State-dependent representation of sound by neural population activity in auditory cortex

By

Charles R. Heller

A DISSERTATION

OHSU

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

July 9, 2021

Contents

Acknowledgements

me. Your passion for science was a huge reason that I chose to pursue graduate school. I thank you for your frequent technical support throughout my project, as well as the conversations about our shared graduate school experiences.

To my parents: From the time I was a kid, you inspired a love for learning and for school that continues today. Thank you for being there for me, not just in graduate school, but for all the years before. It goes without saying, but without your love and support, none of this would have been possible.

Last but not least, to my wife, Katie: You have been with me on every step of this journey. From choosing a graduate school and moving across the country, to starting a new life together in a new city, to putting up with all parts of of my graduate school experience. Your willingness to always be there for me and listen when I need it means the world. Coming to a new city far away from home was something that was new for both of us. Your enthusiasm and ability to form new connections and community in a new place was inspiring. It truly made Portland feel like home. Having you, and this community you helped build for us outside of lab, made all of this possible. Thank you.

Abstract

The auditory system underlies the ability to perceive and respond to sound in the environment. However, perception and behavioral responses to sound are not static, but change over time depending on one's sensory context and internal state. The neural mechanisms underlying this flexibility are not yet well-understood. For example, how is it that by focusing our attention on a single speaker in a crowded, noisy room we are able to improve our ability to perceive their words? Prior work has investigated this by characterizing how the sound driven activity of single neurons in the auditory system depend on behavioral state. Yet, sensory perception and behavior are not mediated by single neurons in isolation. In this dissertation, we build on this work by studying how two aspects of behavioral state, arousal and task engagement, co-modulate the activity of many neurons simultaneously in auditory cortex.

We recorded extracellular spiking activity from neurons in primary auditory cortex (A1) of ferrets using multi-electrode laminar probes. These probes allowed us to measure the activity of tens of neurons simultaneously while animals passively listened to sound stimuli, or engaged in a behavioral task. Pupil size has been shown to track changes in cognitive state, such as alertness, in addition to changes in neural activity. Therefore, we also measured animal's pupil size as an index of global arousal over time.

To analyze the high-dimensional neural population recordings, we first developed a novel dimensionality reduction method which allowed us to reliably measure how accurately populations of neurons in A1 represented sound stimuli. Using this approach, we studied how representation of natural sound stimuli changed between high and low states of arousal. We found that, unlike in previous studies of single neurons, arousal could either improve or reduce the accuracy of sound representations in A1. We showed that this diversity is explained only by considering population level covariability in sensory-evoked responses between cells. Finally, we applied this same method to data collected from animals performing an auditory task. While task engagement improved representations of sound stimuli, the improvement did not correlate with animal's perceptual performance. However, the strength of population level covariability did. These results provide new insight into how the information in A1 may be used to guide behavioral choices. Taken together, our results argue that population covariability plays an important role in determining how accurately sound information is represented in A1 and, ultimately, is used to guide behavior.

1　Introduction

The human auditory system is responsible for mediating the perception of sound. Sound pressure waves generated by speech, music, and other environmental sources enter the system via the ear where they are transduced into electrical signals by the cochlea and routed to the brain via the auditory nerve. At this early stage of processing, sound information is represented with remarkable fidelity by the activity of neurons in the auditory system. At later stages of processing, the signals from the auditory nerve are integrated with information from other sensory modalities as well as aspects of internal state, such as attention and arousal. This results in complex patterns of neural activity that depend not only on sound, but also on the overall context in which the sound was experienced. These context-dependent patterns of activity are thought to help guide appropriate action selection in response to sound by integrating sound information with an animal's overall sensory and behavioral context.

Work over the previous two decades has investigated how different aspects of context influence the activity of single neurons in the auditory system. However, the brain consists of billions of interconnected neurons that act in concert to drive perception and behavior. While informative, studies of single auditory neurons have yet to yield a cohesive theory describing how neural activity translates to sound-dependent behavior. In this dissertation, we take a step towards this larger understanding by developing new computational analysis techniques to determine how the simultaneous activity of many neurons is modulated by two aspects of behavioral state – arousal and task engagement – and investigate how this relates to behavior and perception.

1.1　Overview of the auditory system

General anatomy and physiology of the auditory system

Neural processing of acoustic information begins in the inner ear where sound is decomposed into its constituent frequency components and transduced into an electrical signal by the cochlea. This signal is relayed to the brain via auditory nerve (AN) fibers, each of which contains information about a different sound frequency channel. This organized frequency segregation is referred to as tonotopy and it is a hallmark of the auditory system that is preserved throughout the processing hierarchy. Once in the brain, sound information is routed through multiple brainstem structures before it crosses the midline and ultimately reaches the auditory cortex (AC) (Figure 1.1).
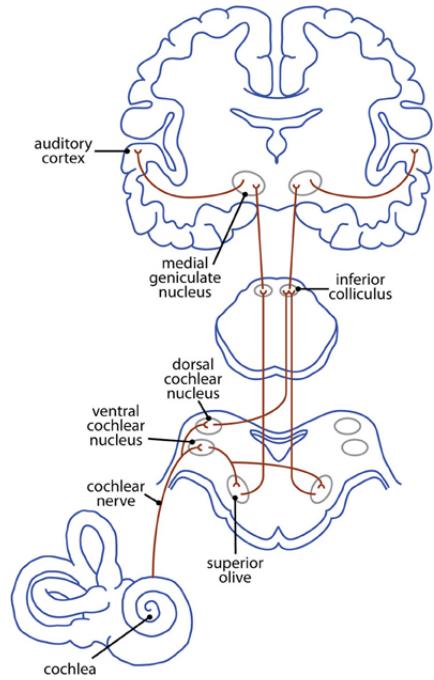
Figure 1.1: **Schematic of auditory pathway.** Simplified schematic of ascending auditory system through auditory cortex. Reproduced from.[1]

Sound information is represented with high precision in early processing stages and underlies, for example, detection of interaural timing and sound level differences[2] that help us to quickly localize sound sources in our environment. Once relayed to the auditory cortex, the peripheral representation of spectrotemporal information is integrated with contextual information about internal state,[3,4] other sensory modalities,[5] and even motor activity.[6] In this dissertation, we will focus exclusively on these later stages of processing and their importance for guiding sound-dependent behavior.

Ferret as a model system

In the work described in this dissertation, we used the ferret as an animal model. The ferret is particularly well-suited for studying higher-order auditory representations, such as those that depend on multimodal sensory and behavioral context. The ferret has a relatively advanced behavioral repertoire[3,7–9] and a well-defined auditory hierarchy within cortex[10,11] that is largely absent from other species, such as mice.[12,13] This hierarchy within auditory cortex is thought to be important for mediating more complex sound-dependent behaviors.[14]

1.2 Behavior state-dependent modulation of coding in auditory cortex

Auditory coding

In what follows, we will discuss how the *neural code* is modulated by behavioral state. In order to do this, we first provide a brief framework for conceptualizing auditory coding.

Sound pressure waves cause neurons throughout the auditory pathway to fire action potentials. The number and timing of action potentials, or spikes, elicited at any moment contain information about the external world that can then be decoded by downstream brain regions to select the appropriate behavioral response. The collection of physical features that elicit spikes in a given neuron comprise its receptive field. For example, an auditory neuron's receptive field is typically composed of some combination of sound frequency, level, and temporal modulation. These features are summarized using a spectrotemporal receptive field (STRF, Figure 1.2B). A neuron is said to be "tuned" to the particular combination of features that cause it to fire the most action potentials. For example, the example neuron shown in Figure 1.2B is tuned to sounds at roughly 800 Hz, with a response latency of approximately 10 ms. Thus, the neuron's pattern of spiking activity forms a neural code for the presence (or absence) of this particular sound feature's presence in the world over time.
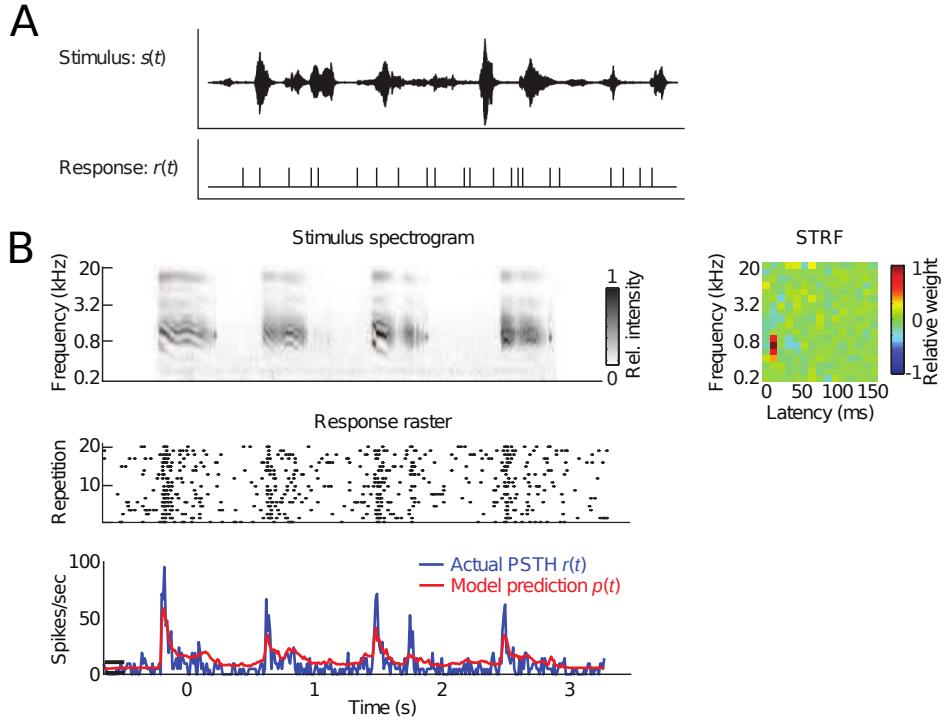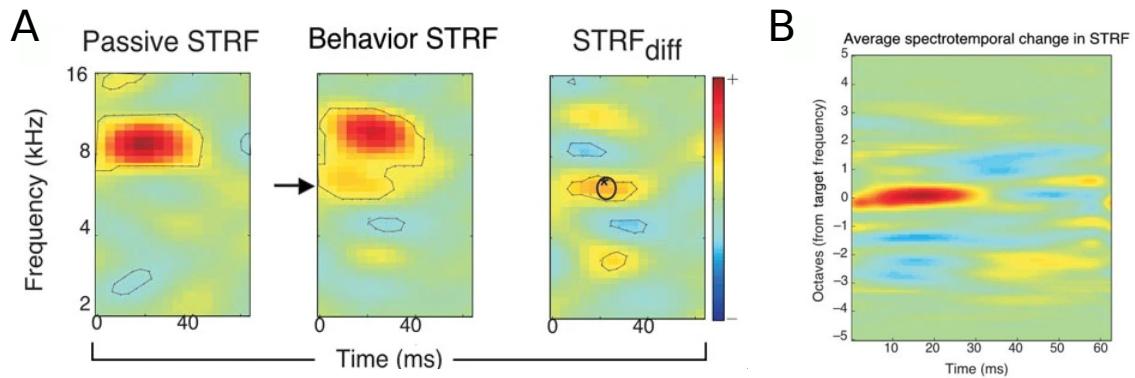
Figure 1.2: **Receptive fields in auditory cortex. A.** Top: Example sound pressure waveform. Bottom: Response of a single auditory neuron. Vertical bars represent single action potentials. **B.** Illustration of receptive field estimation for a single neuron. Top: Stimulus spectrogram for a ferret vocalization. Relative sound level across time and frequency is represented by color. Middle: Spiking response of a single neuron across repeated repetitions (rows) of the stimulus spectrogram. Tick marks represent action potentials. Bottom: Spiking response is summarized by binning spikes and computing the mean in each bin across repetitions (blue). An STRF receptive field model (right), captures most of the sound evoked spiking activity over time (red). In practice, STRFs are estimated using normalized reverse correlation between a neuron's PSTH and the stimulus spectrogram.[15] This can be thought of, roughly, as a spike-triggered average of the stimulus spectrogram. Figure is adapted from.[16]

Plasticity of the neural code in Auditory Cortex

The receptive field coding framework outlined in the previous section has classically been used to describe the physiology of neurons in sensory brain areas. However, sensation, perception, and action selection are not static, fixed processes with respect to physical stimuli. For example, carefully focusing attention on a particular sound source – like a friend's voice in a crowded bar – substantially improves the ability to perceive what they are saying. Furthermore, even if you are able to perceive their voice accurately, you may or may not choose to respond to them depending on internal factors – like your motivation or your understanding of the topic – that are independent of the words they spoke. How do sensory representations dynamically adapt to meet these changing requirements?

Early experiments investigated this by training ferrets to discriminate broad band noise associated

with a water reward from pure tone frequencies which were followed by a mild shock if the animal did not correctly withhold from licking.[3] These studies found that the receptive fields of single auditory neurons were enhanced at the pure tone target frequency relative to receptive fields measured in the same neurons during passive conditions, when the sounds were no longer paired with reward or punishment (Figure 1.3). Thus, the study demonstrated that by shifting their encoding to enhance the representation of negative valence stimuli, AC neurons may help to adaptively guide the appropriate motor response. Similar findings have subsequently been reported for different task designs,[7] animal species,[17] and brain areas.[18]

Figure 1.3: **Receptive field plasticity in auditory cortex during behavior. A.** STRF for single example unit recorded during active tone-detection (behavior STRF) and during passive listening to task stimuli (Passive STRF). Target frequency indicated by arrow in middle panel. **B.** Summary of receptive field plasticity across single units in AC. STRFs are centered at target frequency before pooling across units. Figure adapted from.[3]

While behavior state-dependent modulation of neural coding is well-documented, the neural substrate underlying these phenomenon remains elusive in most cases. During passive conditions, electric shock or reward pairing with specific sound frequencies leads to shifts in frequency tuning near the paired tones.[19, 20] Coupled with findings that punishment and reward are associated with neuromodulator release in cortex,[21–23] these studies suggest that behavior-dependent plasticity may be mediated by neuromodulatory systems. Recent work in anesthetized preps corroborates this hypothesis. Pairing sound presentation with the release of acetycholine, leads to expansion at sound frequency[24] and similar results have been demonstrated for the neuromodulator noradrenaline.[25] Therefore, while the exact mechanisms appear to be complicated have not yet been fully described, neuromodulation seems likely to contribute to the plasticity observed in AC during behavior.

Defining and measuring behavior state

Certain aspects of an animal's behavior state, like goal directed task engagement[3,7] or selective attention,[26] are manipulated by the experimenter and can be inferred with reasonably high levels of accuracy. However, at the same time as an experimenter manipulates experimental conditions, animals also undergo fluctuations in other unobserved and uncontrolled aspects of behavior state, such as arousal and wakefulness. These types of nonspecific changes in alertness occur independent of experimenter defined task conditions *e.g.* animals cycle through high and low periods of arousal every day, such as sleep vs. wakefulness. Importantly, changes in arousal state correlate with striking changes in neural spiking activity and neuromodulator release.[4,27–32] In practice, increased arousal often coincides with an animal's engagement in behavioral tasks. Thus, it is possible that some of the behavior related plasticity observed in AC is not due to the task itself, but to nonspecific changes in global brain state induced by heightened arousal or other uncontrolled state variables.

To address this possibility, recent work coupled measurements of pupil size, a non-invasive readout of arousal state (see below: *Pupil as an index of arousal state*), with neural recordings in ferrets while they performed a standard tone-detection task.[7] Using a novel application of linear regression, this work showed that the effects of arousal and task engagement could be dissociated in the firing rates of single neurons in both the midbrain and the auditory cortex.[33] The study further demonstrated that ignoring arousal state could lead to incorrect conclusions about the effects of task engagement on neural activity, highlighting both the complexity of studying behavior state and the importance of controlling for latent, unobserved state variables in behavioral experiments.

Pupil size as an index of arousal state and neuromodulation

In this dissertation, like in the previously described study,[33] we use pupil size as index of global arousal. Pupil size has been shown to correlate with a number of behavioral readouts that are generally thought to reflect arousal state.[34] For example, pupil size is correlated with listening effort,[35,36] wakefulness,[37] and locomotion.[27,29,38] Although recent studies have shown that more general readouts of arousal, such as movement, seem to explain more variability in neural activity,[29,38] pupil size has the advantage of being tightly coupled to known neuromodulatory activity in the locus coeruleus and basal forebrain which project directly to auditory cortical regions.[4,31,32] Thus, pupil size provides an index for both arousal and neurmodulation that can be measured non-invasively in awake, behaving animals, making it a powerful tool for studying behavioral state. For a more comprehensive review of pupil size as index of arousal state, see.[34,39]

1.3 Neural population coding

Recording the activity of large neural populations

When a neuron fires an action potential, ion channels along the cell membrane open and allow positively charged ions to rush into the negatively charged cell body. This inward flux generates an electrical potential difference (voltage) across the cell membrane that can be measured by placing a high impedance electrode near the soma. The electrodes designed for this purpose are typically made of material like tungsten or glass and are sharp and small in diameter, allowing them to be inserted into the brain while causing minimal damage to neural tissue. Generally speaking, a single electrode allows researchers to measure the activity of one, or maybe up to two or three, neurons at a time. By inserting multiple electrodes, researchers are able to record a handle of neurons simultaneously, however, this quickly becomes technically challenging as the number of electrodes that must me inserted into the brain increases. Until recently, this technique was considered the gold standard in measuring neural activity *in vivo*.

Over the past two decades, many of these challenges have been mitigated by the development of high-density silicon microelectrode arrays *e.g.*[40] (Figure 1.4). These probes are designed to be small, like a tungsten electrode, but each contains many, independent high-impedance recording sites. This allows researchers to measure the spiking activity of tens to hundreds of neurons simultaneously with just a single probe inserted into the brain. While still falling far short of the billions of neurons contained in the brain, this technology offers a dramatic improvement over single electrodes and theoretical work suggests that many key brain computations can be inferred accurately from neural populations of this size.[41]
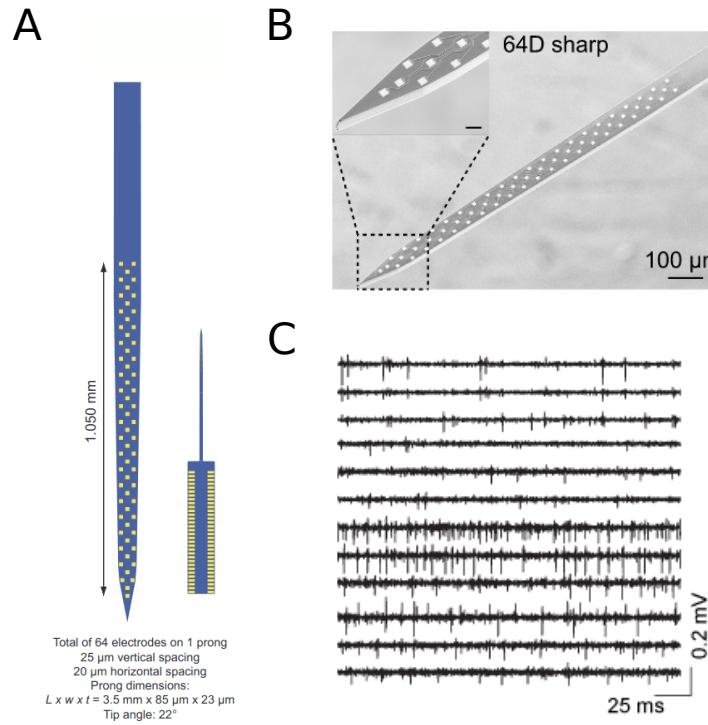
Figure 1.4: **Silicon microelectrode arrays. A.** Cartoon rendering of a 64-channel linear silicon microelectrode array **B.** Scanning electron microscrope image of probe. **C** Example data collected simultaneously across 12 electrodes. Figures adapted from[42] and `https://masmanidislab.neurobio.ucla.edu/technology.html`.

Although the data presented in this dissertation were collected using high-density microelectrode arrays, it is worth also highlighting that imaging techniques and genetically encoded calcium indicators have, likewise, revolutionized neural population recording. In some cases, these imaging techniques now allow researchers to record from thousands to tens of thousands of neurons simultaneously,[43] at least an order of magnitude increase over high-density electrode recordings. Importantly, though, calcium signals are not equivalent to action potentials and there remain key advantages to using electrical recordings.[44] Nevertheless, these studies have yielded important insights about the structure of neural population activity, some of which we describe below.

Spiking activity is correlated across neurons

With the advent of neural population recording, one of the earliest and most ubiquitous findings was that the activity of simultaneously recorded neurons was not independent, but correlated. That is, when one neuron fires more spikes than average, the neurons nearby are likely to fire more than average. These firing rate correlations are stimulus-independent; they are not the result of co-

activation by co-tuned neurons. Therefore, since they cannot be explained by stimulus condition, they have generally been referred to as *noise correlations* (Figure 1.5). Unless otherwise specified, in this dissertation "correlations" or "correlated neural activity" will refer to noise correlations.
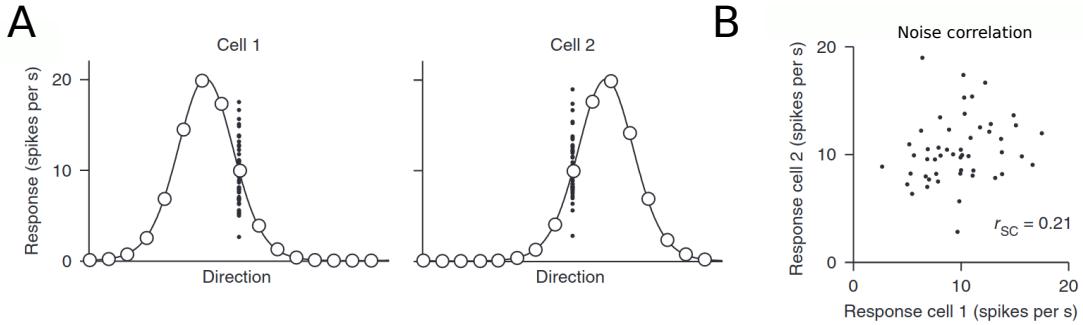
Figure 1.5:   **Noise correlations. A.** Tuning curves for two simultaneously recorded direction selective units. Open circles indicate mean response across all trials for each direction. Closed circles indicate single trial responses for a particular stimulus direction. **B.** Single trial responses (same as closed dots in A) are correlated. This stimulus-independent correlation is called a *noise correlation.* Figure adapted from.[45]

Origin of noise correlations
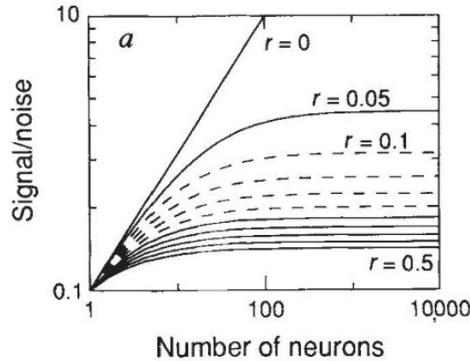
Noise correlations tend to be strongest amongst neurons that are close together, suggesting that correlated activity may be due to recurrent connections between neurons within local networks. Another possibility is that noise correlations reflect noise in the afferent sensory pathway. For example, two neurons in AC that share input from a single upstream auditory never fiber might co-vary according to changes in the shared AN fiber's rate. This hypothesis is supported by findings that noise correlations are strongest in neurons with similar tuning.[46–49] A third hypothesis that has gained support recently is that noise correlations are primarily the result of latent inputs to local circuits that might not not be directly observes *e.g.* top-down co-activation of neurons by a neuromodulator. This explanation is corroborated by recent work showing that noise correlations can be accurately modeled using low-dimensional, latent processes that coherently modulate firing rates[50–53] and that correlate with state variables, such as arousal,[54] which are known to reflect neuromodulatory activity.[31,32] Strikingly, when the firing rates of neurons are "corrected" by regressing out these latent factors, noise correlations are no longer present in their activity.[51]

Neural population (de)coding

A common analysis technique for determining what information is represented in neural activity is neural decoding. Neural decoding uses the (spiking) activity of a single neuron, or population of

neurons, to predict which external stimuli were presented to the system. This approach has provided important insights into how and where information is represented in the brain. For example, classic work demonstrated that the ability to decode visual stimuli give activity of single neurons in area MT closely corresponded to animal's perceptual performance.[55] Thus, decoding can be used to identify a possible neural substrate for perceptual discrimination.[55] However, it has long been acknowledged that behavior likely does not depend solely on single neurons, but on the coordinated activity of many neurons within local neural circuits and communication across different brain areas.

Even before neural population recordings, a large body of theoretical work had explored the question of how information represented by single neurons could be combined into a neural population code.[46, 56–61] Early work demonstrated that pooling the information gleaned from serially recorded, independent single neurons leads to drastic overestimates of decoding accuracy; large neural populations can discriminate between different stimuli much more reliably than animals can perform the same behavioral task.[46] However, simulations demonstrated that if the stimulus-independent activity of single neurons is correlated, information scales sub-linearly with the number of neurons (Figure 1.6) and may more closely resemble animal's behavior,[46] highlighting the importance of neural population recordings for understanding information coding in the brain.

Figure 1.6: **Noise correlations cause stimulus information to scale sub-linearly in neural populations** Signal to noise ratio is plotted as a function of population size when pooling over neurons with different levels of noise correlation ($r = 0$ to $r = 0.5$). Figure adapted from.[46]

Structure of noise correlations and their impact on population coding

While early work demonstrated that noise correlations can limit the accuracy of population coding, subsequent work showed that this only occurred when noise was structured in a very specific way relative to the stimuli being decoded.[62] To illustrate this concept, consider the two-neuron illustration in Figure 1.7. When correlated noise is oriented along the optimal discrimination axis, it fundamentally

limits decoding accuracy. This special type of correlation is referred to as *information limiting*.[62] However, when noise is oriented in any other direction with respect to the sensory discrimination axis, optimal linear decoding is no longer affected, or can even be improved relative to the independent noise case. In large neural populations where the dimensionality of the space is high (N neurons = N dimensions), this latter case becomes more likely. In support of this, a number of experimental studies have since suggested that noise correlations largely have no effect on, or can even improve, neural decoding.[63–67] At the same time, recent data has found that the special case of information limiting correlations do exist,[68–70] however, they are difficult to detect without massive amounts of data and measuring them accurately requires the development of new analytical tools, which we discuss more in chapter 2 of this dissertation.[68]
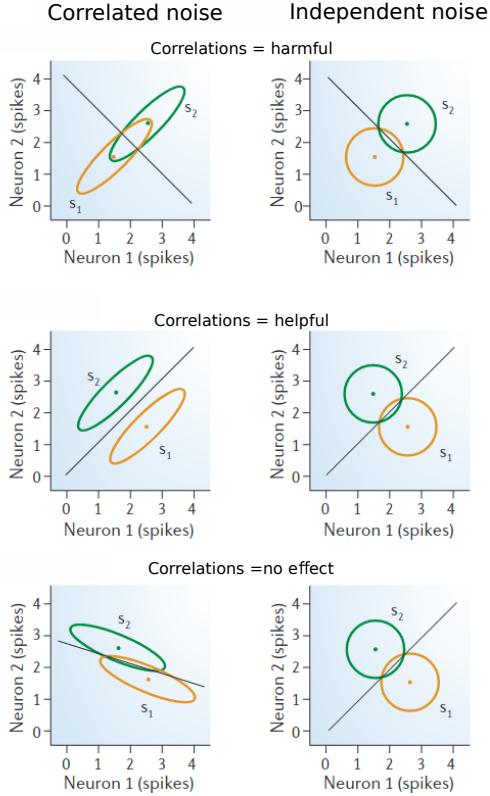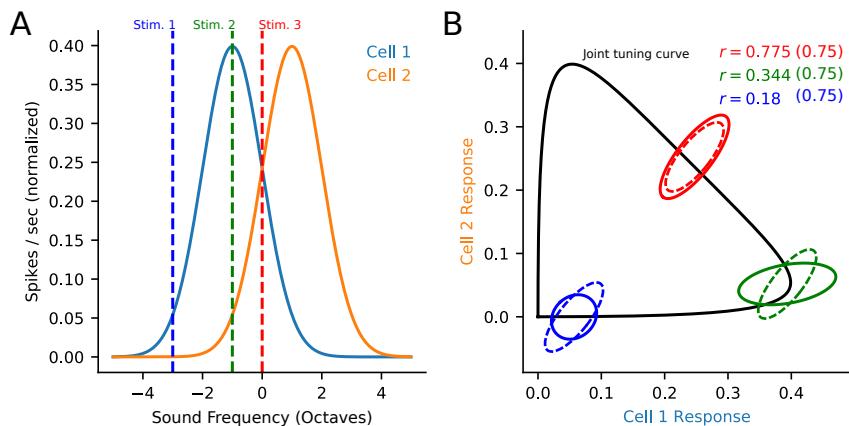
Figure 1.7: **Noise correlations can help or hurt decoding.** Simulated stimulus responses for two neurons for two stimuli (green vs. gold). Ellipses represent the distribution of spike count responses for each stimulus across many repetitions. Left column illustrates the case where response variability is correlated and the right column shows the case where variability is independent between neurons. In the first example (top), correlations are oriented perpendicular to the optimal discrimination boundary (black line) and therefore make decoding difficult relative to the independent noise scenario. Middle panel: correlations are beneficial to decoding. Bottom panel: correlations do not impact decoding. Thus, the interaction between the noise correlation and the optimal decision boundary determine the impact of noise correlation on stimulus decoding. Figure adapted from.[60]

Further complicating the role of correlations in decoding is the challenging, but critical question: How does the brain actually decode information? For example, does the brain perform linear decoding as illustrated in Figure 1.7? If so, is it optimal, or sub-optimal? Further, how do we as experimenters know that the population of neurons we are recording from is actually used by the brain to decode the stimuli that we present? The answers to each of these related questions can lead to dramatically different impacts of correlations on decoding and we discuss the implications of this more in the following section.

