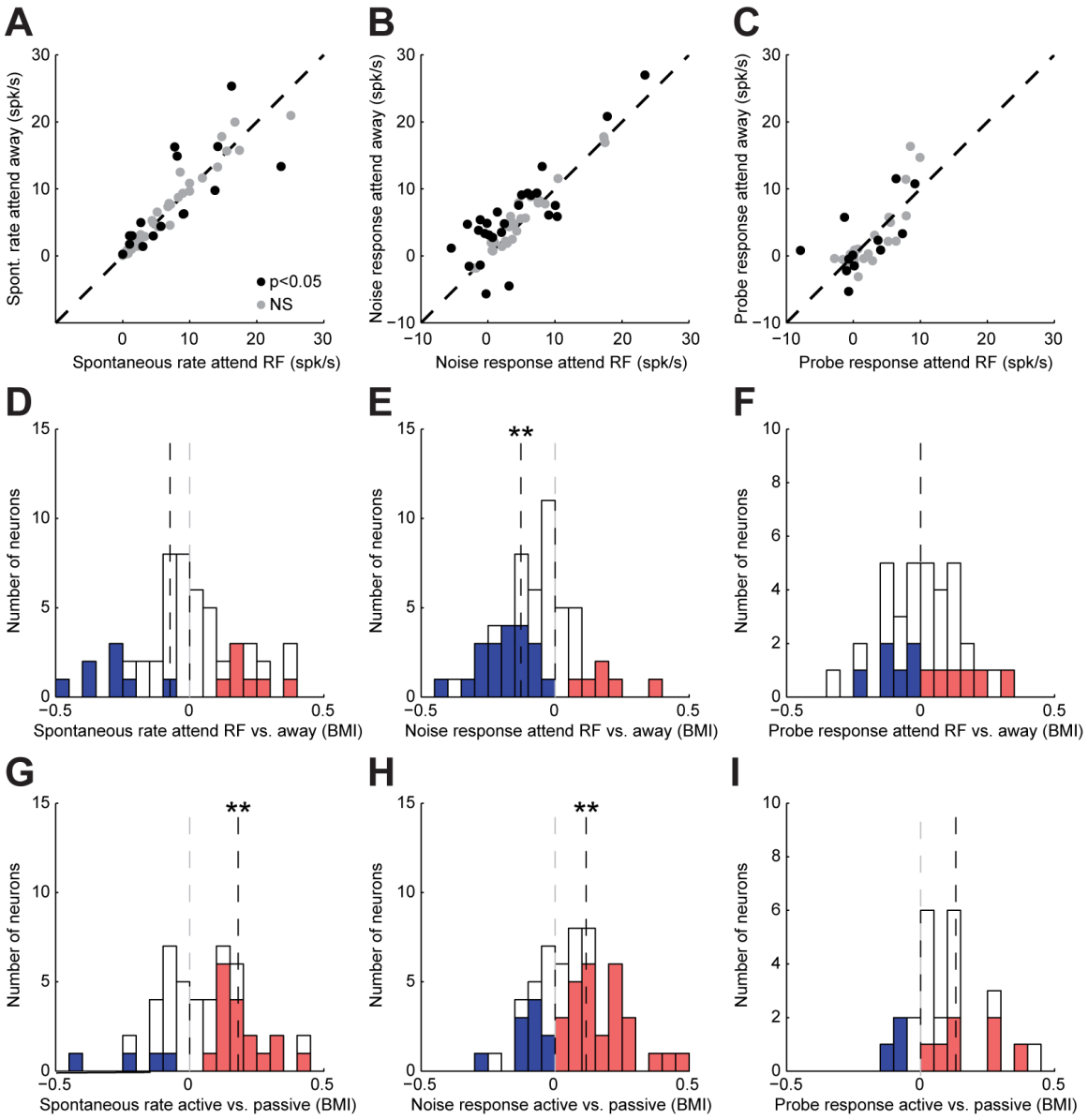

Figure 2.3: Summary of selective attention and task engagement effects on A1 evoked activity. **A-C.** Scatter plot of average spontaneous rate (A), noise-evoked response (spontaneous rate subtracted, B), and target-evoked response (noise-evoked response subtracted, C) for each A1 neuron between attend RF and attend away conditions. Black circles indicate neurons with a significant difference between attention conditions for any of the three statistics ($p<0.05$, jackknifed t-test with Bonferroni correction). Target data

are shown for the subset of neurons with adequate presentations of probe and catch tones (n=36/54). **D-F**. Histograms show behavioral modulation index (BMI, Eq. 2.1), reflecting fractional change in spontaneous rate, noise-evoked response, and target-evoked response between attention conditions for each neuron. Neurons with a significant increase are indicated in red, and those with a significant decrease in blue (p<0.05, jackknifed t-test). Median BMI was significant only for noise-evoked responses; spontaneous rate: -0.14 (p=0.38, n=16/54, sign test), noise-evoked response: -0.12 (**p=0.005, n=26/54), target-evoked response: 0 (p>0.5, n=12/36). **G-I.** Histograms comparing changes between active and passive conditions for the same neurons, plotted as in D-F. For the active condition, data were combined from the two attention conditions. In all comparisons, responses to an identical set of stimuli were used to compute the difference between behavior conditions. Median BMI increased for spontaneous rate and noise-evoked response: spontaneous rate: 0.17 (**p=0.002, n=27/54), noise-evoked response: 0.11 (**p=0.005, n=38/54), target-evoked response: 0.14 (p=0.09, n=13/36).

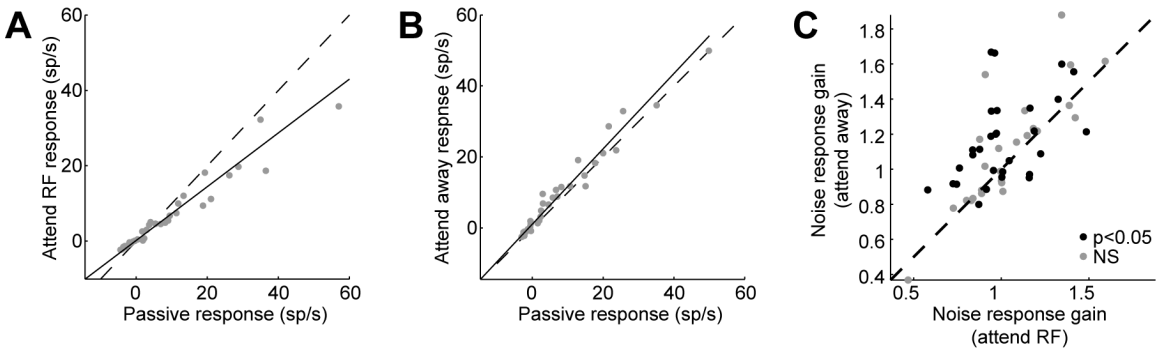

Figure 2.4: Effect of selective attention on A1 response gain. **A.** Example of noise-evoked gain changes between the passive and attend RF conditions (same unit as Fig. 2A). Scatter plot compares responses to identical epochs of the noise stimulus between behavior conditions. The majority of points fall below the line of unity slope (dashed) indicating a general suppression when attention was directed toward a target in the neuron's receptive field. The slope of the line fit to the scatter plot (solid) is 0.76, indicating suppression of 24% during the attend RF condition. **B.** Comparison between passive and the attend away condition for the same neuron reveals a slight gain enhancement (slope 1.04). **C.** Each point compares noise-evoked response gain for a single neuron between attend RF and attend away conditions, computed relative to passive listening. Across all neurons, mean gain was higher during attend away than attend RF (1.11 vs. 0.99, $p<0.001$, sign test). For neurons showing significant changes in mean response (black points, see Fig. 3E), mean gain was also different (1.11 vs. 0.97, $p<0.001$).

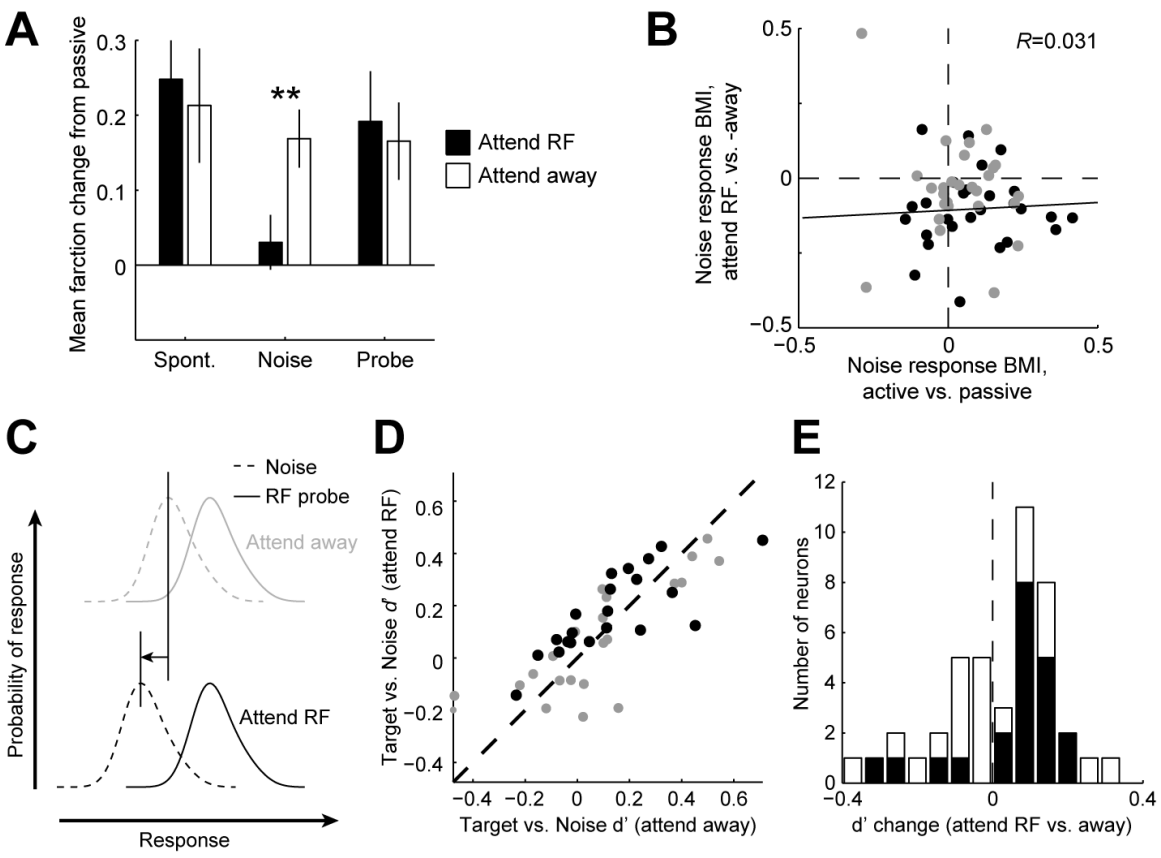

Figure 2.5: Effect of selective attention on A1 neural discriminability **A.** Mean fraction change in spontaneous rate, noise-evoked response and target-evoked response between the passive listening and attend RF (black) or attend away condition (white), for neurons showing significant effects of attention (n=34, Fig. 3D-F). Error bars indicate one standard error (**p=0.005, sign test). **B.** Scatter plot compares BMI for the noise-evoked response of each neuron between active and passive conditions and between attend RF and attend away conditions. Most neurons fall in the lower right quadrant (increased response during behavior and decreased response during attend RF), but there is no correlation between the magnitude of the effects across cells (r=0.031, p>0.5, permutation test). **C.** Model for enhanced discriminability following suppression of noise

response at attended location. Curves represent the distribution of single-trial responses to a tone at the neuronal best frequency (BF) embedded in noise (solid line) and to noise alone (dashed lines). If noise responses are suppressed when attention is directed to BF, then the distributions become more discriminable, as measured by *d*'. **D.** Scatter plot compares *d*' for the discriminability of neural responses to probe or catch tones (the same stimulus, respectively, in attend RF or attend away conditions) and the reference noise during time windows when the target could occur. Only units with at least 5 probe and catch tone presentations were included in order to obtain stable *d*' measures (n=43). Filled circles indicate experimental sessions when DI for probe tones was significantly greater than for catch tones (p<0.05, n=21/43, permutation test). **E.** Histogram of change in *d*' between attend RF and attend away conditions. For sessions in which DI was significantly greater for probe targets (filled bars), mean *d*' increased (0.10, p=0.02, permutation test). The mean change for all sessions was smaller (0.03, p=0.02).

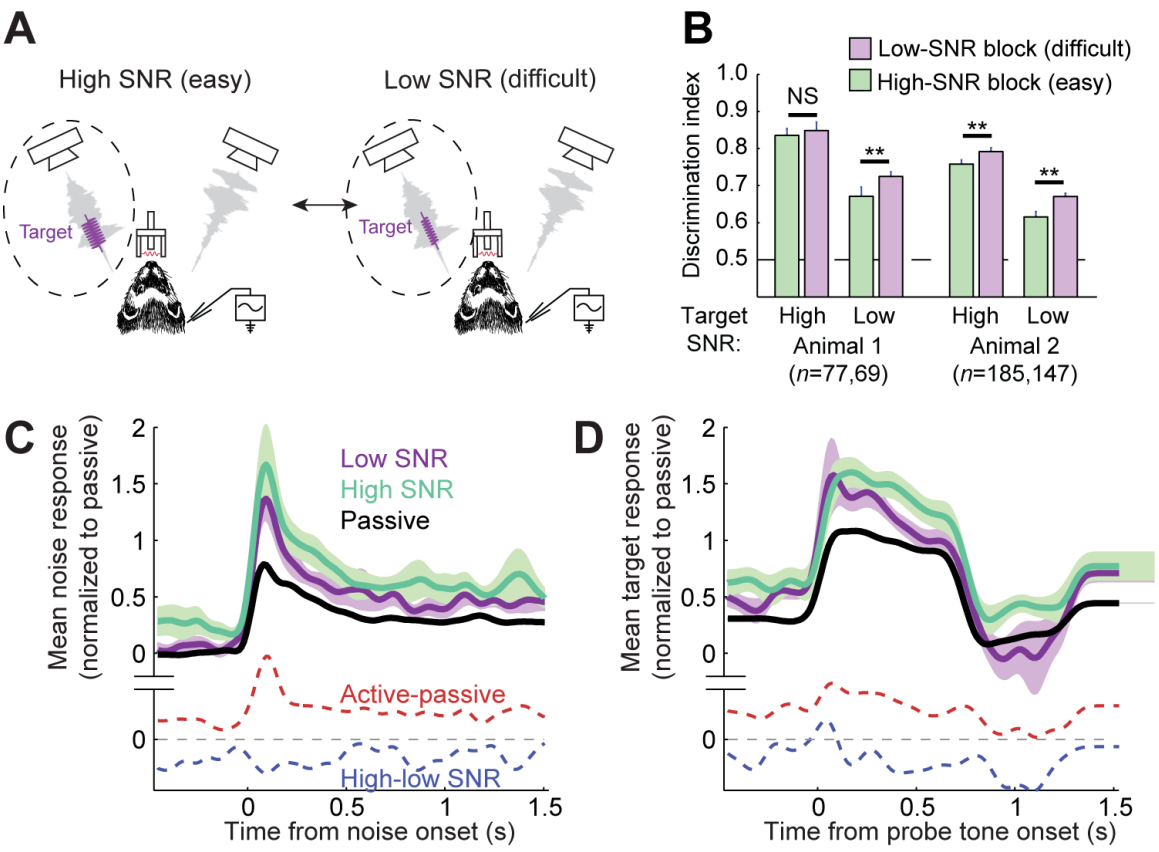

Figure 2.6: Effects of varying target SNR on behavioral effort and A1 evoked activity. **A.** Configuration of variable SNR behavior for manipulating listening effort. Ferrets were presented with vocalization-modulated noise streams, configured as for the selective attention task. A target tone appeared at a random time, always in a stream contralateral to the recording site and centered over BF of the recorded neuron. In high-SNR blocks, the target tone was presented at high SNR (10 to -5 dB relative to the noise) for 80-90% of trials and at low SNR (10 dB lower) for the remaining trials. Conversely, in low-SNR blocks, the target tone had low SNR for 80-90% of trials. Tone levels were adjusted between experiments to span super- and near-threshold behavior, which varied across target frequency and animals. SNR values were always identical during all recordings from a single neuron. **B.** To test for changes in effort, behavioral performance was

compared for identical target tones between behavior conditions. Bar chart compares mean DI for high- and low-SNR targets, broken down between high- and low-SNR blocks and between animals. DI was significantly greater for the low-SNR target in low-SNR blocks when the more difficult target was likely to occur, consistent with an increase in effort (**$p<0.01$, permutation test). DI did not improve for high SNR targets for Animal 1, possibly due to a ceiling effect in performance for the easy target. **C-D.** Average population PSTH for noise-evoked (C) and high-SNR target-evoked responses (D) compared between passive listening, high-SNR and low-SNR behavior conditions, for all neurons showing any significant difference between SNR conditions (n=65/88, $p<0.05$, jackknife t-test, Bonferroni correction).

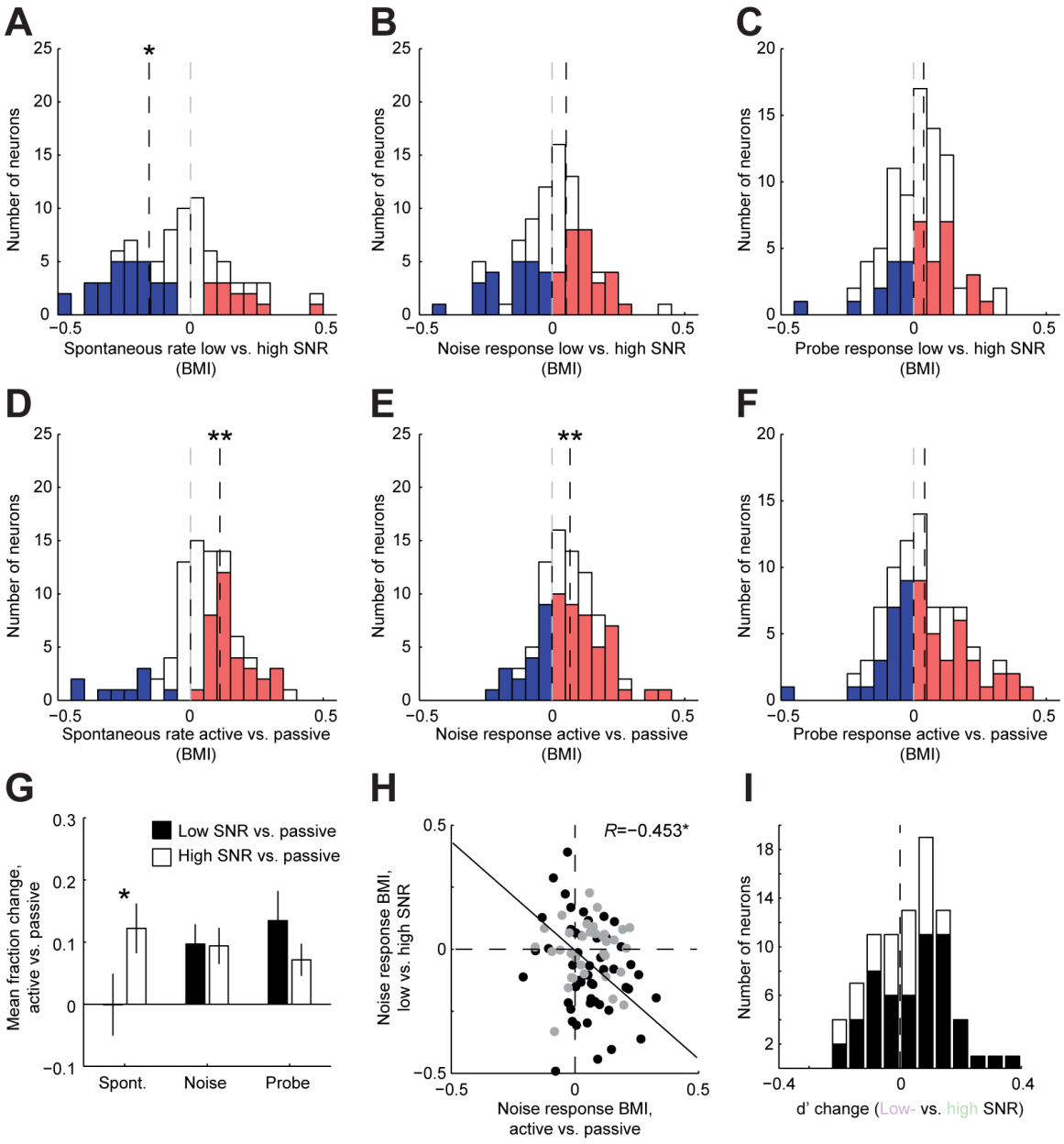

Figure 2.7: Summary of variable SNR effects on A1 evoked responses. **A-C.** Histograms of BMI for spontaneous rate (A), noise-evoked response (B) and target-evoked response (C) between high-SNR and low-SNR blocks, plotted as in Fig. 3D-F. Median spontaneous rate is suppressed (-0.16, p=0.008, n=47/88 significantly modulated neurons, sign test), but there is no change in noise-evoked (0.05, p=0.4, n=51) or target-

evoked responses (0.04, p=0.09, n=35) **D-F.** Histograms comparing active versus passive BMI for the variable SNR task. High- and low-SNR data were combined for the active condition. Task engagement leads to increased spontaneous activity (0.11, p<0.0001, n=43) and noise-evoked responses (0.06, p=0.005, n=63) but no change in target-evoked response (0.04, p=0.11, n=57). **G.** Bar chart shows mean fraction change in activity between passive listening and low-SNR (black) and high-SNR conditions (white), for neurons showing significant change in the corresponding statistic in A-C. **H.** Scatter plot compares change in noise-evoked response between active and passive conditions (horizontal axis) and the change in spontaneous rate between high- and low-SNR behaving conditions (r=-0.53, p<0.001, permutation test). **I.** Histogram comparing the change in d' between low SNR and high SNR conditions. The mean d' across all sessions showed a small increase in the low SNR condition (0.04, p=0.05, n=88, sign test). For sessions in which DI was significantly greater for low SNR targets in low SNR blocks (filled bars), the change was not significant (0.07, p=0.06, n=55).

