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

Daniela Saderi[1,2], Zachary P. Schwartz [1,2], Charlie R. Heller[1,2], Jacob R. Pennington[3], Stephen V. David[2]

[1]Neuroscience Graduate Program, [2]Oregon Hearing Research Center,

Oregon Health and Science University, [3]Washington State University, Vancouver

Correspondence: SVD, 3181 SW Sam Jackson Park Road, MC L335A, Portland, OR 97239. Email: davids@ohsu.edu

DS ORCID iD: https://orcid.org/0000-0002-6109-0367

ZPS ORCID iD: https://orcid.org/0000-0002-1854-5450

CRH ORCID iD: https://orcid.org/0000-0002-9048-1201

JRP ORCID iD: https://orcid.org/0000-0001-9393-0474

SVD ORCID iD: https://orcid.org/0000-0003-4135-3104

Keywords: auditory cortex, behavior, pupil, sound

Acknowledgements: This study was supported by NIH grants F31 DC014888 (DS), R01 DC0495 (SVD), and from the Tartar Trust at Oregon Health & Science University (DS).

Abstract

Both generalized arousal and engagement in a specific task influence sensory neural processing. To isolate effects of these state variables in the auditory system, we recorded single-unit activity from primary auditory cortex (A1) and inferior colliculus (IC) of ferrets during a tone detection task, while monitoring arousal via changes in pupil size. We used a generalized linear model to assess the influence of task engagement and pupil size on sound-evoked activity. In both areas, these two variables affected independent neural populations. Pupil size effects were more prominent in IC, while pupil and task engagement effects were equally likely in A1. Task engagement was correlated with larger pupil; thus, some apparent effects of task engagement should in fact be attributed to fluctuations in pupil size. These results indicate a hierarchy of auditory processing, where generalized arousal enhances activity in midbrain, and effects specific to task engagement become more prominent in cortex.

Introduction

Hearing is a dynamic process that requires integration of the sensory evidence provided by physical attributes of sound with information about the behavioral context in which auditory perception occurs (Bizley and Cohen, 2013; Fritz et al., 2007). As sound information passes through the brain, coordinated activity in auditory areas extracts information necessary to guide behavior. Compared to passive listening, engaging in a task that requires auditory discrimination leads to changes in neural excitability, as well as in spectral and spatial selectivity (Downer et al., 2015; Fritz et al., 2005, 2003; Knudsen and Gentner, 2013; Kuchibhotla et al., 2017; Lee and Middlebrooks, 2011; Otazu et al., 2009; Ryan and Miller, 1977; Yin et al., 2014). These changes are often attributed to the specific task demands, enhancing the representation of important sound features by selectively increasing or decreasing the activity across neural populations (Bagur et al., 2018; David et al., 2012; Kuchibhotla et al., 2017). In some studies, changes have been reported to persist for several minutes after the active behavior, while in others, activity and tuning changes were observed to rapidly return to baseline (Fritz et al., 2003; Slee and David, 2015).

The substantial variability of effects across studies suggests that task engagement activates a constellation of non-auditory signals that comprise a complex internal state. Task-related changes in the activity of auditory neurons have been attributed to temporal expectation (Jaramillo and Zador, 2011), reward associations (Beaton and Miller, 1975; David et al., 2012), self-generated sound (Eliades and Wang, 2008), and non-sound related variables, such as motor planning (Bizley et al., 2013; Huang et al., 2019), motor activity (Schneider et al., 2014; Zhou et al., 2014), degree of engagement (Carcea et al., 2017; Knyazeva et al., 2020), and behavioral choice (Tsunada et al., 2015). Even in the absence of explicit behavior, fluctuations in neural activity are observed throughout the forebrain, including primary sensory regions (Ringach, 2009; Stringer et al., 2019). This activity is related to cognitive states such as arousal and may also modulate sensitivity to sensory inputs (Fu et al., 2014; Wimmer et al., 2015). Therefore, attempting to interpret changes in neuronal activity solely in the light of a single binary

variable, such as being engaged in a task or not, will lead to an incomplete and even incorrect understanding of how sensory information is processed.