Finally, noise correlations are, by definition, stimulus-independent in the sense that they are not directly caused by sensory stimulus activation. However, their orientation and magnitude can still

change depending on the stimulus.[47,63,66] As an example to help understand this phenomenon, imagine that noise correlations are multiplicative. In this case, correlations between two neurons will be strongest when one or both neurons have a high firing rate *i.e.* during presentation of their preferred stimulus. Furthermore, imagine that the two neurons do not have the same tuning preference. This means that their peak firing rates will occur for different stimuli causing the orientation of correlations to rotate in a stimulus dependent manner (Figure 1.8). In diversely tuned population of neurons in the retina, noise correlations seem to follow this structure and contribute to improved decoding accuracy.[63,66] Additionally, even if noise correlations are due to additive noise and are not stimulus-dependent (dashed lines Figure 1.8), correlations can still have stimulus-dependent impacts on coding (correlations interfere with decoding stimulus 1 vs. stimulus 2, but not stimulus 2 vs. stimulus 3 – Figure 1.8). These findings and illustrations highlight the importance of sampling a diverse stimulus space to understand the impact of noise correlation on coding. Chapter 3 of this dissertation addresses this topic.

Figure 1.8: **Stimulus-dependence of noise correlations.** Two frequency tuned auditory neurons were simulated with shared, multiplicative (or additive) noise. **A.** Tuning curves for two simulated auditory neurons. Dashed lines indicate three specific stimuli. **B.** Joint tuning curve for cells 1 and 2 is shown in black. Colored ellipses indicate the responses distribution across many simulated trials for each of the three stimuli indicated in A. Solid lines – distribution for multiplicative noise, dashed lines – distribution for additive noise. Noise correlation coefficient for each stimulus is indicated in the figure legend. Number in parentheses is correlation coefficient for additive noise.

1.4 Behavior-dependent modulation of neural population codes
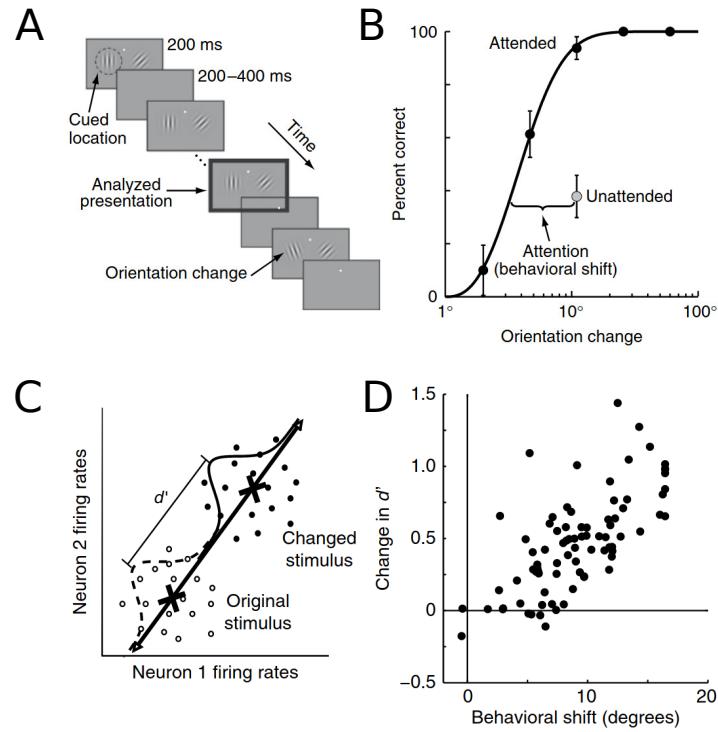
In the final section of this introduction, we introduce the concept of behavior state modulating noise correlations, and the impact this has on stimulus decoding, perception, and behavior. Although the focus of this dissertation is on auditory coding, arousal, and task engagement, most prior work

on the topic of state-dependent population codes, especially as they relate to stimulus coding and perception, has been done using visual selective attention tasks and recording from visual cortex. Therefore, in what follows we focus primarily on this body work.

Behavior state and noise correlations

Given that noise correlations can determine neural population coding accuracy, it is important to understand their relationship to behavior and perception. Early on it was demonstrated that noise correlations, like single cell receptive fields, are not a static property of neural populations. Across cortical brain regions, noise correlation strength is sensitive to a range of state-dependent changes, including fluctuations in pupil-indexed arousal,[27,28,71] locomotion,[27] anesthesia state,[51,72] task engagement,[73] selective attention,[26,74] and learning.[52] Although state-dependent changes in noise correlations are commonly reported, their impact on population coding and behavior remain a topic of debate.

In one of the first studies to investigate this, Cohen et al. trained macaque monkeys to perform a cued change detection task (Figure 1.9).[26] The authors demonstrated that monkey's ability to detect a change in the visual stimulus was improved when they were cued to attend to the location where the change was about to occur (Figure 1.9). While animals performed this task, Cohen et al. recorded neural activity from the visual area V4. They found that the behavioral improvement associated with attention coincided with a selective reduction in noise correlations amongst those neurons whose receptive fields overlapped with the visual stimulus. Furthermore, this change in correlation led to a direct improvement in neural decoding of the visual stimulus which correlated tightly with the animal's performance on single trials (Figure 1.9). Thus, this study offered one of the first examples where the neural population was found to be the key unit of computation and their results argued that noise correlations play an important role in perception and behavior.

Figure 1.9: **Attention improves perception and stimulus decoding. A.** Schematic of cued-change detection task. **B.** Example psychometric curve showing behavioral performance. Behavioral shift represents the improvement in perception attributed to selective spatial attention to the cued location. **C.** Cartoon schematic of procedure for measuring decoding accuracy (d') **D.** Change in d' between attended and unattended conditions is plotted as a function of behavioral shift (shown in B). Figure adapted from.[26]
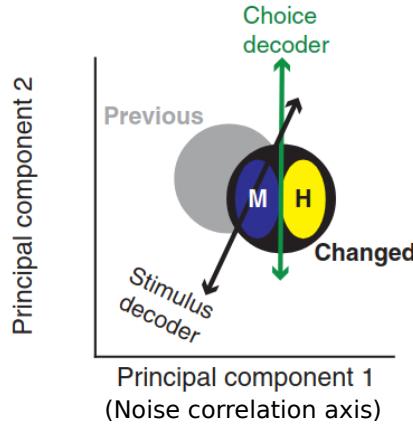
Follow up studies, however, have challenged these findings by demonstrating that the choice of decoding method is critical when determining the impact of correlations.[62,75] These studies have typically argued, as discussed in the prior section, that only under very specific conditions will correlations affect optimal linear decoding of stimulus identity[62,75] *e.g.* Figure 1.7. Yet, these theoretical results remain somewhat paradoxical given the experimental finding that changes in noise correlations, across different task designs and brain regions, are so tightly coupled with animal's perceptual performance.[26,52,73]

Sub-optimal decoding strategies

Ultimately, to understand the effect (or lack thereof) of noise correlations in sensory coding, we must first determine how neural activity informs behavioral choices. More precisely, how is neural activity in sensory areas decoded by downstream brain regions? Up to this point, we have discussed

neural decoding as a method to determine what stimulus information is contained in a neural code. While this approach provides an upper bound on the sensory information encoded by the system, it does not answer the question of whether and/or how it is actually used to guide perception and behavior. To do this, it is necessary to identify features of the neural activity that have an intersection between sensory and choice information.[76]

In a follow up study to,[26] the same group identified the dimension of neural activity in visual area V4 that carried maximal information about the monkey's choice.[52] Interestingly, they found that the largest dimension of noise correlation was aligned with the choice decoding dimension, but not with the optimal stimulus decoding axis. Thus, they concluded that animals seem to decode stimulus information sub-optimally for any particular task, but instead might use a general purpose decoding strategy that applies across many tasks. This hypothesis makes sense, as animal's performance would be perfect if they were to decode optimally from large neural populations.[52,77]

Figure 1.10: **Noise correlations are aligned with choice decoding dimension in V4 populations during cued-change detection.** Cartoon schematic adapted from.[52] Ellipses represent the distribution of spikes counts across trials of each condition: "M" corresponds to miss trials and "H" to hit trials. Grey indicates the stimulus shown prior to the change and black indicates the changed stimulus (see Figure 1.9). The first principal component of noise correlations is aligned with the choice decoding axis, but only partially aligned with the optimal stimulus decoding axis. This leads to sub-optimal stimulus decoding.

Another approach to identifying important dimensions in neural activity is to record simultaneously from multiple brain regions and determine which aspects of upstream activity propagate downstream. Although this approach is new and is not necessarily directly related to perception and behavior, it has been used recently to demonstrate that communication subspaces between cortical areas appear to be low-dimensional,[78] like noise correlations themselves[51–53] (see above: *Origin of noise correlations*). Thus, it appears likely to be a promising avenue for future research.

Origin and dimensionality of state-dependent noise correlation modulation

Both modeling[53] and experimental[79] studies have argued that noise correlations are low-dimensional, meaning that neural activity primarily co-varies along a single axis in state space. However, recent recordings from larger neural populations[68] have discovered that multiple significant dimensions of correlated activity exist. These findings are in line with other recent work showing that multiple dimensions of behavior (*e.g.* running, whisking, face movement, pupil size etc.) can all coherently modulate neural activity.[29] If each of these state variables affects a slightly different subset of neurons (*e.g.* running modulates a different population of neurons than pupil whisking), then correlated activity in the population will be at least two dimensional.

These recent findings raise the question of whether state-dependent decreases in correlated activity, *e.g.* due to selective spatial attention, reflect suppression of all dimensions of noise correlations or if they are restricted to a relatively low-dimensional subspace. In the case of selective attention, this was investigated in a re-analysis of the cued-change detection task data[26] highlighted in the previous section. In this study, Rabinowitz et al. used a Generalized Linear Model (GLM) to predict the time-varying activity of V4 neural populations as a combination of stimulus drive, attention mediated gain of sensory responses (cue), and latent sources of shared variability between neurons (shared modulators) (Figure 1.11).[50] They found that the model performed best with just a small number of latent, shared modulators and, critically, only the variance of a single modulator was sensitive to the attention cue (attend in vs. attend out). Thus, the study concluded that the state variable, attention, seems to modulate just a single dimension of shared noise in V4 populations.
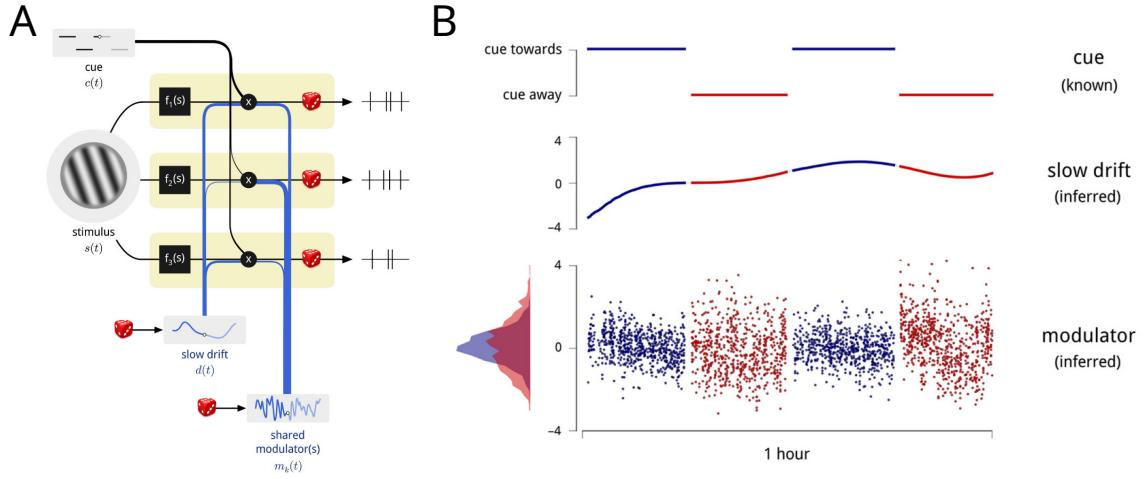
Figure 1.11: **Attentional modulation of noise correlations is low-dimensional. A.** Schematic detailing a GLM designed to predict firing rates in V4 neurons. The slow drift and shared modulators are inferred parameters that are fit by the model, unknown the experimenter. **B.** Results from a single model fit. The variance of the fit shared modulator scales with the attention cue, determined by the experimenter. Figure adapated from.[50]

In the work by Rabinowitz et al., attention was modeled as binary cue that experimenters had precise control over. However, in practice an animal's attention likely wanders from trial to trial. Thus, an alternative interpretation of the results is that animal's attention is more stable during attend-in conditions and the shared modulator Rabinowitz et al. identified is in fact just the trial-to-trial variability of attentional gain itself. This is slightly a nuanced point, but has important implications for both perception and physiological mechanisms. In this latter scenario, perception is improved during the attend-in condition not because noise is suppressed, but because the animal's attention wanders less. That is, they are more certain of where to direct their attention and therefore neural responses vary less and perception improves. Physiologically, in this case, only a single mechanism is required to support both the well-documented attention-dependent gain of sensory responses[80] and the more recently observed reduction in noise correlations.[26]

An elegant study by Denfield et al. recently tested this hypothesis by designing a variant of the cued-change detection task in which they were able to manipulate the stability of an animal's attention between experimental blocks.[81] In agreement with the theory outlined above, noise correlations were strongest when attention was most variable and when perception was the worst,[81] supporting a parsimonious model where changes in single unit sensory gain and noise correlations are mediated by a single neural substrate.

2 Dimensionality reduction for neural population decoding

Charles R. Heller[1,2] & Stephen V. David[1]

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

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

Abstract

Rapidly developing technology for large scale neural recordings has allowed researchers to measure the activity of hundreds to thousands of neurons at single cell resolution *in vivo*. Neural decoding analyses are a widely used tool used for investigating what information is represented in this complex, high-dimensional neural population activity. Most population decoding methods assume that correlated activity between neurons has been estimated accurately. In practice, this requires large amounts of data, both across observations and across neurons. Unfortunately, most experiments are fundamentally constrained by practical variables that limit the number of times the neural population can be observed under a single stimulus and/or behavior condition. Therefore, new analytical tools are required to study neural population coding while taking into account these limitations. Here, we present a simple and interpretable method for dimensionality reduction that allows neural decoding metrics to be calculated reliably, even when experimental trial numbers are limited. We illustrate the method using simulations and compare its performance to standard approaches for dimensionality reduction and decoding by applying it to single-unit electrophysiological data collected from auditory cortex.

2.1   Introduction

Neural decoding analysis identifies components of neural activity that carry information about the external world (*e.g.* stimulus identity). This approach can offer important insights into how and where information is encoded in the brain. For example, classic work by Britten et al. demonstrated that the ability of single neurons in area MT to decode visual stimuli closely corresponds to animal's perceptual performance.[55] Thus, by using decoding the authors identified a possible neural substrate for detection of motion direction.[55] Yet, behavior does not depend solely on single neurons. In the years since this work, many theoretical frameworks have been proposed for how information might be pooled across individual neurons into a population code.[46,56–61] One clear theme that has emerged from this work is that stimulus independent, correlated activity (*i.e.* noise correlations) between neurons may substantially impact information coding.[46,57–61] This has now been confirmed *in vivo* using decoding analysis to measure the information content of large neural populations.[68–70] Therefore, covariability

between neurons must be taken into account when measuring population coding accuracy.

Under most experimental conditions, estimates of pairwise correlation between neurons is unreliable due to insufficient sampling (*e.g.* too few stimulus repeats).[82] In these situations, traditional decoding algorithms are likely to over-fit to noise in the neural data. This issue becomes even more apparent as the number of pairwise interactions that must be estimated increases, a situation that is becoming more common due to the recent explosion in large-scale neurophysiology techniques.[83] In some cases, *e.g.* for chronic recording experiments and anesthetized preps, the number of trials can be increased to circumvent this issue. However, in behavioral experiments, where the number of trials is often fundamentally limited by variables such as animal performance, new analytical techniques for decoding are required.

Here, we present decoding-based dimensionality reduction (*dDR*), a simple and generalizable method for dimensionality reduction that significantly mitigates issues around estimating correlated variability in experiments with a relatively low ratio of observations to neurons. Our method takes advantage of recent observations that population covariability is often low-dimensional[50–52,54] to define a subspace where decoding analysis can be performed reliably while still preserving the dominant mode(s) of population covariability. The *dDR* method can be applied to data collected across many different stimulus and/or behavior conditions, making it a flexible tool for analyzing a wide range of experimental data.

We motivate the requirement for dimensionality reduction by illustrating how estimates of a popular information decoding metric, d'^2,[57,58] can be biased by small experimental sample sizes. Building on a simple two-neuron example, we demonstrate that low-dimensional structure in the covariability of simulated neural activity can be leveraged to reliably decode stimulus information, even when the number of neurons exceeds the number of experimental observations. Finally, we use a dataset collected from primary auditory cortex to highlight the advantages of using *dDR* for neural population decoding over standard principal component analysis.

2.2 Results

2.2.1 Small sample sizes limit the reliability of neural decoding analysis

Linear decoding, a common analytical method in neuroscience, identifies a linear, weighted combination of neural activity along which distinct conditions (*e.g.* different sensory stimuli) can be discriminated. In neural state-space, this weighted combination is referred to as the decoding axis, $\boldsymbol{w}_{opt}$, and it is the line along which the distance between stimulus classes is maximized and trial-trial

variance is minimized (Fig. 2.1a, b). To quantify decoding accuracy, single-trial neural activity is projected onto this axis and a decoding metric is calculated to quantify the discriminability of the two stimulus classes. Here, we use d'^2, the discrete analog of Fisher Information.[57, 58] This discriminability metric has been used in a number of previous studies[59, 62, 68–70] and has a direct relationship to classical signal detection theory.[57, 84]

Looking at the simulated data in Figures 2.1a and b, one can appreciate that an accurate estimate of $\boldsymbol{w}_{opt}$ requires knowledge of both the mean response evoked by each stimulus class ($\boldsymbol{\mu}_a$ vs. $\boldsymbol{\mu}_b$), as well the population covariance, Σ (summarized by the ellipses in Fig. 2.1a and b). Indeed, d'^2, is directly dependent on these features:

$$d'^2 = \Delta\boldsymbol{\mu}^T \boldsymbol{w}_{opt} \tag{2.1}$$

$$\boldsymbol{w}_{opt} = \Sigma^{-1}\Delta\boldsymbol{\mu} \tag{2.2}$$

$$\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_a - \boldsymbol{\mu}_b \tag{2.3}$$

Where $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$ are the Nx1 vectors describing the mean response of an N-neuron population to two stimuli, a vs. b, respectively, and Σ is the average NxN covariance matrix $\frac{1}{2}(\Sigma_a + \Sigma_b)$ (*e.g.* Fig. 2.1c).

In practice, the pairwise spike count covariance between neurons (often referred to as noise correlation, or r_{sc}) is reported to be very small – on the order of 10^{-1} or 10^{-2}.[26, 45, 65] As we can see from the shuffled distribution in Figure 2.1a (bottom), this can pose a problem for accurate estimates of the off-diagonal elements in Σ, and, as a consequence, $\boldsymbol{w}_{opt}$ itself. This difficulty is especially pronounced when sample sizes are relatively small (compare Fig. 2.1a to b). The estimates of covariance and stimulus discriminability improve with increasing sample size, but robust performance is not reached until ≈ 100 stimulus repetitions, even for this case with relatively strong covariance (Fig. 2.1d). The sample sizes (*e.g.* number of trials) in most experiments, especially those involving animal behavior, are typically much lower, raising the question: How can one reliably quantify coding accuracy in large neural populations observed over relatively few trials?
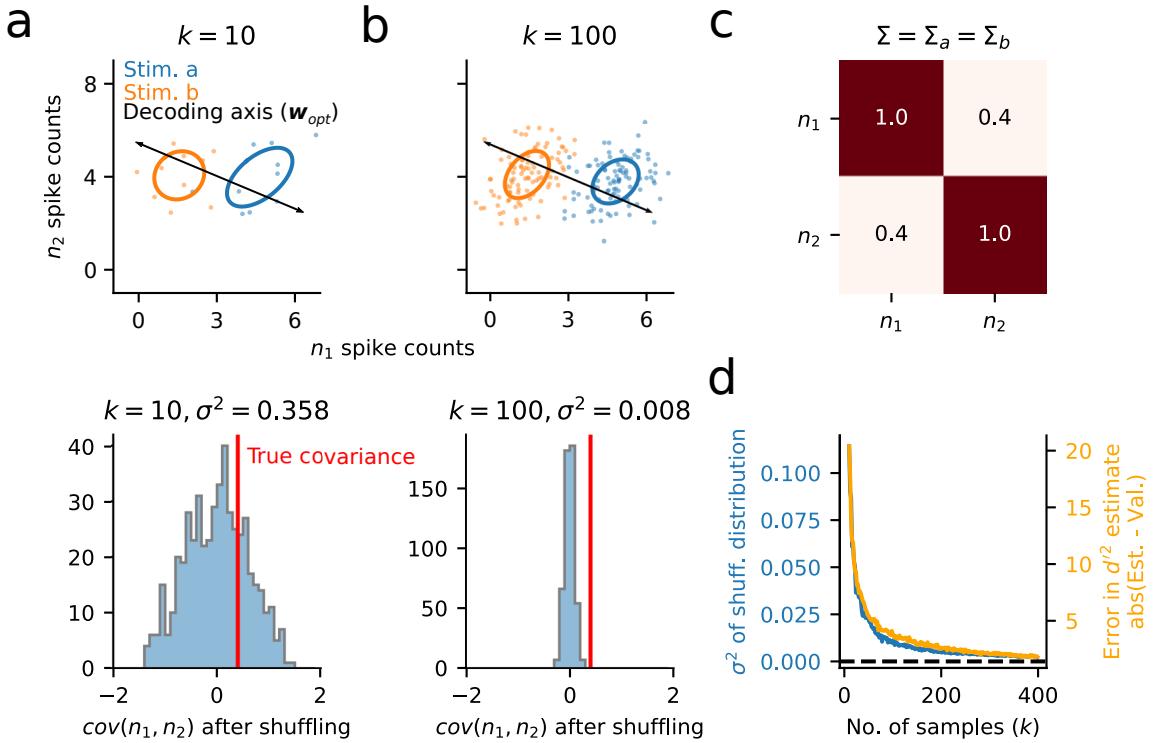
Figure 2.1: **Measurements of pairwise covariance and discriminability are unreliable when sampling is limited. a.** Top: $k = 10$ single trial spike count responses are drawn from standard multivariate Gaussians $\mathcal{N}(\boldsymbol{\mu}_a, \Sigma)$ and $\mathcal{N}(\boldsymbol{\mu}_b, \Sigma)$ corresponding to two different stimulus conditions, a and b. Ellipses show the standard deviation of spike counts across trials. Bottom: Reliability of the pairwise covariance estimate between neuron 1 (n_1) and neuron 2 (n_2) is calculated by shuffling values of n_1 500 times. The true covariance (red line) falls within this distribution, indicating that estimates of covariance are not reliable for $k = 10$. **b.** Same as in (a), but drawing $k = 100$ samples for each stimulus. The narrower distribution of permuted measures indicates a greater likelihood of identifying an accurate estimate of covariance. **c.** The covariance matrix, Σ, used to generate data in (a)/(b). The true pairwise covariance for this pair of simulated neurons has a value of 0.4. **d.** Variance (σ^2) of covariance estimates based on the permutation analysis in (a)/(b) for a range of sample sizes, k (blue). Variance decays as $\mathcal{O}(\frac{1}{k-1})$ (see Appendix). Overlaid is the difference in stimulus discriminability, d'^2 (Eqn. 2.1), between estimation and validation sets (50-50 split) estimated for each sample size (orange). Large values in the d'^2 difference for low k indicate overfitting of $\boldsymbol{w}_{opt}$ to the estimation data. This difference asymptotes toward zero as sample size increases and the estimate of covariance becomes reliable.

2.2.2 Neural activity is low-dimensional
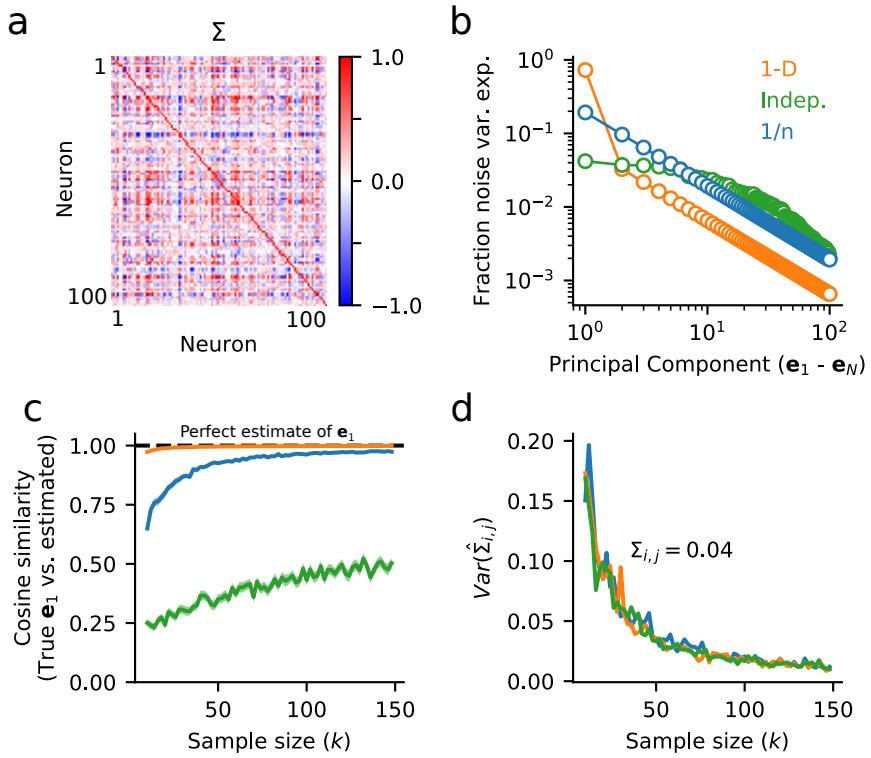
Analysis of neural population data with dimensionality reduction has consistently revealed low-dimensional structure in neural activity.[85] Specifically, recent studies have found that stimulus-independent variability (*i.e.* noise correlations) is dominated by a small number of latent dimensions.[50,51,53,54] Noise correlations are thought to impact stimulus coding accuracy[60] and are known to depend on internal states, such as attention, that affect behavioral task performance.[26,50,52,73]

23

These findings suggest that the space of neural activity relevant for understanding stimulus decoding, and its relationship to behavior, may be small relative to the total number of recorded neurons.

When population data exhibits low-dimensional structure, the largest eigenvector(s) of Σ (*i.e.* the top principal components of population activity) provides a reasonable, low-rank approximation to the full-rank covariance matrix. Importantly, these high variance dimensions of covariability can be estimated accurately even from limited samples. To illustrate this, we simulated population spike counts, X, for $N = 100$ neurons by drawing k samples from a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ (Eqn. 2.4).

$$X = \mathcal{N}(\boldsymbol{\mu}, \Sigma) + \epsilon_{indep.} \tag{2.4}$$

Where in Eqn. 2.4, $\epsilon_{indep.}$ represents a small amount of independent noise added to each neuron, effectively removing any significant structure in the smaller noise modes.

Figure 2.2: **Low-dimensional correlated activity can be estimated reliably for neural populations, even when pairwise covariance cannot. a.** Example covariance matrix, Σ, for a 100-neuron population with low-dimensional covariance structure. **b.** Scree plot shows the fraction of total population variance captured along each noise dimension, computed by PCA, for three different datasets with varying dimensionality. Orange: 1-dimensional noise (1-D), covariance matrix in (a); green: independent noise (Indep.); blue: power law decay ($1/n$). **c.** Surrogate datasets with varying numbers of samples, k, are drawn from the three noise distributions in (b). For each dataset, the cosine similarity between the estimate of the largest noise dimension, $\hat{\boldsymbol{e}}_1$, and the true noise dimension, $\boldsymbol{e}_1$, is plotted as function of sample size. For low-dimensional data, $\boldsymbol{e}_1$ can be estimated very reliably. **d.** Variance in the estimate of covariance, $\Sigma_{i,j}$, for two neurons with a true covariance of 0.04 is plotted as a function of the number of trials, as in Figure 1d. Even at sample sizes > 100, $Var(\hat{\Sigma}_{i,j}) \approx 0.02$, corresponding to a standard deviation of ≈ 0.14. Therefore, estimates of $\Sigma_{i,j}$, may be off by up to an order of magnitude. Note that the amount of uncertainty does not depend on the dimensionality of the data, and results for all three datasets overlap (see Appendix for an analytical derivation).