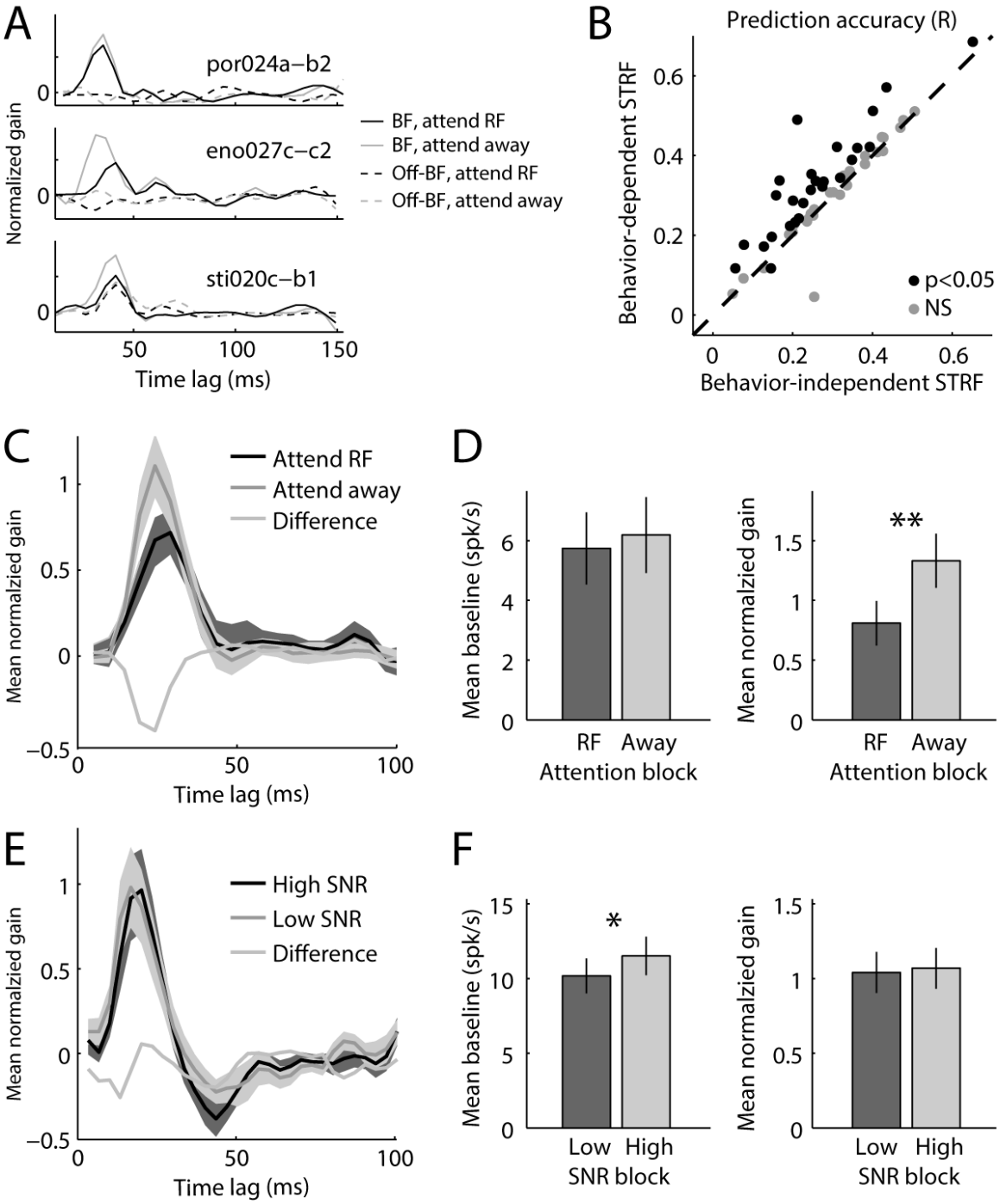

Figure 2.8: Effect of varying target SNR on A1 neural discriminability. **A.** Comparison of behavior-dependent STRFs for three neurons from selective attention experiments, estimated separately using data from attend RF and attend away conditions. In both

attention conditions, the temporal filter for the RF channel (solid lines, for the stimulus channel centered in the neuron's receptive field) showed excitatory tuning with 0-50 ms time lag. Gain was lower for this channel in the attend RF condition (black vs. gray lines). The non-RF channel (dashed lines) typically showed weaker gain, if any, and no consistent change between attention conditions. **B.** Scatter plot comparing prediction accuracy of behavior-independent versus behavior-dependent STRFs across the selective attention dataset. Filled dots correspond to neurons with a significant difference in prediction accuracy between models ($p<0.05$, sign test). Mean prediction accuracy was higher for the behavior-dependent model (mean $r=0.24$ versus 0.28 for behavior-dependent versus -independent, $p<0.001$, $n=54$, sign test). **C.** Average temporal response function at BF for behavior-dependent STRFs in the attend RF vs. attend away condition ($n=26/54$ neurons with significant improvement for behavior-dependent model). **D.** Mean baseline and peak gain for behavior-dependent STRFs in attend RF vs. attend away conditions show a significant difference in gain (**$p=0.007$, sign test). **E.** Average temporal response functions of behavior-dependent STRFs estimated separately for low- and high-SNR conditions ($n=51/88$ with significant benefit for behavior-dependent model). **F.** Mean baseline and peak gain for behavior-dependent STRFs in low-SNR and high-SNR conditions shows a significant difference in response baseline (*$p=0.02$, sign test).

This chapter was adapted from Schwartz et al. (2020) which was published in Journal of Neurophysiology 123(1):191-208. doi: 10.1152/jn.00595.2019.

Chapter 3

Pupil-associated states modulate excitability but not stimulus selectivity in primary auditory cortex

Leah P. Schwartz[1], Brad N. Buran[2], Stephen V. David[2]

[1]Neuroscience Graduate Program, Oregon Health and Science University

[2]Oregon Hearing Research Center, Oregon Health and Science University

Acknowledgements: This work was supported by grants from the National Institute on Deafness and Other Communication Disorders (R01 DC0495, R21 DC016969, F31 DC016204) and the ARCS Foundation Oregon Chapter. The authors would like to thank Daniela Saderi, Luke A. Shaheen, and Sean Slee for training, advice, and assistance with neurophysiological recording, Ulysses Duckler for animal care, and Matthew J. McGinley, Avinash Bala, and Robert Peterka for technical assistance with pupillometry. BNB receives financial compensation for programming, data analysis, experiment design and tutoring services for government agencies, academic institutions and private companies in addition to his part-time work at Oregon Health & Science University. LPS and SVD declare no competing financial interests.

Abstract

Recent research in mice indicates that luminance-independent fluctuations in pupil size predict variability in spontaneous and evoked activity of single neurons in auditory and visual cortex. These findings suggest that pupil is an indicator of large-scale changes in arousal state that affect sensory processing. However, it is not known whether pupil-related state also influences the selectivity of auditory neurons. We recorded pupil size and single-unit spiking activity in the primary auditory cortex (A1) of non-anesthetized male and female ferrets during presentation of natural vocalizations and tone stimuli that allow measurement of frequency and level tuning. Neurons showed a systematic increase in both spontaneous and sound-evoked activity when pupil was large, as well as desynchronization and a decrease in trial-to-trial variability. Relationships between pupil size and firing rate were non-monotonic in some cells. In most neurons, several measurements of tuning, including acoustic threshold, spectral bandwidth, and best frequency, remained stable across large changes in pupil size. Across the population, however, there was a small but significant decrease in acoustic threshold when pupil was dilated. In some recordings, we observed rapid, saccade-like eye movements during sustained pupil constriction, which may indicate sleep. Including the presence of this state as a separate variable in a regression model of neural variability accounted for some, but not all, of the variability and non-monotonicity associated with changes in pupil size.

Introduction

Although the iris primarily functions to regulate the formation of images on the retina, pupil dilation and constriction are also a result of autonomic nervous activity unrelated to visual stimuli . Growing evidence suggests that these autonomic changes in pupil diameter are an indicator of changes in internal state that affect processing in sensory cortex . In mice, pupil dilation is correlated with transitions from stillness to walking (McGinley et al., 2015a; Vinck et al., 2015), which increase firing rates in visual cortex (Niell and Stryker, 2010) and decrease firing rates in auditory cortex (Schneider et al., 2014). Dilated pupil is associated with increases in high-frequency local field potential (LFP) activity, a physiological signature of alertness, in sensory cortex (Harris and Thiele, 2011; Vinck et al., 2015). In mouse auditory cortex, intermediate pupil diameter is also associated with increases in the gain of sound-evoked responses, decreased spontaneous activity, and stable, hyperpolarized subthreshold membrane potential, which have been hypothesized to reflect an “optimal state” for detecting sensory stimuli (McGinley et al., 2015a).

Pupil dilation has also been correlated with changes in sensory selectivity. In visual cortex, orientation tuning becomes more selective when pupil dilates, suggesting that sensory representations are more precise in some internal brain states (Reimer et al., 2014). However, it is not known if pupil-associated state influences the selectivity of neurons in the auditory system. Some population-level associations between pupil size and cortical activity do differ between sensory systems (Shimaoka et al., 2018). Data from mouse auditory cortex – which suggests that pupil size is non-monotonically related

to neural activity (McGinley et al., 2015a) – conflicts with data from mouse visual cortex, which show either monotonic increases in neural activity when pupil dilates (Reimer et al., 2014; Vinck et al., 2015), or simultaneous increases and decreases in different neurons recorded at the same time (Stringer et al., 2019). These observations may reflect functional differences between the auditory and visual systems.

To explore effects of pupil-associated state on sensory coding in auditory cortex, we recorded pupil size and single-unit spiking activity in the primary auditory cortex (A1) of head-restrained, non-anesthetized ferrets. To study effects of internal state on gain and spontaneous activity, we presented a large number of repetitions (up to 120) of ferret vocalizations. To study state effects on sensory selectivity we presented random sequences of tone pips at various frequencies and levels. Each stimulus set therefore allowed in-depth exploration of a distinct feature of auditory responses, as well as comparison to previous work in the visual system (Reimer et al., 2014). We characterized effects of pupil by comparing responses to each distinct sound as pupil size varied. Neurons typically showed an increase in baseline firing rate as well as the gain and reliability of sound-evoked activity when pupil was large, although there was also evidence for non-monotonic effects of pupil-associated state in some neurons. In most neurons, parametric measurements of tuning, including acoustic threshold, spectral bandwidth, and best frequency, remained stable across large changes in pupil size. Across the population, however, we observed a small decrease in acoustic threshold when pupil was dilated.

In some recordings, we observed increased saccadic eye movements during periods of sustained pupil constriction, possibly indicating sleep onset. Including the

presence of this state as a separate variable in a regression model of neural variability accounted for some, but not all, of the variability associated with changes in pupil size. The stability of best frequency across large changes in pupil size (and associated changes in gain and spontaneous activity) contrasts with task-related plasticity of receptive fields in ferret auditory cortex (Fritz et al., 2003; David et al., 2012), suggesting a difference between effects of engaging in tasks that require listening and variation within passive states.

Methods

All procedures were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee (IACUC) and conform to standards of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Neurophysiology

Five young adult ferrets (4 males and 1 female, Marshall Farms) were implanted with stainless steel headpost and acrylic cap to permit head-fixation and access to auditory cortex. After recovery (1-2 weeks), animals were habituated to a head-fixed posture and presentation of auditory stimuli. Following habituation, a small (1-2 mm) craniotomy was opened over primary auditory cortex (A1). In 66 recordings, data was collected using 1-4 independently positioned tungsten microelectrodes (FHC, 1-5 MΩ) advanced by independent microdrives (Electrode Positioning System, Alpha-Omega). In 3 recordings, data was collected using a single-shank 64-channel silicon microelectrode array, spanning 1 mm (Masmanidis Lab, UCLA) (Du et al., 2011).

After a unit was isolated during tungsten recordings, its auditory responsiveness was tested using tones or narrowband noise bursts. If the site showed no response, the electrode was lowered to search for another site. During depth-array recordings, the probe was lowered into auditory cortex until auditory-evoked activity was observed across the span of recorded channels. Amplified and digitized signals were stored using open-source data acquisition software (MANTA: Englitz et al., 2013; Open Ephys: Siegle et al., 2017). Ferrets were observed by video monitor during recordings. If a unit became

indistinguishable from the noise floor following a body movement, the recording was stopped and the 5 trials preceding the putative isolation loss (28 to 40 seconds of recording time) were excluded from analysis. A1 recording sites were identified by characteristic short latency, frequency selectivity, and tonotopic gradients of single or multiunit activity across multiple penetrations (Bizley et al., 2005; Elgueda et al., 2019).

Spike sorting

For tungsten electrode recordings, events were extracted from the continuous electrophysiological signal using a principal components analysis and k-means clustering (Meska-PCA: (David et al., 2009). Single-unit isolation was quantified from cluster overlap as the fraction of spikes likely to be produced by a single neuron rather than another unit. For depth-array recordings, single-unit spikes were sorted using polytrode sorting software (Kilosort: Pachitariu et al., 2016). Only units maintaining stable isolation > 95% through the experiment were considered single neurons for analysis.

Pupillometry

During neurophysiological recordings, infrared video of one eye was collected for offline measurement of pupil size. Video was collected using either (1) a commercial CCTV camera fitted with a lens to provide magnification of the eye and an infrared filter (Heliopan ES 49 RG 780) or (2) an open-source camera (Adafruit TTL Serial Camera) fitted with a lens (M12 Lenses PT-2514BMP 25.0 mm) whose focal length allowed placement of the camera 10 cm from the ferret's 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.

Pupil size was measured using custom MATLAB (version R2016b) software (code available at https://bitbucket.org /lbhb/baphy). For each video, an intensity threshold was selected to capture pupil pixels and exclude the surrounding image. During initial recordings, the threshold was selected manually. We observed that the intensity histogram is multi-modal with the first peak (reflecting the darkest pixels) generally corresponding to the pupil. In later recordings, we therefore automatically updated the threshold of each frame to position it at the first valley in the frame's intensity histogram. Each frame was smoothed by a Gaussian filter before thresholding, then segmented by Moore boundary tracing. The segment with the largest area was identified as pupil. We measured pupil size as the length of the minor axis of an ellipse fit to this region using a direct least-squares method. To avoid identifying shadows at the edge of the eye or other dark regions of the image as pupil, we restricted the search for the largest-area segment to a rectangular region of interest surrounding the pupil identified in the preceding frame (Nguyen and Stark, 1993). Measurements were calibrated by comparing the width of each ferret's eye (in mm) and the width of the ferret eye as it appeared in each video (in pixels), which gave a conversion factor of pixels to mm.

The frame rate of the cameras used in the experiments varied from 10 to 30 frames/second. To compensate for variability in camera frame rate, we recorded a timestamp at the start and end of each trial, then interpolated measurements of pupil size

to match the sampling of the simultaneously recorded neural data. This procedure ensured that the two data streams (video and neural recording) remained synchronized throughout each recording.

Blink artifacts were identified by rapid, transient changes in pupil size (McGinley et al., 2015a). The derivative of the pupil trace was taken and bins with derivatives more than 6 standard deviations from the mean were marked. Blinks were identified within these bins by screening for decreases in pupil size followed by increases. Data during a 1-second period surrounding the blink was then removed from the trace and replaced by a linear interpolation of the pupil size immediately before and after the blink.

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 to allow for comparison with previous research (McGinley et al., 2015a). Setting the offset at values between 0 and 1.5 s did not affect the mean accuracy of the pupil-based regression models of neural activity described below (data not shown).

To measure changes in eye position, we calculated the Euclidean distance between the center of the ellipse fit to the pupil in each video frame, then multiplied by frame rate to find eye speed. We identified putative sleep states by screening for a combination of tonically constricted pupil and a high rate of saccades. Sleep states were identified by selecting parameter values for each recording (maximum pupil size, maximum pupil standard deviation, minimum saccade speed, minimum saccade rate) that marked abrupt state transitions in the traces of pupil size and eye speed. Gaps of less than 15 seconds between sleep episodes were considered artifacts and coded as sleep. We

restricted our analysis to sleep episodes with a duration of at least 30 seconds.

Acoustic stimuli

Stimulus presentation was 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 a free-field speaker (Manger).

During initial recording sessions, ferrets were presented with two 3-second species-specific vocalizations: an infant call and an adult aggression call. Stimuli were recorded in a sound-attenuating chamber using a commercial digital recorder (44-kHz sampling, Tascam DR-400). No animal that produced the recorded vocalizations was used in the study. Each vocalization was repeated up to 120 times. The order of vocalizations was randomized on each repetition. Each repetition was preceded by a 2-second silent period to allow measurement of spontaneous neural activity on each trial.

To improve sampling of internal states, recording was paused for several seconds after the 40^{th} and 80^{th} sound repetitions, during which the ferret was roused by an unexpected sound (e.g., turning the doorknob of the recording booth), which often resulted in pupil dilation. A stepwise regression analysis (see below) using the number of trials since each pause to predict neural activity suggests that the time since this event did not predict neural activity independent of pupil size (data not shown).

During later recording sessions, ferrets were presented with sequences of 10-ms tone pips with varying frequency (4-6 octaves surrounding best frequency, ranging over 32 Hz to 40 kHz) and level (0 to 65 dB SPL, 5 or 10 dB SPL steps). Each pip sequence

was 6 seconds long, preceded by a 2-second silent period to allow measurement of spontaneous neural activity. Tone-pip sequences were repeated multiple times in each recording (mean repetitions +/- SEM: 28 +/- 2). During some recordings, tone-pip sequences were interleaved with ferret vocalizations to improve sampling of pupil states in each data set.

Data analysis

Gain and baseline firing rate

Single-trial neural and pupil data for ferret vocalization recordings was binned at 4 Hz. Peristimulus time histogram (PSTH) responses to each vocalization were calculated by aligning spike activity to stimulus onset and averaging across presentations. To isolate evoked activity, the mean spike rate during the two seconds of silence preceding stimulus onset was subtracted from the PSTH.

Neural and pupil data for responses to tone-pip sequences were binned at 100 Hz. A frequency response area (FRA) was calculated by taking the mean spike rate for each sound during a window 10 ms to 60 ms following pip onset. To isolate evoked activity, the mean spike rate during the two seconds of silence preceding tone-pip sequence onset was subtracted from the FRA.

We used multivariate regression models to examine effects of pupil state on spontaneous and sound-evoked activity. We modeled the response to each presentation of a tone pip or vocalization as a function of average spontaneous rate over the entire

recording, average response to the stimulus across all trials, and pupil diameter during each stimulus presentation.