A straightforward way to monitor some task-independent changes in behavioral state is by measuring pupil size (Kahneman and Beatty, 1966; McGinley et al., 2015; Zekveld et al., 2018). In humans, luminance independence changes in pupil size have been shown to correlate with mental effort (Beatty, 1982; Winn et al., 2015), changes in states of arousal (Kahneman and Beatty, 1966), aspects of decision-making (Gilzenrat et al., 2010), and task performance (Schriver et al., 2018). In animals, fluctuations in pupil size closely track locomotion and neural activity throughout the mouse forebrain (McGinley et al., 2015; Stringer et al., 2019). Pupil size also tracks the degree of neural synchronization observed in local field potentials, which is commonly used to measure arousal (Schwartz et al., 2020; Vinck et al., 2015). Thus, while this single variable is unlikely to completely characterize such a complex phenomenon as arousal, pupil provides a window into a global brain state that is activated during demanding behaviors.

The degree to which pupil size co-varies with task engagement, and whether these variables account for independent changes in neural activity, remains to be determined. To study the interaction of task engagement and pupil size on auditory neural coding, we recorded extracellular spiking activity and pupil size of ferrets during an auditory task and during passive presentation of task stimuli. Single- and multiunit activity was recorded from the primary auditory cortex (A1) and midbrain inferior colliculus (IC). We used a generalized linear model to quantify the unique contributions of task engagement and pupil size to neural activity. Using this approach, we found that independent subpopulations of neurons were modulated uniquely by task or by pupil size, suggesting that these effects are mediated by distinct inputs. Our analysis also revealed that in many neurons, pupil size could account for variability in firing rate that might otherwise be attributed to task engagement. Thus, some previously reported effects of task engagement may in fact be explained by changes in pupil-indexed arousal. In addition, the relative effects of task engagement and pupil size varied between cortex and midbrain. Pupil-related changes occurred in both areas, but task engagement effects were more prominent in A1. This work highlights the value of

accounting for multiple state variables in studies of sensory coding (Musall et al., 2019; Stringer et al., 2019), both in identifying the source of task-related effects and locating where in the sensory processing hierarchy they emerge.

Results

Changes in pupil size track task engagement

To study interactions between pupil-indexed arousal and task engagement, we trained four adult ferrets on a go/no-go auditory detection task (David et al., 2012; Yin et al., 2010). Animals reported the presence of a target tone following a random number of broadband noise reference stimuli (Figure 1A, B). They were required to withhold licking a water spout during the reference sounds and were given liquid reward for licking during the target window (0.1-1.5 s after target onset). Targets were pure tones, either presented alone (tone versus noise discrimination, Ferret L) or embedded in reference stimuli (tone-in-noise detection, Ferrets R, B, and T). After training, all animals performed consistently above chance (Figure 1C).

To track and record changes in pupil size, an infrared video camera was used to image the eye contralateral to the speaker emitting auditory stimuli (Figure 1A). Constant luminance was maintained throughout the experiment so that fluctuations in pupil size reflected only changes in internal state (McGinley et al., 2015; Schwartz et al., 2020). Pupil was recorded continuously as animals passively listened to the task stimuli and actively performed the task. The distribution of pupil size differed between passive and active conditions (Figure 1-figure supplement 1A). Mean pupil size (measured by the major axis diameter) was consistently larger during active behavior ($p = 1.6 \times 10^{-6}$, Wilcoxon signed-rank test, Figure 1D). Within active blocks, pupil size also varied with task performance. Average pupil size was similar for false alarm and hit trials, but it was smaller during miss trials, more closely resembling passive trials (Figure 1-figure supplement 1B). In addition, the short-term dynamics of pupil depended strongly on task condition. During behavior, trial onset evoked a rapid increase in pupil size, which was absent during passive listening ($p = 5.6 \times 10^{-8}$, Wilcoxon signed-rank test, Figure 1E-F). Thus, pupil size tracked slow changes in task-engagement (active versus passive blocks) as well as more rapid changes in trial-by-trial performance during the active condition.