To investigate how different noise structures impact estimates of Σ, we simulated three different surrogate populations. First, we simulated data with just one large, significant noise dimension (Fig. 2.2, 1-D data, orange). In this case, the first eigenvector can be estimated reliably, even from just a few samples (Fig. 2.2c). However, when the noise is independent and shared approximately equally across all neurons, estimates of the first eigenvector are poor (Fig. 2.2, Indep. noise, green). These first two simulations represent extreme examples – in practice, population covariability tends to be spread across at least a few significant dimensions.[86] To investigate a scenario that more closely mirrors this structure, we simulated a third dataset where the noise eigenspectrum decayed as $1/n$, where n goes

from $n = 1$ to N. Recent studies of large neural populations suggest that this power law relationship is a reasonable approximation to real neural data.[86] In this case, by $k \approx 50$ trials, estimates of the first eigenvector are highly reliable, approaching a cosine similarity of ≈ 0.9 between the estimated and true eigenvectors (Fig. 2.2, $1/n$ noise, blue). In all simulations, regardless of dimensionality, we find that estimates of single elements of Σ (*i.e.* single noise correlation coefficients) are highly unreliable (Fig. 2.2d), as we see in the two-neuron example (Fig. 2.1d).

Collectively, these simulations demonstrate that accurate estimates of covariance need not necessarily be limited by uncertainty in estimates of individual noise correlation coefficients themselves. In the following sections we describe a simple decoding-based dimensionality reduction algorithm, dDR, that leverages low-dimensional structure in neural population activity to facilitate reliable measurements of neural decoding.

2.2.3   decoding-based Dimensionality Reduction (dDR)

The dDR algorithm operates on a pairwise basis. That is, given a set of neural data collected over S different conditions, a different dDR projection exists for each of the $\frac{S!}{2!(S-2)!}$ unique pairs. For simplicity, we will describe the case where $S = 2$, and consider these to be two unique stimulus conditions. However, note that the method can be applied in exactly the same manner to handle datasets with many different types and numbers of decoding conditions, where a unique dDR projection would then exist for each pair.

Let us consider the spiking response of an N-neuron population evoked by two different stimuli, S_a and S_b, over k-repetitions of each stimulus. From this data we form two response matrices, A and B, each with shape Nxk. Remembering that our goal is to estimate discriminability (d'^2, Eqn. 2.1), the dDR projection should seek to preserve information about both the mean response evoked by each stimulus condition, $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$, as well as the stimulus-independent noise covariance, Σ. Therefore, we define the first dimension of dDR to be the axis that maximally separate $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_b$. We call this the *signal* axis.

$$signal = \boldsymbol{\mu}_a - \boldsymbol{\mu}_b = \Delta\boldsymbol{\mu} \tag{2.5}$$

Next, we compute the first eigenvector of Σ, $\boldsymbol{e}_1$. This represents the largest noise mode of the neural population activity. Together, *signal* ($\Delta\boldsymbol{\mu}$) and $\boldsymbol{e}_1$ span the plane in state-space that is most optimized for reliable decoding. Finally, to form an orthonormal basis, we define the second dDR dimension as the axis orthogonal to $\Delta\boldsymbol{\mu}$ in this plane. As this second dimension is designed to preserve

noise covariance, we call this the $noise_1$ axis.

$$noise_1 = \boldsymbol{e}_1 - \boldsymbol{e}_1 \Delta \boldsymbol{\mu}^T \tag{2.6}$$

The process outlined above is schematized graphically in Figure 2.3.

Thus, the *signal* and $noise_1$ axes make up a 2xN set of weights, analogous to the loading vectors in standard PCA, for example. By projecting our Nxk data onto this new basis, we capture both the stimulus coding dimension ($\Delta \boldsymbol{\mu}$) and preserve the principal covariance dimension ($\boldsymbol{e}_1$), two critical features for measuring stimulus discriminability. Importantly, because $\boldsymbol{e}_1$ can be measured more robustly than Σ itself (Figure 2.2), performing this dimensionality reduction helps mitigate the issues we encounter due to small sample sizes and large neural datasets.
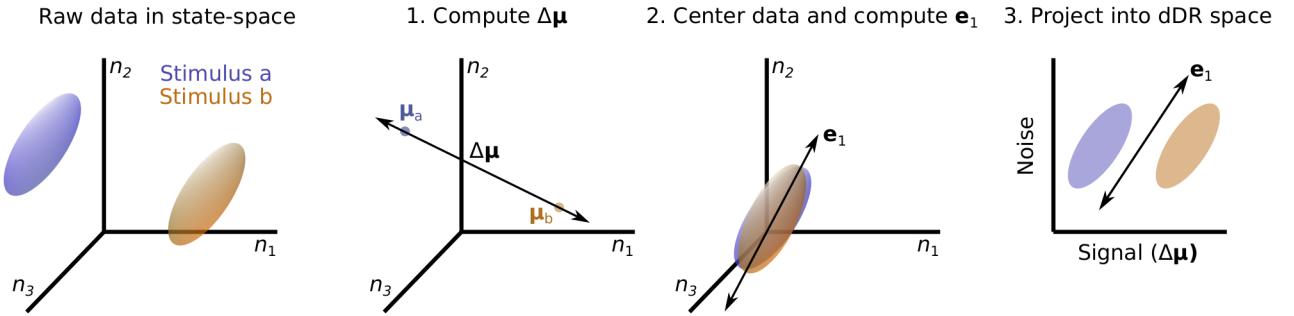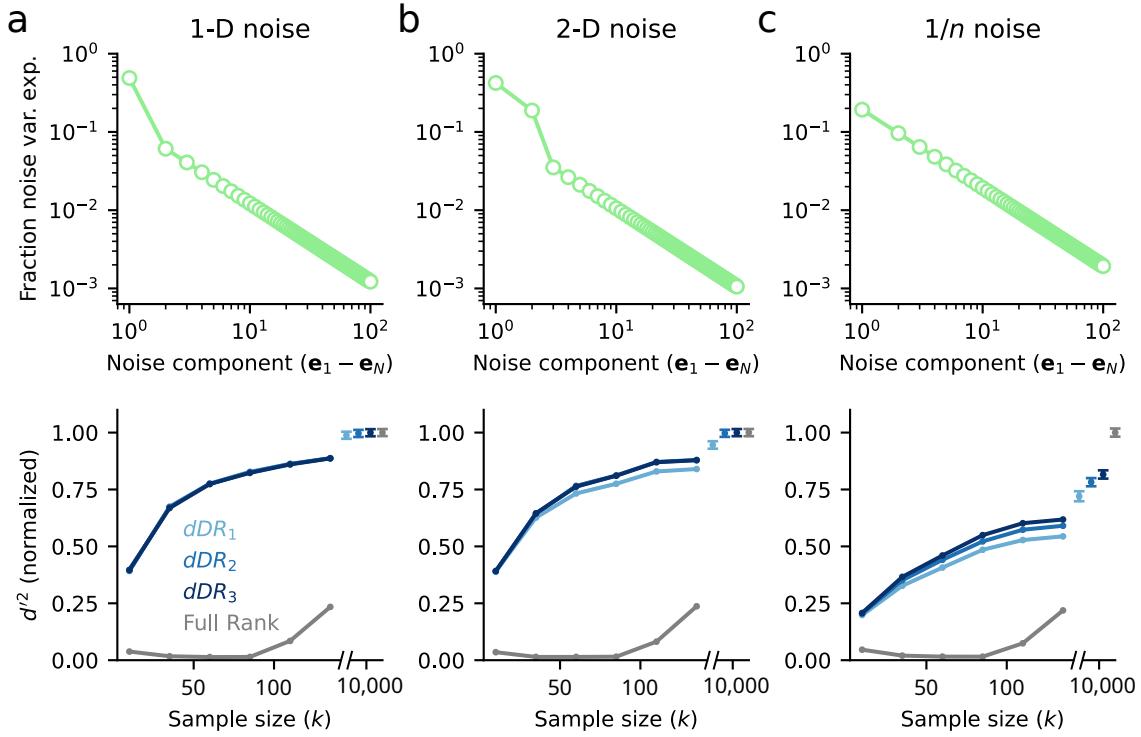
Figure 2.3: **decoding-based Dimensionality Reduction** (dDR). Left to right: Responses of 3 neurons (n_1, n_2, n_3) to two different stimuli are schematized in state-space. Ellipsoids illustrate the variability of responses across trials. **1.** To perform dDR, first the difference is computed between the two mean stimulus responses, $\Delta \boldsymbol{\mu}$. **2.** Next, the mean response is subtracted for each stimulus to center the data around 0, and PCA is used to identify the first eigenvector of the noise covariance matrix, $\boldsymbol{e}_1$ (additional noise dimensions $\boldsymbol{e}_m, m > 1$ can be computed, see text). **3.** Finally, the raw data are projected onto the plane defined by $\Delta \boldsymbol{\mu}$ and $\boldsymbol{e}_1$.

As mentioned in the previous section, neural data often contains more than one significant dimension of correlated trial-trial variability. To account for this, dDR can easily be extended to include more noise dimensions. To include additional dimensions, we deflate the spike count matrix, X, by subtracting out the *signal* and $noise_1$ dimensions identified by standard dDR, then perform PCA on the residual matrix to identify m further *noise* dimensions. Note, however, that for increasing m the variance captured by each dimension gets progressively smaller. Therefore, estimation of these subsequent noise dimensions becomes less reliable and will eventually become prone to over-fitting, especially with small sample sizes. For this reason, care should be taken when extending dDR in this way.

To demonstrate the performance of the dDR method, we generated three sample datasets containing $N = 100$ neurons and $S = 2$ stimulus conditions. Each of the three datasets contained unique noise covariance structure: 1. Σ contained one significant dimension (Fig. 2.4a) 2. Σ contained two significant dimensions (Fig. 2.4b) 3. Noise variance decayed as $1/n$ (Fig. 2.4c). For each dataset, we measured cross-validated d'^2 between stimulus condition a and stimulus condition b using standard dDR with one noise dimension (dDR_1), with two noise dimensions (dDR_2), or with three noise dimensions (dDR_3). We also estimated d'^2 using the full-rank data, without performing dDR. Figure 2.4 plots the decoding performance of each method as a function of sample size (*i.e.* number of stimulus repetitions). In each case, d'^2 is normalized to the asymptotic performance of the full-rank approach, when the number of samples is $>>$ than the number of neurons. This provides an approximate estimate of true discriminability for the population.

In contrast to the full-rank data where overfitting leads to dramatic underestimation of d'^2 on the test data for most sample sizes (Fig. 2.4 grey lines), we find that d'^2 estimates after performing dDR are substantially more accurate and, critically, more reliable across sample sizes. That is, asymptotic performance of the dDR method is reached much more quickly than for the full-rank method.

For the one-dimensional noise case, note that there is no benefit of including additional dDR dimensions (Fig. 2.4a), while for the higher dimensional data shown in Figure 2.4b-c, we see some improvements with dDR_2 and dDR_3. However, these benefits don't begin to appear until k becomes large and they diminish with increasing noise dimensions – the improvement of dDR_2 over dDR_1 is larger than that of dDR_3 to dDR_2 Fig. 2.4b-c. This is because subsequent noise dimensions are, by definition, lower variance and therefore more difficult to estimate reliably from limited sample sizes.
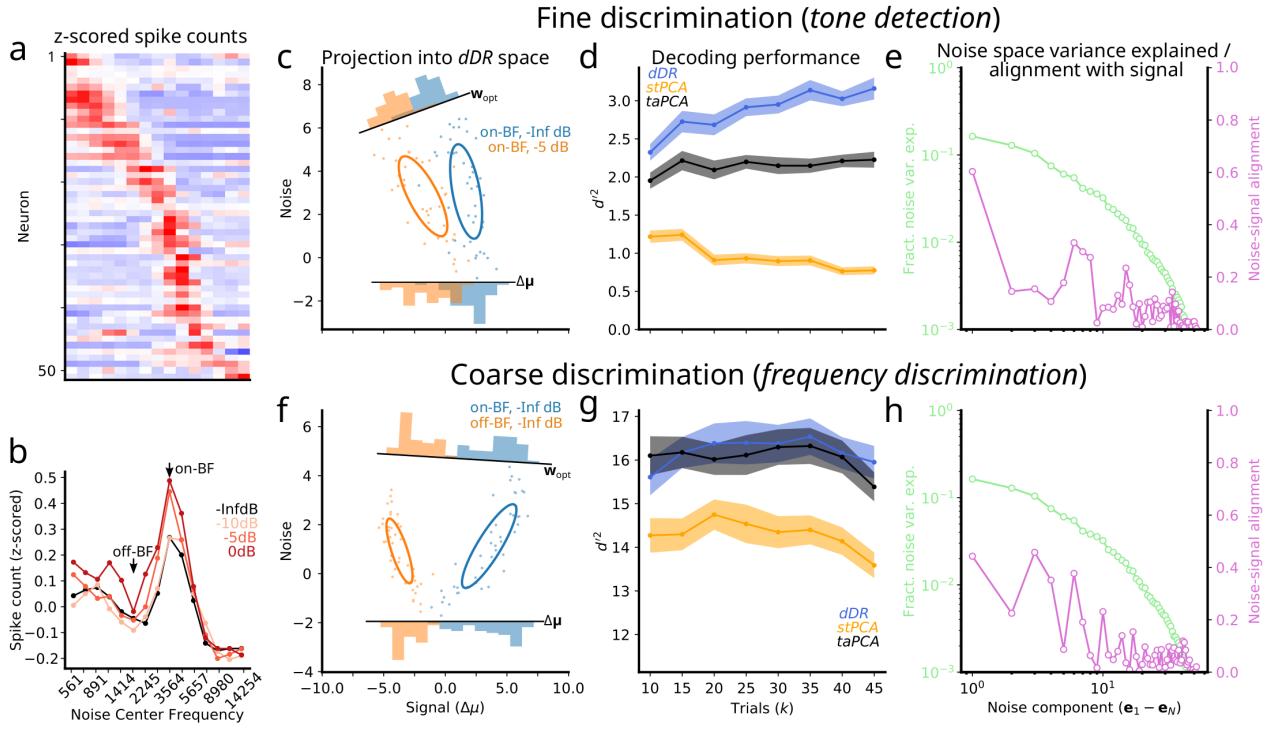
Figure 2.4: **Evaluation of decoding accuracy and reliability with** dDR. **a.** Analysis of data with one-dimensional (1-D) noise covariance. For each sample size, k, 100 datasets were generated from the same multivariate Gaussian distribution (Eqn. 2.4) where Σ was a rank-one covariance matrix and the mean response vector, $\boldsymbol{\mu}$, corresponded to one of two stimulus conditions, a or b. Top: Scree plot of noise covariance. Bottom: Cross-validated discriminability, d'^2, between a and b computed with full-rank data and with dDR using one (dDR_1), two (dDR_2) or three (dDR_3) noise dimensions, as a function of sample size. Mean d'^2 across all 100 surrogate datasets is shown here. For $k \gg N$, the dDR results converge to the asymptotic value of the full-rank d'^2. However, even for small k, the dDR analyses estimates are much more accurate than the full-rank approach. **b.** Same as in (a), but for two-dimensional noise covariance data. In this case, dDR_2 captures the second noise dimension and outperforms the standard 1-D approach (dDR_1) **c.** Same as in (a) and (b), but for $1/n$ noise covariance.

2.2.4 dDR recovers more decoding information than standard principal component analysis

One popular method for dimensionality reduction of neural data is principal component analysis (PCA).[85] Generally speaking, PCA can be implemented on neural data in one of two ways: single trial PCA or trial-averaged PCA. In the single trial approach ($stPCA$), principal components are measured across all single trials and all experimental conditions. The resulting PCs capture variance both across single trials and across different $e.g.$ stimulus conditions. In trial-averaged PCA ($taPCA$), single trial responses are averaged per experimental condition first, and PCs are measured over the resulting N-neuron x S-condition spike count matrix. In this case, for different stimulus conditions, the PCs specifically capture variance of stimulus-evoked activity rather than trial-trial variability,

making this a more logical choice for many decoding applications. In the case of $S = 2$, as we have outlined above for the dDR illustration (Fig. 2.3), $taPCA$ is equivalent to $\Delta\boldsymbol{\mu}$, the first dDR dimension. Thus, dDR can roughly be thought of as a way to combine $taPCA$ and $stPCA$ – $taPCA$ identifies the *signal* dimension and $stPCA$ identifies the *noise* dimension(s).

To demonstrate the relative decoding performance achieved using each method, we applied each to a dataset collected from primary auditory cortex in an awake, passively listening ferret. $N = 52$ neurons were recorded simultaneously using a 64-channel laminar probe[40] as in.[33,87,88] Auditory stimuli consisting of narrowband (0.3 octave bandwidth) noise bursts were presented alone (-Inf dB) or with a pure tone embedded at varying SNRs (0 dB, -5 dB, -10 dB) in the hemifield contralateral to the recording site (see Experimental Methods). Each stimulus was repeated 50 times. For $stPCA$ and dDR, we selected only the top $m = 2$ total dimensions, and for $taPCA$, we selected the single dimension, $\Delta\boldsymbol{\mu}$, that exists for $S = 2$. This dataset allowed us to investigate how each dimensionality reduction method performs for two distinct, behaviorally relevant neural decoding questions: How well can neural activity perform fine discriminations (*tone-in-noise detection*), discriminating noise alone vs. noise with tone? How well can it perform coarse discriminations (*frequency discrimination*), discriminating noise centered at frequency A vs. noise at frequency B?

Figure 2.5: *dDR* **outperforms** *PCA* **for fine sensory discrimination. a.** Heatmap shows mean *z*-scored spike counts of $N = 52$ simultaneously recorded units for 15 different narrowband noise bursts (0.3 octave bandwidth tiling 5 octaves, *x*-axis). Each row shows tuning for one unit, with red indicating higher firing rate response. **b.** Population tuning curve for noise alone (black, data from panel a) and noise plus $-10, -5,$ and 0 dB tones (light to dark red), computed by averaging tuning curves across neurons. **c-e.** Decoding analysis for tone-in-noise detection. **c.** Scatter plot compares single trial responses to noise alone at best frequency (on-BF, blue) vs. noise $+ -5$dB tone (orange), projected into *dDR* space. Ellipses show standard deviation across trials, marginal histograms show projection of data onto optimal decoding axis ($\boldsymbol{w}_{opt}$) or onto $\Delta\boldsymbol{\mu}$ (equivalent to performing trial-averaged *PCA*). **d.** Estimate of d'^2 as a function of sample size (number of trials, k) using each dimensionality reduction method. For each point, d'^2 was averaged over 100 random samplings of k trials, drawn without replacement. Shading indicates standard error. **e.** Fraction variance explained by each noise component (green) computed by performing *PCA* on mean-centered single trial data. The alignment of each noise component with the signal axis is shown in purple. **f-h** Same as panels (c)-(e), for noise alone on-BF vs. noise along off-BF (see panel b).

The A1 dataset displayed a range of frequency tuning (Fig. 2.5a), with the majority of units tuned to ≈ 3.5 kHz. We therefore defined this as the best frequency of the recording site (on-BF, Fig. 2.5b). For *tone detection*, we measured discriminability (d'^2, Eqn. 2.1) between on-BF noise alone (on-BF, -Inf dB) and on-BF noise plus tone (on-BF, -5 dB), which each drove similar sensory responses (Fig. 2.5b-c). For *frequency discrimination*, we measured discriminability between the neural responses to on-BF noise and off-BF noise, where off-BF was defined as ≈ 1 octave away from BF, and drove a very different population response (Fig. 2.5b, f). In both cases, *taPCA* and *dDR* outperformed *stPCA* (Fig. 2.5d, g). This first result is unsurprising due to the fact that *stPCA* is the only method

31

not explicitly designed to capture variability in the sensory response. The top PCs are dominated by dimensions of trial-trial variability that do not necessarily contain stimulus information and thus underestimate d'^2 relative to the other two methods.

We also find that dDR consistently performs as well or better than $taPCA$. For the *tone detection* data, the sensory signal ($\Delta\boldsymbol{\mu}$) is small (*i.e.*, trial-averaged responses to the two stimuli were similar) and covariability is partly aligned with $\Delta\boldsymbol{\mu}$. Under these conditions, dDR makes use of correlated activity to optimize the decoding axis ($\boldsymbol{w}_{opt}$) and improve discriminability. $taPCA$, on the other hand, has no information about these correlations and is therefore equivalent to projecting the single trial responses onto the *signal* axis, $\Delta\boldsymbol{\mu}$. Thus, it underestimates d'^2 (Fig. 2.5c, d). In the *frequency discrimination* example, $\Delta\boldsymbol{\mu}$ is large. The covariability has similar magnitude to the previous example, but it is not aligned to the discrimination axis, and thus has no impact on $\boldsymbol{w}_{opt}$. In this case, dDR and $taPCA$ perform similarly (Fig. 2.5f-g). These examples highlight that under behaviorally relevant conditions, dDR can offer a significant improvement over standard PCA, even with as few as 10 trials.

2.3 Discussion

We have described a new, simple method for dimensionality reduction of neural population data, dDR. This approach combines strategies for both trial-averaged PCA and single-trial PCA to identify important dimensions of population activity that govern neural coding accuracy. Using both simulated and real neural data, we demonstrated that the method performs robustly for neural decoding analysis in low experimental trial count regimes where the performance of full-rank methods break down. Across a range of behaviorally relevant stimulus conditions, dDR consistently performs as well or better than standard principal component analysis.

2.3.1 Applications

dDR is designed to optimize the performance of linear decoding methods in situations where sample sizes are small. This is often the case for neurophysiology data collected from behaving animals, where the number of stimulus and/or behavior conditions are fundamentally limited by task performance. In these situations, using full-rank decoding methods is unfeasible as it leads to dramatic overfitting and unreliable performance.[82] Dimensionality reduction methods, such as PCA, can be used to mitigate overfitting issues. However, the correct implementation of PCA in neural data is often ambiguous, and multiple different approaches to dimensionality reduction have been proposed.[85] We suggest dDR as a simple, standardized alternative that captures the strengths of

different PCA approaches. Unlike conventional PCA, the *signal* and *noise* axes that comprise the dDR space have clear interpretations with respect to neural decoding. Importantly, dDR components explicitly preserve stimulus-independent population covariability. In addition to being important for overall information coding, this covariability is known to depend on behavior state[26, 50, 52, 73, 89] and stimulus condition.[45, 63, 66, 90] Therefore, approaches that do not preserve these dynamics, such as trial-averaged PCA, may not accurately characterize how information coding changes across varying behavior and/or stimulus conditions.

2.3.2 Interpretability and visualization

A key benefit of dDR is that the axes making up the dDR subspace are easily interpretable: The first axis (*signal*) represents the dimension with maximal information about the difference in evoked activity between the two conditions to be decoded, and the second (*noise*) axis captures the largest mode of condition-independent population covariability in the data. Therefore, within the dDR framework it is straightforward to investigate how this covariability interacts with discrimination, an important question for neural information coding. Further, standard dDR (with a single noise dimension) can be used to easily visualize high-dimensional population data, as in Fig. 2.5. For methods like PCA, it can be difficult to dissociate signal and noise dimensions, as the individual principal components can represent an ambiguous mix of task conditions, stimulus conditions, and trial-trial variability.[91] Moreover, with PCA the number of total dimensions is typically selected based on their cumulative variance explained, rather than by selecting the dimensions that are of interest for decoding, as in dDR.

2.3.3 Extensions

Latent variable estimation:

dDR makes the assumption that latent sources of low-dimensional neural variability can be captured using simple, linear methods, such as PCA. While these methods often seem to recover meaningful dimensions of neural variability,[52] a growing body of work is investigating new, alternative methods for estimating these latent dynamics,[50, 51, 92, 93] and this work will continue to lead to important insights about the nature of shared variability in neural populations.

We suggest that dDR can be extended to incorporate these new methods. For example, rather than defining dDR on a strictly per decoding pair basis, a global noise axis could be identified across all experimental conditions using a custom latent variable method. This could then be applied to the decoding-based dimensionality reduction such that the resulting dDR space explicitly preserves

activity in this latent space to investigate how it interacts with coding.

Incorporating additional dDR dimensions:

In this work we have described dDR primarily as a transformation from N-dimensions to two dimensions, *signal* and *noise*, with the exception of Figure 2.4. In our code repository, `https://github.com/crheller/dDR`, we include examples that demonstrate how the dDR method can be extended to include additional dimensions. However, as discussed in the main text, it is important to remember that estimates of neural variability beyond the first principal component may become unreliable as variance along these dimensions gets progressively smaller, especially in low trial regimes. In short, while information may be contained in dimensions $> m = 2$, caution should be used to ensure that these dimensions can be estimated reliably.

2.3.4 Code availability

We provide Python code for dDR which can be downloaded and installed by following the instructions at `https://github.com/crheller/dDR`. We also include a short demo notebook that highlights the basic work flow and implementation of the method to simulated data. All code used to generate the figures in this manuscript is available in the repository.

2.4 Experimental Methods

2.4.1 Surgical procedure

All procedures were performed in accordance with 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). The surgical approach was similar to that described previously.[18] Adult male ferrets were acquired from an animal supplier (Marshall Farms). Head-post implantation surgeries were then performed in order to permit head-fixation during neurophysiology recordings. Two stainless steel head-posts were fixed to the animal along the midline using bone cement (Palacos), which bonded to the skull and to stainless steel screws that were inserted into the skull. After a two-week recovery period, animals were habituated to a head-fixed posture and auditory stimulation. At this point, a small (0.5 - 1 mm) craniotomy was opened above primary auditory cortex (A1) for neurophysiological recordings.

2.4.2   Neurophysiology

Recording procedures followed those described previously.[33, 87]   Briefly, upon opening a craniotomy, 1 - 4 tungsten micro-electrodes (FHC, 1-5 MΩ) were inserted to characterize the tuning and response latency of the region of cortex. Sites were identified as A1 by characteristic short latency responses, frequency selectivity, and tonotopic gradients across multiple penetrations.[10] Subsequent penetrations were made with a 64-channel silicon electrode array.[40] Electrode contacts were spaced 20 μm horizontally and 25 μm vertically, collectively spanning 1.05 mm of cortex. Data were amplified (RHD 128-channel headstage, Intan Technologies), digitized at 30 KHz (Open Ephys[94]) and saved to disk for further analysis.

Spikes were sorted offline using Kilosort2 (`https://github.com/MouseLand/Kilosort2`). Spike sorting results were manually curated in phy (`https://github.com/cortex-lab/phy`). For all sorted and curated spike clusters, a contamination percentage was computed by measuring the cluster isolation in feature space. All sorted units with contamination percentage less than or equal to 5 percent were classified as single-unit activity. All other stable units that did not meet this isolation criterion were labeled as multi-unit activity. Both single and multi-units were included in all analyses.

2.4.3   Acoustic stimuli

Digital acoustic signals were transformed to analog (National Instruments), amplified (Crown), and delivered through a free-field speaker (Manger) placed 80 cm from the animal's head and 30° contralateral to the the hemisphere in which neural activity was recorded. Stimulation was controlled using custom MATLAB software (`https://bitbucket.org/lbhb/baphy`), and all experiments took place inside a custom double-walled sound-isolating chamber (Professional Model, Gretch-Ken).

Auditory stimuli consisted of narrowband white noise stimuli with $\approx$ 0.3 octave bandwidth. In total, we presented fifteen distinct, non-overlapping noise bursts spanning a 5 octave range. Each noise was presented alone (-Inf dB) condition, or with a pure tone embedded at its center frequency for a range of different signal to noise ratios (-10dB, -5dB, 0dB). Thus, each experiment consisted of 60 unique stimuli (4 SNR conditions X 15 center frequencies). Overall sound level was set to 60 dB SPL. Stimuli were 300ms in duration with 200ms ISI and each sound was repeated 50 times per experiment in a pseudo-random sequence.

2.5   Appendix

2.5.1   Variance of parameter estimates

In this work, we approximate the spike counts of a neural population as being drawn from a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance Σ. The accuracy of our estimates of these respective parameters depends on how large the sample size is. That is, if we draw just two samples from the distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, our estimates of $\boldsymbol{\mu}$ and Σ will be highly variable across repeated iterations of this sampling. This means that when sample size is small we can't be certain of the measured parameter values. Here, we provide a brief derivation showing how the uncertainty in each of these parameter values depends on sample size, k.