The *baseline-gain* model accounted for a linear relationship between pupil diameter and both spontaneous and evoked spike rate. According to this model, single-trial activity, $r_i(t, s)$, is

$$r_i(t, s) = \beta_0 m_0 + \beta_1 r_0(t, s) + [\beta_2 m_0 + \beta_3 r_0(t, s)] p_i(t, s)$$

Here, $r_0(t, s)$ is the mean time-varying PSTH response evoked by stimulus *s* (either one vocalization or one tone pip at a single frequency and level), m_0 is the mean spontaneous rate, and $p_i(t, s)$ is pupil size observed simultaneously to $r_i(t, s)$. The model coefficients indicate how the neuron's overall spiking activity (baseline, β_2) and the scaling of the cell's sound-evoked response (gain, β_3) vary with pupil size. To cast all β parameters in units of pupil^{-1}, we normalized the baseline terms by m_0. We therefore refer to this as the "baseline-gain" or "first-order baseline-gain" model.

Unless otherwise specified, model parameters were fit by linear, least-squares regression using 20-fold cross-validation. The median model parameters across all 20 fits was used to compare coefficients across cells. Model accuracy for a given cell was quantified by the squared Pearson correlation between actual single-trial firing rates and firing rates predicted using the cross-validated model coefficients. Significance of model fits was assessed using a permutation test. A cell with a *significant fit* was one in which a fit to real pupil data showed accuracy greater than a model fit to pupil data shuffled across trials for at least 50 of 1000 shuffles (i.e., $p < 0.05$, permutation test).

To determine what aspects of the baseline-gain model were important for the capturing the relationship between pupil and firing rate, we compared it to several variants. The *baseline only* model allowed only the baseline rate, but not the gain of the evoked response, to scale with pupil,

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + \beta_2 m_0 p_i(t,s)$$

The *gain only* model allowed only the gain of the evoked response, and not the baseline rate, to scale with pupil,

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + \beta_2 r_0(t,s) p_i(t,s)$$

We also considered nonlinear relationships between pupil and spike rate. The *second-order baseline* model allowed the baseline rate to scale with the square of pupil size, allowing non-monotonic effects of internal state,

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + [\beta_2 m_0 + \beta_3 r_0(t,s)] p_i(t,s) + \beta_4 m_0 p_i^2(t,s)$$

The *second-order baseline-gain* model allowed both the baseline rate and gain to scale with the square of pupil size,

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + [\beta_2 m_0 + \beta_3 r_0(t,s)] p_i(t,s) + [\beta_4 m_0 + \beta_5 r_0(t,s)]\, p_i^2(t,s)$$

We measured the performance of each model by comparing its ability to predict single-trial data to the PSTH or FRA alone, without any information on pupil size (the *null model*),

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s)$$

For each cell and each model, we measured the percent improvement in accuracy (i.e., the difference in squared Pearson correlation coefficient, R^2, between the model predictions and actual data) over the null model,

$$\frac{R^2(\text{model, data}) - R^2(\text{null, data})}{R^2(\text{null, data})} * 100$$

Each model's performance was quantified as the median improvement in accuracy across all cells that showed a significant fit for any tested model ($p < 0.05$, permutation test). For comparing model performance across collections of neurons, we performed a Wilcoxon signed-rank test (*sign test*) between the distribution of prediction correlations across neurons for each model.

Non-monotonicity

A previous report (McGinley et al., 2015a) observed non-monotonic relationships between pupil size and neural activity in mouse auditory cortex. Following its language, we refer to neurons with a maximum firing rate at intermediate pupil sizes and a lesser

firing rate at small and large pupil sizes as “inverted U”s. We refer to neurons with a minimum firing rate at intermediate pupil sizes as “U”s.

We restricted our analysis of non-monotonicity to neurons that showed a significant fit for the second-order baseline-gain model relative to the same model fit to shuffled pupil data (n = 79/114 neurons). The second-order terms of this model allow for the possibility of a non-monotonic relationships between pupil size and firing rate.

We used a segmented linear model to test for non-monotonic relationships between pupil size and firing rate in single neurons (Simonsohn, Uri, 2018). The segmented model is similar to the baseline-gain model (see above), except that data is partitioned into intervals at a *breakpoint* in pupil size. The fit parameters of the model may differ on either side of the breakpoint. This allows a variety of relationships between pupil and firing rate, including linear increases and decreases, Us, and inverted Us.

The segmented model included separate breakpoints for pupil-related effects on baseline and gain (β_0, β_4). Both breakpoints were fit simultaneously with other free parameters in the model. Baseline and gain effects were combined to estimate the neuron’s firing rate, i.e.,

if $p_i(t,s) < \beta_0$:	$\text{baseline}_i(t,s) = [\beta_1 + \beta_2 p_i(t,s)]m_0$
else:	$\text{baseline}_i(t,s) = [\text{shift}_{base} + \beta_3 p_i(t,s)]m_0$
where:	$\text{shift}_{base} = \beta_1 + \beta_0[\beta_2 - \beta_3]$

if $p_i(t,s) < \beta_4$:	$\text{gain}_i(t,s) = [\beta_5 + \beta_6 p_i(t,s)]r_0(t,s)$
else:	$\text{gain}_i(t,s) = [\text{shift}_{gain} + \beta_7 p_i(t,s)]r_0(t,s)$

where: $\text{shift}_{\text{gain}} = \beta_5 + \beta_4[\beta_6 - \beta_7]$

and

$$r_i(t,s) = \text{baseline}_i(t,s) + \text{gain}_i(t,s)r_0(t,s)$$

The constraints on the two *shift* parameters ensure that the functions will be continuous at their breakpoints. Model parameters were fit by nonlinear least-squares using the trust-region-reflective algorithm and 20-fold cross validation. A change in the sign of pupil coefficients across the breakpoint (β_2 versus β_3 or β_6 versus β_7) indicates a non-monotonic relationship between pupil and firing rate. The direction of the sign change indicates if a cell is a U or inverted U. Data were pre-processed as described (see: "Data analysis: Gain and baseline firing rate").

To identify non-monotonic neurons, we compared the accuracy of the segmented linear model to an identical model that was constrained to be monotonic. That is, the slope coefficients related to baseline (β_2, β_3) and gain (β_6, β_7) were constrained to have the same sign on both sides of the breakpoint. If the non-constrained model showed a sign change as well as an improvement in accuracy over the constrained model, we classified the cell as showing a *non-monotonic trend*.

To identify *significantly non-monotonic* neurons, we performed a permutation test. We randomly shuffled half the firing rate predictions between the constrained and non-constrained models, then calculated the difference in accuracy between the shuffled prediction rates. Significantly non-monotonic neurons were those in which less than 50 of 1000 shuffles showed a greater difference in accuracy than that observed between the constrained and non-constrained models ($p < 0.05$ of a chance improvement).

Neural response variability

We used several measurements (Fano factor, between-trial reliability, and noise correlation) to characterize the variability of neural responses across trials. In all cases, we used data from the repeated vocalizations experiment and compared variability across pupil states by grouping trials based on whether the mean pupil size preceding stimulus onset was above or below the median pupil size observed during the recording. Data for the noise correlation and reliability analyses was binned at 4 Hz for consistency with regression analyses. Data for the Fano factor analysis was binned at 100 Hz to bring out faster temporal dynamics of the neural response to sound. Two neurons with extremely low evoked spike rates produced undefined peristimulus Fano factors (variance of 0 divided by mean of 0) and were therefore excluded from the analysis. Reliability was calculated as the mean trial Pearson correlation of neural responses across trials (McGinley et al., 2015a). Spike rates from the response to each stimulus were z-scored before calculating noise correlations.

Frequency and level tuning

To assess changes in frequency and level tuning associated with pupil state, neural responses to tone-pip sequences were binned at 100 Hz. Data from each tone-pip recording was divided into two bins (*large-pupil* and *small-pupil*) based on the median pupil size during the recording. Data from tone-pip sequences was analyzed using: (1) the 2-second silence preceding each tone-pip sequence (*spontaneous activity*), (2) the interval from 60 ms to 10 ms preceding onset of each tone pip (*prestimulus activity*), and (3) the

interval 10 ms to 60 ms after the onset of each tone pip (*stimulus-evoked activity*). Since the FRA was assessed using a cloud of tone pips presented in rapid succession, the cell's firing rate during the 50 ms preceding sound onset (i.e., prestimulus activity) was a better measure of the cell's baseline response for sub-threshold levels than spontaneous activity.

For display, examples of FRAs were calculated taking the mean stimulus-evoked spike rate for each tone pip, smoothing using a two-dimensional box filter with dimensions equivalent to 15 dB and 1.5 octaves, and subtracting the mean spontaneous activity across all tone-pip sequences in the recording. To measure changes in spontaneous and driven rates in the tone-pip data without exploring changes in tuning, we used the mean spontaneous rate and the difference between the mean stimulus-evoked and prestimulus firing rates. Then to measure possible changes in sensory tuning we fit hierarchical regression models to the prestimulus and stimulus-evoked spike data for each tone pip.

The hierarchical models were implemented in STAN (http://mc-stan.org) and fit using Bayesian inference. In contrast to conventional model fitting, Bayesian analysis allows for simple calculations of credible intervals on derived parameters (e.g., parameters that are mathematical functions of fitted coefficients), is more robust to outliers through the use of Poisson likelihoods for spike count data, and offers simple construction of realistic hierarchical models (Gelman et al., 2003). Each model was fit four times for 1000 samples after a 1000 sample iteration burn-in period. Posterior samples were combined across all chains for inference.

Rate-level functions

To estimate changes in level tuning linked to internal state, we first calculated characteristic frequency (CF) before binning data by pupil size. An FRA for all data from the recording was calculated using data from the stimulus-evoked epoch and smoothed with a two-dimensional box filter with dimensions equivalent to 15 dB and 1.5 octaves. A separate standard error was calculated for each coefficient of the FRA by measuring variance across repetitions of tone pips at that frequency and level. The minimum sound level that evoked a response at least 2 standard errors above the mean spike rate during the 100 ms preceding pip onset, measured across all pips, was calculated for each frequency. CF was defined as the frequency that required the minimum sound level to evoke a response.

For each cell, we went back to the unsmoothed FRA data, binned the tone-evoked responses according to pupil size and generated rate-level functions. Rate-level functions were smoothed by averaging the response across the three frequencies closest to CF. Rate-level functions for individual cells were fit using a function with three parameters, baseline firing rate, b, slope, s, and threshold, θ, for each pupil condition:

$$n_i \sim Poisson(b_{c,p}) \times t_i \qquad l_i \leq \theta_c$$

$$n_i \sim Poisson(b_{c,p} + s_{c,p} \times (l_i - \theta_{c,p})) \times t_i \qquad l_i > \theta_c$$

Here, n is the number of spikes observed in a time interval, t, for cell, c, and pupil condition, p, on the i-th observation of stimulus level, l. Since the time interval varied

across cells, depending on how long we maintained stable recordings, incorporating this information into our model gave greater weight to data points with longer time intervals.

For levels less than threshold, firing rate was expected to be equal to the baseline firing rate. For stimulus levels greater than threshold, the cell could have an excitatory ($s_c > 0$) or inhibitory ($s_c < 0$) response.

The distribution of baseline rates across the cells in the small pupil condition were modeled by a Gamma prior, $b_{c,0} \sim Gamma(b_\alpha, b_\beta)$, with $b_\alpha \sim Gamma(0.5,0.1)$ and $b_\beta \sim Gamma(0.1,0.1)$. Since the expected value of the Gamma distribution is α/β, we can compute the posterior for the average spontaneous rate of the population as b_α / b_β when the pupil is small and $b_\alpha / b_\beta \times bg_\mu$ when the pupil is large. We set the priors for α and β after inspecting the distributions of spontaneous rates in the small and large pupil condition. Although prior work (McGinley et al., 2015a) suggests that we would expect to see a difference in spontaneous rate between the pupil conditions, we chose an unbiased prior for the ratio ($b_{c,1}/b_{c,0}$) such that the mean and standard deviation were $Normal(1,1)$ and $HalfNormal(1)$, respectively.

The distribution of slopes across cells were modeled using a Normal prior, $s_{c,0} \sim Normal(s_\mu, s_\sigma)$, and the priors for the population mean and standard deviation were modeled themselves as $s_\mu \sim Normal(0.1, 0.1)$ and $s_\sigma \sim HalfNormal(0.1)$. The distribution of thresholds across cells was modeled using a Normal prior, $\theta_{c,0} \sim Normal(\theta_\mu, \theta_\sigma)$ with priors for the population mean and standard deviation modeled as $\theta_\mu \sim Normal(40, 5)$ and $\theta_\sigma \sim HalfNormal(5)$. To minimize potential bias, unbiased priors were used for parameters representing the difference between pupil conditions. Specifically, the difference in slope ($s_{c,1} - s_{c,0}$) was modeled with a Normal

prior with mean and standard deviation of $Normal(0, 0.1)$ and $HalfNormal(0.1)$, respectively. The difference in threshold ($\theta_{c,1} - \theta_{c,0}$) was modeled with a Normal prior with mean and standard deviation of $Normal(0, 5)$ and $HalfNormal(5)$, respectively.

Frequency-tuning curves

To quantify changes in frequency tuning linked to internal state, we first found best level (i.e., the level that evoked the maximum rate across all frequencies) using the FRA measured across all pupil-size bins. We then extracted frequency tuning curves (FTCs) at best level from the large-pupil and small-pupil FRAs. FTCs were smoothed by averaging across best level and the two neighboring levels in the FRA.

We assumed that the FTC for individual cells, *c*, and pupil condition, *p*, could be expressed as a Gaussian function with four parameters, baseline firing rate, *b*, gain, *g,* best frequency, *BF,* and bandwidth, *BW*:

$$n_i \sim Poisson\left(\left[b_{c,p} + g_{c,p} \times \exp\left(-0.5 \times \frac{\log_2 f_i - BF_{c,p}}{e^{BW_{c,p}}}\right)\right]\right) \times t_i$$

Here, *n* is the evoked rate for the cell, pupil condition (small vs. large) and stimulus frequency, *f*, on the *i*-th observation. As in the rate-level model, the cell's firing rate during the 50 ms preceding sound onset was a better measure of the cell's baseline response for sub-threshold levels and the spontaneous rate model was also incorporated into this model to improve our estimate of *b*.

The distribution of gains across cells was modeled using a Normal prior, with priors for the population mean and standard deviation modeled as $g_\mu \sim Normal(10, 1)$ and $g_\sigma = HalfNormal(10)$. The distribution of best frequency across cells was modeled using a Normal prior, with priors for the population mean and standard deviation modeled as $BF_\mu \sim Normal(10, 1)$ and $BF_\sigma = HalfNormal(1)$. The distribution of bandwidth across cells was modeled using a Normal prior, with priors for the population mean and standard deviation modeled as $BW_\mu \sim Normal(-0.5, 0.1)$ and $BW_\sigma = HalfNormal(0.25)$.

To avoid bias in estimates of pupil-related changes, uninformative Normal priors were used for parameters representing the difference between small and large pupil conditions. Specifically, the logarithm of the ratio in gain ($g_{c,1}/g_{c,0}$) was a with a mean and standard deviation of $Normal(0, 0.1)$ and $HalfNormal(0.1)$, respectively. The logarithm of the ratio in bandwidth ($BW_{c,1}/BW_{c,0}$) had a mean and standard deviation of $Normal(0, 0.1)$ and $HalfNormal(0.25)$, respectively. For the ratio in gain and ratio in bandwidth, $\exp(0) = 1$, such that the expected value of the prior was no effect of pupil size on gain or bandwidth. The difference in best frequency ($BF_{c,1} - BF_{c,0}$) had a mean and standard deviation of $Normal(0, 0.1)$ and $HalfNormal(0.1)$, respectively.

Sleep states

To test for neural correlates of sleep states, we defined a binary variable that indicated the presence or absence of sleep ($sleep_i(t, s)$) and added it as a term in the baseline-gain model (see above),

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + [\beta_2 m_0 + \beta_3 r_0(t,s)]sleep_i(t,s)$$
$$+ \ [\beta_4 m_0 + \beta_5 r_0(t,s)]p_i(t,s)$$

We also tested a sleep-only model in which baseline and gain were modulated only by sleep state,

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + [\beta_2 m_0 + \beta_3 r_0(t,s)]sleep_i(t,s)$$

Preprocessing and model fitting and evaluation were completed as for the baseline-gain model (see "Data analysis: Gain and baseline firing rate"). The model was fit to neural data from responses to vocalizations.

Local-field potential

The local-field potential (LFP) signals from each recording were bandpass filtered using a forward and reverse filtered second-order Butterworth filter into delta (1-4 Hz), theta (4-7 Hz), alpha (7-14 Hz), beta (15-30 Hz), low gamma (30-60 Hz) and high gamma (60-100 Hz) bands (Yuzgec et al., 2018). The instantaneous amplitude of each bandpass-filtered signal was calculated by computing the magnitude of its Hilbert transform, low-pass filtering with a cut-off at 1 Hz, and taking the mean amplitude across all electrodes.

Sound-evoked pupil dilation

To calculate sound-evoked pupil dilation (Δp), pupil traces were binned at 30 Hz, and the mean pupil size during the 2-second silence preceding each vocalization on each trial (p_0) was subtracted from the raw trace before averaging across repetitions of the same sound. For regression analysis of the effect of sound-evoked dilations, we fit a model that included separate baseline and gain terms for the effect of the mean prestimulus pupil size on each trial (p_i) and the change from prestimulus pupil size in each bin:

$$r_i(t,s) = \beta_0 m_0 + \beta_1 r_0(t,s) + m_0[\beta_2 p_0(t,s) + \beta_3 \Delta p_0(t,s)] + r_0[\beta_2 p_0(t,s) + \beta_3 \Delta p_0(t,s)]$$

We compared the accuracy of this model's results with the first-order gain model, using the model comparison procedure described above.

Experimental design and statistical analyses

Statistical tests used to assess whether model fits were significantly better than chance (i.e., whether they accounted for any auditory response, tuning parameter, or state-related changes in neural activity), and to compare predictions across models are described above, together with other aspects of the models. Tests for correlations between data across conditions are reported as the unsquared Pearson correlation coefficient (r) and a t-test for significance. Tests for differences between conditions are reported using two-tailed Wilcoxon rank-sum tests (*rank-sum test*), Wilcoxon signed-rank

tests (*sign test*), or *t*-tests (*paired t-test* or *unpaired t-test*).

Code

Custom software for stimulus presentation and pupil analysis is available at https://bitbucket.org /lbhb/baphy. MATLAB and Python code for statistical analyses is available from the corresponding author on request.

Results

Dilated pupil is correlated with increases in spontaneous activity and gain

To examine the relationship between internal brain state and neural representation of natural sounds, we simultaneously recorded infrared (IR) video of the eye and single-unit spiking activity in the primary auditory cortex (A1) of head-restrained, non-anesthetized ferrets (Fig 3.1A, n = 114 neurons from 46 recording sites in 3 ferrets). Pupil size could be extracted from the video, providing a measure of internal state (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015). In order to gather data on the response to the same sounds across a wide range of states, our stimulus consisted of a small number of sounds (two ferret vocalizations) repeated many times (up to 120 repetitions per recording, mean +/- SD = 77 +/- 40 repetitions). To further increase our sampling of internal states, we paused this presentation at set times in the recording to rouse the ferret using auditory stimuli (see Methods). Although lighting conditions were held static during recordings, we also observed large changes in pupil size across timescales of tens of seconds to minutes (Fig. 3.1B).

Visual inspection of spike rasters suggested that trial-to-trial variability in neural activity sometimes tracked changes in pupil size (Fig. 3.1C). To quantify the association between pupil size and firing rate, we fit a linear model that allowed baseline spike rate and gain of sound-evoked responses to depend on pupil size (baseline-gain model, see Methods, Fig. 3.1D). In a majority of neurons (n = 75/114, 66%) this model was significantly more accurate at predicting single-trial neural activity than a control model where pupil was shuffled in time (i.e., the prediction was based only on the neuron's peristimulus time histogram [PSTH] response to each vocalization) (Fig. 3.1E, $p < 0.05$,

permutation test). The degree of improvement was not correlated with the prediction accuracy of the control model alone (r = -0.08, p = 0.5, t-test), suggesting that the influence of pupil on trial-to-trial variability did not depend on a neuron's overall auditory responsiveness.

The baseline-gain model included separate terms for effects of internal state on gain and baseline firing rate, both of which tended to be positive, indicating an increase in firing rate with increasing pupil size (Fig. 3.1F). Both the baseline and gain terms contributed to model prediction accuracy, compared to models that included only one term (Fig. 3.1G, Table 3.1). Adding second-order terms improved median prediction accuracy further, and increased the number of neurons that showed a significant fit from 75 to 79 (Fig. 3.1G, Table 3.1, median increase in accuracy over PSTH only = 0.7% [gain model], 1.3% [baseline], 2.5% [gain and baseline], 2.4% [gain, baseline, and baseline squared], 3.7% [gain, baseline, baseline squared, and gain squared]).