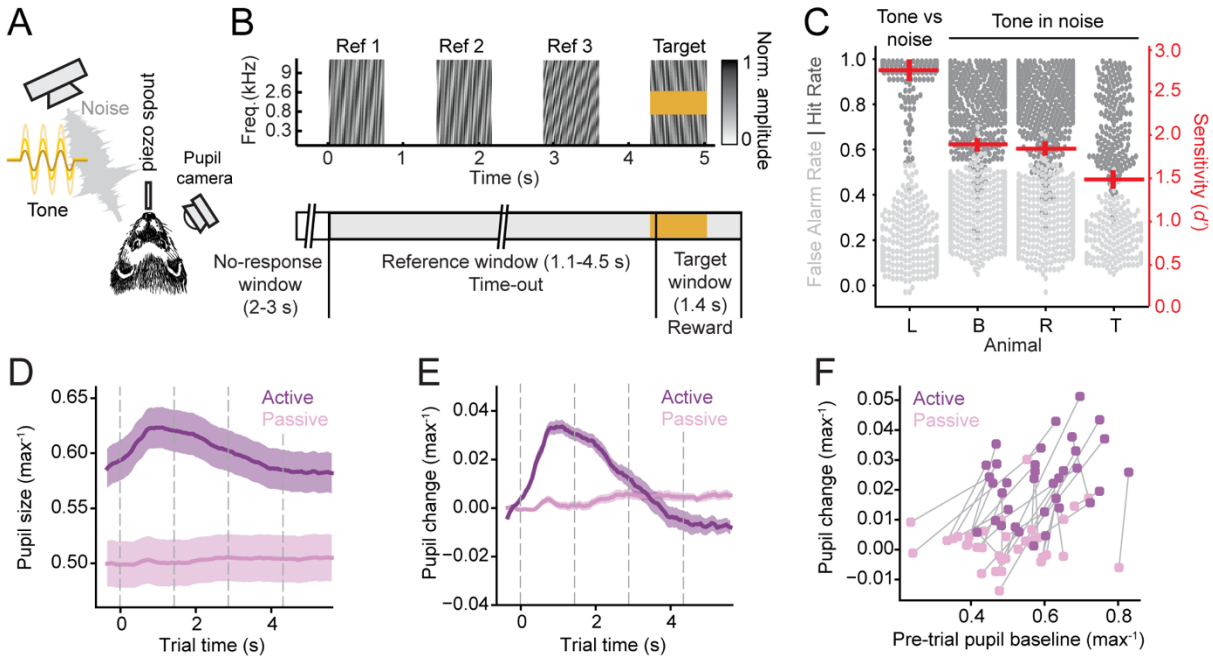


Figure 1: Pupil size correlates with task engagement. **A.** Schematic of behavioral setup, including free-field speaker for sound presentation, piezo spout to register licks and deliver reward, and infrared video camera for pupillometry**. B.** Spectrogram of example trial (top) and task structure (bottom). False alarms before the target window were punished with a timeout and hits resulted in a liquid reward. **C.** Swarm plot showing behavioral performance (hit rate or false alarm rate) for each animal. Each point corresponds to a behavioral block. Red lines indicate mean and standard error of the mean (SEM) of the sensitivity (*d*') for each animal. All animals performed above chance on average (*d*' > 1). **D.** Mean timecourse of pupil size aligned to trial onset for active and passive blocks, averaged across all animals and sessions and normalized to maximum size per session. Shading indicates SEM. Grey dashed lines indicate sound onset. **E.** Average timecourse of change in pupil size, normalized to 0.35-s pre-trial period for active and passive behavioral blocks. Shading indicates SEM. Grey dashed lines indicate sound onset. **F.** Scatter plot compares pre-trial pupil size and mean change per trial (3-s post-onset window). Each dot represents a behavioral block. Passive and active blocks within the same session are connected by a line.

Distinct task engagement and pupil size effects in A1 and IC neurons

We recorded extracellular spiking activity from the primary auditory cortex (A1, n = 129 units) and the inferior colliculus (IC, n = 66 units) of ferrets, while they switched between active engagement and passive listening. In the IC, we recorded from central (ICC, n = 18 units) and non-central regions (NCIC, n = 48 units), distinguished by their functional topography (Slee and David, 2015). Results from the IC are presented jointly between the two sub-regions, unless otherwise specified.

To study how behavior state modulated auditory neural activity over time, we recorded activity in the same units over multiple passive and active blocks and recorded pupil size throughout each

experiment as a measure of the animal's state of arousal. For analysis, the peri-stimulus time histogram (PSTH) response to the reference stimuli was averaged across passive and correct (hit) active trials, and data from target periods were excluded. On hit trials, animals did not lick during the reference stimuli; thus this approach reduced the possibility that lick-related motor signals confounded our results (Brosch et al., 2005; Schneider et al., 2014; Stringer et al., 2019).