Mean (μ):

We will investigate the mean of just a single neuron, μ, for simplicity. Here, and in the following cases, we assume the data has been centered such that the mean response across all trials for each neuron is zero. Consider repeated samples of a random variable, x_i, drawn from $\mathcal{N}(0, \sigma^2)$. Let us define the variable Y to be the mean of a random sequence of i.i.d. numbers, $x_i...x_n$ with $E[x_i] = \mu$ and $Var(x_i) = \sigma^2$.

$$Y = \frac{1}{k} \sum_{i=1}^{k} x_i$$

Next, we can ask how *variable* our estimates of Y are with increasing sample size.

$$Var(Y) = Var\left(\frac{1}{k} \sum_{i=1}^{k} x_i\right)$$

$$Var(Y) = \frac{1}{k^2} \sum_{i=1}^{k} Var(x_i)$$

$$Var(Y) = \frac{1}{k^2} \sum_{i=1}^{k} \sigma^2$$

$$Var(Y) = \frac{\sigma^2}{k}$$

Thus, estimates of the mean spike count for a single neuron, μ, decay with increasing sample size as:

$$\mathcal{O}\left(\frac{1}{k}\right) \tag{2.7}$$

Single neuron variance (Σ_{diag}):

For the variance of single neurons, *i.e.* the diagonal elements of Σ, we can similarly derive their uncertainty as a function of k by defining Y as:

$$Y = \frac{1}{k-1} \sum_{i=1}^{k} x_i^2$$

$$Var(Y) = Var\left(\frac{1}{k-1} \sum_{i=1}^{k} x_i^2\right)$$

$$Var(Y) = \frac{1}{(k-1)^2} \sum_{i=1}^{k} Var(x_i^2)$$

$$Var(Y) = \frac{1}{(k-1)^2} \sum_{i=1}^{k} 2\sigma^4$$

$$Var(Y) = \frac{2\sigma^4}{k-1}$$

Thus, the uncertainty in single neuron variance depends the neuron's true variance σ^2, and decays as a function of sample size k.

$$\mathcal{O}\left(\frac{1}{k-1}\right) \tag{2.8}$$

Covariance (Σ):

And finally, for uncertainty of the covariance between two correlated neurons x and y, *i.e.* the off-diagonal elements of Σ, we define Y as:

$$Y = \frac{1}{k-1} \sum_{i=1}^{k} x_i y_i$$

As above, can write:

$$Var(Y) = Var\left(\frac{1}{k-1} \sum_{i=1}^{k} x_i y_i\right)$$

$$Var(Y) = \frac{1}{(k-1)^2} \sum_{i=1}^{k} Var(x_i y_i)$$

Then, using the three following identities:

$$Var(xy) = E[x^2 y^2] - E[xy]^2$$

$$E[x^2y^2] = cov(x^2, y^2) + E[x^2]E[y^2]$$

$$E[XY]^2 = \big(cov(x, y) + E[x]E[y]\big)^2$$

We can write the following expression for the $Var(Y)$, taking $E[x] = E[y] = 0$:

$$Var(Y) = \Big(\frac{2(\Sigma_{x,y}^2)^2 + \sigma_x^2\sigma_y^2 - \Sigma_{x,y}}{k - 1}\Big)$$

where $\Sigma_{x,y}$ is the true covariance between neurons x and y, and σ_x^2 and σ_y^2 represent each neuron's respective independent variance. Thus, as for single neuron variance, the uncertainty in covariance decays with sample size, k (Eqn. 2.9). Note, though, that typical covariance values are much smaller than single neuron variance, making this a much more difficult parameter to estimate given a particular sample size.

$$\mathcal{O}\Big(\frac{1}{k-1}\Big) \tag{2.9}$$

3 Arousal can either improve or reduce population coding accuracy of natural sounds in auditory cortex

Charles R. Heller[1,2], Mateo López Espejo[1,2], Leah P. Schwartz[2], Daniela Saderi[2], & Stephen V. David[1]

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

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

Abstract

The ability to discriminate between complex natural sounds is critical for survival. Changes in arousal and other aspects of behavioral state can impact the accuracy of sensory coding, affecting both the reliability of single neuron responses and the degree of correlated noise between neurons. However, it is unclear how these effects interact to influence coding of diverse natural stimuli. We recorded the spiking activity of neural populations in primary auditory cortex (A1) evoked by a large library of natural sounds while monitoring changes in pupil size as an index of arousal. Heightened arousal increased response magnitude and reduced noise correlations between neurons. Despite improving on average, arousal reduced population coding accuracy for a significant portion of sensory discriminations. Using a normative modeling approach, we demonstrated that this diversity can be explained by coherent, low-dimensional suppression of response variability in A1. The degree to which this modulation was aligned with high-dimensional natural sound-evoked activity was variable, resulting in stimulus-dependent changes in coding accuracy.

3.1   Introduction

Humans and other animals are able to discriminate between a multitude of natural sounds. This ability is not static, as the precision of sensory representations by neural activity fluctuates with changes in behavioral state.[95] Arousal, task engagement, and attention have all been reported to modulate the magnitude and selectivity of single neuron auditory responses,[3,17,18,28,96–102] as well as correlated variability across neural populations, often referred to as noise correlations.[26,71,73,74] In general, increased arousal and focused attention are associated with increased response magnitude and decreased noise correlations which are believed to enhance the accuracy of sensory coding.[26,27,71,95] However, the mechanisms that produce these changes, and the consistency of their effects between different behavioral contexts, are not fully understood.

Recent studies have argued that attention-driven changes in both single neuron responses and correlated activity can be modeled as fluctuations in a single, latent signal that coherently modulates the activity of a subset of neurons. These findings suggest that state-dependent neural population activity occurs in a low-dimensional subspace,[50,81] supporting theoretical models in which a single mechanism accounts for changes in single neuron responses and correlated variability.[53,103] Fluctua-

tions in arousal, measured by luminance-independent changes in pupil size, modulate neural activity in similar ways to attention,[27, 28, 71] yet these changes can occur independent of attention.[54] Previous work has not specifically investigated the dimensionality of arousal-dependent signaling and it remains uncertain whether, like other behavioral contexts, it can be explained by a low-dimensional process.

Most studies of population coding accuracy rely on relatively small, simple stimulus sets that drive neural activity in stereotyped ways.[26, 72, 104] Yet, theoretical work predicts that noise correlations can either enhance or impair coding accuracy, depending on their alignment with the stimulus-evoked activity in the neurons being studied.[43, 59, 62, 63, 66, 103, 105] If the effects of arousal are relatively high-dimensional, meaning that they suppress noise along many different dimensions of neural activity, they should improve coding accuracy of most sensory stimuli equally. Alternatively, if the effects of arousal are confined to a low-dimensional subspace of neural activity, their alignment with sensory-evoked responses should be variable, resulting in stimulus-dependent changes in coding accuracy.
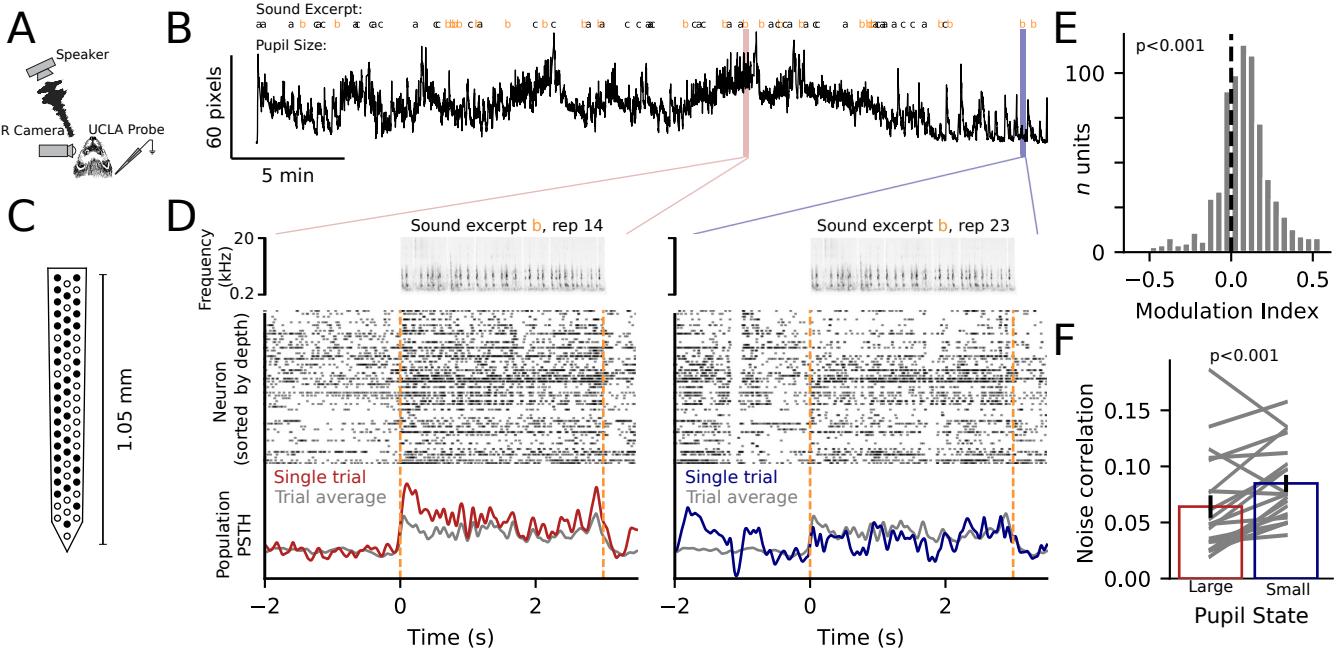
In the present study, we investigated the dimensionality of arousal-dependent signaling and its impact on coding accuracy by recording population activity from primary auditory cortex while presenting a large library of natural sounds. We simultaneously monitored arousal level using pupil size.[4, 28] Arousal could either improve, or reduce, neural discriminability of natural sounds measured using an optimal linear decoder.[106] This diversity is consistent with the hypothesis that arousal acts on a low-dimensional subspace rather than providing a generalized improvement in coding accuracy. Using a normative latent variable model, we demonstrate that low-dimensional, pupil-dependent modulation of correlated activity is required to accurately reproduce these effects. In contrast with attention, modulation of single neuron gain and noise correlations by arousal were distinct. These processes operated on different neural populations and timescales. Thus, our results demonstrate that arousal drives robust, but selective changes in population coding accuracy across diverse sound stimuli and that these changes act through at least two distinct mechanisms.

3.2   Results

We recorded simultaneous single- and multi-unit activity from primary auditory cortex (A1) of awake, head-fixed ferrets using 64- or 128-channel linear silicon probes[40] ($n = 729$ sorted-units, 20 recording sites, eight animals, Figure 3.1A, C). During each recording session, we presented a diverse set of randomly interleaved natural sound excerpts[107] (*e.g.* Figure 3.2B) in the acoustic field contralateral to the recording hemisphere (Figure 3.1A). Pupil size was measured continuously using infrared video during neural recordings to track spontaneous fluctuations in arousal[4, 28] (Figure 3.1A,

B).

In ferret A1, changes in pupil size are associated with mostly monotonic changes in neural firing rate.[28] Therefore, to assess overall pupil-related changes in activity, we split the neural data in half based on the median pupil size during each experiment (large pupil/high arousal vs. small pupil/low arousal). When pupil was large, responses to the same sound were stronger and more reliable than when pupil was small. We measured the pupil-dependent modulation index of sound evoked activity and confirmed that average responses were significantly enhanced during high arousal ($MI = 0.083 \pm 0.007, p < 0.001$ Bootstrap test; 3.1E). Additionally, during large pupil trials population activity was desynchronized relative to the small pupil state. Stimulus-independent fluctuations in the population PSTH were often observed during small pupil trials, while they were absent during large pupil trials (Figure 3.1D). Consistent with this observation, pairwise noise correlations were significantly reduced in the high arousal state ($\rho_{small} = 0.084 \pm 0.008$ vs. $\rho_{large} = 0.064 \pm 0.010$, $p = 0.005$, Bootstrap test, $n = 20$ recording sessions, 10423 ± 533 stimulus x unit pairs per session; Figure 3.1F). This pattern of desynchronization is consistent with previous reports in ferret,[28] mouse,[4,27,71] and primate.[108]

Figure 3.1: **Pupil-indexed arousal modulates neural responses to natural sound stimuli.**
A. Single- and multi-unit activity was recorded from A1 of awake, head-fixed ferrets using laminar
electrode arrays during presentation of natural sound stimuli. Pupil size, an index of arousal, was
measured simultaneously using infrared video. **B.** Pupil trace from one recording session. Pupil
size varied substantially over the course of the session, indicating spontaneous transitions between
high and low arousal states. Top: Three distinct natural sound excerpts were randomly interleaved
during the session (a, b, and c). **C.** Schematic of 64-channel laminar array (UCLA probe) used
to record neural activity. Filled circles represent electrode channels on which at least one unit was
detected during the same session ($n = 55$ total units for this example). **D.** Example data is shown
for two example trials, during which the same natural sound excerpt (excerpt b) was presented.
Example trials are highlighted in panel B. Top: Spectrogram of 3 s natural sound excerpt presented
during each trial. Middle: Population raster plot of spiking activity by all simultaneously recorded
units during a single stimulus presentation when pupil was large (left, red inset) and when pupil
was small (right, blue inset). Bottom: Population peri-stimulus time histogram (PSTH) response,
averaged across units during the single trial (red / blue) and averaged over all repetitions of sound
excerpt b (gray). **E.** Pupil modulation index for all recorded units ($n = 729$ from $n = 20$ recording
sessions). Sound evoked responses were generally larger during high arousal states, when pupil was
large ($MI = 0.083 \pm 0.007, p < 0.001$ Bootstrap test) **F.** Means within recording session are shown by
individual grey lines. Error bars represent standard error of mean across sites and bars represent the
mean. Pairwise noise correlations were reduced during high arousal states ($\rho_{small} = 0.084 \pm 0.007$ vs.
$\rho_{large} = 0.064 \pm 0.010$, $p < 0.001$, Bootstrap test, $n = 20$ recording sessions, 10423 ± 533 stimulus X
neuron pairs per session).

3.2.1 Arousal can either improve, or reduce, neural decoding accuracy of natural stimuli

Previous work suggests that the larger sound evoked responses and reduced variability associated

with high arousal states enhances the accuracy of neural coding.[27,71] To measure the effects of

arousal on coding accuracy in the current dataset, we measured neural discriminability of natural

sounds using d'^2, a well-established metric of neural population discriminability.[57, 59, 62, 68, 69, 109] d'^2 describes the ability to discriminate between two stimuli using an optimal linear decoder trained on population responses in state-space. When sound-evoked activity between stimulus classes is distinct, and response variability within class is low, d'^2 is large. To prevent overfitting of the decoding axis to noise in the high-dimensional population data, we projected single trial activity into a reduced dimensionality space that was designed for robust estimation of d'^2.[106] The dimensionality of this space was selected by determining where cross-validated estimates of d'^2 plateaued, indicating overfitting (Figure S2). For a detailed description of the decoding procedure, see methods and Figure S1.

For this analysis, we defined a "stimulus" by binning responses to natural sound excerpts into non-overlapping 250 ms bins (*e.g.* Figure 3.2A). Pairwise discriminability was then measured for all stimulus pairs within a recording session (mean number of stimulus pairs per recording: 1167 ± 256). Examples taken from our data suggest that the interaction between arousal and stimulus discriminability is heterogeneous (Figure 3.2B). For example, arousal-dependent suppression of response variability can be aligned with (Figure 3.2B - Stim. a_2 vs. d_1) or orthogonal to (Figure 3.2B - Stim. a_1 vs. a_2) the sensory discrimination axis. Thus, increased arousal either helps (Figure 3.2B - Stim. a_2 vs. d_1), or hurts (Figure 3.2B - Stim. a_1 vs. a_2), stimulus discriminability. Given these observations, we hypothesized that arousal may exert more diverse effects on stimulus discrimination than previously reported.
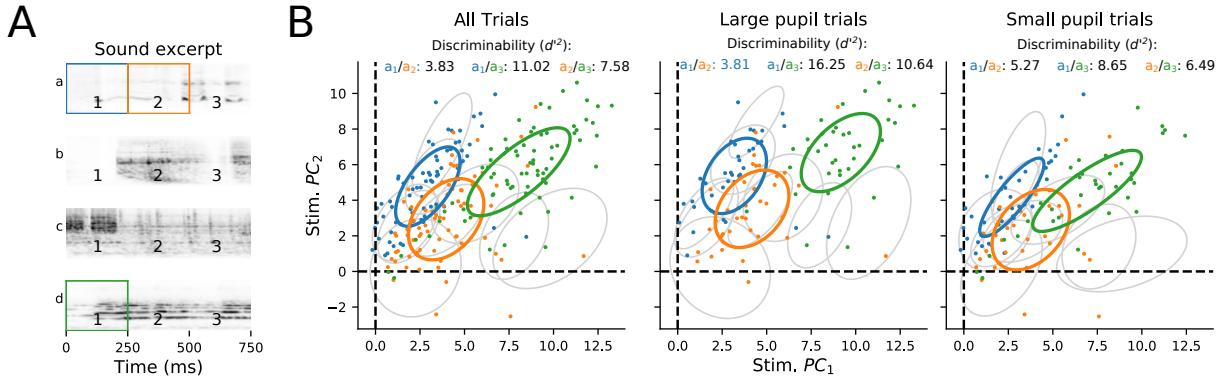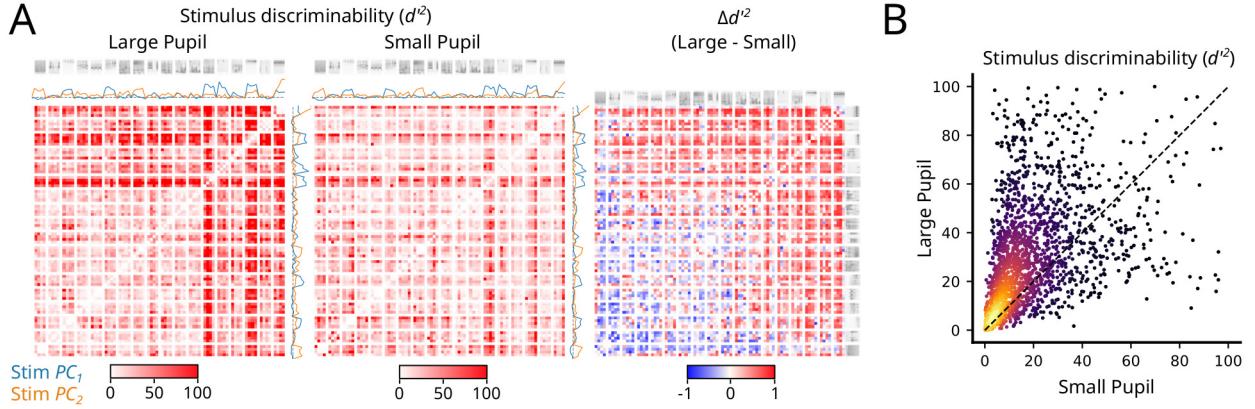
Figure 3.2: **Natural sounds evoke diverse sensory responses. A.** Spectrograms of four natural sound excerpts repeated during a single experiment. "Stimuli" were defined by binning excerpts into 12 non-overlapping 250 ms bins (*e.g.* three example stimuli are highlighted). Thus, d'^2 was measured for 66 stimulus pairs in this experiment. **B**. Population responses for each stimulus were projected into the space defined by the first two principal components of trial-averaged activity across all 12 stimuli (Stim. PC_1 and Stim. PC_2). Pairwise measurements of d'^2 are indicated for the three highlighted stimulus pairs. Left: Single trial data collected over the entire experiment. Dots represent the projection of a single trial. Ellipses are centered on the mean response across trials and their size represents the standard deviation across trials. Grey stimuli represent the nine other stimuli not specifically highlighted in A. Middle/Right: Same, divided by stimulus trials where pupil was large/small.

Across all stimulus pairs within a single experiment, arousal-dependent changes in pairwise stimulus discriminability varied substantially. While many stimulus pairs showed increased d'^2, others did not change and many showed a decrease (Figure 3.3A). Across all recording sessions, stimulus discriminability was enhanced during high arousal states on average ($d'^2_{large} = 32.33 \pm 0.26$ vs. $d'^2_{small} = 18.99 \pm 0.16$, $p < 0.001$, Bootstrap test, $n = 20$ recording sessions, 1167 ± 256 stimulus pairs per session). However, a substantial fraction of stimulus pairs were more discriminable during low arousal states ($29.2 \pm 0.8\%$). The sign of d'^2 changes was not dependent on baseline stimulus discriminability (Figure S3). Furthermore, the same amount of diversity was observed when excluding stimuli with unreliable and/or weak responses (Figure S4) arguing that this diversity is a real feature of the data and is not due to "noisy" estimates of trial-trial variability and/or stimulus discriminability.

Figure 3.3: **Arousal can increase or decrease neural discriminability of natural sounds. A.** Left: Heatmap intensity indicates d'^2 measured during large pupil trials for each pair of 0.25 s sound segments in one example experiment. The spectrogram and the first two principal-component (PC) projections of average evoked activity for each stimulus are shown along the heatmap axes. Middle: Same, for d'^2 computed across small pupil trials. Right: Normalized change in $d'2$, $\Delta d'^2$, between large and small pupil. Positive values indicate better stimulus discriminability during high arousal. All three heatmaps are sorted according to the $\Delta d'^2$ matrix. **B**: Scatter plot of d'^2 for large/small pupil conditions across all experimental sessions. Color indicates the density of points at each location. Stimulus pairs ($n = 23,352$ total) were randomly subsampled for visualization ($n = 2500$ displayed here).

3.2.2 A low-dimensional latent variable model accounts for diversity of arousal-dependent coding changes

Based on examples (*e.g.* Figure 3.2B), we hypothesized that pupil-dependent changes in response variance could cause the heterogeneity observed in $\Delta d'^2$ measurements. To test this, we designed a normative latent variable model (Figure 3.4A, Eqn. 3.6) to determine whether pupil-dependent changes in covariance alone could account for the effects of arousal. The model was fit in two stages. In the first stage, the mean sensory evoked response of each neuron (PSTH) was scaled and offset according to pupil size in order to minimize the mean squared error between the predicted and actual time varying firing rate response (Eqn. 3.5).[28,33] In the second stage, the pupil weights for each neuron were frozen. Pupil-dependent single neuron (independent) variance and pupil-dependent shared variance was then added to the predicted firing rate of each neuron. Weights for independent and shared variance were optimized for each neuron such that the pupil-dependent covariance matrix of the resulting response matched the actual data (*e.g.* Figure 3.4B, Eqn. 3.6). Because the model required a large amount of data to fit, we restricted our analysis to a subset of recordings ($n = 7$ recordings from three animals) in which each stimulus was repeated at least 80 times during the

experiment.

A

B Difference Covariance
(Small - Large)
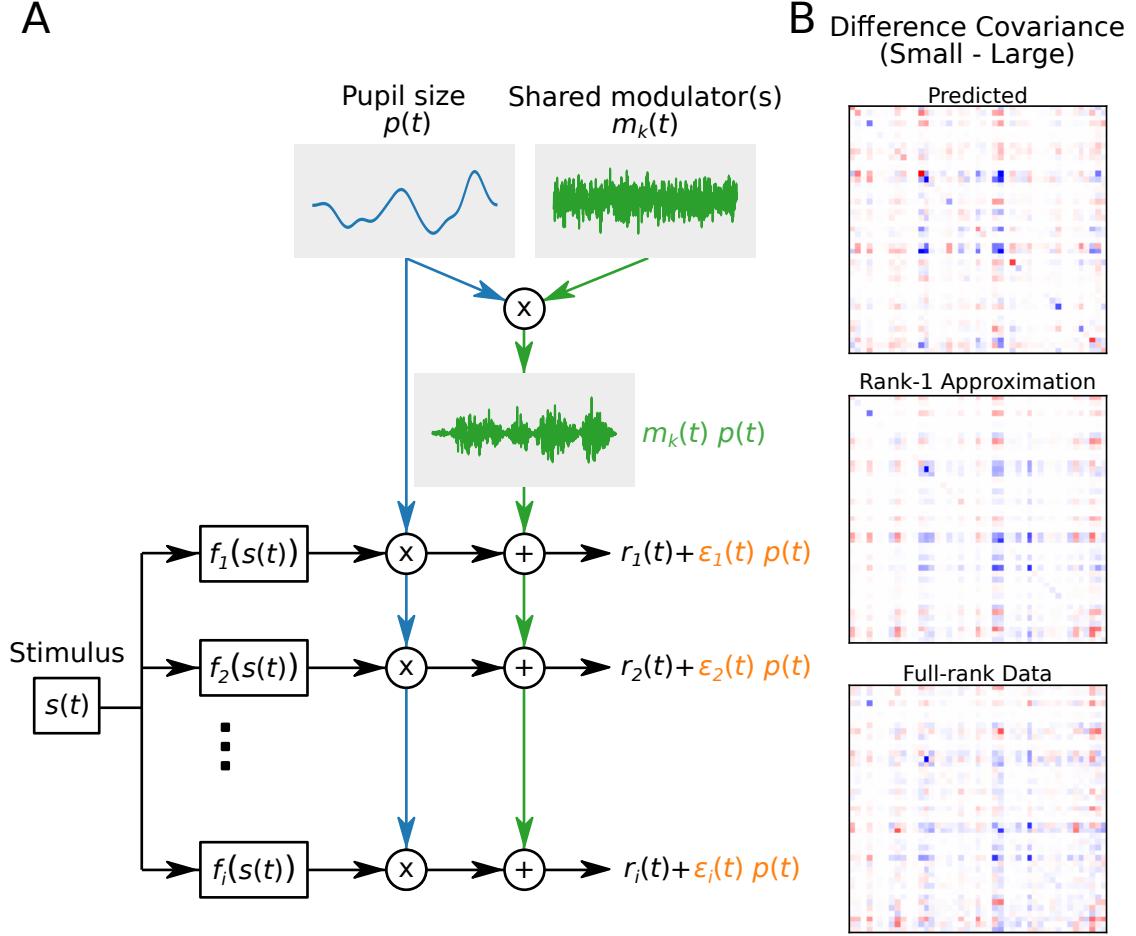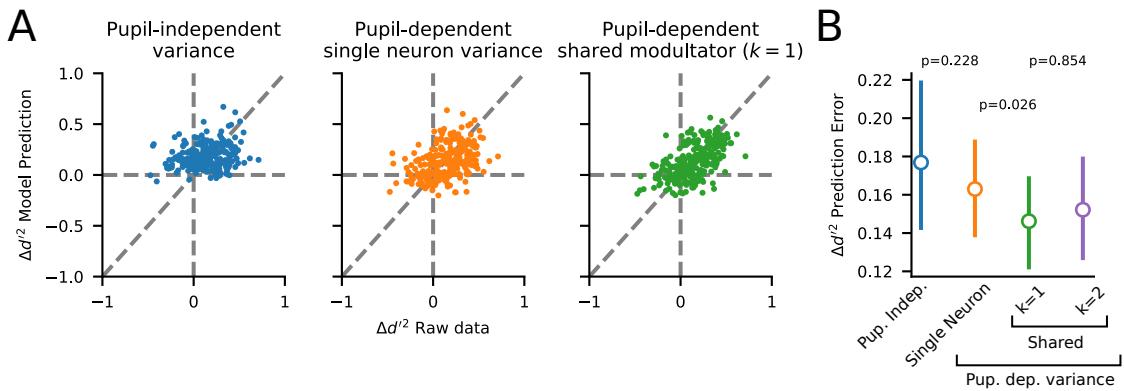
Figure 3.4: **Normative latent variable model architecture. A.** Schematic of model architecture. Sensory responses, $f(s(t))$ were estimated for each neuron by computing the peri-stimulus time histogram (PSTH) across all repetitions of each stimulus. The PSTH for each neuron, i, was provided as input to the model. Model fitting was split into two stages. During the first stage, pupil-dependent sensory responses of single neurons was predicted by minimizing the mean-square error between predicted firing rate and true firing rate. In this first stage, only the pupil weights (blue) were fit (Eqn. 3.5). In the second model fitting stage, single neuron pupil weights were frozen and pupil-dependent single neuron variance ($\epsilon(t)p(t)$, orange) and pupil-dependent shared variance ($m_k(t)p(t)$, green) were added to responses. During the second stage of fitting pupil-dependent variance weights were optimized to minimize the mean squared error between predicted and actual large/small pupil covariance matrices (Eqn. 3.6). Models were fit to the rank-1 approximation of the covariance matrix to prevent overfitting. **B.** The neuron-by-neuron difference covariance matrix (small pupil covariance minus large pupil covariance) for a single recording site is plotted for the model output, the rank-1 approximation to the raw data, and the full-rank raw data.