Given that pupil size was more variable in some recordings than others, we wondered whether we were unable to detect effects of internal state in some neurons simply because the animal's pupil size remained fairly stable during the recording. Variability of pupil size during the recording was greater for neurons that showed a significant effect of pupil (median SD = 0.4 mm, n = 79 neurons) than for neurons that did not (median SD = 0.2 mm, n = 35 neurons, p = 1e-5, rank-sum test). Among neurons that showed a significant effect of pupil on firing rate, there was a correlation between the variance of pupil during the recording and the degree of improvement of the model over a model based on PSTH response alone (n = 79 neurons, r = 0.37, p = 4e-4, t- test). Thus, the measurement of 66% of neurons showing a significant effect of state on activity

represents a lower bound on the frequency of neurons showing pupil-related effects. To confirm this point, we split the dataset into high-variability and low-variability recordings based on the variance of pupil within a recording relative to the median variance across all recordings. A greater percentage of neurons in the high-variance subset were modulated by pupil (number of modulated neurons: 31/35 modulated in high-variance dataset vs. 48/79 in the low-variance dataset, p=3e-03, chi-square test). The median improvement in variance explained by the pupil model over a PSTH-based model was greater for the high-variance than low-variance model (change in R^2: 0.01 [low-variance] vs. 0.13 [high-variance], p=2e-04, rank-sum test). These results suggest that the level of pupil modulation would be greater if all neurons were drawn from recordings with the greatest possible range of pupil states.

The models were fit to data that included sound-evoked activity as well as activity during the silent intervals before and after each vocalization. The baseline term of the models could reflect either modulation of spontaneous activity or a stimulus-dependent modulation of sound-evoked firing rate that did not scale with stimulus strength. To distinguish between these possibilities, we fit a model with only a baseline term to data from the prestimulus epoch alone, ignoring sound-evoked activity. We then fit a model that included terms for both baseline and gain to the peristimulus epoch alone, ignoring spontaneous activity. The magnitude of the baseline terms in the two models was correlated (n = 75 neurons, r = 0.55, p = 4e-08, t-test), suggesting that the baseline term of the model reflected primarily modulation of spontaneous activity rather than (or in addition to) stimulus-evoked activity.

Auditory neurons encode sensory stimuli more reliably when pupil is large

Previous research in mouse auditory and visual cortex suggests that the degree to which neural activity is dominated by sensory input depends on pupil-indexed state (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015). To replicate and extend these results, we split the data from each recording into two groups depending on whether pupil size was greater or less than the median pupil size observed in the recording. We then calculated several metrics of neural variability across stimulus repetitions (Fig. 3.2-4). When pupil was large, local-field potential (LFP) showed a consistent shift towards greater power at high-frequency (Fig. 3.2A), and pairs of simultaneously-recorded neurons showed a small decrease in stimulus-independent correlated activity (Fig. 3.2B, n = 88 neurons forming 430 pairs, mean noise correlation = 0.03 [large pupil], 0.04 [small pupil], p = 1.1e-03, paired t-test).

Previous research also indicates that the level of pupil-indexed arousal is associated with changes in the reliability of neurons in auditory cortex, where reliability is defined as the mean cross correlation between spiking responses to the same stimulus in the same neuron across repetitions (McGinley et al., 2015a). To test for this effect, we calculated the correlation between the spiking responses of each vocalization (Fig. 3.3A-B). The mean correlation was greater during trials when pupil was large (Fig. 3.3C, n = 114 neurons, p = 5e-8, paired t-test). This difference in reliability was preserved in the subset of cells that showed modulation of neural activity by pupil-indexed state under our second-order baseline-gain model (Fig. 3.3D, n = 79 neurons, p = 5e-8, paired t-test), but disappeared when comparing only cells that did not show pupil-associated modulation under this model (Fig. 3.3D, n = 35 neurons, p = 0.2, paired t-test). These results suggest

that pupil-associated state affects both the mean rate at which auditory neurons respond to sound and the variability of their responses, and that these effects occur in the same population of cells.

Finally, dilated pupil was also associated with a decrease in a third metric of neural variability, the Fano factor (Fig 3.4). We noted a decrease in Fano factor upon stimulus onset and following large increases in sound amplitude during the stimulus (Fig. 3.4A), the latter of which was not reported in previous studies that examined the effect of static stimuli on neural variability (Churchland et al., 2010). Interestingly, the decrease in Fano factor associated with pupil size was present in both spontaneous and sound-driven activity (Fig. 3.4B, n = 112 neurons, p = 3e-05 [prestimulus Fano factor], p = 7e-08 [peristimulus Fano factor], paired t-test). Moreover, the effects of stimulus onset on Fano factor was independent of pupil size (Fig. 3.4B, n = 112 neurons, p = 0.9 [prestimulus Fano factor – peristimulus Fano factor], paired t-test), suggesting differences between the effects of pupil-indexed arousal and previously-studied effects of stimulus onset on neural variability (Churchland et al., 2010).

Non-monotonic effects of pupil on firing rate are observed in some A1 neurons

Arousal can have a non-monotonic effect on behavior: both learning (Yerkes and Dodson, 1908; Diamond et al., 2007) and task performance (Aston-Jones and Cohen, 2005; Murphy et al., 2011; van Kempen et al., 2019) are sub-optimal in minimally and maximally aroused states. A previous study found that multiunit spiking activity in mouse auditory cortex showed similarly non-monotonic relation between pupil and firing rate: on average, spontaneous activity was lowest, and gain highest, at intermediate pupil

diameter, suggesting that this range of pupil sizes was optimal for detection of auditory signals (McGinley et al., 2015a). We found examples of non-monotonic pupil-firing rate relationships in some cells (Fig. 3.5A). To test for their prevalence, we fit data on neural responses to ferret vocalizations with a segmented regression model (Simonsohn, Uri, 2018) that predicted a linear effect of pupil on neural activity but allowed the slope of the fit to take two different values depending on pupil size (Fig. 3.5B). A difference in the sign of the slope between segments indicated a non-monotonic relationship between pupil size and spiking activity. We then compared the accuracy of this model to a similar model in which slope could vary between segments, but both line segments were constrained to have the same sign (i.e., the fit could not assume the shape of a U or inverted U).

Although a large proportion of neurons that showed effects of pupil-associated state also showed a trend towards non-monotonicity in either baseline firing rate or gain (Fig. 5c, $n = 38/79$ [baseline], $n = 41/79$ [gain]), allowing for a non-monotonic segmented fit significantly improved the accuracy of the model for a smaller fraction of cells ($n = 15/79$ [baseline], 14/79 [gain], $p < 0.05$, permutation test). Among neurons that showed a trend towards non-monotonicity, more showed trends towards U in baseline firing rate than inverted U, and more showed an inverted U than U in gain (Fig. 3.5C), suggestive of previous results comparing spontaneous and sound-evoked activity in mouse A1 (McGinley et al., 2015a).

Frequency and level tuning show no or small dependence on pupil size

To examine the effect of changes in internal state on stimulus selectivity, we recorded the responses of A1 neurons to tone pips at a variety of frequencies and levels (Fig. 3.6A-C, 114 neurons from 40 recording sites in 4 ferrets). To test for changes in frequency and level selectivity, we divided the data from each cell into large-pupil and small-pupil bins based on the median pupil size observed during the recording (Fig. 3.6A-B), and constructed frequency response areas (FRAs) from the data in each bin (Fig. 3.6C). FRAs showed a variety of patterns typical of the auditory system (Bizley et al., 2005), including a sound-level response threshold and broadening of spectral bandwidth at higher sound levels (Fig. 3.6C).

Comparing data from the large and small-pupil bins showed a systematic increase in spontaneous activity and the response to tones across all levels and frequencies (Fig. 3.6B, D-E, n = 114 neurons, p = 1e-4 [spontaneous rate], p = 0.03 [driven rate], sign test). To confirm that this result did not depend on our division of the data into large and small-pupil bins, and to further examine changes in driven rate, we used linear regression to predict the response to tones based on pupil size and the average FRA of the neuron (see Methods). As it did for vocalizations, pupil size predicted trial-to-trial variability in a subpopulation of cells (Fig. 3.6F, n = 57/114 neurons, 50%, $p < 0.05$, permutation test). Most state-modulated neurons showed enhanced gain and baseline firing rate when pupil was large (Fig. 3.6G), suggesting that pupil-indexed state acted to shift and multiplicatively scale the neuron's auditory tuning curve.

Does the stimulus selectivity of auditory neurons change depending on pupil-indexed state? To test for changes in level tuning, we again divided data at the median

pupil value in each recording, then fit a hinge function (see Methods) to the characteristic-frequency rate-level function in each pupil condition (Fig. 3.7A-B). Across the population, there was a significant decrease in threshold when pupil was large (mean change: -2.65 dB, 90% credible interval: -4.00 to -1.33 dB), but no change in the slope of the rate-level function (mean change: 0.05 [spikes/s]/dB, 90% credible interval: -0.07 to 0.15 [spikes/s]/dB SPL). Consistent with previous analyses, we also observed an increase in the baseline firing rate of the cells when pupil was large (mean change: 1.68 spikes/s, 90% credible interval: 0.80 to 2.59 spikes/s). At the single-cell level, most neurons showed a significant change in baseline firing rate (n = 68/114), but few showed significant changes in slope (n = 12/114) or threshold (n = 8/114). The low number of cells showing significant changes in threshold may be due to the small effect size (-2.65 dB average across the population) and the uncertainty in model fits for individual cells.

To test for changes in frequency tuning, we fit a Gaussian function to the best-level frequency tuning curve (FTC) in each pupil condition (Fig. 3.7C-D). To isolate changes in frequency tuning (i.e., bandwidth) from nonspecific changes in baseline firing rate or gain, our model included multiplicative gain term and additive offset terms. Consistent with data from previous analysis, the gain and baseline parameters of the FTC showed a systematic increase across the population when pupil was large (gain: mean large/small ratio = 1.12, 90% credible interval = 1.06 to 1.18, baseline: mean change = 2.02 spikes/s, 90% credible interval = 1.20 to 2.75). There was no change in the spectral bandwidth of the FTC when pupil was large (mean large/small ratio = 1.01, 90% credible interval = 0.94 to 1.07). There was also no change in mean best frequency across pupil conditions (mean change = -0.02 octaves, 90% credible interval = -0.09 to 0.05). The

number of neurons that showed a significant change in each parameter was greater for gain (n = 27/114) and baseline firing rate (n = 78/114) than it was for bandwidth (n = 14/114) or best frequency (n = 4/114). Thus, while A1 neurons did sometimes show pupil-dependent changes in response threshold or other tuning parameters, these changes were relatively small compared to the changes in baseline activity and response gain.

Sleep states account for additional neural variability

Although we did not explicitly seek to study sleep state, during some recordings we observed periods of tonically constricted pupil accompanied by an increase in saccade-like eye movements (Fig. 3.8A-B, Movies 3.1 and 3.2). We initially speculated that these bouts represented rapid eye-movement (REM) sleep, based on a previous report correlating constricted pupil and REM sleep in mice (Yuzgec et al., 2018). However, the delta/theta ratio of local-field potential (LFP) – a signature of REM sleep (Yuzgec et al., 2018) – did not show a systematic change between putative REM bouts and recording segments with pupil size in the same range (p = 0.8, n = 26 recordings, sign-rank test). Instead, the increase in eye movements was accompanied by a decrease in alpha power (Fig. 3.8C, p = 6e-09, n = 26 recordings, sign-rank test), suggesting that the eye movements indicated sleep onset (Silber et al., 2007).

Given that neural responses to sound in auditory cortex are preserved during natural sleep states, but sometimes suppressed or enhanced (Brugge and Merzenich, 1973; Issa and Wang, 2008), we wondered if these brief sleep episodes might show changes in firing rate distinct from those associated with changes in pupil size. We therefore fit linear regression models that included putative sleep state, pupil size, or both

sleep state and pupil size as predictors of neural activity (Fig. 3.8D-E and Table 3.2). Sleep state significantly predicted neural activity in 81% of neurons recorded in experiments that included one or more sleep episodes (n = 38/47 neurons). Across the population of neurons with a significant fit for any tested model (n = 44), sleep state predicted less variability than pupil, which in turn predicted less variability than both pupil and sleep state (Fig. 3.8D and Table 3.2, median change in accuracy over PSTH-only model: 1.5% [sleep], 8.4% [pupil], 11% [pupil and sleep]). There was no correlation between the duration of sleep during the recording and the improvement in accuracy of the sleep-based model (r = 0.2, p = 0.16, t-test), suggesting that sleep effects vary across neurons.

In some neurons, visual inspection of spiking activity and pupil state indicated that sleep episodes were associated with a non-monotonic change in firing rates (Fig. 3.8D). To quantify this effect, we compared trials recorded during sleep episodes to non-sleep trials falling in an equivalent range of pupil sizes (Fig. 3.9A-B). Across the population, there was no significant difference in spontaneous rate, sound-evoked rate, or reliability between these two conditions (Fig. 3.9C, p = 0.16 [spontaneous rate], p = 0.38 [evoked rate], p = 0.17 [reliability], sign test, n = 47 neurons with recordings that included sleep episodes). However, a subpopulation of neurons did show a difference in spontaneous or sound-evoked rates at the single-cell level (Fig. 3.9C, n = 7/47 [spontaneous rate], n = 15/47 [evoked rate], $p < 0.05$, unpaired t-test). To further examine the effect of sleep state, we removed data from trials recorded during sleep episodes from the dataset and refit the segmented regression model initially used to test for non-monotonic effects. Fitting to data that excluded sleep state reduced the number of neurons

that showed non-monotonic effects in both baseline and gain (Fig. 3.9D), indicating that sleep accounted for non-monotonic effects in some neurons (n = 7/42 [change in non-monotonicity of baseline effect], n = 6/42 [change in non-monotonicity of gain effect], n = 42 neurons with recordings that included sleep episodes and a significant fit for second-order baseline-gain model).

Sound-selective pupil dilations

To test whether the presentation of sound stimuli themselves affected the animals' internal state, we aligned pupil traces to sound onset and calculated the mean pupil dilation in response to ferret vocalizations (Fig. 3.10A). We observed a sound-selective pupil dilation in response to one vocalization during the 2.5 seconds of silence following sound offset (Fig. 3.10B-C, n = 46 recordings, p = 0.1 [vocalization-related difference in peristimulus pupil size], p = 6e-03 [poststimulus pupil size], paired t-test). These sound-evoked pupil dilations did not predict neural activity with greater accuracy than the raw pupil trace, suggesting that they do not influence neural activity in a distinct way from slower, intrinsically generated fluctuations in pupil-diameter (Fig. 3.10D, n = 81/114 neurons, p = 0.5, sign test).

Discussion

Pupil size is an indicator of central neuromodulatory processes related to arousal that affect processing in sensory cortex (McGinley et al., 2015b; Reimer et al., 2016). Our results show that pupil size is correlated with changes in the gain and baseline firing rate of neurons in the primary auditory cortex (A1) of non-anesthetized ferrets. Non-monotonic effects of changes in pupil-indexed state were observed in some neurons, but the majority of effects were monotonic and showed a positive correlation between pupil size and spike rate. Across our population of recorded neurons, pupil size was also correlated with increases in the reliability of responses to sound, small decreases in acoustic threshold, and no change in spectral bandwidth or best frequency. The changes in gain that we observed suggest that pupil size tracks the gross level of activity evoked by auditory stimuli: sounds become more salient to the rest of the brain when pupil is large. Changes in threshold and decreased neural variability may also support a more sensitive or precise representation of sound in high-arousal states.

Stability of sensory tuning across pupil states

In mouse visual cortex, single neurons' orientation selectivity increases when pupil is dilating rather than constricting: the neurons' response to their preferred direction, but not the orthogonal direction, increases (Reimer et al., 2014). Although the impact of this change in orientation selectivity on tuning bandwidth has not been quantified, it suggests a narrowing of orientation tuning in the mean response across the population (Reimer et al., 2014). Our observation of stable tuning bandwidth in A1 may

therefore indicate differences between how pupil-indexed states affect sensory selectivity in visual and auditory cortex.

Our results are also relevant to previous work on behavioral modulation of sensory receptive fields. Studies of task engagement effects on A1 reveal enhancement or suppression of neural responses to task-relevant sound features, including frequency (Fritz et al., 2003; David et al., 2012), amplitude modulation (Niwa et al., 2012), and spatial position (Lee and Middlebrooks, 2010), as well as generalized changes in excitability (Miller et al., 1972; Otazu et al., 2009). We found that best frequency is stable across large changes in pupil size, in contrast to previous studies of task-related plasticity in ferret primary auditory cortex that show frequency-specific effects (Fritz et al., 2003; David et al., 2012), suggesting a key difference between tasks that involve manipulation of the behavioral relevance of specific sound features and uncontrolled variation within passive states. Our data therefore suggest that effects of task engagement on frequency tuning in primary auditory cortex are not the result of non-specific increases in arousal, but instead involve separate mechanisms, such as feedback from frontal cortex (Fritz et al., 2010) or pairing activation of neuromodulatory systems involved in pupil dilation with specific auditory stimuli (Froemke et al., 2007; Martins and Froemke, 2015). The random sequences of tones used to characterize tuning would not pair release of acetylcholine with specific tone frequencies, and therefore would not be expected to gate shifts in tuning (Weinberger, 2004).

Changes in neural excitability and variability

We found an increase in both spontaneous and sound-evoked firing rate when pupil was large. In addition, we found shifts in multiple metrics of neural variability, including Fano factor, reliability, and noise correlations, as well as a shift towards desynchronized neural activity associated with high-arousal states (Harris and Thiele, 2011). Our results therefore support previous studies that found an influence of pupil-indexed arousal on encoding of stimuli in early sensory cortex (McGinley et al., 2015a; Vinck et al., 2015).

A previous study in mouse auditory cortex found that intermediate pupil diameter was associated with maximum evoked responses to sound and minimal spontaneous activity (McGinley et al., 2015a). Although we observed some cells with non-monotonic relationships between pupil size and spike rate, this was the not the predominant pattern in our sample. In addition, we found that the effect of pupil-associated state usually had the same sign for both spontaneous and sound-evoked activity within a single neuron, in contrast to earlier results suggesting that intermediate pupil sizes were associated with opposite changes in spontaneous and evoked activity (McGinley et al., 2015a).

Several factors could explain this inconsistency between studies. The difference could reflect sampling of cortical layers. The earlier study targeted layers 4/5. We did not target any particular layer, but our method of recording auditory cells as the electrode advanced through cortex may have biased our sample towards more superficial layers. It is also possible that, compared to the earlier study, we tended to sample a different range of cortical states, either due to differences in experimental technique (the earlier study recorded head-restrained animals on a treadmill, while we did not use a treadmill) or

differences in the behavioral patterns of ferrets and mice. Ferrets were not able to run and rarely showed substantial motor activity during the recordings. Thus, they may not have achieved the very high arousal state observed during bouts of running and other motor activity in mice.

Pupil as an index of arousal or cognitive engagement

We used absolute pupil size as a measurement of cortical state. Because the variations in pupil size we observed occurred outside a controlled behavior, we have characterized them as a measurement of physiological arousal rather than inferring changes in a particular cognitive state. However, mechanisms like those we observed may underlie correlations between absolute pupil size and the efficiency of responses to sensory stimuli in some tasks (Beatty, 1982b; Murphy et al., 2011; van Kempen et al., 2019). Our work complements human pupillometry's traditional focus on small, rapidly-decaying changes in pupil size that coincide with behavioral events (Kahneman and Beatty, 1966; Beatty, 1982a). The amplitude of these task-evoked dilations depends on a variety of cognitive variables (Zekveld et al., 2018), some of which may be related to the sound-selective pupil dilation we observed in response to pup cries over adult aggression calls. Work on sound-evoked pupil dilation in owls indicates that pupil adapts to repeated sounds (Bala and Takahashi, 2000). Studies that use a larger number of stimuli with fewer sound repetitions per session may therefore provide more insight into neural correlates of sound-evoked pupil dilation, as may studies that use controlled behaviors rather than passive listening.

Pupil as an index of noradrenergic tone

Activity in neuromodulatory centers, particularly the noradrenergic and cholinergic systems, has been proposed as a mechanism underlying correlations between pupil size and the level of neural activity in sensory cortex (McGinley et al., 2015b; Reimer et al., 2016). Evidence from multiple labs, species, and experimental techniques indicates that central release of noradrenaline from locus coeruleus is causally involved in pupil dilation and that noradrenaline release in sensory cortex accompanies pupil dilation (Aston-Jones and Cohen, 2005; Murphy et al., 2014; Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017; Liu et al., 2017; Lovett-Barron et al., 2017; Larsen et al., 2018).