For many units, activity appeared to be modulated by changes in task engagement, as would be expected from previous behavioral studies (Fritz et al., 2003; Kuchibhotla et al., 2017; Niwa et al., 2012; Otazu et al., 2009). During active behavior, responses to the noise stimuli either increased or decreased, and then returned to their baseline passive state during subsequent passive periods (Figure 2A-B). For other units, however, no consistent change in activity was observed between active and passive conditions. In these cases, firing rate could change between behavioral blocks, but the effect was not consistent across engagement conditions (Figure 2D-E). Thus, despite our controls for sound acoustics and motor activity, the firing rates of some units varied over time in a way that did not appear to be related to task engagement.

We wondered if changes in activity that could not be attributed to task engagement could instead be explained by fluctuations in pupil-indexed arousal (Lin et al., 2019; McGinley et al., 2015; Schwartz et al., 2020). A simple way to investigate if pupil size predicts changes in neural firing rate is to divide all trials (pooled across active and passive) into two groups, based on the mean pupil size, and compute a separate PSTH for each group (Figure 2C, F). In the example A1 units, activity was often enhanced when pupil was larger than median (large versus small pupil PSTH, Figure 2C, F), indicating that pupil size was positively correlated with firing rate.

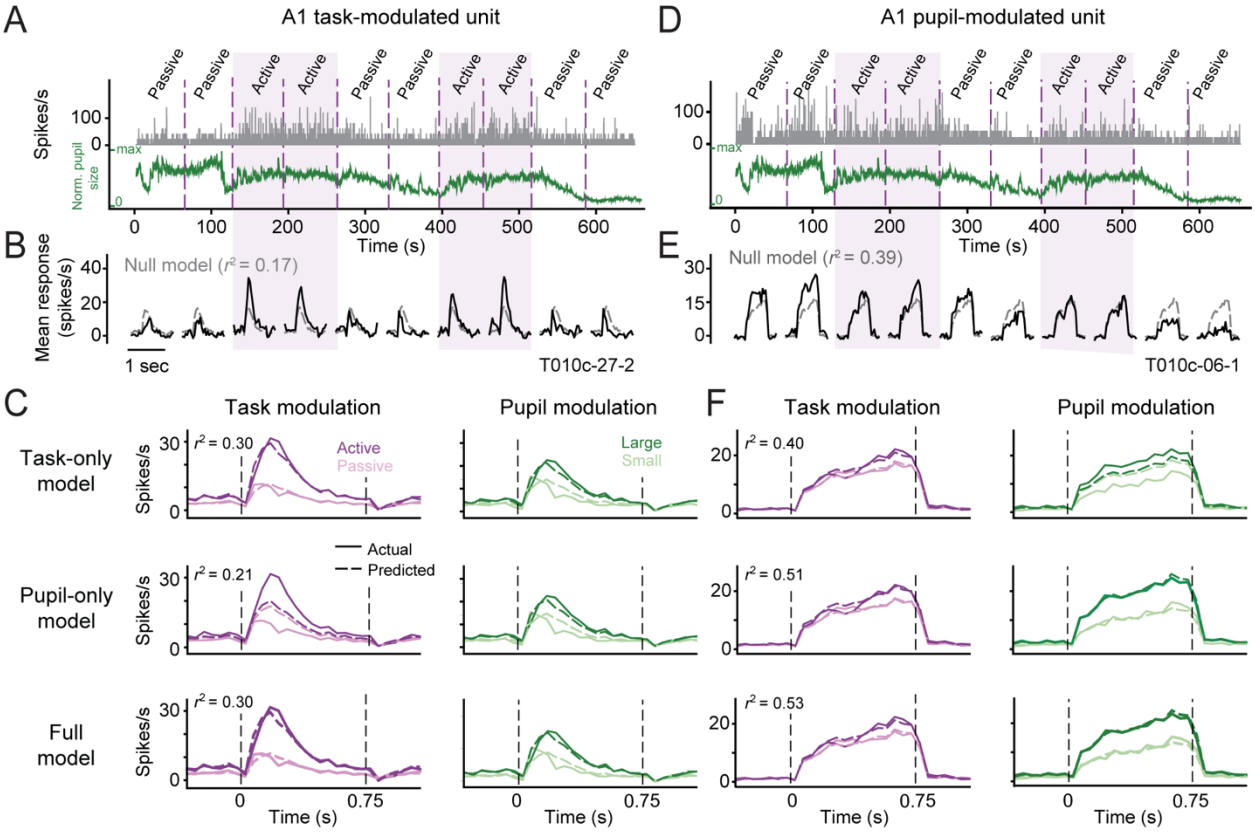


Figure 2: Simultaneously recorded units showing task versus pupil-related changes. **A.** Firing rate (grey) of one A1 unit during the reference period and time-aligned pupil size (green) over the duration of one experiment (total recording time ~1h, inter-trial and target periods not shown). Purple shading highlights active blocks. Dashed lines delineate boundaries between half of each behavioral block. **B.** Peristimulus time histogram (PSTH) responses to reference noise averaged separately across the first and second half of each behavior block (black). Dashed gray lines are responses predicted by the null model, *i.e.*, the PSTH averaged across all blocks. Model prediction accuracy is indicated by fraction variance explained (r^2). **C.** Top: PSTH response for the example unit averaged across passive and active blocks (left, dark and light purple, solid lines) and averaged across large- and small pupil size trials (right, dark and light green, solid lines). Predictions of the task-only model are overlaid (same color pattern, dashed lines). Middle: Active/passive and large/small PSTH responses plotted with predictions by the pupil-only model, which only accounts for pupil-related changes. Bottom: Active/passive and large/small PSTH responses plotted with predictions by the full model, accounting for both task and pupil-related changes. **D.** Firing rate and pupillometry for a second A1 unit, recorded simultaneously to the first example. **E-F.** PSTH responses for the second unit, plotted as in B-C. Example units from IC appear in Figure 8.