To test whether the model generated responses recapitulated the diversity observed in $\Delta d'$, we computed $\Delta d'^2$ for the model responses and compared the result to the actual $\Delta d'^2$ for each stimulus pair. We found that a model using just a single pupil-dependent shared modulator and pupil-dependent sin-

gle neuron variance closely replicated the actual arousal-dependent changes in discriminability (Figure 3.5A, right). To determine if this result required pupil-dependent shared and/or single neuron variance, we tested two additional models (Eqns. 3.7, 3.8). In the first model, pupil was shuffled for both shared and single neuron variance terms, making both terms independent of pupil (Figure 3.5A, left: pupil-independent variance). In the second model, pupil was shuffled only for the shared modulator(s) but preserved for single neuron variance (Figure 3.5A, middle: pupil-dependent single neuron variance). Neither model was able to perform as well as the full model containing pupil-dependent shared modulation (Figure 3.5B). Indeed, without pupil-dependent noise, $\Delta d'^2$ was virtually always positive for the model generated response (Figure 3.5A, left), in contrast to our finding that arousal can also reduce stimulus discriminability. This is in line with previous work arguing that pupil-dependent changes in response gain alone consistently improve sensory coding.[27, 28]

Finally, we found that adding additional pupil-dependent shared modulators did not improve $\Delta d'^2$ predictions (Figure 3.5B), suggesting that pupil-dependent modulation of variance occurs in a low-dimensional subspace. This result was consistent even when fitting our model to higher rank estimations of the pupil-dependent covariance matrices, arguing that it is not an artifact of the model fitting procedure (Figure S5).
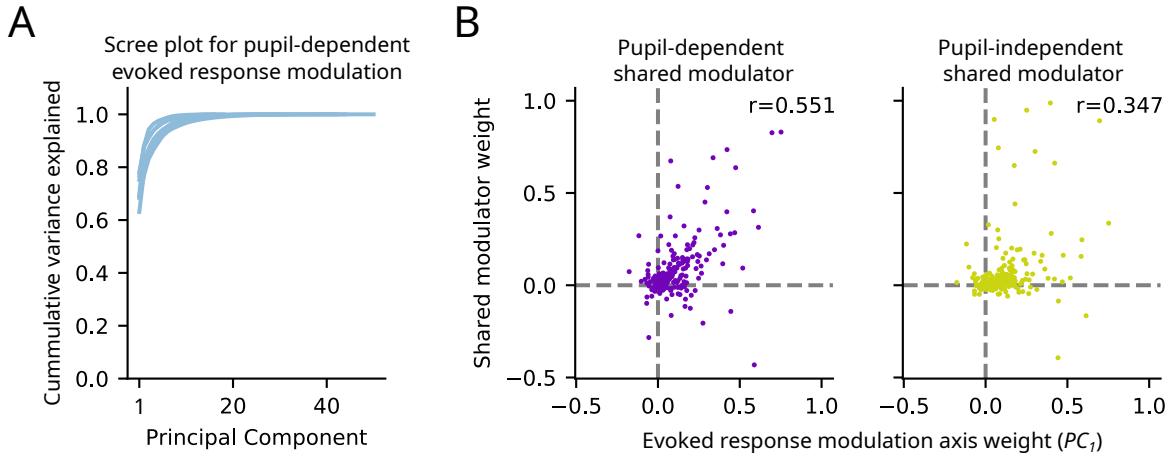
Figure 3.5: **Pupil-dependent response variability accounts for diversity in state-dependent coding. A.** $\Delta d'^2$ measured using responses output by normative latent variable models (y-axis) are compared with actual $\Delta d'^2$ for individual stimulus pairs across $n = 7$ recording sites. Three different model results are shown: A model with pupil-independent variance (blue), pupil-dependent single neuron variance only (orange), and a model with both pupil-dependent single neuron variance and shared variance (green). **B.** Model performance is quantified by computing the absolute error in $\Delta d'^2$ prediction for each stimulus pair. Lines represent bootstrapped 95% confidence intervals. p-values were computed using paired bootstrap samples between the error of two respective models (see Methods: Statistical Methods).

3.2.3 Arousal-dependent changes in gain and correlated variability operate on highly overlapping populations

Studies of selective attention have shown that attention-related changes in response gain of single neurons and the strength of correlated variability are related. In this framework, changes in correlated variability simply reflect changes in the variance of single neuron's sensory gain across trials.[50,81] Thus, a single, shared mechanism appears to mediate changes in both gain and correlation. We wondered if a single mechanism also produced effects of arousal, or if changes in gain and correlated variability were mediated by separate mechanisms.

To investigate this question we first characterized pupil-dependent modulation of sensory evoked gain. To do this, we subtracted the mean sensory response across all trials (PSTH) from the first stage model prediction (Eqn. 3.5) in our latent variable model. At this stage of the model fit, the only trial-to-trial variability in responses was entirely attributable to pupil-dependent gain (Figure 3.4A). We then performed PCA on the residual responses which, by definition, identified the dimensions where pupil-dependent changes in gain were largest and accounted for the most response variability. Across recording sites, this space was low-dimensional; the first PC accounted for at least 65% of total gain variability in all cases (Figure 3.6A).

Next, we compared this first principal component to the pupil-dependent shared modulator weights that were fit during the second stage of model fitting (Figure 3.4A, Eqn. 3.6). For the pupil-dependent shared modulator, weights were highly correlated with the first PC indicating that changes in stimulus-independent response variance overlapped with first order changes in sensory evoked response gain (Figure 3.6B). We confirmed that this correlation was not a trivial consequence of neural activity being low dimensional by also comparing the model weights for a pupil-independent shared modulator. Although these were also correlated with gain, the relationship was much weaker than for pupil-dependent noise (Figure 3.6B).

Figure 3.6: **Arousal-dependent changes in evoked response magnitude and correlated variability operate on overlapping neural populations. A.** Cummulative fraction variance explained by each prinicipal component of pupil-dependent gain modulation. Each line represent a single recording session. **B.** Left: Loading weights for the first PC in A are plotted against the pupil-dependent shared modulator weights from the normative latent variable model. One data point per neuron. r indicates the correlation coefficient for all neuron weights across all recording sites. Right: Same, but for pupil-independent model weights. Thus, this correlation offers a noise floor for the relationship between first and second order model weights.

3.3 Discussion

Previous studies have suggested that behavior-dependent modulation of neural population coding operates in a low-dimensional space.[29, 50, 51, 53, 54, 81] That is, signals reflecting behavioral state are well-described by processes that modulate the activity of many neurons coherently and produce correlated variability in sensory responses. However, most previous work has utilized relatively small, focused stimulus sets. This raises questions about whether the observed low-dimensional processes are a consequence of the stimuli tested, or if they are a general feature of state-dependent modulation. These questions are critical for understanding population coding of sensory stimuli. Theoretical studies have long shown that correlated variability can impact coding accuracy, but only if it aligns with the sensory tuning of neurons in the population.[59] Thus, the dimensionality of the mechanisms driving correlated variability and how they interact with sensory selectivity is critical for understanding their impact on sensory processing.

In the case of pupil-indexed arousal, we found that correlated activity is modulated in a low-dimensional subspace of primary auditory cortex (A1) that was highly overlapping with the arousal-dependent changes in single neuron responses. These modulations were present across a diverse set of natural sound stimuli and their effect on neural discrimination of sounds varied substantially with

the sound stimulus; some stimulus discriminations were improved by arousal and others were reduced. This heterogeneous relationship between coding and arousal is predicted for a low-dimensional state signal interacting with relatively high-dimensional stimulus-evoked activity.

3.3.1 Effects of shared intrinsic variability on discriminability are stimulus-dependent

Correlated, intrinsic variability within neural populations is ubiquitous in cortex. Even before this phenomenon was observed experimentally, substantial efforts were made to develop a theoretical understanding of how correlated activity might affect coding by neural populations.[46, 55, 57, 60, 62, 110–112] This early work established that correlated variability can interfere with the brain's ability to accurately discriminate sensory stimuli. Therefore, experimental characterization of this phenomenon is critical to fully understand neural population codes.

Although evidence for intrinsic correlation is widespread, experimental studies have provided conflicting evidence as to whether or not it does in fact interfere with population coding.[43, 63, 66, 68, 69, 103, 105, 109] There are at least two reasons why the reported effects of correlated variability might vary across studies. First, the dimensions containing interfering noise appear to often be very low variance.[68, 69, 109] Thus, detecting this noise could require recording large amounts of data, both over many neurons and over many trials, a methodology that has only recently become feasible.[68, 69, 109] A second possibility is related to the fact that the impact of correlated noise depends on the tuning of neurons in the population being read out and their relationship with the noise space.[29, 62, 105] Therefore, discrepancies in previous work might be explained by differences in the neural populations that were sampled and/or in the stimuli that were tested. Because the effects of correlated noise may depend on the stimuli that are presented, it is important to characterize coding accuracy across a wide array of sensory space. Indeed, our results showed that the effects of correlated variability on coding are highly variable and only benefit specific sensory discriminations.

Because there is a trade-off between the number of stimuli that can be presented and the number of times that each can be repeated during a single recording session, questions about stimulus-dependent changes in population coding have been difficult to completely address in a single study. Unlike recent work,[68, 69, 109] we measured neural responses to a large set of stimuli over a relatively small number of repeats and neurons. Thus, we could not measure low-variance dimensions and draw strict conclusions about the presence (or absence) of information-limiting noise.[62] Instead, we leveraged a new dimensionality reduction approach (dDR[106]) with which we were able to reliably estimate the interaction between the dominant, high-variance noise dimensions and sensory discrimination across a large acoustic stimulus space. This approach revealed substantial variability of arousal dependent

changes in coding within each recording site, highlighting the practical benefit of dimensionality reduction techniques for studying neural population dynamics across a diversity of stimulus and behavioral contexts.[85]

3.3.2 State-dependent coding in auditory cortex

It is increasingly clear that neural activity in primary sensory regions of the brain is modulated by non-sensory variables, including arousal.[27–30, 71] Arousal here refers to spontaneous changes in alertness as measured by pupil diameter, even in the absence of a behavioral task.[30] Similar to previous work, we find that increased arousal is associated with enhanced excitability and reduced noise correlations in A1.[4, 28, 71] These effects boost the neural signal to noise ratio in V1[27] and improve population coding accuracy of tonal stimuli in A1.[71]

Building on this previous work, we explored the effects of arousal on population coding accuracy across a large space of natural sounds. Consistent with prior results, increases in pupil-indexed arousal led to improved discriminability between sounds on average. However, the relative magnitude of this improvement varied substantially across stimuli and many sensory discriminations were even harmed by increased arousal. Using a normative modeling approach, we demonstrated that this variability could be explained by low-dimensional modulation of shared variance.

The strong stimulus dependence highlights the importance of a broad exploration of state-dependent changes in neural coding across the sensory response space. Parametric stimuli might be used in the future to more systematically probe sound feature representations across a range of behaviorally relevant stimuli. For example, one study of auditory processing began to address this question in anesthetized animals.[72] In this work, Kobak et al. measured population coding in A1 of sounds that varied along two dimensions: Inter-aural level (ILD) and absolute binaural level (ABL). By inducing different states of cortical activation with urethane anaesthesia, the authors demonstrated that in the awake (desynchronized) state, noise and signal subspaces shift to become orthogonal, thereby facilitating accurate encoding across both ILD and ABL. Extending this approach to spectro-temporally varying and ethologically relevant naturalistic stimuli,[113] especially in the context of behavior, will be critical for a complete understanding of state-dependent population coding.

3.3.3 Mechanisms driving arousal-dependent changes in sensory-evoked activity and correlated neural variability

Recent studies of selective attention have suggested that correlated variability results from the coherent modulation of many neurons by an intrinsic behavioral state variable.[50, 81, 103] In a study

by Denfield et al., macaques were trained on a visual change-detection task in which the stability of spatial selective attention was manipulated between behavioral blocks.[81] Because the gain of evoked responses in visual cortex is known to be coherently modulated by attention,[80] the authors proposed that the magnitude of correlated variability should be highest when attention itself is most variable, as changes in gain are shared across neurons within the receptive field. Indeed, when animals were required to switch attention between multiple locations within a behavioral block, noise correlations were strongest. This idea of correlations produced by a shifting spotlight of attention is consistent with previous characterizations of neural population activity and attention[50] and agrees well with theoretical work.[53,103] These findings offer a parsimonious explanation for why gain changes are accompanied by a reduction in noise correlations during traditional cued change-detection tasks, where attention is focused stably on a single spatial location.[26]

Similar to the case of visual selective attention, we found that arousal-dependent modulation of evoked rates and shared variability occurred in largely overlapping neural populations. However, a model with only pupil-dependent gain modulation did not accurately reproduce the discriminability changes observed in our data, while incorporating an additional source of pupil-dependent shared modulation did. Therefore, whether or not these processes reflect independent mechanisms of modulation is difficult to tease apart in our data.

Pupil-indexed changes in arousal are slow, happening on the order of seconds (*e.g.* Figure 3.1A) while, by design, the shared modulator in our model acted on a faster timescale ($\approx$ 4 Hz). One possibility is that pupil offers only an approximate read out of arousal-associated modulation of evoked responses in auditory cortex. In this case, a model containing only pupil-dependent gain would fail to replicate the precise variability in sensory responses across time. Therefore, adding a second shared modulator to our model may simply be providing this added resolution, allowing the variance of the pupil-dependent gain signal to vary with arousal, as well as its magnitude. This explanation is consistent with the mechanism of visual selective attention. However, because we did not experimentally manipulate the variability of arousal, as Denfield et al.[81] did for attention, we are not able to conclusively test this hypothesis.

An alternative hypothesis is that the two effects are caused by distinct cortical mechanisms. Although it was not feasible to directly isolate the circuitry underlying these distinct effects in the current study, we propose that they may arise through a combination of neuromodulation and intra-cortical feedback. Several studies have shown a strong correlation between slow fluctuations in pupil diameter and brain-wide release of norepinephrine and acetlycholine,[32] making these modulators good candidates for mediating the slow changes in response baseline and gain across individual neurons.

The decrease in correlated activity, on the other hand, may arise due to modulation of feedback from other cortical areas that are themselves targeted by the same neuromodulatory signals.

Intracortical pathways to auditory cortex have been identified from multiple areas, including visual,[5] motor[6] and prefrontal cortex.[114] These inputs can activate inhibitory networks that desynchronize local network activity, and modulating their strength could produce the de-correlation effects observed in the current study. Given the diversity of these intracortical signals, it might seem surprising that the arousal-related changes reported here should occur in such a low-dimensional space, shared by slow, pupil-indexed gain modulation. However, if the same inhibitory networks are activated by slow changes in arousal and introcortical signalling, it's possible that the gain and correlation effects would appear in overlapping population of neurons, as we observed. Further investigation with selective control of feedback from different cortical areas will determine if, in fact, the impact of signals from these different cortical areas can be dissociated in A1.

3.4 Materials and Methods

3.4.1 Surgical procedure:

All procedures were performed in accordance with 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). The surgical approach was similar to that described previously.[7, 18, 28, 115] Five young adult male ferrets were acquired from an animal supplier (Marshall Farms). Head-post implantation surgeries were then performed in order to permit head-fixation during neurophysiology recordings. Two stainless steel head-posts were fixed to the animal along the midline using UV-cured dental composite (Charisma) or bone cement (Palacos), which bonded to the skull and to stainless steel screws that were inserted into the skull. After a two-week recovery period, animals were habituated to a head-fixed posture and auditory stimulation. At this point, a small (0.5 - 1 mm) craniotomy was opened above primary auditory cortex (A1) for neurophysiological recordings.

3.4.2 Acoustic stimuli:

Digital acoustic signals were transformed to analog (National Instruments), amplified (Crown), and delivered through a free-field speaker (Manger) placed 80 cm from the animal's head and 30°contralateral to the the hemisphere in which neural activity was recorded (Figure 3.1). Stimulation was controlled using custom MATLAB software (`https://bitbucket.org/lbhb/baphy`), and all experiments took

place inside a custom double-walled sound-isolating chamber (Professional Model, Gretch-Ken).

Natural sounds stimuli were presented in four different configurations. Set 1 consisted of 93, 3-sec samples (2.5 sec ISI, $n = 3$ sites), set 2 consisted of 306, 4-sec samples (1 sec ISI, $n = 14$ sites), set 3 consisted of 306, 1-sec samples (0.5 sec ISI, $n = 6$ sites), and set 4 consisted of 2, 3-sec samples of ferret vocalizations (2.5 sec ISI, $n = 2$ sites). In sets 1-3, the stimulus sets contained species conspecific and heterospecific vocalizations, speech, music, and environmental sounds chosen to sample diverse spectro-temporal statistics.[107] All stimuli were presented at 65 dB SPL. During every experimental session, a subset of samples were repeated at least ten times (set 1: 3 samples, set 2: 18 samples, set 3: 18 samples, set 4: all samples), while the remainder were played only once. In order to study the trial-to-trial variability in neural responses, only the high-repeat sounds were included in this study. The order in which stimuli were presented was generated pseudo-randomly. Stimuli were played continuously until all sound samples in the library had been presented. In the case of set 1, the entire stimulus set was repeated 2-3 times. This meant that experiments lasted approximately 40 minutes. The full sound library can be accessed at `https://bitbucket.org/lbhb/baphy`). Some of data used in this study has been published previously.[28, 88]

3.4.3 Neurophysiology:

Upon opening a craniotomy, 1 - 4 tungsten micro-electrodes (FHC, 1-5 MΩ) were inserted to characterize the tuning and response latency of the region of cortex. Sites were identified as A1 by characteristic short latency responses, frequency selectivity, and tonotopic gradients across multiple penetrations.[10, 116] Subsequent penetrations were made with a single (64-channel) or dual shank (128-channel) silicon electrode array.[40] Electrode contacts were spaced 20 μm horizontally and 25 μm vertically, collectively spanning 1.05 mm of cortex. On each consecutive recording day, we changed the location of the electrode penetration to access fresh cortical tissue, expanding the craniotomy as necessary. Data were amplified (RHD 128-channel headstage, Intan Technologies), digitized at 30 KHz (Open Ephys[94]) and saved to disk for further analysis.

Spikes were sorted offline using Kilosort[117] or Kilosort2 (`https://github.com/MouseLand/Kilosort2`). Spike sorting results were manually curated in phy (`https://github.com/cortex-lab/phy`). For all sorted and curated spike clusters, a contamination percentage was computed by measuring the cluster isolation in feature space. All sorted units with contamination percentage less than or equal to 5 percent were classified as single-unit activity. All other stable units that did not meet this isolation criterion were labeled as multi-unit activity.

3.4.4 Pupillometry:

During neurophysiological recordings, video of the ipsilateral pupil (relative to the recording hemisphere) was collected using 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 camera 10 cm from the eye. Contrast was increased using infrared illumination. Ambient light levels were fixed for each experiment at roughly 1500 lux to provide maximum dynamic range of pupil size.[28] Pupil size was measured offline by fitting an ellipse to each video frame using using a custom machine learning algorithm (Python and Tensorflow). The area of the ellipse was extracted and saved for analysis with neurophysiological data. Pupil data was shifted by 750 ms relative to spike times in order to account for the lagged relationship between changes in pupil size and neural activity in auditory cortex and to allow for comparison with previous research.[4]

The pupil tracking algorithm itself utilized a deep learning approach. Our model architecture was based on DenseNet201,[118] which is available through Keras (`https://keras.io/`). In order to transform the output of the model to pupil ellipse predictions, we added a single global pooling layer and a final prediction layer in which five pupil ellipse parameters (x-position, y-position, minor axis, major axis, and rotation), and eight eyelid keys points (x-y positions of the top / bottom eyelid and the corners of the eye) were fit to each video frame. In order to initialize model weights, the model was pre-trained on ImageNet,[119] then fine-tuned using roughly 500 previously analyzed, nonconsecutive frames from video of the pupil of multiple different ferrets (data from[28]). Qualitatively, after this first round of training the model performed well, even on novel video frames of pupil from new animals. However, in cases where the pupil video quality was poor, or differed substantially from the video frames in our training data set, we noticed failures in the model predictions. To further improve the model, we employed an active learning procedure. For each new analyzed video, pupil ellipse fits were analyzed qualitatively by experimenters. If the fit quality was deemed poor, predictions for these frames were manually corrected and added to the training data set. The model was then retrained and the analysis rerun. The network became robust to varying levels of video quality and performed consistently without the need for user intervention. The code for this analysis is available at `https://github.com/LBHB/nems_db`.

3.4.5 Movement exclusion:

During experiments, we observed that blinks were often associated with periods of high arousal and more global facial / body movements. As this motor activity is known to modulate cortical

activity,[29, 38] we used our videos of pupil to exclude blinks from our analysis to isolate pupil-indexed changes in arousal state. To this, we measured the variance of the animal's eyelid over time using a sliding window and excluded periods where this variance trace exceed 0.25 time the standard deviation of eyelid movement across the entire experiment. Visual inspection showed that this procedure was effective at removing most periods of blinking and movement.

3.4.6 Pupil modulation index:

To quantify overall pupil-dependent modulation in single neurons without differentiating between baseline and gain, we measured an overall pupil modulation index (MI). MI was defined by as the mean sound-evoked response when pupil was large minus the mean response when pupil was small, normalized by the sum of these two quantities. Large and small trials were defined based on a median split of pupil size across the entire recording session.

$$MI = \frac{\bar{r}_{large} - \bar{r}_{small}}{\bar{r}_{large} + \bar{r}_{small}} \tag{3.1}$$

3.4.7 Noise correlations:

Pairwise noise correlations were measured by grouping spike counts into 250 ms bins, extracting only evoked periods (epochs when sound stimuli were playing), and computing Pearson's correlation between all pairwise combinations of z-scored spike counts. Z-scores were calculated for each stimulus independently, as in Eqn. 3.2, where $r(t)$ is the single trial response, r_0 is the trial averaged response, and σ is the standard deviation of spike counts across repetitions. Therefore, the z-scored spike counts $Z(t)$ of each neuron i for each stimulus s had mean zero and standard deviation one.

$$Z_i(t) = \frac{r_i(t) - r_{0,i}(s)}{\sigma_i(s)} \tag{3.2}$$

3.4.8 Pairwise stimulus discrimination:

Natural sound excerpts were broken into non-overlapping 250 ms segments, similar to the procedure followed by Pachitariu et al., 2015[120] (illustrated in Figure 3.2A). For each pair of stimulus segments, we extracted the N neuron X k trial response matrices, A and B. Because the number of recorded neurons was greater than the number of stimulus repetitions, we first performed decoding-based dimensionality reduction (dDR[106]), method similar to PCA, but designed to specifically preserve only significant dimensions of the population response relevant for decoding. We selected the

dimensionality of this space using cross-validation (Figure S2). This procedure allowed us to reliably estimate population statistics and coding accuracy without overfitting to single-trial noise in the data. We quantified decoding accuracy in this reduced-dimensionality space by measuring neural stimulus discriminability, d'^2, the discrete analog of Fisher information:[57, 59, 62, 68, 69, 109]

$$d'^2 = \Delta\boldsymbol{\mu}^T \Sigma^{-1} \Delta\boldsymbol{\mu} = \Delta\boldsymbol{\mu}^T \mathbf{w}_{opt} \tag{3.3}$$

where $\Delta\boldsymbol{\mu}$ represents the vector connecting the mean ensemble responses to stimulus A and stimulus B, $\Sigma = \frac{1}{2}(\Sigma_A + \Sigma_B)$ represents the mean noise-covariance matrix, and $\mathbf{w}_{opt}$ is the optimal decoding axis, *i.e.* the vector orthogonal to the optimal linear discrimination hyperplane in state-space. In practice, we estimated $\mathbf{w}_{opt}$ using 50-percent of trials (training data) then projected the held out 50-percent of trials (test data) onto this vector and measured discriminability. For a detailed schematic of this procedure, see Figure S1. Pupil-dependent measurements of d'^2 followed an identical procedure, but before measuring discriminability, the test data was first split in half based on median pupil size.

Pupil-dependent changes in stimulus discriminability:

d'^2, measured across pupil states, could vary greatly across the sensory responses. Therefore, in order to measure pupil-dependent changes in coding accuracy across recording sites, we used a normalized metric, $\Delta d'^2$. For each stimulus pair, $\Delta d'^2$ was defined as d'^2 measured during large pupil trials minus d'^2 for small pupil trials, normalized by the sum of these two quantities.

$$\Delta d'^2 = \frac{d'^2_{large} - d'^2_{small}}{d'^2_{large} + d'^2_{small}} \tag{3.4}$$

3.4.9 Normative latent variable model:

In order to determine if pupil-dependent modulation of response variance predicted the effect of arousal on stimulus discriminability, we designed a normative latent variable model (Eqn. 3.6. The model was fit in two stages. Only the second stage of model fitting used the normative framework.

Stage 1: Pupil-dependent modulation of evoked response magnitude:

During the first stage of model fitting, we fit a state-dependent generalized linear encoding model similar to those described in previous work.[28, 33] For each recorded unit, i, the input to this model was defined as the peri-stimulus time histogram (PSTH) response averaged over all stimulus repetitions $(r_{0,i}(t))$. The predicted firing rate was calculated by scaling the PSTH by a pupil-dependent multi-

plicative and additive factor to model pupil-dependent changes in gain and baseline firing rate over time (Eqn. 3.5). To account for a possible nonlinear relationship between pupil size and neuromodulation, the pupil signal was first passed through a static sigmoid nonlinearity, F (double exponential[16]). The parameters of this nonlinearity, as well as the gain (β_0) and baseline (β_1) coefficients for each neuron were fit simultaneously using 10-fold jackknifed cross validation in order to minimize the mean-squared-error between the predicted and actual time varying firing rate response of each neuron.

$$\hat{r}_i(t) = \beta_0\Big(1 + F\big(p(t)\big)r_{0,i}(t)\Big) + \beta_1 F\big(p(t)\big) \tag{3.5}$$

Stage 2: Normative approach to fitting pupil-dependent response variability

In the second stage of model fitting, we used a normative modeling framework. Thus, rather than predict time-varying response rate of each neuron, we optimized the model to predict the state-dependent population covariance matrix. Thus, the model did not seek to predict actual neural activity, but to generate a population responses whose state-dependent covariance matched that of the real data.

The input to this model was defined by the predicted response of each neuron from stage 1, $\hat{r}_i(t)$. Its output, for each neuron, $z_i(t)$, was generated by adding pupil ($p(t)$) dependent single neuron ($\epsilon_i(t)$) and/or shared ($m_k(t)$) variance to the first stage prediction. A pupil-independent, neuron-independent baseline term, d was included to account for state-independent correlations (Eqn. 3.6). Noise signals (ϵ and m_k) were generated with Gaussian white noise ($\mu = 0$, $\sigma = 1$).

$$z_i(t) = \hat{r}_i(t) + \sum_{k=1}^{K} w_{i,k}p(t)m_k(t) + \big(d + u_i p(t)\big)\epsilon_i(t) \tag{3.6}$$

On each iteration of the fit, the stimulus-dependent large / small pupil covariance matrices were computed for the model generated responses and compared with the actual covariance matrices. Model weights were optimized to minimize the mean-square-error between the model and actual data. Thus, if five neurons were recorded for three different sensory stimuli, the model was tasked with predicting 150 covariance coefficients (five neurons x five neurons x three stimuli x two pupil conditions). To reduce model fit times, we fit to a randomly selected of 10 stimuli for each recording site.

Additionally, we found that models trained to predict the full-rank covariance matrices seems to suffer from overfitting in some instances. Therefore, we fit models on reduced rank approximations to the true covariance matrix using singular value decomposition. The simplest model that maximized performance was fit using the rank-1 covariance matrix approximation (Figure S5).

Finally, to test the requirement of pupil-dependent noise, we tested two additional models. One in which pupil was shuffled ($p0(t)$) for the shared modulator term (Eqn. 3.7) and one where pupil was shuffled for both the shared modulator and the independent noise term (Eqn. 3.8). Shuffling pupil breaks the correlation between the variance modulation and arousal state while still preserving the overall number of model parameters and theoretical rank of model generated covariance, allowing us to determine if pupil was necessary to explain the effects in our data.

$$z_i(t) = \hat{r}_i(t) + \sum_{k=1}^{K} w_{i,k} p0(t) m_k(t) + \big(d + u_i p(t)\big) \epsilon_i(t) \tag{3.7}$$

$$z_i(t) = \hat{r}_i(t) + \sum_{k=1}^{K} w_{i,k} p0(t) m_k(t) + \big(d + u_i p0(t)\big) \epsilon_i(t) \tag{3.8}$$

All models / stages of model fitting were performed using the Neural Encoding Model System (NEMS, `https://github.com/LBHB/NEMS`).

3.4.10 Comparing evoked-response modulation to shared variability model weights

To determine if the effects of arousal on sensory response gain overlapped with modulation of correlated variability, we compared the evoked-response modulation measured in the first stage of the latent variable model fit to the shared modulator weight(s) (W_k) in the second stage of model fitting.

While the sigmoid transformation applied to the pupil during the first model fitting stage improved model performance, it made it difficult to interpret the gain (β_0) and baseline (β_1) parameters directly. Therefore, rather than extract them directly we instead measured the pupil-dependent modulation using PCA. To do this, we computed the difference between the model prediction $\hat{r}_i(t)$ and $r_{0,i}(t)$, the state-independent PSTH. This difference summarized the pupil-dependent modulation of sensory evoked responses. Therefore, we then performed PCA on the difference and defined the first principal component as the axis where most modulation occurred. We then compared the loading weights of this PC with the W_k weights extracted from the latent variable model (Eqn. 3.6).