Despite the substantial evidence linking pupil size to noradrenergic tone, the gain increases we observed during pupil dilation do not directly match in vivo measurements of the effects of noradrenaline in auditory cortex. Iontophoresis of noradrenaline increases the signal-to-noise ratio of some auditory cortex neurons via a decrease in spontaneous, but not sound-evoked, activity (Foote et al., 1975; Manunta and Edeline, 1997, 1999). The gain increases we observe persist after subtracting mean spontaneous activity and are present when analyzing data from the evoked period alone. It is possible that iontophoresis does not replicate the spatial distribution or temporal dynamics of noradrenaline release in non-anesthetized animals. It is also possible that the gain increases we observed were a result of multiple modulatory systems acting on auditory cortex, both directly and indirectly. Activity in the cholinergic and dopaminergic systems are also correlated with changes in pupil size (Reimer et al., 2016; de Gee et al., 2017; Larsen et al., 2018). Given that pupil size is correlated with activity in multiple

neuromodulatory systems, our failure to directly replicate results from spatially-localized, single-neuromodulator experiments is unsurprising, and indicates the complexity of neuromodulation under more natural conditions.

Involvement of other cortical and subcortical brain regions

Some of the trial-to-trial variability we observed in auditory cortex may also be due to feedback from motor cortex (Schneider et al., 2014). Pupil size varies even when mice are still, and explains more variability in neural activity than locomotion (McGinley et al., 2015a). However, in non-anesthetized mice motor activities such as running, whisking, and licking a water reward are associated with pupil dilation (Lee and Margolis, 2016; Stringer et al., 2019). Detailed tracking of face and body movements explains more variability than pupil alone in mouse visual cortex (Stringer et al., 2019) and dorsal cortex (Musall et al., 2019). We did not track detailed motor behavior in the current study, and thus the extent to which motor activity predicts trial-to-trial variability in our data is not known. However, our results add to the evidence that pupil is one of several variables related to movement that can be used to predict cortical activity.

Pupil-related changes in neural activity are not restricted to the cortex. The same study that showed evidence of a causal relationship between pupil dilation and LC activity showed a similar relationship between pupil dilation and activity in superior and inferior colliculus (IC), including spiking preceding pupil dilation by hundreds of milliseconds and pupil dilation induced by microstimulation (Joshi et al., 2016). Some of the effects of arousal state we observed are likely to be inherited from auditory signals which must pass through IC before reaching cortex.

Sleep states during head-restrained recordings

Recent results in head-restrained mice showed fluctuations between awake, rapid eye-movement (REM) sleep, and non-REM sleep states associated with changes in pupil size, with constricted pupil associated with REM sleep (Yuzgec et al., 2018). We observed epochs of tonically constricted pupil accompanied by an abrupt increase in saccade rate. Although alpha power consistently increased as pupil constricted, this trend reversed itself during these epochs. We speculate that over the course of lengthy, head-restrained passive recordings, the ferrets may become progressively more drowsy, and that the high-saccade, low-alpha epochs indicate sleep onset.

The pattern of constricted pupil, saccades, and lack of eyelid closure we observed may also be an artifact of head-restrained recording. Methods for recording the activity of single neurons in freely-moving animals can be supplemented with head-mounted cameras to track pupil (Meyer et al., 2018). It is likely that these techniques will yield data on changes in internal state more relevant to natural environments.

Our observation of putative sleep states, like previous data from mice (Yuzgec et al., 2018), complicates the suggestion that the pupil size of head-restrained animals reveals a continuum of awake arousal states analogous to, but distinct from, sub-states of sleep (McGinley et al., 2015b). Thus, a full characterization of arousal states must incorporate both pupillometry and other physiological measurements that can distinguish between drowsy waking states and sleep.

Figures

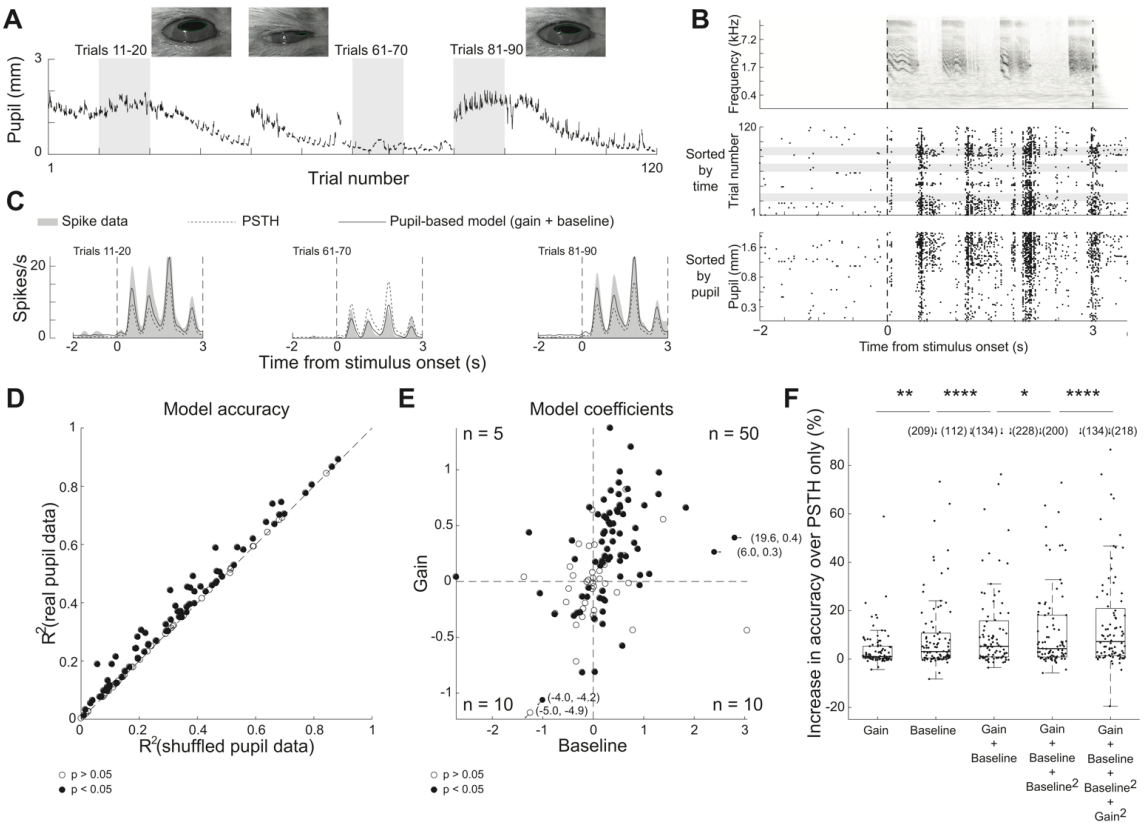

Fig. 3.1: Trial-to-trial variability in neural responses to sound tracks changes in pupil size. **A.** Schematic of experiment, illustrating ferret, free-field speaker, camera recording video of eye, and contralateral extracellular electrode. **B.** Pupil trace recorded across 120 repetitions of one ferret vocalization, with example video frames. **C.** Spectrogram of one ferret vocalization and spike raster of one neuron's response to the vocalization, sorted by time (top) and prestimulus pupil size (bottom). **D.** Predicted and actual pupil-dependent changes in spiking activity for the neuron shown in 1b-c. Panels show peristimulus time histogram (PSTH) responses averaged across blocks of 10 stimulus presentations selected

from epochs with different pupil size (see 1b. PSTHs predicted by the null model (dashed) and baseline + gain model (solid) are overlaid on the actual PSTH (gray shading). **E.** Accuracy of pupil-based baseline-gain regression model for neural responses to ferret vocalizations (n = 114 neurons), plotted against accuracy of a control, PSTH-only model, fit to temporally shuffled pupil data from each cell. Filled dots indicate neurons with a significant improvement over the control model (n = 75/114, $p < 0.05$, permutation test). **F.** Coefficients of baseline-gain model for all data from vocalization recordings (n = 114 neurons). Positive values indicate an increase in baseline spike rate (horizontal axis) or response gain (vertical axis) with larger pupil. Numbers in corners of each quadrant indicate the count of neurons with parameters with a significant improvement in prediction accuracy over the control, pupil-independent model in that quadrant. **G.** Improvement in accuracy over PSTH-only model for various models for population of neurons with a significant fit for any model (n = 92/114, *: $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$, sign test). Boxplot indicates median, interquartile range, and 1.5 times the interquartile range.

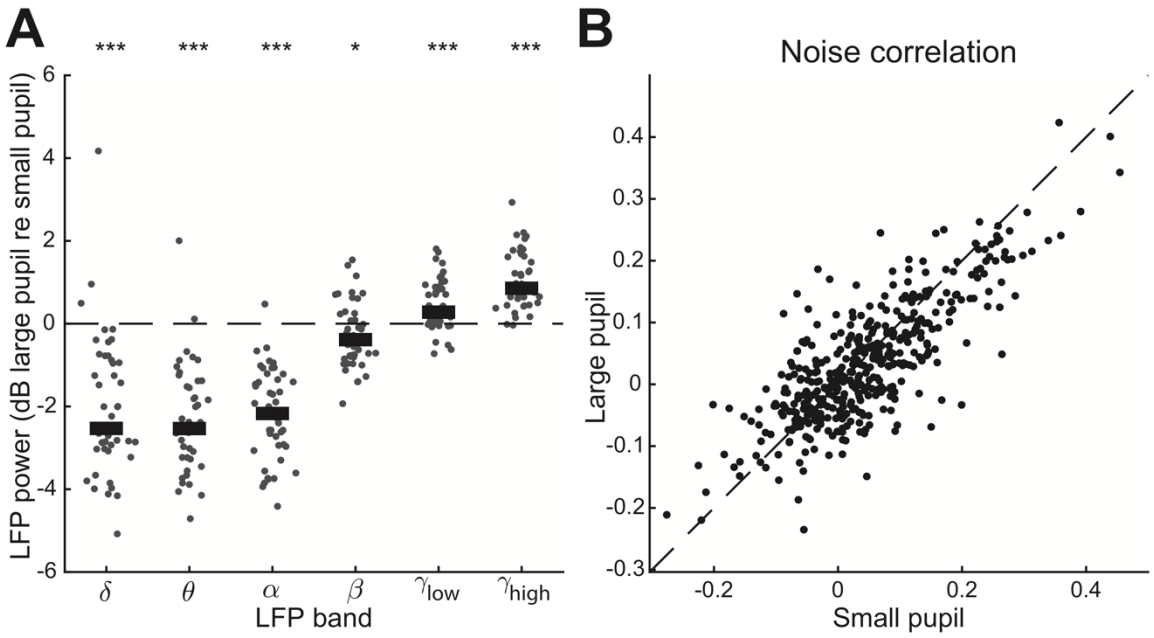

Fig. 3.2: Dilated pupil is associated with neural desynchronization. **A.** Log ratio change (dB) in local field potential (LFP) power between recording segments with large and small pupil (n = 46 recordings). Thick line indicates population median (*: $p < 0.05$; ***: $p < 0.001$, rank-sum test on hypothesis that median is equal to 0). Putative sleep states (see Fig. 8) have been excluded from the comparison. **B.** Change in noise correlation across large and small pupil trials (n = 429 neuronal pairs, mean noise correlation = 0.03 [large pupil], 0.04 [small pupil], p = 1.1e-03, paired t-test). Trials are classified as "large pupil" and "small pupil" based on whether the mean pupil size preceding the vocalization is greater or less than the median pupil size during the recording.

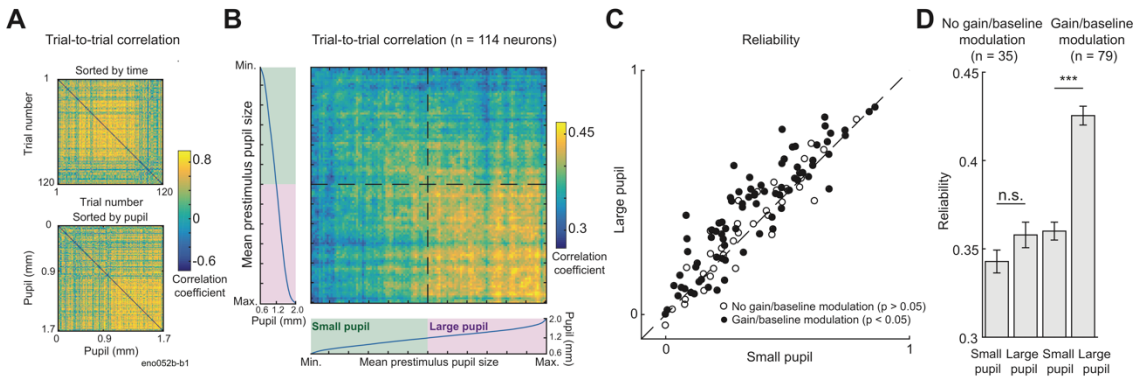

Fig. 3.3: Reliability of neural response to sound increases when pupil is dilated. **A.** Correlation between time-varying activity across trials evoked by one ferret vocalization in one neuron, sorted by trial order (top) and mean pupil size before stimulus onset (bottom). **B.** Mean trial-to-trial correlation for all neural responses to ferret vocalizations. This heat map was constructed by computing the average correlation matrix sorted by pupil size across all neurons (n = 114). For display, the correlation of each trial with itself was replaced by the mean correlation for the two trials with the most similar pupil size before averaging. Marginal plots indicate the mean pupil size in each trial, across all neurons. Dashed line indicates median pupil size. **C.** Comparison of reliability (mean trial-to-trial correlation) for each neuron's response to vocalizations for small versus large pupil (n = 114, p = 5e-8, paired t-test). Trials are classified as "large pupil" and "small pupil" based on whether the mean pupil size preceding the vocalization is greater or less than the median pupil size during the recording. Filled circles indicate cells with a significant fit under a second-order baseline-gain regression model (n = 79, $p < 0.05$, permutation test, p = 5e-8, paired t-test). **D.** Mean reliability (+/- SEM) in each condition for subpopulations of cells that do or do not show effect of pupil-associated state under the regression model (***: $p < 0.001$, paired t-test).

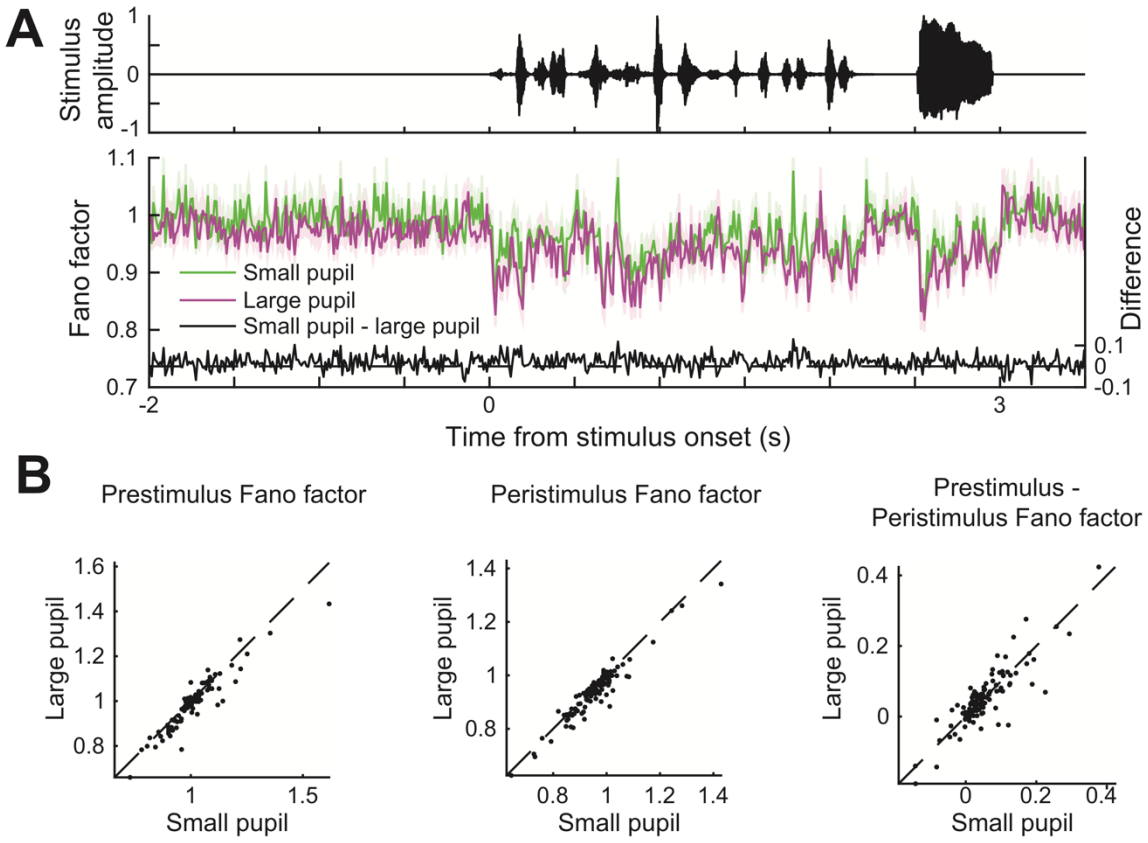

Fig. 3.4: Fano factor decreases at stimulus onset and when pupil is dilated. **A.** Waveform of one ferret vocalization (top) and time-varying Fano factor for neural responses (bottom). Purple and green lines indicate mean across 112 neurons +/- SEM. **B.** Mean Fano factor preceding and during each vocalization (n = 112 neurons, p = 3e-05 [prestimulus Fano factor], p = 7e-08 [peristimulus Fano factor], paired t-test).

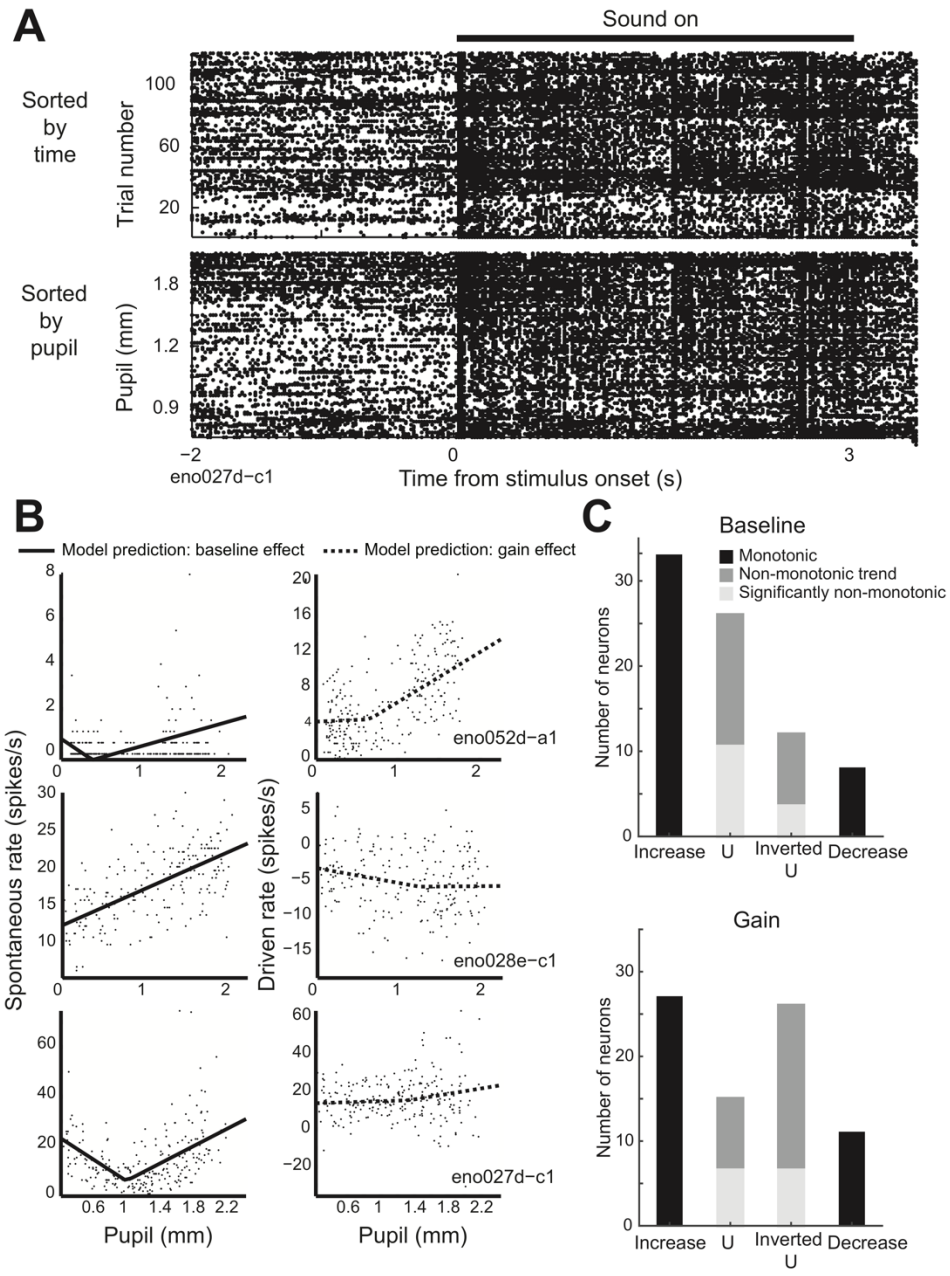

Fig. 3.5: Non-monotonic effects of pupil-related state. **A.** Spike raster of one neuron's response to single ferret vocalization sorted by time (top) and pupil size preceding the stimulus (bottom). A non-monotonic relationship is evident between pupil size and spontaneous firing rate. **B.** Examples of segmented regression model fits to spontaneous and driven activity for three neurons. Each point represents spike rate for one neuron on

one trial (spontaneous activity: 2 s, driven activity: 3 s). The bottom row displays data from the neuron in panel A. **C.** Results of test for non-monotonicity using segmented regression model, counting the number of neurons with significant or trends toward non-monotonic changes in baseline or gain.