Because changes in pupil size were correlated with transitions between active and passive blocks (Figure 1E), we could not dissociate the effects of task engagement and pupil size using just the raw PSTH. Therefore, to test our hypothesis that pupil size accounted for changes in neural activity following task engagement, we fit a generalized linear model (*Eqn. 1*) in which the response to each presentation of

the reference stimuli was modulated by task engagement and pupil size. Each of these variables could additively modulate the baseline (mean) firing rate or multiplicatively modulate response gain. Task engagement was modeled as a discrete regressor (active or passive) and pupil size as a continuous regressor. This approach allowed us to dissociate effects of pupil size and task engagement on firing rate.

Model performance was measured by the fraction variance explained (squared correlation coefficient, r^2) between the predicted and actual single-trial spike rates, using nested cross validation. To test for significant contributions of each regressor, we introduced these variables stepwise (Fritz et al., 2010; Musall et al., 2019). The full model included both task engagement and pupil size variables. We compared performance of the full model to the performance of a state-independent model, for which both task and pupil variables were shuffled in time (null model), and of two partial models, which accounted for effects of a single state variable while shuffling the values of the other variable in time (task- or pupil partial models).

The regression analysis revealed substantial variability in effects of task engagement and pupil size on auditory responses. For one example A1 unit, the modulation of firing rate between active and passive blocks was almost completely accounted for by the task variable, with little contribution of pupil size (Figure 2A-C). When we incorporated the task variable, predictions significantly improved with respect to the state-independent null model (null model, $r^2 = 0.12$; task partial model, $r^2 = 0.30$; jackknifed *t*-test, $p < 0.05$; Figure 2C, top). Pupil size was also able to account for some changes in firing rate (pupil partial model, $r^2 = 0.21$; Figure 2C, middle), but the full model incorporating both task engagement and pupil size showed no additional improvement over the task partial model ($r^2 = 0.30$, $p > 0.05$; Figure 2C, bottom). Thus, all behavior-dependent changes for this unit were accounted for by task engagement. The apparent pupil-related effects were in fact explained by correlations between pupil size and task engagement.