3.4.11 Statistical methods

Our data followed a nested structure; multiple cells were recorded from the same animal and many different stimuli were presented during each experimental session. Therefore, it is possible our results could be biased by differences between animals and/or experimental recording session. To account for this, in all of our statistical tests we took one of the following two approaches: (1) Averaged metrics

across cells (or pairs of cells) and sound stimuli within a recording session before performing statistical tests or (2) Performed statistical tests using hierarchical bootstrapping.[121] Although each approach reduces statistical power relative to treating each individual measurement as independent, they provide a more conservative estimate of p-values and reduce the chance of detecting false positives.[121]

For all statistical tests measuring large vs. small pupil effects where we first averaged results within recording session, we performed a two-tailed Wilcoxon signed-rank test. For each test, we report the test statistic, W, the p-value, and the exact n number of recording sessions used to perform the test. In cases where we performed a hierarchical bootstrap, we report the direct bootstrap probability of the null hypothesis.[121] In both cases, we also provide the mean and standard error of the number of measurements per recording session.
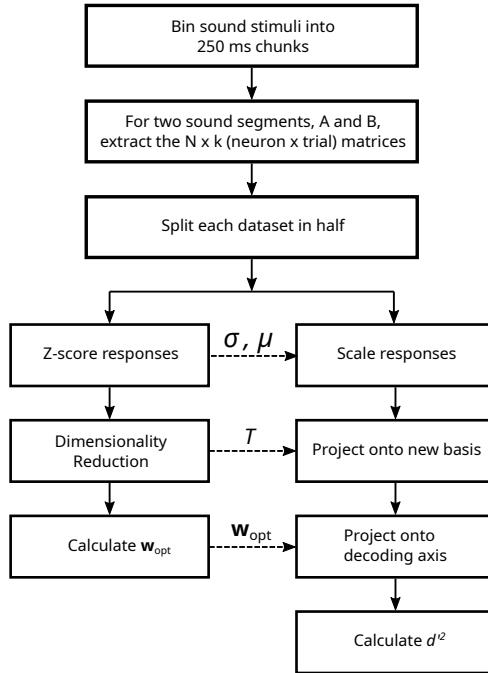
All bootstrapped confidence intervals (*e.g.* in Figure 3.5) were also generated using the hierarchical bootstrap resampling procedure. For paired tests, *e.g.* testing the difference in model performance for different architectures (Figure 3.5, S5), the paired statistic was first computed for each observation (in this case, for the difference in model performance for each stimulus pair), resampled using hierarchical bootstrapping, and the direct probability of this resampled distribution being different than zero was reported.[121]

3.5   Data Availability
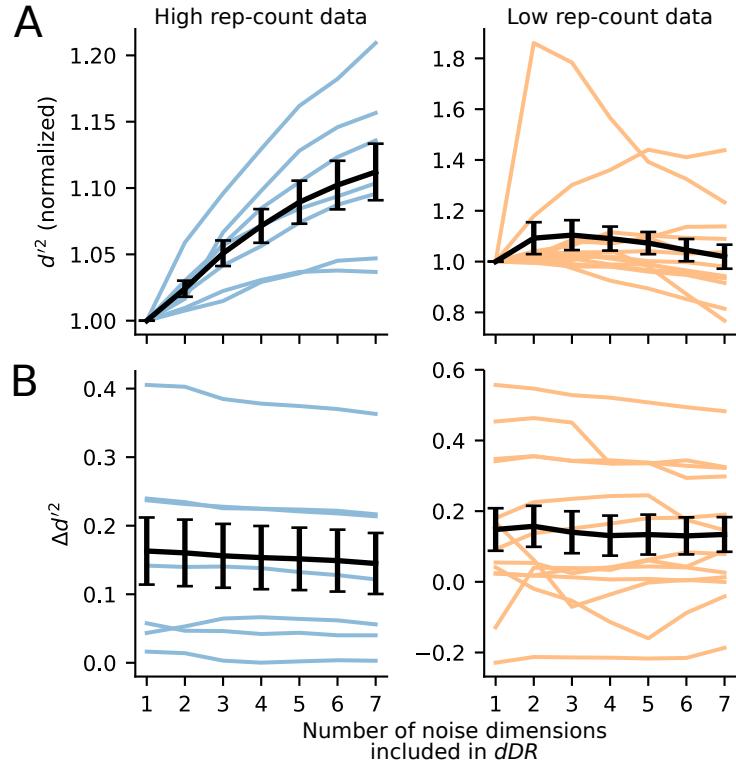
The datasets analyzed in this study are available from the corresponding author upon reasonable request.

3.6   Author contributions
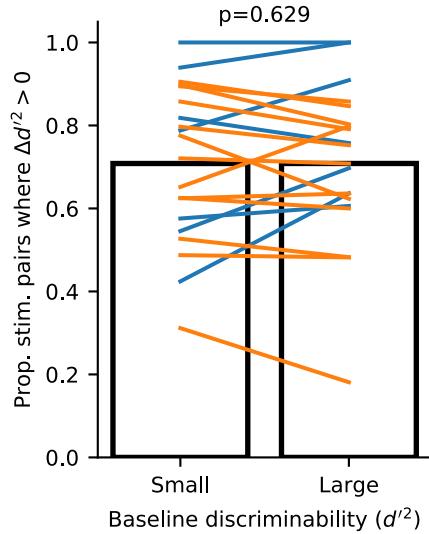
CRH and SVD designed experiments. CRH performed experiments, analyzed data, and developed pupil tracking software. MLE, DS, and LPS assisted with data collection. LPS assisted with software development for pupil tracking. CRH and SVD wrote the manuscript. All authors edited the manuscript.
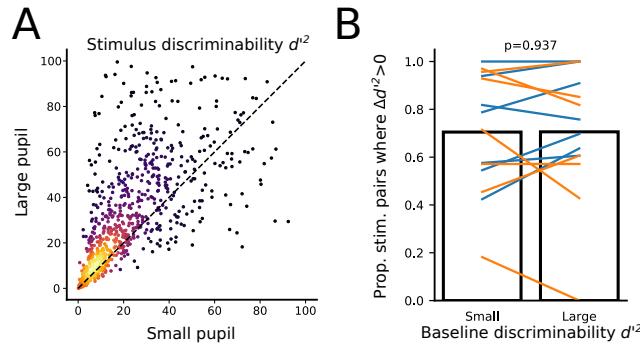
Figure S1:   **Calculating neural discriminability.** Procedure for measuring discriminability of natural sound segments by neural populations in primary auditory cortex (A1). Stimuli were binned into 250 ms chunks, and a discrimination index, d'^2, was calculated between each pairwise combination of segments. To prevent over-fitting to noise in the data, we took two steps: First, we performed cross-validation, estimating the optimal decoding axis using 50% of the data and evaluating d'^2 along this axis with the held out 50%. Second, because we were in a trial-limited regime where the number of stimulus repetitions, k, was often less than the number of neurons, N, we performed dimensionality reduction which allowed us to reliably estimate neural discriminability for each pair of stimuli. The dimensionality of this reduced space was chosen using cross-validated estimates of d'^2 (Figure S2).
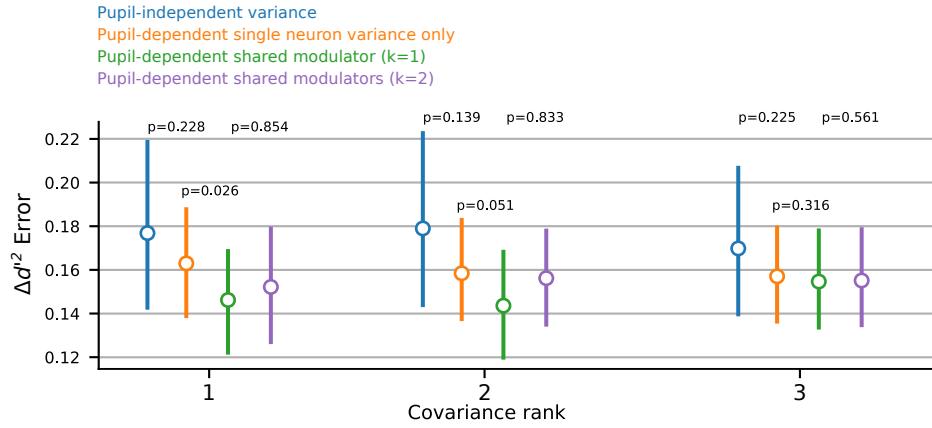
Figure S2: **Choosing number of dimensions for decoding-based dimensionality reduction.** To reduce the dimensionality of our data, we used dDR, a method designed specifically for measuring decoding accuracy from trial-limited data.[106] dDR is composed of the stimulus discrimination axis, $\Delta\mu$ (Eqn. 4.4), as well as a user-defined number of noise dimensions. **A.** We measured cross-validated decoding accuracy (d'^2) using dDR with between one and seven noise dimensions. Results are subdivided into two groups: high rep-count data, where the number of stimulus repetitions exceeded the number of recorded neurons and low rep-count data, where the number of stimulus repetitions was less than the number of recorded neurons. Individual lines (orange / blue) represent d'^2 averaged across all stimulus pairs for a single recording session. Black lines are mean $\pm$ the standard error of the mean across recordings. For the high rep-count data, we began to observe saturation of d'^2 in a couple of recording sites at $n = 7$ noise dimensions. Therefore, we chose to use seven noise dimensions to analyze the high rep-count data. For the low rep-count data, information saturated much more quickly, with some sites already showing no improvement between $n = 1$ and $n = 2$ dimensions. Thus, for the low rep-count data we used only one noise dimension. **B.** Mean change in d'^2 ($\Delta d'^2$) across stimulus pairs is plotted in the same fashion as in A. Choice of dimensionality had little effect on the overall impact of arousal on decoding accuracy.

Figure S3: **Heterogeneity of $\Delta d'^2$ is not explained by baseline discriminability.** We split $\Delta d'^2$ based on the median absolute stimulus discriminability across all trials (d'^2) and measured the fraction of stimulus pairs that showed an increase in coding accuracy in each group for each recording site. Orange lines indicate low rep-count sites and blue lines indicate high rep-count sites, as in Figure S2. We found no significant difference between small and large baseline d'^2 stimulus pairs ($p = 0.629, W = 83$, Wilcoxon sign test, $n = 20$ recording sessions.)

Figure S4: **Heterogeneity of $\Delta d'^2$ is not explained by unreliable response estimates.** Using cross validation, we identified the stimuli whose noise distributions could be estimated reliably. That is, sites where the first PC of single trial responses could be estimated above chance performance (see Methods). **A.** Same as in Figure 3.3B, but only for reliable stimulus pairs. d'^2 increased significantly overall ($d'^2_{large} = 33.88 \pm 0.96$ vs. $d'^2_{small} = 23.54 \pm 0.79$, $p < 0.001$, Bootstrap test, $n = 20$ recording sessions, 39.6 ± 6.1 stimulus pairs per session). Across sites the mean percentage of stimulus pairs with $\Delta d'^2 < 0$ was $29.8 \pm 1.4\%$. Almost identical to larger dataset. **B.** Same as in Figure S3, for reliable stimulus pairs. We observed no significant different between pairs with large and small baseline discriminability ($p = 0.937, W = 38$, Wilcoxon sign test, $n = 20$ recording sites).

Figure S5: **Normative latent variable model selection.** Normative models were fit to optimize predictions of either the rank-1, rank-2, or rank-3 approximations to pupil-dependent covariance matrices. Relaxing model constraints to fit rank-2 / rank-3 covariance matrices did not improve predictions of $\Delta d'^2$. Colors indicate model architecture specified in the figure legend. As in Figure 3.5, lines represent bootstrapped 95% confidence intervals and p-values were measured using paired bootstrapped samples of differences in model error for the different architectures.

4 Task-related suppression of correlated variability in A1 predicts behavior performance but not changes in neural discrimination

Charles R. Heller[1,2], Daniela Saderi[1,2] & Stephen V. David[1]

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

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

Abstract

Sound-evoked activity of single neurons in primary auditory cortex (A1) is modulated by active task engagement, enhancing the neural discriminability of task-relevant sound features. At the neural population level, task engagement also suppresses correlated variability (*i.e.*, noise correlations) between pairs of simultaneously recorded neurons. Theoretical work suggests that a reduction in noise correlations can improve sensory coding accuracy, but it is unknown how these task related changes actually impact coding in A1 or behavioral performance. We analyzed data from A1 of ferrets performing a go / no-go tone detection task. Multi-channel neural data was collected using linear microelectrode arrays during alternating passive and active (engaged) listening blocks. As expected, transitions from passive to active listening were associated with modulation of evoked activity and decreased noise correlation. Using an optimal linear decoder, we found that discrimination of task-relevant sound categories by neural populations was improved during active listening blocks. Notably, this effect could be almost entirely attributed to modulation of single neuron responses; discrimination did not benefit from reduced noise correlations. Furthermore, we found that noise correlation strength could predict trial-trial behavior performance, while neural decoder accuracy could not. The decoupling of decoder accuracy from behavioral performance suggests that downstream brain areas utilize a non-optimal readout of activity in A1.

4.1 Introduction

Task engagement is known to modulate the gain and variance of single neuron sensory responses in auditory cortex (AC)[3, 7, 11, 17, 122, 123] as well as suppress correlated trial-to-trial variability across neurons *i.e.* noise correlations.[73, 74] While neural discrimination of task-relevant sounds by single neuron activity is improved during engagement relative to passive listening,[7, 11, 122] it is currently unknown how noise correlations in auditory cortex affect discrimination. Furthermore, the precise relationship between perception and task-dependent plasticity in noise correlations has not been explored. A recent study found that bouts of correlated activity preceded detection failures of tones embedded in background noise,[124] suggesting that increased noise correlations may lead to perceptual deficits in AC for at least some auditory tasks.

Work in the visual system recently demonstrated that neural population activity is the key computational unit underlying perception in a change detection task.[52] This work found that reductions in noise correlations predict the improvements in task performance associated with learning and selective attention but do not predict improvement of an optimal decoder trained to discrimination stimuli from neural activity. Effects also vary between visual areas. In visual area V4, noise correlation changes are restricted to a low-dimensional neural state-space that overlaps with choice related information.[52] This pattern was not observed for neurons recorded in primary visual cortex (V1) during the same task; noise correlations in V1 did not predict task performance.[52] Thus data from the visual system indicate that stimulus information shifts to a sub-optimal, choice-dominated read-out in later processing stages.
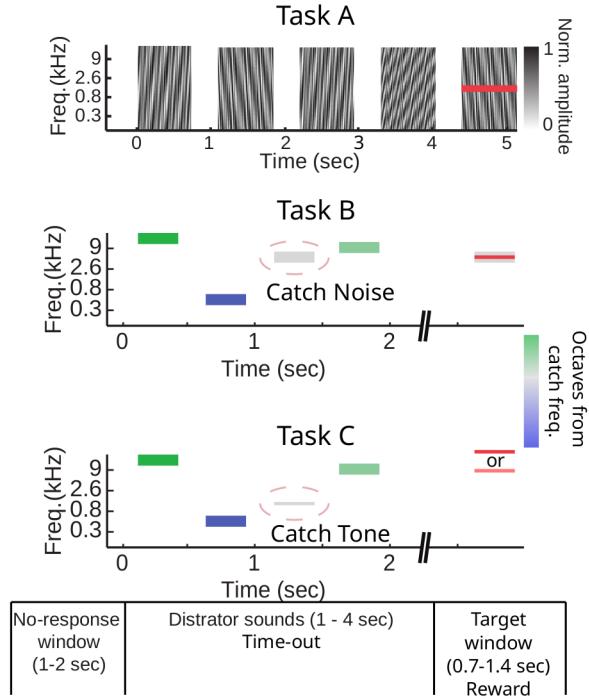
Here, we studied the relationship between behavior and neural population coding in primary auditory cortex (A1) of ferrets engaged in three simple go / no-go tone detection tasks. Across three different task designs, we observed a robust decrease in noise correlations during active task engagement compared to during passive presentation of sound stimuli. Optimal linear decoding of go vs. no-go task stimuli by population activity was also improved during engagement. However, decoding of go vs. no-go cues was not impacted by changes in noise correlation and was consistently found to be higher than animal's psychometric performance, even during passive listening. This decoupling of decoder performance and behavior indicates that sound information is not read out optimally in A1 for these tasks. In contrast, changes in noise correlations were tightly coupled to animal's task performance. Thus, it appears that A1 may adopt a more general purpose decoding strategy that benefits from reduced variability in population activity.

4.2   Results

We trained nine young adult male ferrets to perform one of three variants of a go / no-go tone vs. noise discrimination task (Tasks A, B, and C). Schematics of a single behavioral trial are shown for each task in Figure 4.1. Despite small differences in the precise trial structure and sound stimuli (Methods), all three tasks required animals to lick in response to a target tone in order to receive reward. Licking during reference noise burst stimuli was penalized with a time out. Task C also required animals to perform frequency discrimination in addition to tone vs. noise discrimination. For details on task structure and sound stimuli, see Methods: Behavioral training.

Animal's task performance was measured using a behavior discrimination index, DI, the area under the receiver operating characteristic curve for responses to target tone stimuli vs. reference noise

bursts.[125] This metric ranged from 0 to 1, with 0.5 indicating chance performance and 1 indicating perfect performance. Although closely related to the standard signal detection theory metric, d', the discrimination index also takes into account reaction time. In our task, most target stimuli were significantly above animal's psychometric threshold and miss trials were rare. Therefore the discrimination index helped to improve resolution of behavioral task-engagement and performance.

Figure 4.1: **Go / no-go task paradigms.** Nine ferrets were each trained on one of three different go / no-go tone detection tasks. All tasks were similar in that they required animals to detect the presence of a target tone in a sequence of distractor noises. Tones were either presented in in isolation or masked by noise. The precise trial structure and acoustic stimuli varied between tasks. Most notably, in task A distractor stimuli were broadband noise, while narrowband noise (0.5 octave bandwidth) was used in tasks B and C. Additionally, tasks B and C contained explicit catch stimuli that could occur in place of a target during the a subset of trials. This feature was absent from task A. For more details on each of the tasks, see methods: Behavioral training. Finally, task C is unique in that it required animals to perform frequency discrimination in addition to tone vs. noise discrimination.

Once animals reached the behavioral criterion of stable, above chance performance, a craniotomy was opened over primary auditory cortex (A1) and neural activity was recorded using single shank 64-channel or dual shank 128-channel silicon microelectrode arrays[40] inserted perpendicular to the brain's surface. Target tone stimuli for each session were chosen to match the most common best frequency of neurons recorded at that site. In tasks B and C, narrowband noise bursts were chosen to span a three octave range around the target frequency, while broadband noise in task A was fixed across

69

recording sessions. To measure the effect of task-engagement on activity in A1, neural responses to task stimuli were compared between active and passive listening conditions. Passive blocks consisted of identical sound stimuli, but licking was no longer rewarded or penalized. For all analyses presented in the following sections, we did not observe substantial differences between data collected from the three tasks, therefore all data was pooled.

Transitions between active task-engagement and passive listening are known to be correlated with changes in overall levels of arousal, as indexed by pupil size.[33] As arousal is known to independently modulate neural activity in A1[28, 33, 71] and more generally throughout cortex,[27, 38, 86] we simultaneously recorded pupil size during physiology recordings in order to control for pupil-associated changes in firing rates (see Methods, Eqn. 4.3). These controls had little effect on the main conclusions presented here, therefore we conclude that our results are driven primarily by changes in neural activity specific to task-engagement.
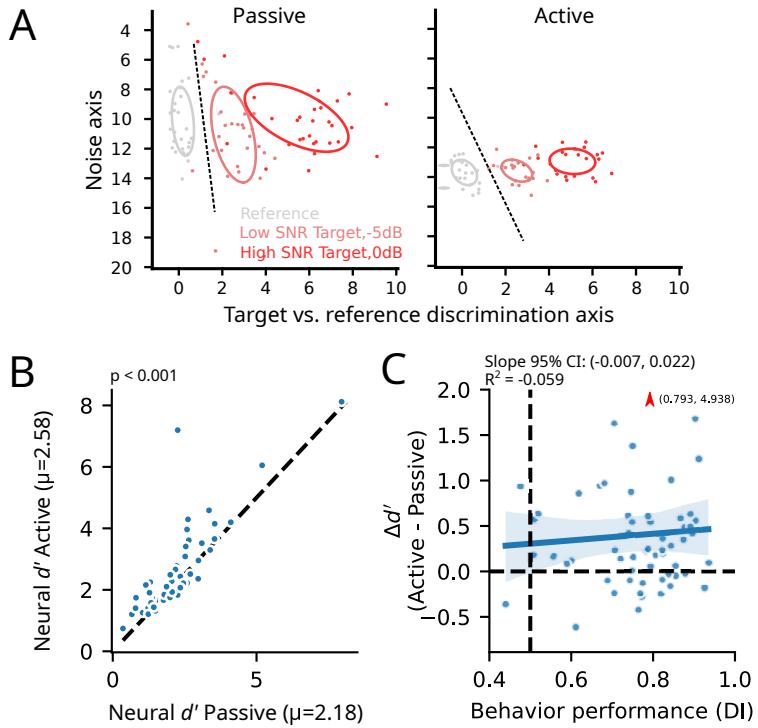
4.2.1 Task-dependent improvement of target vs. reference discriminability in A1 does not predict behavior performance

Behavioral performance was measured as the ability to discriminate reference noise bursts from target tones, either presented in isolation or embedded in noise. To investigate if neural activity in A1 supported task performance, we asked how well reference and target stimuli could be discriminated based on neural population activity. We measured the pairwise decoding accuracy between each target and all references using optimal linear decoding.[57, 58] Decoding performance was quantified with the signal detection theory metric, d'.[84] Therefore, a unique neural d' was associated with each target (*e.g.* each SNR) within a behavioral session (60 total target vs. reference discriminations across seven animals and 28 recording sessions).

It is known that trial-to-trial variability in neural population activity can critically impact decoding accuracy.[46, 57–61] However, this variability cannot be measured reliably in neural populations using the relatively small number of trials available in a typical behavioral experiment,[68] and standard measures of decoding accuracy like neural d' are likely to be biased by overfitting to experimental variability in population responses. Therefore, we performed a decoding-based dimensionality reduction[106] to project population activity into a 2-D space optimized for reliable measurement of trial-trial variability and d'. The first dimension of this space was defined as the axis in neural activity space passing through the mean reference response and mean target response (sensory discrimination axis) and the second axis captured the first principal component of trial-to-trial variability (noise axis,[106]). The decoding space was defined using data from both active and passive listening conditions, allowing unbiased

comparison of responses between the two states.

During active task engagement, responses to target and reference stimuli were more separated along the discrimination axis, and trial-trial variability was strongly suppressed (Figure 4.2A), reflecting task-dependent suppression of noise correlations (Figure S1). Neural d' was measured in this space separately for active and passive conditions. Neural discriminability of target and reference sounds was significantly improved during task-engagement (d' active: 2.58 vs. neural d' passive: 2.18, $p < 0.001$ hierarchical bootstrap test) (Figure 4.2B). However, the improvement was not significantly correlated with the animal's behavioral performance (Figure 4.2C), indicating that animals do not utilize an optimal linear readout of activity in A1 to guide behavior in this task. These results were consistent even after controlling for pupil-indexed changes in sensory responses (Figure S2), which are known to modulate excitability of single units in A1.[28]
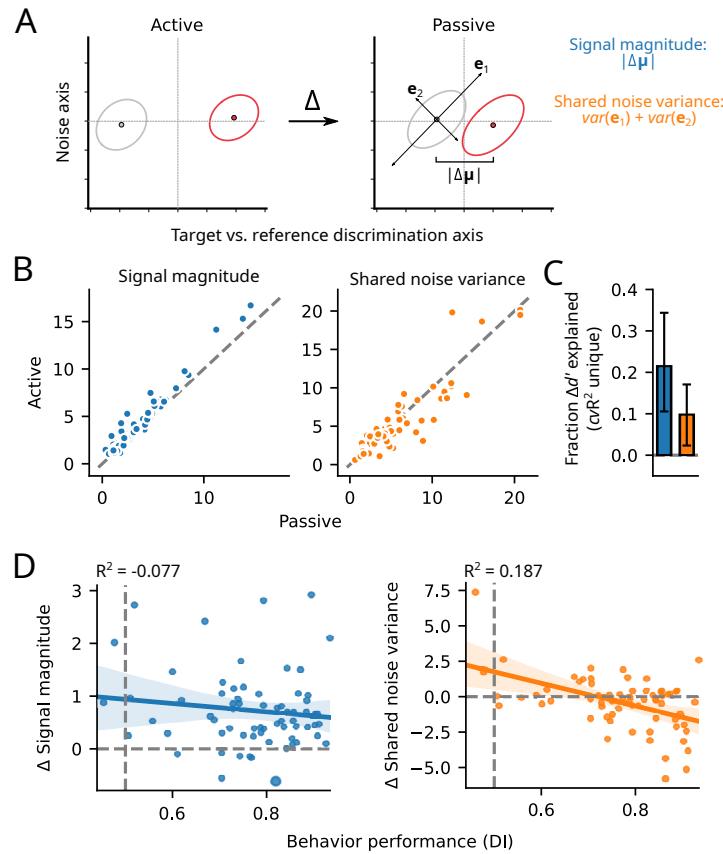
Figure 4.2: **Task-dependent improvements in neural discriminability are not predicted by behavior performance. A.** Example distributions of target and catch responses for a single recording session collected during task B. Single trial population spike count responses (dots) are projected into a reduced dimensionality decoding space.[106] Ellipses represent one standard deviation of responses across all trials. Dashed line represents the optimal linear decision boundary. **B.** For each target / recording site combination ($n = 60$ total combinations), active target vs. reference discriminability is plotted against passive discriminability. Discriminability (d') is significantly greater during active task engagement ($p < 0.001$, hierarchical bootstrap test). **C.** The change in neural d' between active and passive conditions is plotted as a function of behavioral performance. There is no significant relationship between the two (cross-validated $R^2 = -0.059$, 95% bootstrapped confidence interval of slope: $(-0.007, 0.022)$.

4.2.2 Task-dependent suppression of noise correlations explains behavior performance but not neural decoding accuracy

Improvements in neural d' could be explained by increased separation along the discrimination axis and/or a reduction of variance in the decoding space. We observed that both of these features of neural population activity were modulated by task-engagement, but the changes in variance occurred mostly along an axis that was orthogonal to the sensory discrimination axis (*e.g.* Figure 4.2A). Thus, we hypothesized that changes in variance contribute little to the task-dependent improvement in decoding (Figure 4.2B). To test this prediction, we defined two population level metrics: signal magnitude and shared noise variance (illustrated in Figure 4.3A) which, respectively, captured sensory discrimination

separation and correlated response variability. As expected, both were consistently modulated by task-engagement (Figure 4.3B), however, the modulation of signal magnitude explained the bulk of changes in decoding accuracy (95% confidence interval of R^2 signal magnitude (0.11, 0.34) vs. shared noise variance: (0.02, 0.17)) (Figure 4.3C). These results did not change after controlling for pupil size (95% confidence interval of R^2 signal magnitude (0.19, 0.29) vs. shared noise variance: (-0.005, 0.13)) (Figure S3A, B).

Recent work in the visual system has suggested that changes in correlated variability, but not response magnitude, are tightly coupled to perceptual performance.[52] Therefore, we next determined the extent to which behavioral performance could predict changes in signal magnitude and shared noise variance. In line with this previous work, we found that behavioral performance and shared noise variance were significantly correlated, while signal magnitude was not (95% confidence interval for R^2 signal magnitude vs. behavior: (-0.18, -0.006) and shared noise variance vs. behavior: (.125, 0.224)) (Figure 4.3D). Thus, animals' task performance is best when shared response variability is smallest. Again, these results could not be explained by spontaneous fluctuations in pupil-indexed arousal (95% confidence interval for R^2 signal magnitude vs. behavior: (-0.109, 0.023) and shared noise variance vs. behavior: (0.061, 0.164)) (Figure S3C).

Figure 4.3: **Changes in shared variability explain behavior, but not stimulus discriminability. A.** Cartoon schematic. Ellipses represent response distributions for a reference (grey) vs. target (red) stimulus. Two metrics were defined to describe population activity in the decoding space: Signal magnitude and shared noise variance (see Methods) **B.** For each target / recording site combination ($n = 60$ total combinations), the two metrics are compared for active and passive conditions. **C.** The contribution of each metric to task-dependent changes in discriminability (d') were determined by performing multiple linear regression. Cross-validated, unique R^2 is plotted for each metric. Error bars represent 95% bootstrapped confidence intervals (signal magnitude: (0.11, 0.34), shared noise variance: (0.02, 0.17)). **D.** The change in signal magnitude and noise variance are plotted against behavioral performance. Shaded regions represent 95% confidence intervals of best fit regression line. Signal magnitude is not significantly correlated with behavior (R^2 95% confidence interval: (-0.18, -0.006)). Shared noise variance is significantly correlated with behavior (R^2 95% confidence interval: (0.125, 0.224)).

4.3 Discussion

Go / no-go auditory discrimination tasks have been used extensively to investigate the impact of task engagement and attention on auditory coding. Prior work has found that engaging in auditory tasks improves neural discrimination of sound by single A1 neurons.[3,7,11,17,123] A recent study expanded on this work by pooling single neurons recorded serially across multiple go / no-go tasks and, in line with single cell results, demonstrated that pseudo-populations could more accurately discrimi-

nate target from reference stimuli during engaged conditions relative to passive listening.[122] However, studies that have recorded population activity in A1 have found that activity is correlated between neurons. Further, the strength of correlated activity is sensitive to behavioral state[28,71,73,74] and may even limit the ability to detect pure tones embedded in noise.[124] These findings raise questions about utility of using single neurons to understand population coding in auditory cortex.

We addressed outstanding questions about the relationship between behavior and population coding by recording activity in A1 of ferrets performing three different go / no-go tasks. In line with previous work, we found neural discrimination of task stimuli is improved during active task engagement and correlated activity is suppressed. We were able to investigate how each of these effects related to the level of behavioral performance. While behavioral performance does not track improvements in neural discrimination, task-dependent reductions in correlated activity do.