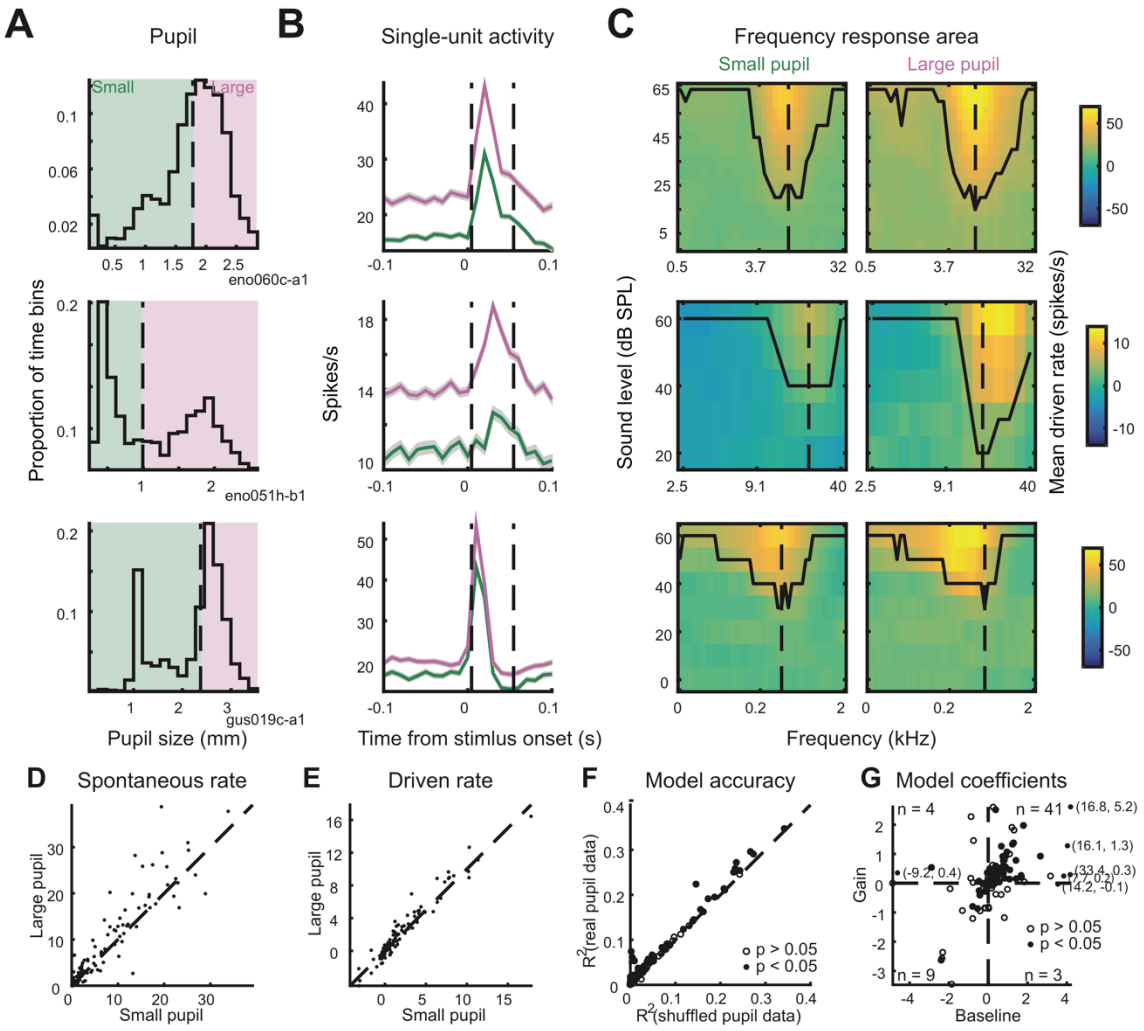

Fig. 3.6: Spontaneous activity and evoked response to tones increases when pupil is dilated. **A.** Distribution of pupil size during three recordings of neural responses to tone pips, indicating division into small-pupil and large-pupil bins based on median pupil size during the recording. **B.** Mean peristimulus time histograms (PSTHs) of neural responses for the three example cells, computed separately for data in the small-pupil (green) and large-pupil (purple) bins and averaged across all tone pip levels and frequencies. Shading indicates SEM. Dashed lines indicate window for measurement of response to sound.

C. Frequency-response areas for the three examples cells, computed separately for small and large pupil bins. Contour indicates lowest level that shows an auditory response at each frequency. Dashed line indicates characteristic frequency. **D-E.** Spontaneous and driven rate for all neurons' responses to tone pips (n = 114 neurons). **F.** Single-trial prediction accuracy of linear baseline-gain model and control, pupil-independent model for all neurons' response to tone pips. Filled dots indicate neurons with a significant improvement for the baseline-gain model (n = 57/114, $p < 0.05$, permutation test). **G.** Coefficients of baseline-gain model fit for all data from tone-pip recordings. Numbers indicate count of neurons with significant improvement for the pupil-dependent model in each quadrant.

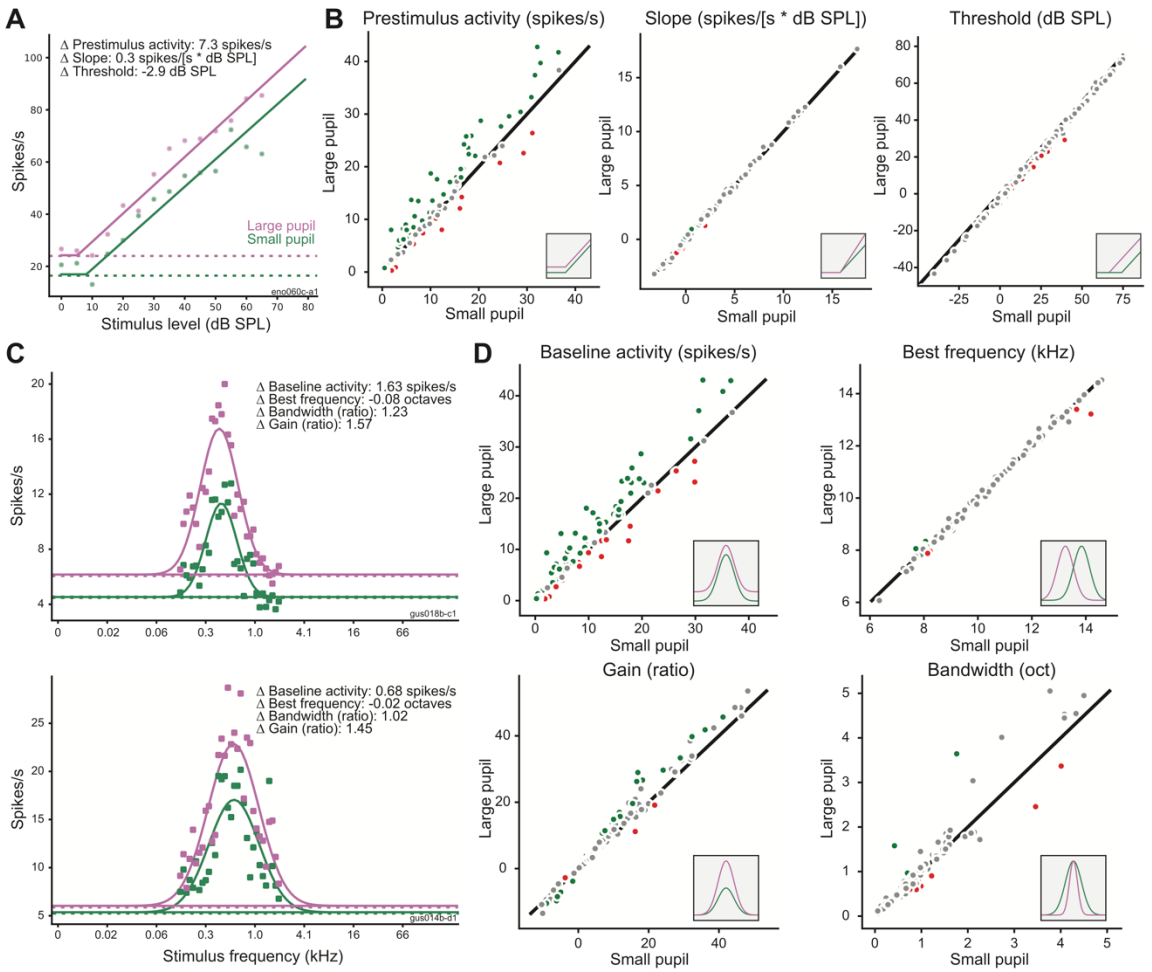

Fig. 3.7: Effects of pupil-related state on frequency and level tuning. **A.** Example of hinge function fit to neural response at characteristic frequency for large-pupil (purple) and small-pupil (green) conditions. Dashed lines indicate prestimulus firing rate in each condition. **B.** Comparison of each parameter of rate-level function between pupil conditions (large pupil vs. small pupil) for all neurons (n = 114). Inset cartoons illustrate effect of changing each parameter fit. Green and red dots indicate neurons with a significant change in the tuning parameter (as defined by the 90% credible interval for difference between conditions not bracketing 0). Gray dots indicate neurons with changes that were not significant at the single-cell level. **C.** Examples of Gaussian functions fit to

neural responses at best level for large-pupil (purple) and small-pupil conditions (green). **D.** Comparison of each parameter of the frequency-tuning curve across pupil conditions for all neurons (n = 114). Insets and color coding as in panel B.

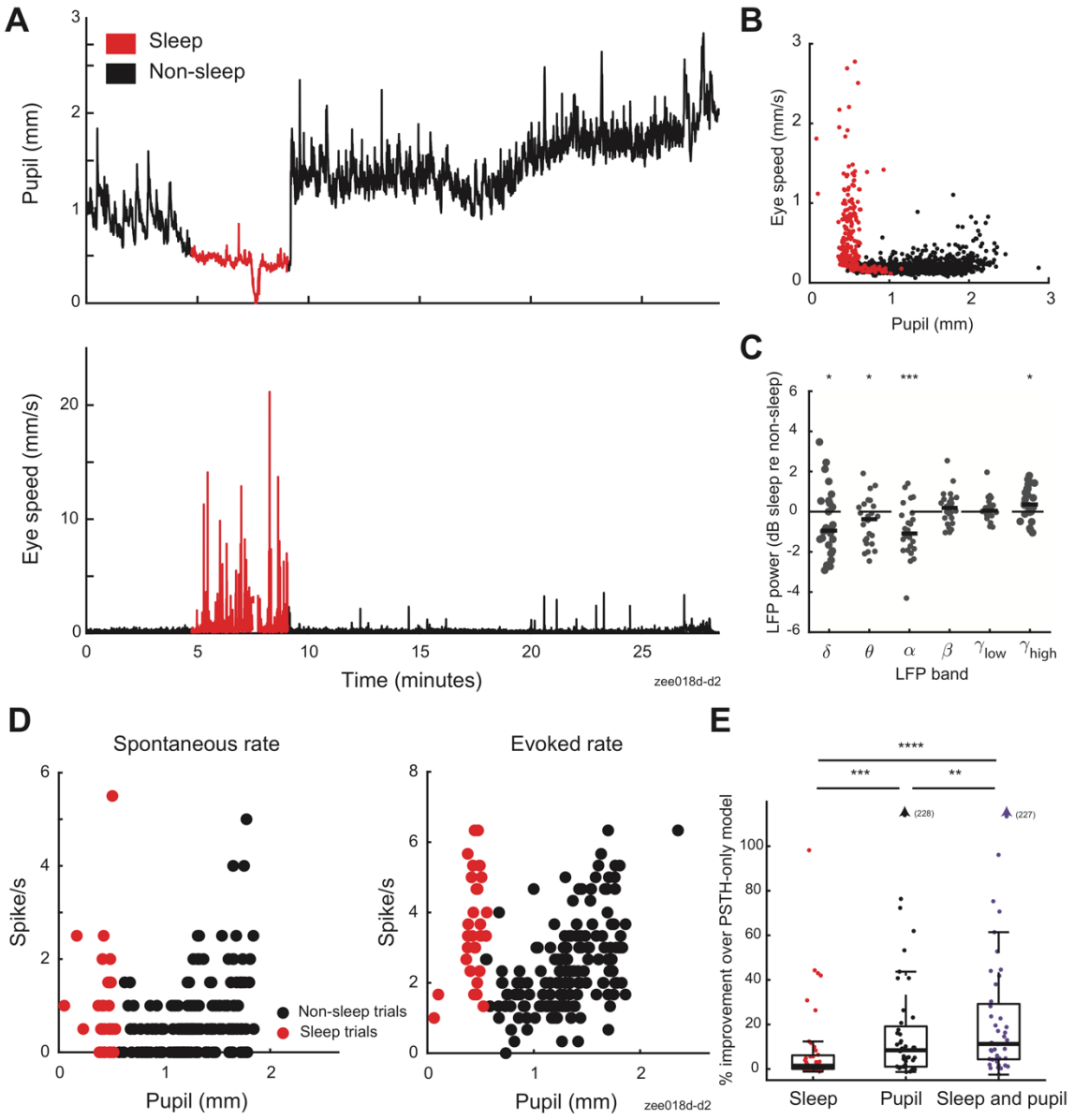

Fig. 3.8: Sleep states account for additional variability in neural responses. **A.** Pupil size and eye-speed traces from one recording. **B.** Pupil size and eye speed data from one ferret (13 recordings). Each point illustrates average pupil size and eye speed during one trial (5.5 s). Color indicates whether a trial was classified as sleep (red) or non-sleep (black), using criteria based on eye movement and pupil size data (see Methods). **C.** Relative

change in LFP power between recording segments with and without sleep, matched for pupil size (n = 26 recordings). Bold line indicates population median (*: $p < 0.05$, ***: $p < 0.001$, rank-sum test on hypothesis that median is equal to 0). **D.** Spontaneous (left) and sound-evoked neural data (right) from the recording shown in A. Each point indicates spike rate before or during one vocalization (prestimulus period = 2 s, stimulus period = 3 s). **E.** Stepwise regression results for models including sleep state, pupil size or both sleep state and pupil size as predictors of spike rate (n = 44 neurons with significant fit for any model and recorded during experiments that include sleep episodes, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$, sign test). Boxplot indicates median, interquartile range, and 1.5 times the interquartile range.

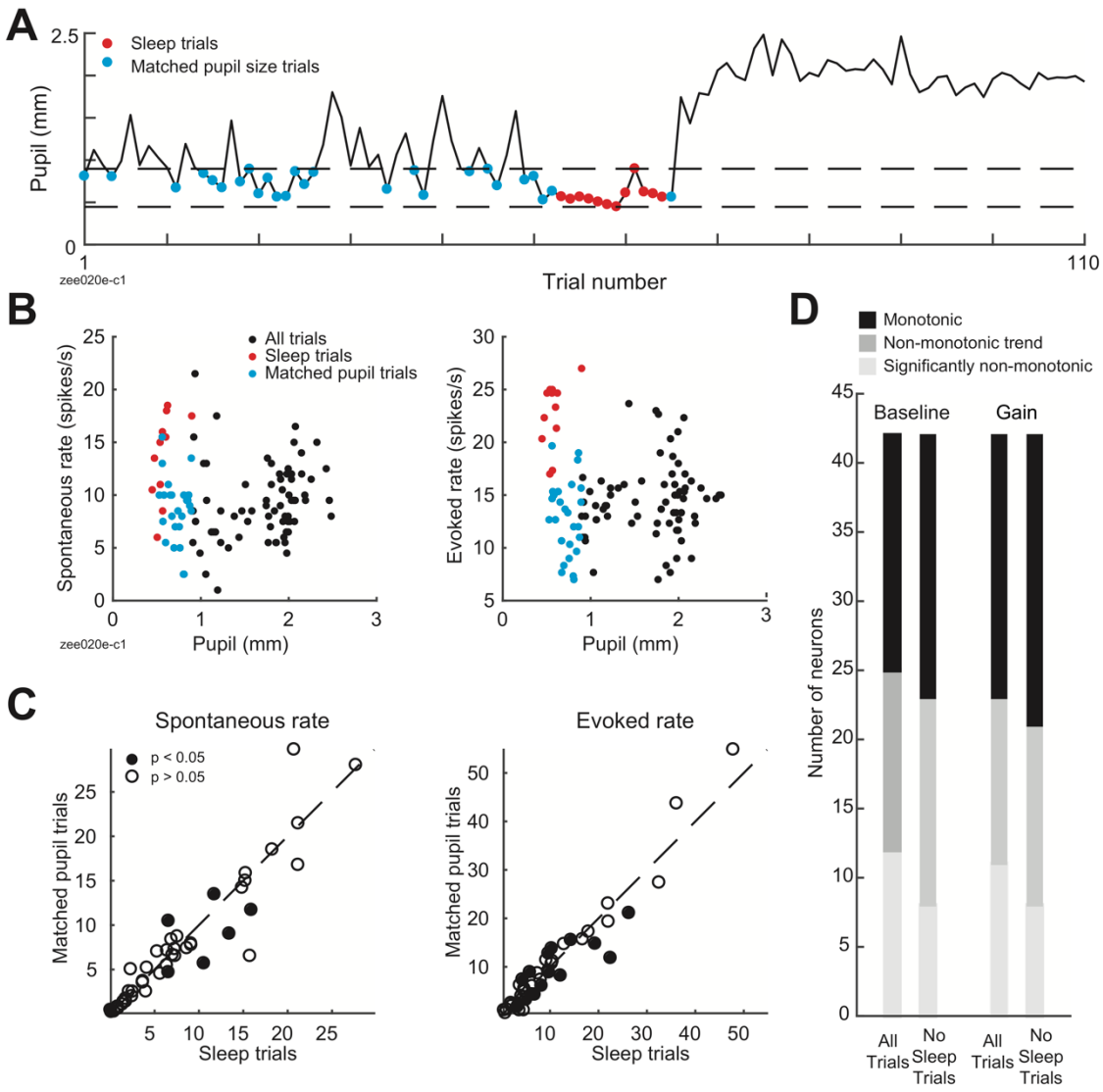

Fig. 3.9: Sleep state accounts for non-monotonicity in some neurons. **A.** Mean prestimulus pupil size for responses to one ferret vocalization, highlighting sleep trials (red) and trials with matched prestimulus pupil size (blue). **B.** Spontaneous (i.e. prestimulus) and stimulus-evoked neural activity from the recording in A, illustrating greater spike rate in sleep trials as compared to trials with matched pupil size. **C.** Average spontaneous (left) and sound-evoked spike rate (right) during sleep trials versus trials

with matching pupil size ($n = 47$ neurons recorded in experiments including sleep episodes, filled circles: $p < 0.05$, unpaired t-test). **D.** Results of test for non-monotonicity using segmented regression model fit to all data or data with sleep trials removed ($n = 42$ neurons recorded in experiments including sleep episodes, and showing significant fit to second-order baseline-gain regression model).

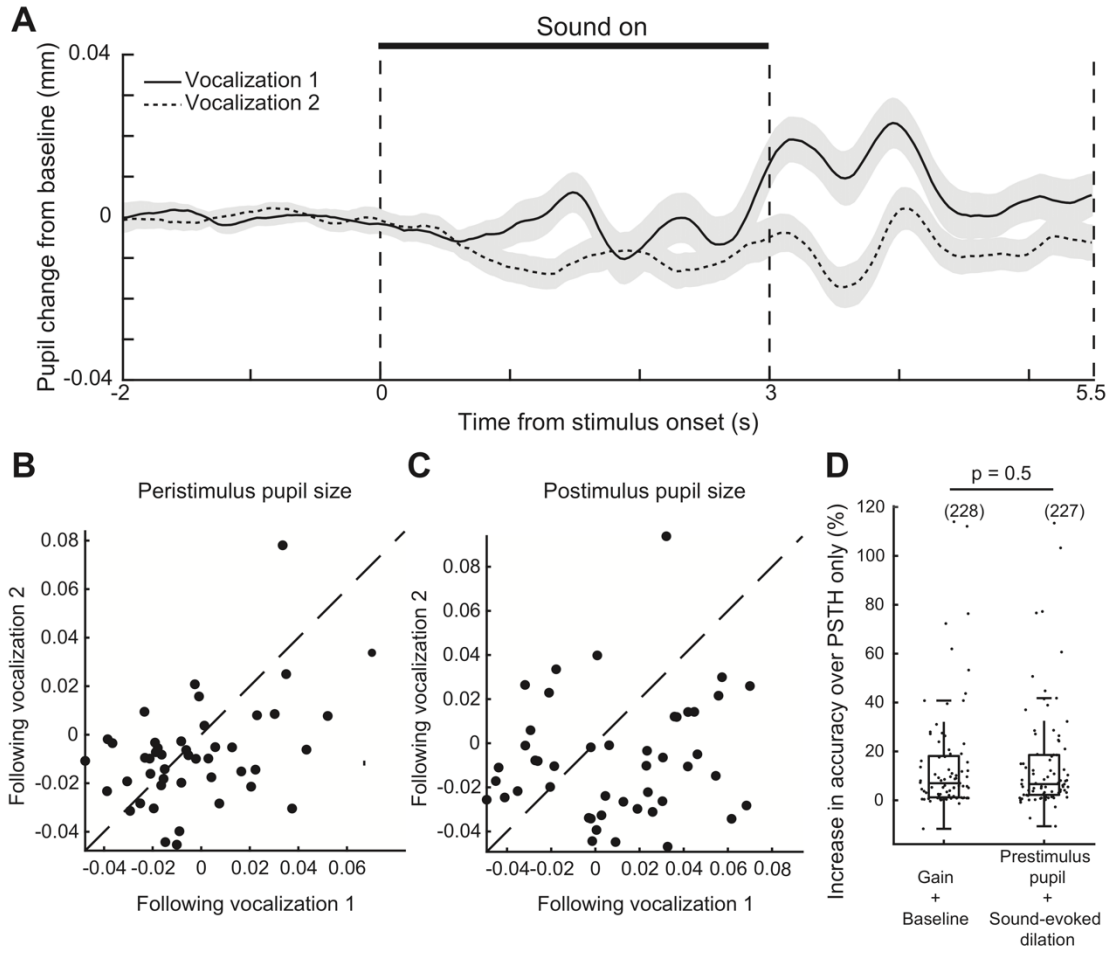

Fig. 3.10: A sound-selective pupil dilation does not affect neural activity independent of other variation in pupil size. **A.** Mean pupil dilation in response to ferret vocalizations (n = 46 recordings). Gray shading indicates SEM. **B.** Mean pupil size during each 3-second ferret vocalization. **C.** Mean pupil during the 3 seconds following each vocalization. **D.** Accuracy of regression models fit to the full pupil timecourse (baseline-gain model, see Fig. 2) or to the prestimulus pupil size on each trial and the sound-evoked dilation or constriction in each time bin. Filled dots indicate neurons with a significant fit to either model (n = 79/114, $p < 0.05$, permutation test).

Table 3.1. Stepwise regression of pupil-based models of neural activity

Baseline	Gain	Baseline-gain	2nd-order baseline	2nd-order baseline-gain	
-0.009****	**-0.009******	**-0.02******	**-0.01******	**-0.02******	Null
	0.0001**	**-0.006******	**-0.002*****	**-0.01******	Baseline
		-0.006****	**-0.002****	**-0.01******	Gain
			-0.004*	**-0.004*****	Baseline-gain
				-0.008****	2nd-order baseline

Table 3.1: Stepwise comparison of accuracy of different pupil-based models of neural activity applied to vocalization dataset. Numbers indicate difference in median accuracy across the population of neurons with significant fit for any model (n = 92/114). Bold text indicates comparisons for which the difference is significant at alpha level 0.05 with Bonferroni correction for multiple comparisons (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$, sign test).

Table 3.2. Stepwise comparison of pupil- and/or sleep-based models of neural activity

Sleep	Pupil (baseline-gain)	Sleep and pupil	
-0.2****	**-0.03******	**-0.04******	Null
	-0.01***	**-0.02******	Sleep
		-0.006**	Pupil (baseline-gain)

Table 3.2: Stepwise comparison of accuracy of pupil and/or sleep-state based models of neural activity (see Methods) applied to all vocalization data including sleep episodes (n = 47 neurons). Text indicates median accuracy for population of neurons with significant fit for any model (n = 44/47 neurons). Bold text indicates that the difference is significant at alpha level 0.05 with Bonferroni correction for multiple comparisons (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$, sign test).

Movie 3.1. Real-time example of pupil size and eye movement dynamics during awake state. Video available at https://doi.org/10.6084/m9.figshare.9859358

Movie 3.2. Real-time example of pupil size and eye movement dynamics during sleep state. Video available at https://doi.org/10.6084/m9.figshare.9859358

Chapter 4: Conclusions

Sensory processing is traditionally studied by mapping the relationship between the physical features of stimuli and neural activity (Hubel and Wiesel, 1959; deCharms et al., 1998). In auditory cortex (AC), like other sensory systems, neural activity is also modulated by non-sensory variables, including the demands of tasks that require responding to sound (Fritz et al., 2003; David et al., 2012; Niwa et al., 2012; Osmanski and Wang, 2015). Such task-related plasticity is rapid, reversible, and encodes information not present at the auditory periphery. Thus, it suggests that the response properties of single neurons in AC can be modified to some degree by the large-scale network state of the brain at a given time, supporting behavioral flexibility.

This dissertation explored how two aspects of behavioral state affect the brain's representation of sound: attention and arousal. We recorded extracellular single-unit and population activity in primary auditory cortex (A1) of ferrets during passive listening as well as a behavioral task that involved attention to frequency, spatial, and envelope cues. In Chapter Two, we compared neuronal responses across attention conditions in ferret AC. We recorded single-neuron activity in A1 of ferrets as they were rewarded for attending to tones in bandpass noise, but not tones in bandpass noise at a different frequency. The neural response to the noise at the attended frequency decreased when ferrets attended to frequencies to which the neuron was tuned. Because neural responses to the tones did not differ between attention conditions, suppressing the neural responses to noise increased the discriminability between neural responses to tones and noise. Comparing neural responses when ferrets performed the task to those in which they

passively listened to the same sounds showed an increase in spontaneous activity as well as an increase in neural responses to noise.

In Chapter Three, we compared neuronal responses in ferret AC across arousal states. Building on recent findings in mice (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015), we used pupil size as a measurement of moment-to-moment fluctuations in arousal (Satpute et al., 2019). As in prior studies in mice, we found that arousal modulates the gain and baseline firing rate of A1 neurons' response to ferret vocalizations and tones. Although neurons showed a variety of relationships between pupil size and firing rate, the most commonly observed relationship in our dataset was a monotonic increase in gain, spontaneous activity, and the reliability of sound-evoked responses when pupil dilated. Frequency tuning (best frequency and bandwidth) was stable across arousal states, but there was a small decrease in auditory threshold when pupil was large. During some recordings, we noted an abrupt increase in saccades and tonically constricted pupil that may have indicated sleep onset. In some neurons, this state was accompanied by distinctive changes in firing rate.

A recent study that used 2-photon Ca2+ imaging to examine frequency tuning in mouse A1 across pupil size observed no change in best frequency, but a small increase in spectral bandwidth when pupil was large (Lin et al., 2019). Surprisingly, the bandwidth increase selectively affected frequencies greater than best frequency. It is not clear to what extent the divergence between this result and our own is due to species differences or different analytic methods for separating changes in gain and bandwidth.

Separating out effects of attention, arousal, and task engagement

Plasticity in the responses of auditory neurons has been demonstrated following behavioral training, physiological manipulations, and during behavioral tasks (Weinberger, 2004; Schreiner and Polley, 2014; Elgueda et al., 2019). In order to understand the role of cortical plasticity in hearing, it is important to distinguish effects of task structure on neural activity from non-stimulus-specific variables such as task engagement (Otazu et al., 2009), motor activity (Schneider et al., 2014; Musall et al., 2019; Stringer et al., 2019), and cortical network state (Harris and Thiele, 2011). In particular, because most reports of task-related plasticity compare neural activity during behavior and passive listening, it is difficult to completely dissociate the effect of attention to particular acoustic features from nonspecific arousal and task engagement. Arousal and both attention both desynchronize neurons and increase stimulus-evoked responses (Harris and Thiele, 2011). Therefore, failure to distinguish these effects could result in falsely attributing attentional effects to brain regions that are affected by sensory input and arousal.

Our results suggest that attention and task structure produce localized changes in frequency tuning in A1, dependent on task-relevant sound features, while arousal produces non-task-specific changes in gain and baseline activity. In addition, we found that task engagement modulates spontaneous activity independent of selective attention. However, these experiments have not directly compared effects of arousal to task engagement and attention by recording pupil size and neural activity in the same neurons as animals attend to and passively listen to sound. Data from our lab indicates that arousal and task engagement both modulate firing rate in A1 of ferrets performing a tone-in-noise

detection task (Saderi et al., 2020). Some neurons show greater modulation by task engagement, while others show greater modulation by pupil-linked arousal, suggesting that different top-down circuits may be involved in task engagement and arousal. Activity in noradrenergic and cholinergic nuclei is correlated with changes in pupil size, and provides a possible circuit linking this measurement of arousal state to activity in cortex (Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017; Liu et al., 2017; Larsen et al., 2018). Task-evoked pupil dilations in humans performing a challenging speech-recognition task is associated with activity in cortical regions outside A1, including superior temporal gyrus, anterior cingulate cortex, and several areas of frontal cortex (Zekveld et al., 2014). Comparisons of the effects of these corticocortical and subcortical areas on A1 could aid in understanding task engagement and arousal.

Analyzing receptive field changes and arousal in behaving animals could answer additional questions about the degree to which arousal, task engagement, and other behavioral variables interact. For example, it is possible that arousal does not uniformly increase the gain on all sensory stimuli in all cases, but that its effects instead depend on training history or task structure, such that arousal boosts responses to sounds associated with reward.

Selective attention and arousal in nonprimary auditory cortex

Future experiments could explore whether effects of selective attention or arousal are quantitatively larger in nonprimary auditory cortex than they are in A1. In ferrets, sounds that differ in behavioral meaning also show progressively greater contrast in A1 and two regions of nonprimary AC (Fritz et al., 2003; Atiani et al., 2014; Elgueda et al.,

2019). If effects of selective attention studied in animal models follow the same pattern, then we would expect to see a quantitatively larger suppression of the response to noise distractors in nonprimary AC compared to A1. This effect may offer insight into mechanisms of human selective attention to pitch or location, both of which modulate activity in nonprimary AC (Alho et al., 2014).

It is possible that the behavioral paradigm used in previous ferret experiments does not generalize to selective attention. These experiments used two categories of sounds differing in bandwidth (tones vs. noise) or modulation rate (click-rate discrimination). One category of sound signaled that ferrets would receive water for licking a water spout, and the other that ferrets would receive an aversive stimulus for making the same movement. Our selective attention task involved two streams of tones in noise that differed in frequency and location. Ferrets were rewarded for licking in response to tones at one frequency, but punished with a time-out for responding to tones at the other frequency, or for responding to the noise at any time. We found that in A1 the neural response to noise in the receptive field of the neuron decreased when ferrets attended to the receptive field, increasing the discriminability of the rewarded tone and noise in the same band. This strategy may only be applicable in A1 due to the relatively narrow bandwidth of neurons in this region (Elgueda et al., 2019). In addition, previous experiments show that behavioral training itself enhances the contrast between the neural response to sound categories in passively-listening animals (Elgueda et al., 2019). Given that the same tone can signal reward or a time-out depending on the block of the attention task in which it appears, AC may use different mechanisms to learn the task.

Although correlations between neural activity and pupil size have been shown in several cortical and subcortical areas (McGinley et al., 2015b), it is not known if effects of arousal differ between primary and nonprimary sensory cortex. Data from our lab shows that compared to A1, a greater portion of variability in inferior colliculus (IC) can be explained by pupil-linked arousal rather than task engagement (Saderi et al., 2020), suggesting that arousal effects differ across the ascending auditory pathway. Non-selectively increasing gain and spontaneous activity in areas outside primary sensory cortex would increase the metabolic cost of achieving increased contrast between different types of stimuli during behavior. Therefore there could be functional advantages to decreasing arousal effects further in nonprimary auditory cortex. The neuromodulatory centers whose activity is correlated to pupil size project diffusely, and arousal is traditionally considered a global state that simultaneously affects multiple brain areas. However, differences in the distribution of postsynaptic receptors could create functional differences between primary and nonprimary auditory cortex, as could differentiation of coeruleus and basal forebrain into subpopulations that project to distinct cortical targets (Záborszky et al., 2018; Chandler et al., 2019).

Effects of arousal state on auditory perception

Our data indicate that arousal state affects sensory coding in A1. When pupil was large, we noted desynchronized population activity, decreases in correlated variability between pairs of neurons, and decreased single-neuron variability across stimulus repetitions. These results are consistent with previous research indicating that during high-arousal states, sensory cortex is dominated by sensory input rather than self-

generated, correlated activity (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015). In addition, although frequency tuning remained unchanged, we found small decrease in auditory thresholds when pupil was large.

Future experiments could explore how these changes in sensory coding affect auditory perception. Baseline pupil size is correlated with performance on tasks that require auditory and somatosensory discrimination in mice and detection of visual stimuli with social significance in macaques (Ebitz et al., 2014; McGinley et al., 2015a; Schriver et al., 2018). In general, these experiments show that optimal performance occurs at intermediate pupil diameter. In mice, one study found that effects of arousal on neural gain and variability similarly peak at intermediate pupil diameters (McGinley et al., 2015a). Although we did not find a distinctive effect of arousal state on auditory coding at intermediate pupil diameter in most cells in our dataset, many of the coding changes we noted could support enhanced sound detection or discrimination in high-arousal states. Future experiments could explore whether there is a correlation between behavior and the degree of pupil modulation of neural activity on a given trial. In addition, behavioral experiments could test whether there is a change in detection thresholds for sound in line with the decrease in neural thresholds we observed when pupil is dilated.

An understanding of how auditory coding and perception changes across arousal states could aid in determining the function of cognitive pupil dilations. It is not known whether these changes in pupil size are epiphenomena or a part of active vision. Pupil dilation increases visual sensitivity at large spatial scales at the expense of discriminating fine detail. It has been hypothesized that shifting vision towards sensitivity rather than acuity is useful during orienting or exploration (Mathôt, 2018; Ebitz and Moore, 2019).

The threshold decrease we observed is consistent with the idea that the auditory system similarly increases its sensitivity during high-arousal states. Further behavioral experiments could explore whether there is a trade-off between sound detection and discrimination when pupil is dilated. In particular, it would be interesting to know whether pupil dilation has an effect on the discriminability of natural sounds.

During some recordings, we observed periods of tonically constricted pupil accompanied by saccades. In some A1 neurons, these epochs were accompanied by increases in firing rate. Given that mouse pupils are constricted during sleep (Yuzgec et al., 2018), we speculated that these epochs were period of rapid eye movement sleep. Although local-field potential did not reveal the increase in high-frequency neural activity that is a marker of rapid eye movement sleep, we observed a decrease in alpha power, indicating that the epochs may involve sleep onset. Future experiments that use pupil size as a measure of arousal state, particularly during protocols that involve passive listening, should consider the possibility that some recordings may be contaminated by sleep episodes. Recording other established markers of sleep state, such as electromyography, could help to distinguish sleep onset and low-arousal states more clearly.

In humans, pupil dilation has been linked to the degree of cognitive load required by many tasks, including those that require auditory processing (Beatty, 1982a; Zekveld et al., 2018). To reveal these pupil dilations, human studies generally examine the change from baseline pupil size evoked by particular task events, such as listening to a segment of distorted speech (Winn et al., 2015). Although there is not a similarly developed literature exploring behavioral correlates of changes in baseline pupil size, a number of

studies have noted that spontaneous pupillary movements vary with task engagement or wakefulness. At rest, the human pupil undergoes slow oscillations known as pupillary unrest or hippus that can be abolished by asking subjects to engage in mental arithmetic (Bouma and Baghuis, 1971). In subjects that are fatigued, progressive constrictions of the pupil lasting several minutes also occur, and can be reversed by sudden sensory stimuli (Lowenstein et al., 1963; Lowenstein and Loewenfeld, 1964; Loewenfeld and Lowenstein, 1999). Baseline pupil size is also related to both time awake and circadian rhythms, although it is not a direct proxy for subjective sleepiness in humans (Daguet et al., 2019). The pupillary movements reported in these studies have a greater dynamic range and slower dynamics than task-evoked dilations, raising the question of whether they are controlled by different neural circuits. For example, it is possible that areas of frontal cortex associated with task-evoked pupil dilation (Zekveld et al., 2014) are not involved in regulating all changes in baseline pupil size in humans that are fatigued, drowsy, or bored, or in animals engaged in passive listening. Data from our lab (Saderi et al., 2020) shows that sound-evoked dilations occur in behaving ferrets, but not in ferrets passively listening to same sounds, suggesting that different processes may be involved. To test whether task-evoked and baseline changes in pupil size have different effects on auditory processing, future experiments could directly compare changes in neural activity in sensory areas that depend on the degree of task-evoked pupil dilation to those that are sensitive to any change in pupil size. Ideally, these experiments would also use behavioral manipulations that vary the cognitive load imposed by the task in order to provide an animal model for human listening under conditions that impose effort, including hearing loss.

References

Aertsen AMHJ, Johannesma PIM (1981) The spectro-temporal receptive field: a functional characteristic of auditory neurons. Biol Cybern 42:133–143.

Alho K, Rinne T, Herron TJ, Woods DL (2014) Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. Hear Res 307:29–41.

Aston-Jones G, Cohen JD (2005) An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. Annu Rev Neurosci 28:403–450.

Atiani S, David SV, Elgueda D, Locastro M, Radtke-Schuller S, Shamma SA, Fritz JB (2014) Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. Neuron 82:486–499.

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

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366.

Bajo VM, Leach ND, Cordery PM, Nodal FR, King AJ (2014) The cholinergic basal forebrain in the ferret and its inputs to the auditory cortex. Eur J Neurosci 40:2922–2940.

Bajo VM, Nodal FR, Bizley JK, Moore DR, King AJ (2006) The ferret auditory cortex: descending projections to the inferior colliculus. Cereb Cortex 17:475–491.

Bala AD, Takahashi TT (2000) Pupillary dilation response as an indicator of auditory discrimination in the barn owl. J Comp Physiol [A] 186:425–434.

Baruni JK, Lau B, Salzman CD (2015) Reward expectation differentially modulates attentional behavior and activity in visual area V4. Nat Neurosci 18:1656–1663.

Beaton R, Miller JM (1975) Single cell activity in the auditory cortex of the unanesthetized, behaving monkey: Correlation with stimulus controlled behavior. Brain Res 100:543–562.

Beatty J (1982a) Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol Bull 91:276–292.

Beatty J (1982b) Phasic not tonic pupillary responses vary with auditory vigilance performance. Psychophysiology 19:167–172.

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

Bizley JK, Bajo VM, Nodal FR, King AJ (2015) Cortico-cortical connectivity within ferret auditory cortex: Connectivity in auditory cortex. J Comp Neurol 523:2187–2210.

Bizley JK, King AJ (2009) Visual influences on ferret auditory cortex. Hear Res 258:55–63.

Bizley JK, Nodal FR, Nelken I, King AJ (2005) Functional organization of ferret auditory cortex. Cereb Cortex 15:1637–1653.