4.3.1 Implications for auditory coding in A1

Our finding that task-dependent suppression of correlated activity predicts behavioral performance while neural discriminability does not is at odds with the idea that sound information is decoded optimally from A1. We propose two possible explanations for this result. First, A1 is not necessary for performing these behaviors or, second, activity in A1 is decoded sub-optimally by downstream brain regions.

Whether or not auditory cortex is required for different auditory behaviors is currently the topic of much debate, with a number of results providing contradictory conclusions.[17,126–133] Some of the disparity is likely to be explained by the method and efficacy of cortical inactivation and/or lesion used in each study. In addition, some of the variability in results may be explained by the hypothesis that the necessity of AC depends on how perceptually demanding the task is. For example, processing in A1 seems to be dispensable for performing very simple tasks, such as pure tone detection or discrimination of dissimilar tones,[134] but might be required for tasks that come closer to animal's perceptual threshold.[17] All of the behavioral tasks in the manuscript contained at least one pure tone target presented in isolation, while only certain blocks and/or task paradigms required the more difficult task of tone-in-noise detection. And even then, most target SNRs were well above perceptual threshold. Therefore, it is possible that auditory cortex is dispensable for the simple go / no-go discrimination tasks tested here.

This conclusion is challenged, however, by the striking modulation of A1 activity observed during engagement both for our tasks as well as across a range of similar task designs.[3,7,11,33] Specifically, if A1 was completely irrelevant it seems unlikely that we would observe such a robust association

between correlated variability and behavioral performance. Therefore, we argue that our results favor the second hypothesis, that activity in A1 is decoded sub-optimally. Work in the visual system has demonstrated a finding that is consistent with this idea.[52] Noise correlation strength is strongly suppressed by attention and learning in V4. Using a choice decoder trained on V4 activity, the study found that the dimensions of neural activity that are informative about choice are aligned with noise correlations, but not with stimulus information.[52] This suggests that learning and attention improve perception by suppressing variability along a sub-optimal stimulus readout in V4. Unfortunately, performance in our tasks was high, so that animals performed very few miss trials. Therefore, we did not have sufficient data to build a choice decoder and test this hypothesis in A1.

4.3.2 Underlying mechanisms responsible for state-dependent modulation of correlated activity A1

The origin of stimulus-independent, correlated variability (*i.e.* noise correlations) has been debated over the past two decades. However, emerging evidence gleaned from recent large scale neural population recordings have helped provide some clarity. In most cases, correlated activity is accurately modeled using low-dimensional, latent processes that co-modulate neurons.[50, 51] In addition, changes in behavioral state, such as fluctuations in arousal and motivation,[29, 38, 54] co-modulate activity.

To isolate correlated activity related specifically to task demands, we controlled for changes in firing rate that could be explained using pupil size, a common index of global arousal.[4, 28] Controlling for pupil-indexed arousal reduced task-related changes in noise correlations significantly but did not remove them completely. One potential explanation for this is residual shared variability in task-dependent gain of sensory responses. Prior work has demonstrated that co-tuned neurons in A1 exhibit consistent gain changes during task-engagement.[3, 11] Though the sign of these changes depends on task-reward structure,[7] it is typically consistent across neurons within a task condition. Thus, variable gain from trial to trial is likely to produce shared response variability. Furthermore, if gain is more stable during task-engagement (*e.g.* because animals know the target frequency to attend), then correlated activity would be reduced during active task engagement as we observe in our data. A similar explanation was recently proposed for spatial attention-dependent reductions of noise correlations in visual cortex.[81] Future studies in auditory cortex could test this hypothesis conclusively by alternating between task blocks where animals must attend to many different sound frequencies simultaneously and task blocks where the target frequency is fixed throughout a session.

4.4 Materials and Methods

4.4.1 Surgical procedure

All procedures were performed in accordance with 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). The surgical approach was similar to that described previously.[7, 18, 28, 115] Nine young adult male ferrets were acquired from an animal supplier (Marshall Farms). Head-post implantation surgeries were then performed in order to permit head-fixation during neurophysiology recordings. Two stainless steel head-posts were fixed to the animal along the midline using UV-cured dental composite (Charisma) or bone cement (Palacos), which bonded to the skull and to stainless steel screws that were inserted into the skull. After a two-week recovery period, animals were habituated to a head-fixed posture and auditory stimulation. They were then trained on one of three go / no-go tone detection tasks (see below). Once they reached consistent above chance task performance, a small (0.5 - 1 mm) craniotomy was opened above primary auditory cortex (A1) for neurophysiological recordings.

4.4.2 Behavioral training and acoustic stimuli

Stimulus delivery:

Digital acoustic signals were transformed to analog (National Instruments), amplified (Crown), and delivered through a free-field speaker (Manger) placed 80 cm from the animal's head and 30°contralateral to the the hemisphere in which neural activity was recorded (Figure 3.1). Stimulation was controlled using custom MATLAB software (`https://bitbucket.org/lbhb/baphy`), and all experiments took place inside a custom double-walled sound-isolating chamber (Professional Model, Gretch-Ken).

Behavioral training:

Animals were trained on one of three go / no-go task paradigms, each of which is described in detail below and also briefly in the results. During training, animals were provided access to water ad libitum on weekends, but were placed on water restriction during the weekdays (Monday through Friday), allowing them to maintain >90% of their baseline body weight long term. On weekdays, they were given the opportunity to receive liquid reward during behavioral training. In each of the three task paradigms, animals were trained to lick a fluid spout in response to a target tone in order to receive a sweetened, nutrient rich fluid reward (Ensure). Licking during presentation of any other (distractor) sounds was punished with a brief time out. In all tasks, behavioral performance was measured using a

discrimination index (DI), which is closely related to the signal detection theory metric, d',[84] but also includes information about reaction time.[7, 135] The DI was defined as the area under the receiver operating characteristic curve for responses to target tone stimuli vs. reference noise bursts.[125] Thus, DI ranged between 0 and 1, with 1 indicating perfect performance and 0.5 indicating chance performance.

Task A: Tone in broadband noise This task was described in a previous publication.[33] Data from 4 animals trained on this task is included here. Briefly, each behavioral trial consisted of two to five broadband noise bursts (TORCs - Temporally Orthogonal Ripple Combinations[136]) followed by a pure tone target presented in isolation (Inf dB, pure tone detection) or embedded in a TORC (tone-in-noise detection, SNR between tone and noise could range from +5dB to -15dB, depending on the experimental condition). TORCs were selected from a library of 30 total samples. Each spanned five octaves. All sounds were presented for 0.75s duration with a 0.7s inter-stimulus interval. Animals were rewarded for licking 0.2 to 1.5s following the target onset (hit trials) and were penalized with a 5-8s timeout for failure to lick (miss trials). Licking at any other point during the trial was also penalized with a 5-8s timeout. Regardless of the trial outcome, the next trial began after a random delay of 2 - 3 sec.

A single behavioral block typically consisted of 60-100 trials with the same target frequency (100-20,000 Hz) and same distribution of target SNRs. Between each block, target frequency could change. Over the course of training, target frequency was chosen randomly in order to span the range of ferret hearing[137] while during neurophysiology the target frequency was chosen to match the most common best frequency in the recorded population of neurons.

Task B: Tone in narrowband noise Two animals were trained on this task. Neurophysiology data is included for one animal. On each trial, a random number of narrowband white noise bursts (1-8) were followed by a pure tones presented in isolation (+Inf dB) or embedded in noise (SNRs: 0dB, -5dB, and -10dB). Noise bursts had 0.5 octave bandwidth and spanned 3 octaves around the target frequency. Sounds were 300 ms in duration with 200 ms ISI. On 25% of trials, a catch noise (-Inf SNR at the target frequency) was played at a time when the target could occur. On catch trials, a pure tone target always appeared at the end of the trial so that correct reject trials could be dissociated from lapse trials. Licking responses during the catch noise, or during any other noise burst, were treated as a false alarm and penalized with a timeout ranging from 4-8 sec while licking 0.2 to 0.8 sec following a target onset resulted in a liquid reward as in Task A. Miss trials were not penalized. Target tone frequency was fixed across behavioral blocks within a day, but changed randomly between

days to span the hearing range for ferrets.[137] Behavioral blocks typically lasted between 80 and 200 trials.

Task C: Frequency discrimination Two ferrets were trained on this task. Neurophysiology data is included for one of these animals. Similar to task B, animals were trained to discriminate pure tone targets from narrowband noise. However, in this task each behavioral block consisted of up to two targets, each +Inf dB and separated by < 0.3 octaves. In addition, the catch noise was replaced by a catch tone (also +Inf dB) separated by at least 0.5 octaves from all target tones. Noise bursts and trial structure were equivalent to task B.

4.4.3 Neurophysiology

Upon reaching behavioral criterion, neurophysiological recordings were completed for seven of the nine animals. A small, 0.5 to 1 mm craniotomy was opening above auditory cortex and 1 - 4 tungsten micro-electrodes (FHC, 1-5 MΩ) were inserted to characterize the tuning and response latency of the region of cortex. Sites were identified as A1 by characteristic short latency responses, frequency selectivity, and tonotopic gradients across multiple penetrations.[10,116] Subsequent penetrations were made with a single (64-channel) or dual shank (128-channel) silicon electrode array.[40] Electrode contacts were spaced 20 μm horizontally and 25 μm vertically, collectively spanning 1.05 mm of cortex. On each consecutive recording day, we changed the location of the electrode penetration to access fresh cortical tissue, expanding the craniotomy as necessary. Data were amplified (RHD 128-channel headstage, Intan Technologies), digitized at 30 KHz (Open Ephys[94]) and saved to disk for further analysis.

Spikes were sorted offline using Kilosort[117] or Kilosort2 (`https://github.com/MouseLand/Kilosort2`). Spike sorting results were manually curated in phy (`https://github.com/cortex-lab/phy`). For all sorted and curated spike clusters, a contamination percentage was computed by measuring the cluster isolation in feature space. All sorted units with contamination percentage less than or equal to 5 percent were classified as single-unit activity. All other stable units that did not meet this isolation criterion were labeled as multi-unit activity. In all analyses presented here, both multi units and single units are included.

4.4.4 Pupillometry

During neurophysiological recordings, video of the ipsilateral pupil (relative to the recording hemisphere) was collected using 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 camera 10 cm from the eye. Contrast was increased using infrared illumination. Ambient light levels were fixed for each experiment at roughly 1500 lux to provide maximum dynamic range of pupil size.[28] Pupil size was measured offline by fitting an ellipse to each video frame using using a custom machine learning algorithm (Python and Tensorflow `https://github.com/LBHB/nems_db`). The minor axis of the fit ellipse was extracted and saved for analysis with neurophysiological data. Blinks were detected and excluded as in[28] and pupil data was shifted by 750 ms relative to spike times in order to account for the lagged relationship between changes in pupil size and neural activity in auditory cortex and to allow for comparison with previous research.[4]

Pupil correction:

In this work, we were interested in studying task-dependent changes in correlated neural activity. Because spontaneous changes in arousal are known to coherently modulate firing rates in A1[39] and fluctuations in arousal may be different between active and passive states (*e.g.* depending on task performance,[33] it was important to control for any neural variability not directly attributable to task-engagement.

To remove changes in firing rate that were associated with pupil, we regressed out all stimulus-independent variability that could be explained with pupil size. We performed this correction on a per-neuron, per-stimulus basis to completely remove pupil-associated changes in firing.

For each neuron (i) and stimulus (s) the mean response across all trials was measured ($\bar{r}_{i,s}$) and subtracted from the true response ($r_{i,s}$, Eqn. 4.1). A linear regression model was used to predict to the residual variability in firing rate with pupil size ($\hat{r}_{resid_{i,s}}$, Eqn. 4.2). The prediction of this model was then subtracted, and the remaining, pupil-corrected, trial to trial variability was added back to the mean stimulus response (Eqn. 4.3).

$$r_{resid_{i,s}} = r_{i,s} - \bar{r}_{i,s} \tag{4.1}$$

$$\hat{r}_{resid_{i,s}} = \beta_1 p(t) + \beta_0 \tag{4.2}$$

$$r_{corr_{i,s}} = \bar{r}_{i,s} + \left(r_{resid_{i,s}} - \hat{r}_{resid_{i,s}}\right) \tag{4.3}$$

4.4.5 Neural population decoding of target vs. reference stimuli

In order to measure the neural discriminability of target and reference stimuli, we used the information theoretic measure, d'^2, the discrete analog of Fisher Information,[57,58] to measure the discriminability between each target stimulus and all reference stimuli. References were structured slightly differently between tasks, but always had equivalent behavioral meaning. Therefore, we grouped them into a single stimulus class in order to allow comparison of results across task paradigms. Thus, d'^2 was defined as:

$$d'^2 = \Delta\boldsymbol{\mu}^T \Sigma^{-1} \Delta\boldsymbol{\mu} = \Delta\boldsymbol{\mu}^T \mathbf{w}_{opt} \tag{4.4}$$

where $\Delta\boldsymbol{\mu}$ represents the vector connecting the mean ensemble responses to a single target stimulus (TAR) and the mean response across all references (REF), $\Sigma = \frac{1}{2}(\Sigma_{REF} + \Sigma_{TAR})$ represents the mean noise-covariance matrix, and $\mathbf{w}_{opt}$ is the optimal decoding axis, *i.e.* the vector orthogonal to the optimal linear discrimination hyperplane in state-space. Ensemble responses for a given sound stimulus were neuron X trial matrices of spike counts elicited in the first 200 ms following sound onset for each neuron on each sound repetition. Throughout the manuscript we report d', rather than d'^2, in order to provide a more interpretable metric in light of classic signal detection theory.[84]

Estimation of optimal decoding accuracy is known to be unstable in the regime where the number of neurons is large and the number of trials is relatively small.[68] Therefore, in order to reliably measure d'^2, we first projected our data into a 2-dimensional space using decoding-based dimensionality reduction (dDR).[106] Though similar to standard implementations of PCA, dDR explicitly preserves the dimensions of neural population activity that are most important for performing optimal linear decoding. The first axis in dDR space is defined as $\Delta\mu$, the sensory discrimination axis, and the second axis preserves the maximal dimension of correlated activity *i.e.* the first principal component of stimulus-independent variability.

4.4.6 Population decoding metrics

In order to quantify which features of the population activity were modulated by task-engagement, and how each contributed to decoding accuracy, we defined two metrics: signal magnitude and shared noise variance. Signal magnitude was defined as the vector magnitude of $\Delta\mu$, the axis connecting the mean ensemble response to the target vs. reference stimuli. Thus, this metric quantifies how separable two stimuli are based only on their trial-averaged sensory evoked activity. Shared noise variance, in contrast, summarizes the amount of overall trial-to-trial variability across repeated stimulus presenta-

tions. To measure shared noise variance, response distributions in the dDR space (*e.g.* Figure 4.2A) were mean-centered and total variance of the distribution was measured.
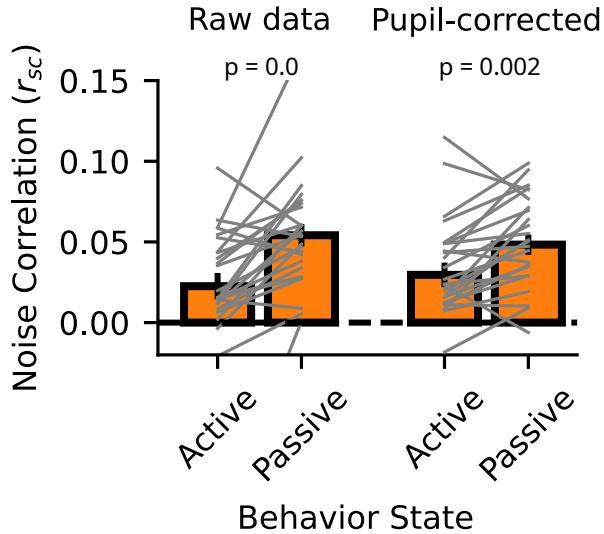
4.4.7 Statistical Methods

Our data followed a multi-level hierarchical structure. That is, for each recording session, we measured the decoding accuracy and associated metrics (see above) for multiple different target vs. reference pairs. To prevent outlier recording sessions from biasing statistical conclusions, we used hierarchical bootstrapping.[121]

For all linear regression models (*e.g.* testing for correlation between decoding accuracy and behavioral performance) we re-sampled by bootstrapping and performed 10-fold nested cross-validation in order to predict all data within each bootstrap sample. The 95% confidence interval for R^2, measured across 100 bootstraps, is reported in all regression analysis. We concluded that the relationship between two variables is significant when the confidence interval is in the range $R^2 > 0$.

4.5 Supplemental Material

Figure S1: **Task-engagement suppresses noise correlations.** Noise correlations were measured for all simultaneously recorded cell pairs during active and passive conditions. Correlations were grouped per site by computing the mean across all cell pairs. Each grey line represents one site, orange bars represent means across sites. Both before (left) and after (right) controlling for effects of pupil-indexed (Methods) we observe a significant reduction in correlation strength during active listening. Raw data: median active = 0.16, median passive = 0.05, $p = 0.0$, $W = 28$, Wilcoxon signed-rank test. Pupil-corrected responses: median active = 0.24, median passive = 0.05, $p = 0.002$, $W = 50$, Wilcoxon signed-rank test.
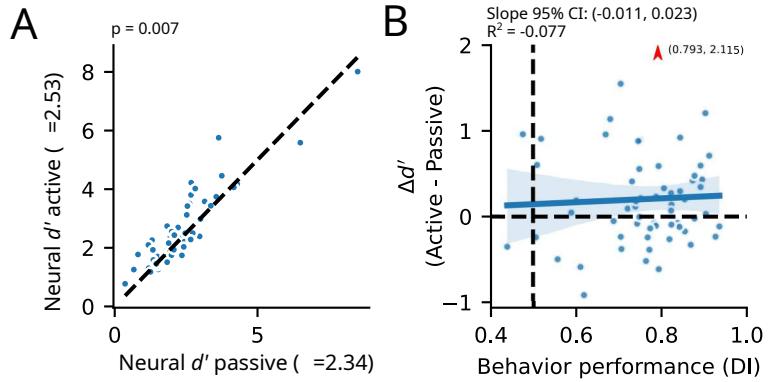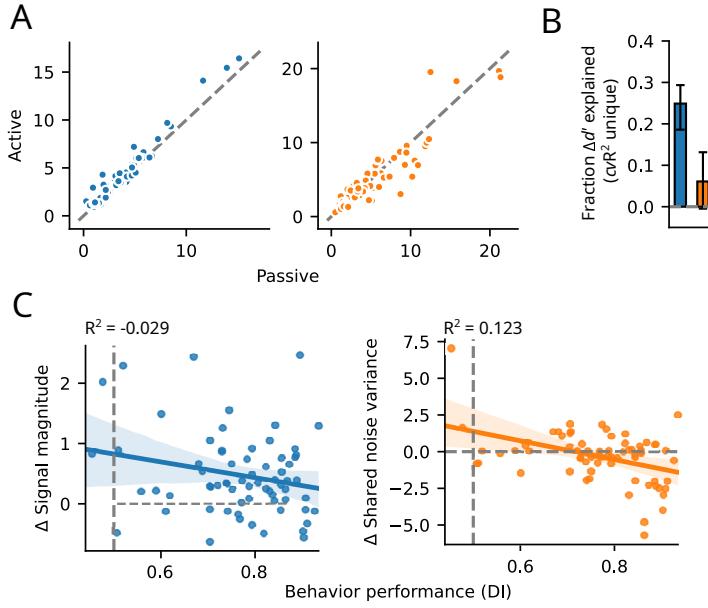
Figure S2: **Task-dependent improvements in neural discriminability are not pupil-dependent.** Same as in Figure 4.2 but after controlling for changes in response magnitude that could be explained by pupil-indexed arousal (see Methods) **B.** For each target / recording site combination ($n = 60$ total combinations), active target vs. reference discriminability is plotted against passive discriminability. Discriminability (d') is significantly greater during active task engagement ($p = 0.007$, hierarchical bootstrap test). **C.** The change in neural d' between active and passive conditions is plotted as a function of behavioral performance. There is no significant relationship between the two (cross-validated $R^2 = -0.077$, 95% bootstrapped confidence interval of slope: $(-0.011, 0.023)$.

Figure S3: **Changes in shared variability explain behavior, but not stimulus discriminability.** Same as Figure 4.3, but after correcting for pupil-dependent variability in spike rates (see Methods) **B.** For each target / recording site combination ($n = 60$ total combinations), the two metrics are compared for active and passive conditions. **C.** The contribution of each metric to task-dependent changes in discriminability (d') were determined by performing multiple linear regression. Cross-validated, unique R^2 is plotted for each metric. Error bars represent 95% bootstrapped confidence intervals (signal magnitude: (0.19, 0.29), shared noise variance: (-0.005, 0.13)). **D.** The change in signal magnitude and noise variance are plotted against behavioral performance. Shaded regions represent 95% confidence intervals of best fit regression line. Signal magnitude is not significantly correlated with behavior (R^2 95% confidence interval: (-0.109, 0.023)) Shared noise variance is significantly correlated with behavior (R^2 95% confidence interval: (0.061, 0.164)).

5 Conclusions

In addition to encoding acoustic information, activity in AC is modulated by many aspects of behavioral state such as a locomotion,[6] arousal,,[28,33,71] and task-engagement.[3,7,11,17,33,122] These state changes significantly affect the response properties of single neurons, but their effect on neural population codes is less well understood.

This dissertation builds on previous studies of state-dependent neural coding by investigating how the collective activity of many simultaneously recorded neurons in auditory cortex are modulated by two aspects of behavioral state, arousal and task-engagement. In chapter two, we present a novel method for dimensionality reduction of neural data that facilitates reliable estimates of population decoding accuracy when data are trial limited. In chapters three and four, we apply this method to neural population data collected from awake, behaving ferrets. Chapter three shows that the effects of increased arousal on neural decoding accuracy of natural sound stimuli are much more heterogeneous than previously thought. This study highlights the importance of studying population coding accuracy using rich, ethologically relevant sensory stimuli. Finally, in chapter four we studied the effect of task engagement on neural discrimination of task relevant sound cues. Our results indicate that activity in primary auditory cortex is decoded suboptimally by downstream brain regions. Thus simultaneous recordings of population activity and behavior led to new insights about how sound information in A1 contributes to behavior.

5.1 Methods for dimensionality reduction of neural data

In chapter two, we present a new method for dimensionality reduction of neural population data, dDR, adding to a growing library of existing tools that have been applied to neural data.[85] We developed dDR specifically to facilitate measurements of population coding accuracy when data are trial limited – that is, when the number of recorded neurons exceeds the number of stimulus repetitions during a given experimental condition. This is an increasingly common situation thanks to the explosion in technology for neural population recordings.[83] During behavioral experiments, the problem is even more apparent because the number of times a stimulus is repeated is often fundamentally limited by animal performance and motivation.

The studies presented in chapters three and four highlight the benefits of dDR as a method for understanding population coding accuracy. In chapter three, we studied state-dependent coding accuracy across an extremely wide range of natural sound stimuli. Due to the large number of

different stimuli, we were fundamentally constrained in the number of times each could be presented. Therefore, without dDR reliable estimates of coding accuracy would have been unfeasible. In chapter four, we apply the method to data collected from animals actively engaged in a tone vs. noise detection task. Even though stimulus repetitions were limited by animal's motivation, the method allowed us to measure discrimination of task cues by neural populations reliably during active task engagement. Critically, dDR preserves information of correlated population variability. In both studies, we found that characterizing the impact of correlated neural activity on discrimination was critical for understanding how state-dependent coding relates to behavioral performance.

5.2   State-dependent coding of natural sensory stimuli

In chapter three, we study how arousal affects the representation of natural sounds in primary auditory cortex. Prior work has found that increased arousal leads to increased signal-to-noise ratio of single unit responses in visual cortex,[27] increased gain and reliability in ferret A1,[28] and changes in frequency selectivity in mouse.[71] Furthermore, population activity is decorrelated in high arousal states.[27, 28, 71] Collectively, these changes are believed to enhance the reliability of sensory coding. Unlike previous work, we find that within a cortical column, arousal can either improve, or reduce, population coding accuracy of natural sounds. We demonstrate that this diversity is the result of a stimulus-dependent relationship with low-dimensional correlated activity in A1; when correlations are suppressed along the axis of sensory discrimination, sound discrimination improves. We find that this only happens for a small subset of natural sound stimuli, explaining why previous studies that used limited stimulus sets have not observed the same phenomenon.

In this study, we defined coding accuracy as the ability to discriminate between randomly selected, unique 250 ms natural sound excerpts. Though there is some precedent for this choice,[120] it is fundamentally arbitrary. While our results highlight the theoretical diversity of state-dependent sound discrimination in A1, we cannot say whether or not the neurons we recorded from are actually responsible for making said discriminations. Therefore, it is our hope that future work will build on this study by applying the methods we have developed to targeted recordings of specific populations of neurons and/or sensory stimuli that are more ethologically relevant. Additionally, the diversity of state-dependent coding changes that we observed help to contextualize prior conflicting reports on the impact that correlated variability has on population coding.[43, 59, 62, 63, 66, 103, 105]

Finally, to characterize state-dependent modulation of correlated variability we utilized a normative latent variable modeling framework. In more standard modeling approaches such as,[50] models are

typically fit to the actual time varying neural response. The normative framework is unique in that model weights are not optimized to predict the actual neural data, but to generate simulated neural activity that match high-level features of the actual data – in our case, model weights were fit to match the state-dependent population covariance matrices. Thus, without worrying about "overfitting" to single trial variability in the true responses, were able to easily test hypotheses about the impact of state-dependent modulation of covariability on neural decoding accuracy. This study explored only a relatively small space of latent variable model architectures. Future work testing alternative models under the normative framework will provide further insight into the precise nature of state-dependent modulation and its impact on sensory coding and behavior.

5.3 State-dependent changes in correlated population activity as a window into behavior-relevant sensory coding

In chapter four, we investigate the impact of task-engagement on tone vs. noise discrimination by A1 populations. Studies of serially recorded, single neurons have repeatedly shown that neural discriminability is improved during active task-engagement relative to passive listening.[3,7,11,17,122] These changes are rapid and reversible,[3] and are thought to provide behavioral flexibility. Yet, changes in neural selectivity underlying these discrimination improvements have been shown to persist for a short time after task engagement, and their direct relationship to behavioral performance has not been demonstrated. Therefore, questions remain regarding the link between state-dependent neural discriminability and behavior. In one study, it was found that suppression of trial-to-trial response variability, not sensory selectivity, more closely followed behavioral performance and predicted changes in neural discriminability,[123] though this doesn't always seem to be the case.[138]

A separate line of work has demonstrated that correlated variability between neurons is significantly suppressed during task engagement,[73,74] raising the question of whether and how correlated noise is related to sound discriminability and behavior. To address this, we used multi-channel recordings from primary auditory cortex during behavior. In line with previous results, we found that single neuron sound-evoked responses were enhanced by task engagement and correlated variability was suppressed. Leveraging our method for dimensionality reduction, dDR, we measured discriminability of task cues by A1 populations. Although task engagement improved decoding accuracy, in agreement with previous work, we demonstrated that the suppression of correlated variability played no role and that the improvement in discriminability could not predict animal's behavior performance. Thus, these results indicate that sound information is decoded suboptimally from A1.

Unlike stimulus discrimination, we discovered that the changes in correlated variability were tightly coupled to animal's task performance. Theoretical work has demonstrated that when correlated variability is suppressed along coding dimensions, accuracy of the population code is improved. Thus, one possible explanation for our results is that the change in correlated variability is oriented along a suboptimal decoding axis which guides animal's behavioral choice. Although we were not able to test this hypothesis with our data, recent work suggest that, as least in macaque visual area V4, this is the case.[52] Additionally, the fact that the relationship between correlated variability and task performance was consistent across all three behavioral tasks we tested suggests that it may be a general processing mechanism employed by animals when making perceptual judgements. Thus, the structure of changes in correlated variability may offer a window into how information in a given brain area is read out by downstream regions.

5.4   State-dependent modulation across cortical layers and celltypes

Previous work exploring state-dependent modulation of sensory coding has shown that the effects of behavioral state vary between celltypes and cortical layers. For example, work in macaque visual cortex has shown that attention-dependent modulation of response gain and correlated variability predominate in supragranular layers.[81] In mouse visual cortex the impact of locomotion and arousal differs between fast-spiking putative interneurons and regular spiking pyramidal neurons.[27] More generally, GABA-ergic interneurons are reported to be the targets of neuromodulation and mediators of state-dependent gain changes across cortical areas.[139] In auditory cortex specifically, optogentic manipulation of PV+ interneurons can modulate tone responses and behavior[140] and models of AC circuits suggest that modulation of inhibition can affect network level properties of neural activity.[140] Thus, even from these few examples it is clear that not all cells are equal in their role for state-dependent modulation of sensory coding. Unfortunately, in the data we present in this dissertation, we were unable to reliably distinguish between celltypes and laminar location. Future work that incorporates this information through the use of new genetic and viral tools coupled with population recordings will be critical for advancing our understanding of cortical circuits and state-dependent modulation.

References

[1] Blake E. Butler and Stephen G. Lomber. Functional and structural changes throughout the auditory system following congenital and early-onset deafness: implications for hearing restoration. *Frontiers in Systems Neuroscience*, 7, 2013. Publisher: Frontiers.

[2] Tom C. T. Yin. Neural Mechanisms of Encoding Binaural Localization Cues in the Auditory Brainstem. In Donata Oertel, Richard R. Fay, and Arthur N. Popper, editors, *Integrative Functions in the Mammalian Auditory Pathway*, Springer Handbook of Auditory Research, pages 99–159. Springer, New York, NY, 2002.

[3] Jonathan Fritz, Shihab Shamma, Mounya Elhilali, and David Klein. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6(11):1216–1223, November 2003. Number: 11 Publisher: Nature Publishing Group.

[4] Matthew J. McGinley, Stephen V. David, and David A. McCormick. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, 87(1):179–192, July 2015.

[5] Jennifer K. Bizley and Andrew J. King. Visual influences on ferret auditory cortex. *Hearing Research*, 258(1):55–63, December 2009.

[6] David M. Schneider, Anders Nelson, and Richard Mooney. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–194, September 2014.

[7] S. V. David, J. B. Fritz, and S. A. Shamma. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences*, 109(6):2144–2149, February 2012.

[8] Jennifer K. Bizley, Kerry M.M. Walker, Fernando R. Nodal, Andrew J. King, and Jan W.H. Schnupp. Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. *Current Biology*, 23(7):620–625, April 2013.

[9] Victoria M Bajo, Fernando R Nodal, David R Moore, and Andrew J King. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature neuroscience*, 13(2):253–260, February 2010.

[10] Jennifer K. Bizley, Fernando R. Nodal, Israel Nelken, and Andrew J. King. Functional organization of ferret auditory cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 15(10):1637–1653, October 2005.

[11] Serin Atiani, Stephen V. David, Diego Elgueda, Michael Locastro, Susanne Radtke-Schuller, Shihab A. Shamma, and Jonathan B. Fritz. Emergent Selectivity for Task-Relevant Stimuli in Higher-Order Auditory Cortex. *Neuron*, 82(2):486–499, April 2014.

[12] Sharba Bandyopadhyay, Shihab A. Shamma, and Patrick O. Kanold. Dichotomy of functional organization in the mouse auditory cortex. *Nature Neuroscience*, 13(3):361–368, March 2010.

[13] Troy A. Hackett, Tania Rinaldi Barkat, Barbara M. J. O'Brien, Takao K. Hensch, and Daniel B. Polley. Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(8):2983–2995, February 2011.

[14] Jennifer K. Bizley and Yale E. Cohen. The what, where and how of auditory-object perception. *Nature Reviews. Neuroscience*, 14(10):693–707, October 2013.

[15] Stephen V. David, Nima Mesgarani, and Shihab A. Shamma. Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems*, 18(3):191–212, January 2007.

[16] Ivar L. Thorson, Jean Liénard, and Stephen V. David. The Essential Complexity of Auditory Receptive Fields. *PLOS Computational Biology*, 11(12):e1004628, December 2015.

[17] Kishore V. Kuchibhotla, Jonathan V. Gill, Grace W. Lindsay, Eleni S. Papadoyannis, Rachel E. Field, Tom A. Hindmarsh Sten, Kenneth D. Miller, and Robert C. Froemke. Parallel processing by cortical inhibition enables context-dependent behavior. *Nature Neuroscience*, 20(1):62–71, January 2017. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Auditory system;Sensory processing Subject_term_id: auditory-system;sensory-processing.

[18] S. J. Slee and S. V. David. Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in the Auditory Midbrain. *Journal of Neuroscience*, 35(38):13090–13102, September 2015.

[19] Norman M. Weinberger. Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, 5(4):279–290, April 2004. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Reviews Publisher: Nature Publishing Group.

[20] Richard G. Rutkowski and Norman M. Weinberger. Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. *Proceedings of the National Academy of Sciences*, 102(38):13664–13669, September 2005.

[21] Jacques Mirenowicz and Wolfram Schultz. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, 379(6564):449–451, February 1996. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6564 Primary_atype: Research Publisher: Nature Publishing Group.

[22] Masayuki Matsumoto and Okihide Hikosaka. Representation of negative motivational value in the primate lateral habenula. *Nature Neuroscience*, 12(1):77–84, January 2009.

[23] Shih-Chieh Lin and Miguel A. L. Nicolelis. Neuronal Ensemble Bursting in the Basal Forebrain Encodes Salience Irrespective of Valence. *Neuron*, 59(1):138–149, July 2008.

[24] Michael P. Kilgard and Michael M. Merzenich. Cortical Map Reorganization Enabled by Nucleus Basalis Activity. *Science*, 279(5357):1714–1718, March 1998. Publisher: American Association for the Advancement of Science Section: Report.

[25] Ana Raquel O. Martins and Robert C. Froemke. Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. *Nature Neuroscience*, 18(10):1483–1492, October 2015. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 10 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Olfactory bulb;Sensory processing Subject_term_id: olfactory-bulb;sensory-processing.

[26] Marlene R. Cohen and John H. R. Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12):1594–1600, December 2009.

[27] Martin Vinck, Renata Batista-Brito, Ulf Knoblich, and Jessica A. Cardin. Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*, 86(3):740–754, May 2015.

[28] Zachary P. Schwartz, Brad N. Buran, and Stephen V. David. Pupil-associated states modulate excitability but not stimulus selectivity in primary auditory cortex. *Journal of Neurophysiology*, 123(1):191–208, November 2019. Publisher: American Physiological Society.

[29] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D. Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437):eaav7893, April 2019.

[30] Matthew J. McGinley, Martin Vinck, Jacob Reimer, Renata Batista-Brito, Edward Zagha, Cathryn R. Cadwell, Andreas S. Tolias, Jessica A. Cardin, and David A. McCormick. Waking

State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron*, 87(6):1143–1161, September 2015.

[31] Jacob Reimer, Emmanouil Froudarakis, Cathryn R. Cadwell, Dimitri Yatsenko, George H. Denfield, and Andreas S. Tolias. Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. *Neuron*, 84(2):355–362, October 2014.

[32] Jacob Reimer, Matthew J. McGinley, Yang Liu, Charles Rodenkirch, Qi Wang, David A. McCormick, and Andreas S. Tolias. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, 7:13289, 2016.

[33] Daniela Saderi, Zachary P Schwartz, Charles R Heller, Jacob R Pennington, and Stephen V David. Dissociation of task engagement and arousal effects in auditory cortex and midbrain. *eLife*, 10:e60153, February 2021. Publisher: eLife Sciences Publications, Ltd.

[34] Adriana A. Zekveld, Thomas Koelewijn, and Sophia E. Kramer. The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, 22, September 2018.

[35] Matthew B. Winn, Jan R. Edwards, and Ruth Y. Litovsky. The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and hearing*, 36(4):e153–e165, 2015.

[36] Matthew B. Winn, Dorothea Wendt, Thomas Koelewijn, and Stefanie E. Kuchinsky. Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, 22:2331216518800869, January 2018. Publisher: SAGE Publications Inc.

[37] Ajay B. Satpute, Philip A. Kragel, Lisa Feldman Barrett, Tor D. Wager, and Marta Bianciardi. Deconstructing arousal into wakeful, autonomic and affective varieties. *Neuroscience Letters*, 693:19–28, February 2019.

[38] Simon Musall, Matthew T. Kaufman, Ashley L. Juavinett, Steven Gluf, and Anne K. Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, 22(10):1677–1686, October 2019.

[39] Leah Phillips Schwartz. *Effects of attention and arousal on neural activity in auditory cortex*. PhD thesis, :unav, 2020. Version Number: 4.0.

[40] Jiangang Du, Timothy J. Blanche, Reid R. Harrison, Henry A. Lester, and Sotiris C. Masmanidis. Multiplexed, High Density Electrophysiology with Nanofabricated Neural Probes. *PLOS ONE*, 6(10):e26204, October 2011.

[41] Eric M. Trautmann, Sergey D. Stavisky, Subhaneil Lahiri, Katherine C. Ames, Matthew T. Kaufman, Daniel J. O'Shea, Saurabh Vyas, Xulu Sun, Stephen I. Ryu, Surya Ganguli, and Krishna V. Shenoy. Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron*, 103(2):292–308.e4, July 2019.

[42] Long Yang, Kwang Lee, Jomar Villagracia, and Sotiris C Masmanidis. Open source silicon microprobes for high throughput neural recording. *Journal of Neural Engineering*, 17(1):016036, January 2020.

[43] Carsen Stringer, Michalis Michaelos, and Marius Pachitariu. High precision coding in visual cortex. *bioRxiv*, page 679324, November 2019.

[44] Ziqiang Wei, Bei-Jung Lin, Tsai-Wen Chen, Kayvon Daie, Karel Svoboda, and Shaul Druckmann. A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLOS Computational Biology*, 16(9):e1008198, September 2020. Publisher: Public Library of Science.

[45] Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819, July 2011.

[46] Ehud Zohary, Michael N. Shadlen, and William T. Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485):140–143, July 1994.

[47] Adam Kohn and Matthew A. Smith. Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. *Journal of Neuroscience*, 25(14):3661–3673, April 2005.

[48] Wyeth Bair, Ehud Zohary, and William T. Newsome. Correlated Firing in Macaque Visual Area MT: Time Scales and Relationship to Behavior. *Journal of Neuroscience*, 21(5):1676–1697, March 2001.

[49] Marlene R. Cohen and William T. Newsome. Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron*, 60(1):162–173, October 2008. Publisher: Elsevier.

[50] Neil C Rabinowitz, Robbe L Goris, Marlene Cohen, and Eero P Simoncelli. Attention stabilizes the shared gain of V4 populations. *eLife*, 4:e08998, November 2015.

[51] Alexander S. Ecker, Philipp Berens, R. James Cotton, Manivannan Subramaniyan, George H. Denfield, Cathryn R. Cadwell, Stelios M. Smirnakis, Matthias Bethge, and Andreas S. Tolias. State Dependence of Noise Correlations in Macaque Primary Visual Cortex. *Neuron*, 82(1):235–248, April 2014.

[52] A. M. Ni, D. A. Ruff, J. J. Alberts, J. Symmonds, and M. R. Cohen. Learning and attention reveal a general relationship between population activity and behavior. *Science*, 359(6374):463–465, January 2018.

[53] Robbe L T Goris, J Anthony Movshon, and Eero P Simoncelli. Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–865, June 2014.

[54] Benjamin R. Cowley, Adam C. Snyder, Katerina Acar, Ryan C. Williamson, Byron M. Yu, and Matthew A. Smith. Slow Drift of Neural Activity as a Signature of Impulsivity in Macaque Visual and Prefrontal Cortex. *Neuron*, 108(3):551–567.e8, November 2020. Publisher: Elsevier.

[55] K. H. Britten, M. N. Shadlen, W. T. Newsome, and J. A. Movshon. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 12(12):4745–4765, December 1992.

[56] Michael N. Shadlen and William T. Newsome. The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation, and Information Coding. *Journal of Neuroscience*, 18(10):3870–3896, May 1998.

[57] L. F. Abbott and Peter Dayan. The Effect of Correlated Variability on the Accuracy of a Population Code. *Neural Computation*, 11(1):91–101, January 1999.

[58] Peter Dayan and L. F. Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Computational neuroscience. Massachusetts Institute of Technology Press, Cambridge, Mass, 2001.

[59] Bruno B. Averbeck and Daeyeol Lee. Effects of noise correlations on information encoding and decoding. *Journal of Neurophysiology*, 95(6):3633–3644, June 2006.

[60] Bruno B. Averbeck, Peter E. Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, May 2006.

[61] Xaq Pitkow, Sheng Liu, Dora E. Angelaki, Gregory C. DeAngelis, and Alexandre Pouget. How Can Single Sensory Neurons Predict Behavior? *Neuron*, 87(2):411–423, July 2015.

[62] Rubén Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, and Alexandre Pouget. Information-limiting correlations. *Nature Neuroscience*, 17(10):1410–1417, October 2014.

[63] Joel Zylberberg, Jon Cafaro, Maxwell H. Turner, Eric Shea-Brown, and Fred Rieke. Direction-Selective Circuits Shape Noise to Ensure a Precise Population Code. *Neuron*, 89(2):369–383, January 2016.

[64] Joel Zylberberg. The role of untuned neurons in sensory information coding. *bioRxiv*, 2018.

[65] Alexander S. Ecker, Philipp Berens, Georgios A. Keliris, Matthias Bethge, Nikos K. Logothetis, and Andreas S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science (New York, N.Y.)*, 327(5965):584–587, January 2010.

[66] Felix Franke, Michele Fiscella, Maksim Sevelev, Botond Roska, Andreas Hierlemann, and Rava Azeredo da Silveira. Structures of Neural Correlation and How They Favor Coding. *Neuron*, 89(2):409–422, January 2016. Publisher: Elsevier.

[67] Kiersten Ruda, Joel Zylberberg, and Greg D. Field. Ignoring correlated activity causes a failure of retinal population codes under moonlight conditions. *bioRxiv*, page 2019.12.18.881201, December 2019.

[68] Oleg I. Rumyantsev, Jérôme A. Lecoq, Oscar Hernandez, Yanping Zhang, Joan Savall, Radosław Chrapkiewicz, Jane Li, Hongkui Zeng, Surya Ganguli, and Mark J. Schnitzer. Fundamental bounds on the fidelity of sensory cortical coding. *Nature*, March 2020.

[69] Ramon Bartolo, Richard C. Saunders, Andrew R. Mitz, and Bruno B. Averbeck. Information limiting correlations in large neural populations. *The Journal of Neuroscience*, pages 2072–19, January 2020.

[70] MohammadMehdi Kafashan, Anna Jaffe, Selmaan N. Chettih, Ramon Nogueira, Iñigo Arandia-Romero, Christopher D. Harvey, Rubén Moreno-Bote, and Jan Drugowitsch. Scaling of information in large neural populations reveals signatures of information-limiting correlations. *bioRxiv*, page 2020.01.10.902171, January 2020.

[71] Pei-Ann Lin, Samuel K. Asinof, Nicholas J. Edwards, and Jeffry S. Isaacson. Arousal regulates frequency tuning in primary auditory cortex. *Proceedings of the National Academy of Sciences*, November 2019.

[72] Dmitry Kobak, Jose L Pardo-Vazquez, Mafalda Valente, Christian K Machens, and Alfonso Renart. State-dependent geometry of population activity in rat auditory cortex. *eLife*, 8:e44526, April 2019.

[73] J. D. Downer, M. Niwa, and M. L. Sutter. Task Engagement Selectively Modulates Neural Correlations in Primary Auditory Cortex. *Journal of Neuroscience*, 35(19):7565–7574, May 2015.

[74] Joshua D. Downer, Brittany Rapone, Jessica Verhein, Kevin N. O'Connor, and Mitchell L. Sutter. Feature-Selective Attention Adaptively Shifts Noise Correlations in Primary Auditory Cortex. *The Journal of Neuroscience*, 37(21):5378–5392, May 2017.

[75] Adam Kohn, Ruben Coen-Cagli, Ingmar Kanitscheider, and Alexandre Pouget. Correlations and Neuronal Population Information. *Annual Review of Neuroscience*, 39(1):237–256, July 2016.

[76] Stefano Panzeri, Christopher D. Harvey, Eugenio Piasini, Peter E. Latham, and Tommaso Fellin. Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. *Neuron*, 93(3):491–507, February 2017.

[77] Carsen Stringer, Michalis Michaelos, Dmitri Tsyboulski, Sarah E. Lindo, and Marius Pachitariu. High-precision coding in visual cortex. *Cell*, 184(10):2767–2778.e15, May 2021.

[78] João D. Semedo, Amin Zandvakili, Christian K. Machens, Byron M. Yu, and Adam Kohn. Cortical Areas Interact through a Communication Subspace. *Neuron*, 102(1):249–259.e4, April 2019.

[79] Alexander S. Ecker and Andreas S. Tolias. Is there signal in the noise? *Nature Neuroscience*, 17(6):750–751, June 2014.

[80] Carrie J. McAdams and John H. R. Maunsell. Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *Journal of Neuroscience*, 19(1):431–441, January 1999. Publisher: Society for Neuroscience Section: ARTICLE.

[81] George H. Denfield, Alexander S. Ecker, Tori J. Shinn, Matthias Bethge, and Andreas S. Tolias. Attentional fluctuations induce shared variability in macaque primary visual cortex. *Nature Communications*, 9(1), December 2018.

[82] Ingmar Kanitscheider, Ruben Coen-Cagli, Adam Kohn, and Alexandre Pouget. Measuring Fisher Information Accurately in Correlated Neural Populations. *PLOS Computational Biology*, 11(6):e1004218, June 2015. Publisher: Public Library of Science.

[83] Ian H. Stevenson and Konrad P. Kording. How advances in neural recording affect data analysis. *Nature Neuroscience*, 14(2):139–142, February 2011. Number: 2 Publisher: Nature Publishing Group.

[84] David M. Green and John A. Swets. *Signal detection theory and psychophysics.* Signal detection theory and psychophysics. John Wiley, Oxford, England, 1966. Pages: xi, 455.

[85] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, November 2014.

[86] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D. Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571(7765):361–365, July 2019.

[87] Charles R. Heller, Zachary P. Schwartz, Daniela Saderi, and Stephen V. David. Selective effects of arousal on population coding of natural sounds in auditory cortex. *bioRxiv*, page 2020.08.31.276584, December 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[88] Jacob Pennington and Stephen David. Complementary effects of adaptation and gain control on sound encoding in primary auditory cortex. preprint, Neuroscience, January 2020.

[89] Martina Valente, Giuseppe Pica, Caroline A. Runyan, Ari S. Morcos, Christopher D. Harvey, and Stefano Panzeri. Correlations enhance the behavioral readout of neural population activity in association cortex. preprint, Neuroscience, April 2020.

[90] Douglas A. Ruff and Marlene R. Cohen. Stimulus Dependence of Correlated Variability across Cortical Areas. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(28):7546–7556, July 2016.

[91] Dmitry Kobak, Wieland Brendel, Christos Constantinidis, Claudia E Feierstein, Adam Kepecs, Zachary F Mainen, Xue-Lian Qi, Ranulfo Romo, Naoshige Uchida, and Christian K Machens. Demixed principal component analysis of neural population data. *eLife*, 5:e10989, April 2016. Publisher: eLife Sciences Publications, Ltd.

[92] Matthew R. Whiteway, Bruno Averbeck, and Daniel A. Butts. A latent variable approach to decoding neural population activity. *bioRxiv*, page 2020.01.06.896423, January 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[93] Byron M. Yu, John P. Cunningham, Gopal Santhanam, Stephen I. Ryu, Krishna V. Shenoy, and Maneesh Sahani. Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *Journal of Neurophysiology*, 102(1):614–635, July 2009.

[94] Joshua H Siegle, Aarón Cuevas López, Yogi A Patel, Kirill Abramov, Shay Ohayon, and Jakob Voigts. Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. *Journal of Neural Engineering*, 14(4):045003, August 2017.

[95] Nima Mesgarani and Edward F. Chang. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236, May 2012. Number: 7397 Publisher: Nature Publishing Group.

[96] Daniela Saderi, Zachary P. Schwartz, Charlie R. Heller, Jacob R. Pennington, and Stephen V. David. Dissociation of task engagement and arousal effects in auditory cortex and midbrain. *bioRxiv*, page 2020.06.16.155432, June 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[97] A. Ryan and J. Miller. Effects of behavioral performance on single-unit firing patterns in inferior colliculus of the rhesus monkey. *Journal of Neurophysiology*, 40(4):943–956, July 1977.

[98] Gonzalo H. Otazu, Lung-Hao Tai, Yang Yang, and Anthony M. Zador. Engaging in an auditory task suppresses responses in auditory cortex. *Nature Neuroscience*, 12(5):646–654, May 2009.

[99] Chen-Chung Lee and John C. Middlebrooks. Auditory cortex spatial sensitivity sharpens during task performance. *Nature Neuroscience*, 14(1):108–114, January 2011.

[100] Daniel P. Knudsen and Timothy Q. Gentner. Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. *Journal of Neurophysiology*, 109(7):1690–1703, April 2013.

[101] Pingbo Yin, Jonathan B. Fritz, and Shihab A. Shamma. Rapid Spectrotemporal Plasticity in Primary Auditory Cortex during Behavior. *The Journal of Neuroscience*, 34(12):4396–4408, March 2014.

[102] Zachary P Schwartz and Stephen V David. Focal Suppression of Distractor Sounds by Selective Attention in Auditory Cortex. *Cerebral Cortex*, 28(1):323–339, January 2018.

[103] Alexander S. Ecker, George H. Denfield, Matthias Bethge, and Andreas S. Tolias. On the Structure of Neuronal Population Activity under Fluctuations in Attentional State. *Journal of Neuroscience*, 36(5):1775–1789, February 2016.

[104] Jalal K. Baruni, Brian Lau, and C. Daniel Salzman. Reward expectation differentially modulates attentional behavior and activity in visual area V4. *Nature Neuroscience*, 18(11):1656–1663, November 2015. Number: 11 Publisher: Nature Publishing Group.

[105] Alexander S. Ecker, Philipp Berens, Andreas S. Tolias, and Matthias Bethge. The Effect of Noise Correlations in Populations of Diversely Tuned Neurons. *Journal of Neuroscience*, 31(40):14272–14283, October 2011.

[106] Charles R. Heller and Stephen V. David. Dimensionality reduction for neural population decoding. *bioRxiv*, page 2021.04.18.440336, April 2021. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[107] Josh H. McDermott, Michael Schemitsch, and Eero P. Simoncelli. Summary statistics in auditory perception. *Nature Neuroscience*, 16(4):493–498, April 2013. Number: 4 Publisher: Nature Publishing Group.

[108] Tatiana A. Engel, Nicholas A. Steinmetz, Marc A. Gieselmann, Alexander Thiele, Tirin Moore, and Kwabena Boahen. Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144, December 2016.

[109] MohammadMehdi Kafashan, Anna W. Jaffe, Selmaan N. Chettih, Ramon Nogueira, Iñigo Arandia-Romero, Christopher D. Harvey, Rubén Moreno-Bote, and Jan Drugowitsch. Scaling of sensory information in large neural populations shows signatures of information-limiting correlations. *Nature Communications*, 12(1):473, January 2021. Number: 1 Publisher: Nature Publishing Group.

[110] P. L. E. van Kan, R. P. Scobey, and A. J. Gabor. Response covariance in cat visual cortex. *Experimental Brain Research*, 60(3):559–563, November 1985.

[111] M. N. Shadlen, K. H. Britten, W. T. Newsome, and J. A. Movshon. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16(4):1486–1510, February 1996.

[112] A. J. Parker and W. T. Newsome. SENSE AND THE SINGLE NEURON: Probing the Physiology of Perception. *Annual Review of Neuroscience*, 21(1):227–277, March 1998.

[113] Shi Tong Liu, Pilar Montes-Lourido, Xiaoqin Wang, and Srivatsun Sadagopan. Optimal features for auditory categorization. *Nature Communications*, 10(1):1302, March 2019. Number: 1 Publisher: Nature Publishing Group.

[114] Daniel E. Winkowski, Sharba Bandyopadhyay, Shihab A. Shamma, and Patrick O. Kanold. Frontal Cortex Activation Causes Rapid Plasticity of Auditory Cortical Processing. *Journal of Neuroscience*, 33(46):18134–18148, November 2013. Publisher: Society for Neuroscience Section: Articles.

[115] Daniela Saderi, Brad N. Buran, and Stephen V. David. Streaming of Repeated Noise in Primary and Secondary Fields of Auditory Cortex. *Journal of Neuroscience*, 40(19):3783–3798, May 2020. Publisher: Society for Neuroscience Section: Research Articles.

[116] Diego Elgueda, Daniel Duque, Susanne Radtke-Schuller, Pingbo Yin, Stephen V. David, Shihab A. Shamma, and Jonathan B. Fritz. State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. *Nature Neuroscience*, 22(3):447–459, March 2019.

[117] Marius Pachitariu, Nicholas Steinmetz, Shabnam Kadir, Matteo Carandini, and Harris Kenneth D. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, page 061481, June 2016.

[118] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. *arXiv:1608.06993 [cs]*, January 2018. arXiv: 1608.06993.

[119] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. ISSN: 1063-6919.

[120] M. Pachitariu, D. R. Lyamzin, M. Sahani, and N. A. Lesica. State-Dependent Population Coding in Primary Auditory Cortex. *Journal of Neuroscience*, 35(5):2058–2073, February 2015.

[121] Varun Saravanan, Gordon J. Berman, and Samuel J. Sober. Application of the hierarchical bootstrap to multi-level data in neuroscience. *arXiv:2007.07797 [q-bio]*, July 2020. arXiv: 2007.07797.

[122] Sophie Bagur, Martin Averseng, Diego Elgueda, Stephen David, Jonathan Fritz, Pingbo Yin, Shihab Shamma, Yves Boubenec, and Srdjan Ostojic. Go/No-Go task engagement enhances population representation of target stimuli in primary auditory cortex. *Nature Communications*, 9(1):1–16, June 2018.

[123] G. von Trapp, B. N. Buran, K. Sen, M. N. Semple, and D. H. Sanes. A Decline in Response Variability Improves Neural Signal Detection during Auditory Task Performance. *Journal of Neuroscience*, 36(43):11097–11106, October 2016.

[124] Jennifer Resnik and Daniel B. Polley. Cochlear neural degeneration disrupts hearing in background noise by increasing auditory cortex internal noise. *Neuron*, February 2021.

[125] Eric I. Knudsen. Fundamental Components of Attention. *Annual Review of Neuroscience*, 30(1):57–78, 2007. _eprint: https://doi.org/10.1146/annurev.neuro.30.051606.094256.

[126] Irving T. Diamond and W. D. Neff. ABLATION OF TEMPORAL CORTEX AND DISCRIMI-NATION OF AUDITORY PATTERNS. *Journal of Neurophysiology*, 20(3):300–315, May 1957.

[127] Tyler L. Gimenez, Maja Lorenc, and Santiago Jaramillo. Adaptive categorization of sound frequency does not require the auditory cortex in rats. *Journal of Neurophysiology*, 114(2):1137–1145, August 2015. Publisher: American Physiological Society.

[128] Ian A. Harrington, Rickye S. Heffner, and Henry E. Heffner. An investigation of sensory deficits underlying the aphasia-like behavior of macaques with auditory cortex lesions. *NeuroReport*, 12(6):1217–1221, May 2001.

[129] Santiago Jaramillo and Anthony M Zador. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience*, 14(2):246–251, February 2011.

[130] Frank W. Ohl, Wolfram Wetzel, Thomas Wagner, Alexander Rech, and Henning Scheich. Bilateral Ablation of Auditory Cortex in Mongolian Gerbil Affects Discrimination of Frequency Modulated Tones but not of Pure Tones. *Learning & Memory*, 6(4):347–362, July 1999. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[131] Shraddha S. Pai, Jeffrey C. Erlich, Charles Kopec, and Carlos D. Brody. Minimal Impairment in a Rat Model of Duration Discrimination Following Excitotoxic Lesions of Primary Auditory and Prefrontal Cortices. *Frontiers in Systems Neuroscience*, 5, 2011. Publisher: Frontiers.

[132] Natalia Rybalko, Daniel Šuta, Fidel Nwabueze-Ogbo, and Josef Syka. Effect of auditory cortex lesions on the discrimination of frequency-modulated tones in rats. *European Journal of Neuroscience*, 23(6):1614–1622, 2006. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-9568.2006.04688.x.

[133] Sanjiv K. Talwar and George L. Gerstein. Reorganization in Awake Rat Auditory Cortex by Local Microstimulation and Its Effect on Frequency-Discrimination Behavior. *Journal of Neurophysiology*, 86(4):1555–1572, October 2001. Publisher: American Physiological Society.

[134] Sebastian Ceballo, Zuzanna Piwkowska, Jacques Bourg, Aurélie Daret, and Brice Bathellier. Targeted Cortical Manipulation of Auditory Perception. *Neuron*, 104(6):1168–1179.e5, December 2019. Publisher: Elsevier.

[135] Pingbo Yin, Jonathan B. Fritz, and Shihab A. Shamma. Do ferrets perceive relative pitch? *The Journal of the Acoustical Society of America*, 127(3):1673–1680, March 2010.

[136] D. J. Klein, D. A. Depireux, J. Z. Simon, and S. A. Shamma. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *Journal of Computational Neuroscience*, 9(1):85–111, August 2000.

[137] Jack B. Kelly, Gerard L. Kavanagh, and James C. H. Dalton. Hearing in the ferret (Mustela putorius): Thresholds for pure tone detection. *Hearing Research*, 24(3):269–275, January 1986.

[138] Melissa L. Caras and Dan H. Sanes. Top-down modulation of sensory cortex gates perceptual learning. *Proceedings of the National Academy of Sciences*, 114(37):9972–9977, September 2017.

[139] Katie A. Ferguson and Jessica A. Cardin. Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, pages 1–13, January 2020.

[140] Jennifer M. Blackwell and Maria N. Geffen. Progress and challenges for understanding the function of cortical microcircuits in auditory processing. *Nature Communications*, 8(1):1–9, December 2017.