Bizley JK, Walker KMM, Nodal FR, King AJ, Schnupp JWH (2013) Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Curr Biol 23:620–625.

Bouma H, Baghuis LCJ (1971) Hippus of the pupil: Periods of slow oscillations of unknown origin. Vision Res 11:1345–1351.

Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45:602–607.

Bronkhorst AW (2000) The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acust United Acust 86:117–128.

Brugge JF, Merzenich MM (1973) Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. J Neurophysiol 36:1138–1158.

Butler BE, Lomber SG (2013) Functional and structural changes throughout the auditory system following congenital and early-onset deafness: implications for hearing restoration. Front Syst Neurosci 7:92.

Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SMN (2011) A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PloS One 6:e16104.

Chandler DJ, Jensen P, McCall JG, Pickering AE, Schwarz LA, Totah NK (2019) Redefining noradrenergic neuromodulation of behavior: impacts of a modular locus coeruleus architecture. J Neurosci 39:8239–8249.

Chen Y, Martinez-Conde S, Macknik SL, Bereshpolova Y, Swadlow HA, Alonso J-M (2008) Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. Nat Neurosci 11:974–982.

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979.

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Netw Bristol Engl 12:199–213.

Churchland MM et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat Neurosci 13:369–378.

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

Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. J Neurophysiol 75:1306–1308.

Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M (2013) Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. J Neurosci 33:1858–1863.

Daguet I, Bouhassira D, Gronfier C (2019) Baseline pupil diameter is not a reliable biomarker of subjective sleepiness. Front Neurol 10:108.

Dai H, Scharf B, Buus S (1991) Effective attenuation of signals in noise under focused attention. J Acoust Soc Am 89:2837–2842.

David SV, Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc Natl Acad Sci 109:2144–2149.

David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59:509–521.

David SV, 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 29:3374–3386.

David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. Netw Bristol Engl 18:191–212.

David SV, Shamma SA (2013) Integration over multiple timescales in primary auditory cortex. J Neurosci 33:19154–19166.

de Cheveigné A (1993) Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing. J Acoust Soc Am 93:3271–3290.

de Gee JW, Colizoli O, Kloosterman NA, Knapen T, Nieuwenhuis S, Donner TH (2017) Dynamic modulation of decision biases by brainstem arousal systems. eLife 6.

deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. Science 280:1439–1443.

Deschenes M, Moore J, Kleinfeld D (2012) Sniffing and whisking in rodents. Curr Opin Neurobiol 22:243–250.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.

Diamond DM, Campbell AM, Park CR, Halonen J, Zoladz PR (2007) The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. Neural Plast 2007:60803.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci 109:11854–11859.

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 37:5378–5392.

Du J, Blanche TJ, Harrison RR, Lester HA, Masmanidis SC (2011) Multiplexed, high density electrophysiology with nanofabricated neural probes. PLOS ONE 6:e26204.

Durlach NI (1963) Equalization and cancellation theory of binaural masking-level differences. J Acoust Soc Am 35:1206–1218.

Ebitz RB, Moore T (2017) Selective modulation of the pupil light reflex by microstimulation of prefrontal cortex. J Neurosci 37:5008.

Ebitz RB, Moore T (2019) Both a gauge and a filter: cognitive modulations of pupil size. Front Neurol 9:1190.

Ebitz RB, Pearson JM, Platt ML (2014) Pupil size and social vigilance in rhesus macaques. Front Neurosci 8.

Elgueda D, Duque D, Radtke-Schuller S, Yin P, David SV, Shamma SA, Fritz JB (2019) State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. Nat Neurosci 22:447–459.

Englitz B, David SV, Sorenson MD, Shamma SA (2013) MANTA--An open-source, high density electrophysiology recording suite for MATLAB. Front Neural Circuits 7:69.

Festen JM, Plomp R (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. J Acoust Soc Am 88:1725–1736.

Foote SL, Freedman R, Oliver AP (1975) Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. Brain Res 86:229–242.

Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci 6:1216–1223.

Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA (2010) Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat Neurosci 13:1011–1019.

Fritz JB, Elhilali M, David SV, Shamma SA (2007a) Auditory attention—focusing the searchlight on sound. Curr Opin Neurobiol 17:437–455.

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

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

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

Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian Data Analysis, Second Edition. Boca Raton, FL: Taylor & Francis.

Greenberg GZ, Larkin WD (1968) Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. J Acoust Soc Am 44:1513–1523.

Hackett TA (2011) Information flow in the auditory cortical network. Hear Res 271:133–146.

Hafter ER, Sarampalis A, Loui P (2007) Auditory attention and filters. In: Auditory Perception of Sound Sources (Yost WA, Popper AN, Fay RR, eds), pp 115–142 Springer Handbook of Auditory Research. Boston, MA: Springer US.

Hakerem G, Sutton S (1966) Pupillary response at visual threshold. Nature 212:485–486.

Harris KD, Thiele A (2011) Cortical state and attention. Nat Rev Neurosci 12:509.

Hartline HK (1938) The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. Am J Physiol 121:400–415.

Heffner HE, Heffner RS (1995) Conditioned Avoidance. In: Methods in Comparative Psychoacoustics (Klump GM, Dooling RJ, Fay RR, Stebbins WC, eds), pp 79–93. Basel: Birkhäuser Basel.

Hess EH, Polt JM (1960) Pupil size as related to interest value of visual stimuli. Science 132:349–350.

Hess EH, Polt JM (1964) Pupil size in relation to mental activity during simple problem-solving. Science 143:1190–1192.

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 117:51–68.

Hubel DH, Henson CO, Rupert A, Galambos R (1959) "Attention" units in the auditory cortex. Science 129:1279–1280.

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

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

Issa EB, Wang X (2008) Sensory responses during sleep in primate primary and secondary auditory cortex. J Neurosci 28:14467–14480.

Jaramillo S, Borges K, Zador AM (2014) Auditory thalamus and auditory cortex are equally modulated by context during flexible categorization of sounds. J Neurosci 34:5291–5301.

Jepma M, Nieuwenhuis S (2011) Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. J Cogn Neurosci 23:1587–1596.

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 89:221–234.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci 97:11793–11799.

Kahneman D (1973) Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman D, Beatty J (1966) Pupil diameter and load on memory. Science 154:1583–1585.

Kahneman D, Beatty J (1967) Pupillary responses in a pitch-discrimination task. Percept Psychophys 2:101–105.

Kahneman D, Beatty J, Pollack I (1967) Perceptual deficit during a mental task. Science 157:218–219.

Kidd G, Arbogast TL, Mason CR, Gallun FJ (2005) The advantage of knowing where to listen. J Acoust Soc Am 118:3804–3815.

Kilgard MP, Merzenich MM (1998) Cortical map reorganization enabled by nucleus basalis activity. Science 279:1714–1718.

Koelewijn T, Shinn-Cunningham BG, Zekveld AA, Kramer SE (2014) The pupil response is sensitive to divided attention during speech processing. Hear Res 312:114–120.

Koelewijn T, Zekveld AA, Festen JM, Kramer SE (2012) Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. Ear Hear 33:291–300.

Kuchibhotla KV, Gill JV, Lindsay GW, Papadoyannis ES, Field RE, Sten TAH, Miller KD, Froemke RC (2017) Parallel processing by cortical inhibition enables context-dependent behavior. Nat Neurosci 20:62–71.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.

Larsen RS, Turschak E, Daigle T, Zeng H, Zhuang J, Waters J (2018) Activation of neuromodulatory axon projections in primary visual cortex during periods of locomotion and pupil dilation. bioRxiv:502013.

Larsen RS, Waters J (2018) Neuromodulatory correlates of pupil dilation. Front Neural Circuits 12:21.

Lee C-C, Middlebrooks JC (2010) Auditory cortex spatial sensitivity sharpens during task performance. Nat Neurosci 14:108.

Lee CR, Margolis DJ (2016) Pupil dynamics reflect behavioral choice and learning in a go/nogo tactile decision-making task in mice. Front Behav Neurosci 10:200.

Lehmann SJ, Corneil BD (2016) Transient pupil dilation after subsaccadic microstimulation of primate frontal eye fields. J Neurosci 36:3765.

Liao H-I, Kidani S, Yoneya M, Kashino M, Furukawa S (2016) Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. Psychon Bull Rev 23:412–425.

Lin P-A, Asinof SK, Edwards NJ, Isaacson JS (2019) Arousal regulates frequency tuning in primary auditory cortex. Proc Natl Acad Sci 116:25304–25310.

Liu Y, Rodenkirch C, Moskowitz N, Schriver B, Wang Q (2017) Dynamic lateralization of pupil dilation evoked by locus coeruleus activation results from sympathetic, not parasympathetic, contributions. Cell Rep 20:3099–3112.

Loewenfeld I, Lowenstein O (1999) Reflex integration: pupillary consequences. In: The Pupil: Anatomy, Physiology, and Clinical Applications, 2nd ed., pp 480–518. Boston, MA: Butterworth-Heinemann.

Lovett-Barron M, Andalman AS, Allen WE, Vesuna S, Kauvar I, Burns VM, Deisseroth K (2017) Ancestral circuits for the coordinated modulation of brain state. Cell 171:1411-1423.e17.

Lowenstein O, Feinberg R, Loewenfeld I (1963) Pupillary movements during acute and chronic fatigue : a new test for the objective evaluation of tiredness. Invest Ophthalmol Vis Sci 2:138–157.

Lowenstein O, Loewenfeld I (1964) The sleep-waking cycle and pupillary activity. Ann N Y Acad Sci 117:142–156.

Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol 77:24–42.

Luo TZ, Maunsell JHR (2015) Neuronal modulations in visual cortex are associated with only one of multiple components of attention. Neuron 86:1182–1188.

Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. J Neurosci Off J Soc Neurosci 24:1089–1100.

Manunta Y, Edeline JM (1997) Effects of noradrenaline on frequency tuning of rat auditory cortex neurons. Eur J Neurosci 9:833–847.

Manunta Y, Edeline JM (1999) Effects of noradrenaline on frequency tuning of auditory cortex neurons during wakefulness and slow-wave sleep. Eur J Neurosci 11:2134–2150.

Martins ARO, Froemke RC (2015) Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. Nat Neurosci 18:1483–1492.

Mathôt S (2018) Pupillometry: psychology, physiology, and function. J Cogn 1:16.

Mauss IB, Robinson MD (2009) Measures of emotion: A review. Cogn Emot 23:209–237.

McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441.

McDermott JH (2009) The cocktail party problem. Curr Biol 19:R1024–R1027.

McGinley MJ, David SV, McCormick DA (2015a) Cortical membrane potential signature of optimal states for sensory signal detection. Neuron 87:179–192.

McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, Cadwell CR, Tolias AS, Cardin JA, McCormick DA (2015b) Waking state: Rapid variations modulate neural and behavioral responses. Neuron 87:1143–1161.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236.

Meyer AF, Poort J, O'Keefe J, Sahani M, Linden JF (2018) A head-mounted camera system integrates detailed behavioral monitoring with multichannel electrophysiology in freely moving mice. Neuron 100:46-60.e7.

Miller JM, Sutton D, Pfingst B, Ryan A, Beaton R, Gourevitch G (1972) Single cell activity in the auditory cortex of Rhesus monkeys: behavioral dependency. Science 177:449–451.

Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782–784.

Moray N (1959) Attention in dichotic listening: affective cues and the influence of instructions. Q J Exp Psychol 11:56–60.

Motter BC (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J Neurophysiol 70:909–919.

Murphy PR, O'Connell RG, O'Sullivan M, Robertson IH, Balsters JH (2014) Pupil diameter covaries with BOLD activity in human locus coeruleus. Hum Brain Mapp 35:4140–4154.

Murphy PR, Robertson IH, Balsters JH, O'connell RG (2011) Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. Psychophysiology 48:1532–1543.

Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK (2019) Single-trial neural dynamics are dominated by richly varied movements. Nat Neurosci 22:1677–1686.

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN (2015) Complementary control of sensory adaptation by two types of cortical interneurons. eLife 4.

Nelson PC, Carney LH (2007) Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. J Neurophysiol 97:522–539.

Nguyen AH, Stark LW (1993) Model control of image processing: Pupillometry. Comput Med Imaging Graph Off J Comput Med Imaging Soc 17:21–33.

Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. Neuron 65:472–479.

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

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 33:8378–8395.

O'Connell MN, Barczak A, Schroeder CE, Lakatos P (2014) Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. J Neurosci 34:16496–16508.

Oertel D, Doupe A (2013) The auditory central nervous system. In: Principles of Neural Science, 5th ed. (Kandel E, Schwartz J, Jessell T, Siegelbaum S, Hudspeth AJ, eds), pp 682–711. New York, NY: McGraw Hill.

Osmanski MS, Wang X (2015) Behavioral dependence of auditory cortical responses. Brain Topogr 28:365–378.

O'Sullivan J, Herrero J, Smith E, Schevon C, McKhann GM, Sheth SA, Mehta AD, Mesgarani N (2019) Hierarchical encoding of attended auditory objects in multi-talker speech perception. Neuron 104:1195-1209.e3.

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

Pachitariu M, Steinmetz N, Kadir S, Carandini M, Kenneth D. H (2016) Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv:061481.

Pashler HE (1998) Selective attention. In: The Psychology of Attention, pp 37–101. Cambridge, MA: MIT Press.

Pi H-J, Hangya B, Kvitsiani D, Sanders JI, Huang ZJ, Kepecs A (2013) Cortical interneurons that specialize in disinhibitory control. Nature 503:521–524.

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10:59–63.

Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS (2014) Pupil fluctuations track fast switching of cortical states during quiet wakefulness. Neuron 84:355–362.

Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tolias AS (2016) Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat Commun 7:13289.

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

Rinne T, Muers RS, Salo E, Slater H, Petkov CI (2017) Functional imaging of audio-visual selective attention in monkeys and humans: How do lapses in monkey performance affect cross-species correspondences? Cereb Cortex N Y N 1991 27:3471–3484.

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 82:1157–1170.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373.

Rutkowski RG, Weinberger NM (2005) Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. Proc Natl Acad Sci 102:13664–13669.

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

Saderi D, Schwartz ZP, Heller CR, Pennington JR, David SV (2020) Dissociation of task engagement and arousal effects in auditory cortex and midbrain. bioRxiv:2020.06.16.155432.

Satpute AB, Kragel PA, Barrett LF, Wager TD, Bianciardi M (2019) Deconstructing arousal into wakeful, autonomic and affective varieties. Neurosci Lett 693:19–28.

Scharf B, Quigley S, Aoki C, Peachey N, Reeves A (1987) Focused auditory attention and frequency selectivity. Percept Psychophys 42:215–223.

Scheich H, Brechmann A, Brosch M, Budinger E, Ohl FW (2007) The cognitive auditory cortex: Task-specificity of stimulus representations. Hear Res 229:213–224.

Schneider DM, Nelson A, Mooney R (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513:189–194.

Schnupp J, Nelken I, King A (2011a) The ear. In: Auditory Neuroscience: Making Sense of Sound, pp 51–92. Cambridge, MA: The MIT Press.

Schnupp J, Nelken I, King A (2011b) Neural basis of sound localization. In: Auditory Neuroscience: Making Sense of Sound, pp 177–223. Cambridge, MA: The MIT Press.

Schnupp J, Nelken I, King A (2011c) Development, learning, and plasticity. In: Auditory Neuroscience: Making Sense of Sound, pp 269–295. Cambridge, MA: The MIT Press.

Schreiner CE, Polley DB (2014) Auditory map plasticity: diversity in causes and consequences. Curr Opin Neurobiol 24:143–156.

Schreiner CE, Winer JA (2007) Auditory cortex mapmaking: principles, projections, and plasticity. Neuron 56:356–365.

Schriver BJ, Bagdasarov S, Wang Q (2018) Pupil-linked arousal modulates behavior in rats performing a whisker deflection direction discrimination task. J Neurophysiol 120:1655–1670.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18.

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870–3896.

Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34:114–123.

Shamma SA, Fleshman JW, Wiser PR, Versnel H (1993) Organization of response areas in ferret primary auditory cortex. J Neurophysiol 69:367–383.

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Shimaoka D, Harris KD, Carandini M (2018) Effects of arousal on mouse sensory cortex depend on modality. Cell Rep 22:3160–3167.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182–186.

Shinn-Cunningham BG, Best V (2008) Selective attention in normal and impaired hearing. Trends Amplif 12:283–299.

Siegle JH, Lopez AC, Patel YA, Abramov K, Ohayon S, Voigts J (2017) Open ephys: An open-source, plugin-based platform for multichannel electrophysiology. J Neural Eng 14:045003.

Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, Kapen S, Keenan SA, Kryger MH, Penzel T, Pressman MR, Iber C (2007) The visual scoring of sleep in adults. J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med 3:121–131.

Simonsohn, Uri (2018) Two-lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. Adv Methods Pract Psychol Sci 1:538–555.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am 114:3394–3411.

Slee SJ, David SV (2015) Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. J Neurosci 35:13090–13102.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90.

Spence CJ, Driver J (1994) Covert spatial orienting in audition: exogenous and endogenous mechanisms. J Exp Psychol Hum Percept Perform 20:555–574.

Spitzer H, Desimone R, Moran J (1988) Increased attention enhances both behavioral and neuronal performance. Science 240:338–340.

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brainwide activity. Science 364:255.

Sundberg KA, Mitchell JF, Reynolds JH (2009) Spatial attention modulates center-surround interactions in macaque visual area v4. Neuron 61:952–963.

Telenczuk B, Nikulin VV, Curio G (2010) Role of neuronal synchrony in the generation of evoked EEG/MEG responses. J Neurophysiol 104:3557–3567.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Netw Bristol Engl 12:289–316.

Thorson IL, Liénard J, David SV (2015) The essential complexity of auditory receptive fields. PLoS Comput Biol 11:e1004628.

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

Tsunada J, Liu ASK, Gold JI, Cohen YE (2016) Causal contribution of primate auditory cortex to auditory perceptual decision-making. Nat Neurosci 19:135–142.

van Kempen J, Loughnane GM, Newman DP, Kelly SP, Thiele A, O'Connell RG, Bellgrove MA (2019) Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. eLife 8:e42541.

Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron 86:740–754.

Wang C-A, Boehnke SE, White BJ, Munoz DP (2012) Microstimulation of the monkey superior colliculus induces pupil dilation without evoking saccades. J Neurosci 32:3629–3636.

Weinberger NM (2004) Specific long-term memory traces in primary auditory cortex. Nat Rev Neurosci 5:279–290.

Willmore BDB, Prenger RJ, Gallant JL (2010) Neural representation of natural images in visual area V2. J Neurosci 30:2102–2114.

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

Winn MB, Edwards JR, Litovsky RY (2015) The impact of auditory spectral resolution on listening effort revealed by pupil dilation. Ear Hear 36:e153-165.

Winn MB, Wendt D, Koelewijn T, Kuchinsky SE (2018) Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. Trends Hear 22:233121651880086.

Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nat Neurosci 9:1156–1160.

Woods DL, Alain C, Diaz R, Rhodes D, Ogawa KH (2001) Location and frequency cues in auditory selective attention. J Exp Psychol Hum Percept Perform 27:65–74.

Wu MC-K, David SV, Gallant JL (2006) Complete functional characterization of sensory neurons by system identification. Annu Rev Neurosci 29:477–505.

Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 18:459–482.

Yin P, Fritz JB, Shamma SA (2010) Do ferrets perceive relative pitch? J Acoust Soc Am 127:1673–1680.

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

Yuzgec O, Prsa M, Zimmermann R, Huber D (2018) Pupil size coupling to cortical states protects the stability of deep sleep via parasympathetic modulation. Curr Biol 28:392-400.e3.

Záborszky L, Gombkoto P, Varsanyi P, Gielow MR, Poe G, Role LW, Ananth M, Rajebhosale P, Talmage DA, Hasselmo ME, Dannenberg H, Minces VH, Chiba AA (2018) Specific basal rorebrain–cortical cholinergic circuits coordinate cognitive operations. J Neurosci 38:9446–9458.

Zekveld AA, Heslenfeld DJ, Johnsrude IS, Versfeld NJ, Kramer SE (2014) The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. NeuroImage 101:76–86.

Zekveld AA, Koelewijn T, Kramer SE (2018) The pupil dilation response to auditory stimuli: current state of knowledge. Trends Hear 22:2331216518777174.

Zekveld AA, Kramer SE, Festen JM (2010) Pupil response as an indication of effortful listening: the influence of sentence intelligibility. Ear Hear 31:480–490.

Zhao S, Chait M, Dick F, Dayan P, Furukawa S, Liao H-I (2019) Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. Nat Commun 10:1–16.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron 77:980–991.