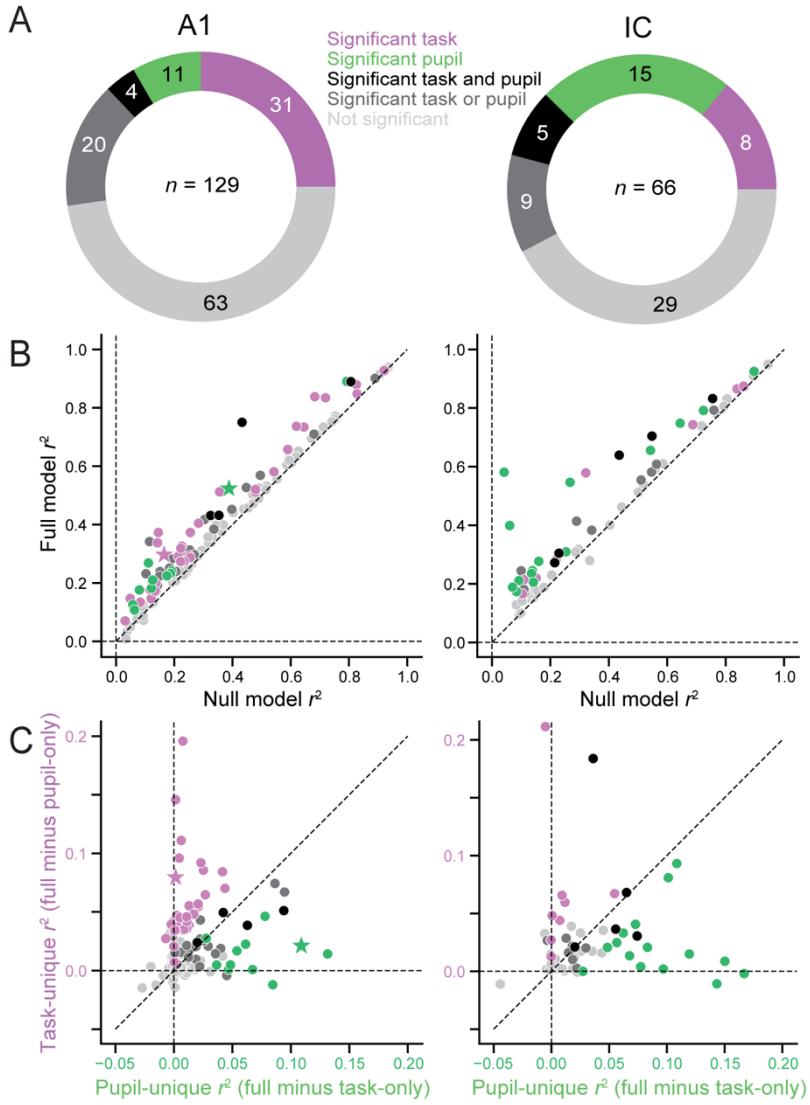


Figure 3: Task and pupil-related modulation of firing rates in A1 and IC. **A.** Doughnut plots indicate number of units significantly modulated by task engagement, pupil size, or both in A1 (left) and IC (right, $p < 0.05$, jackknifed t-test). Total number of recorded units reported in the center. Purple and green: significant unique modulation by task or pupil, respectively. Black: unique modulation by both task and pupil. Dark grey: ambiguous task or pupil modulation. Light gray: no significant change in accuracy between full and null models. **B.** Scatter plot of variance explained (r^2) in single-trial activity by full model versus null model. Each symbol represents a unit in AC (left) or IC (right). Colors as in A. **C.** Unique variance explained by pupil size (horizontal axis) plotted against unique variance explained by task engagement (vertical axis) for each unit. Stars correspond to examples in Figure 2A (purple) and D (green).

Conversely, for other units, changes in firing rate were uniquely accounted for by pupil size. For a second example A1 unit (Figure 2D-F), the task partial model ($r^2 = 0.39$) indicated a small improvement over the null model ($r^2 = 0.40$, $p > 0.05$, jackknifed t-test). However, the pupil partial model

showed a significant improvement over the null model, and it was able to account for changes in the PSTH between passive and active conditions ($r^2 = 0.51$, $p < 0.05$, Figure 2F, middle). The full model showed no additional improvement over the pupil partial model ($r^2 = 0.53$, $p > 0.05$; Figure 2F, bottom). Thus, we found evidence for A1 units whose activity could be modulated either by task engagement or by pupil size.

Next, we investigated the prevalence of task- and pupil-related modulation across the population of A1 and IC units. To quantify the relative contribution of each variable to model performance, we compared the variance in single-trial spike rate explained by each of the four models (r^2). Across the entire dataset, 66/129 (51%) A1 units and 37/66 (56%) IC units showed a significant increase in r^2 for the full model over the null model, indicating an effect of either task engagement or pupil size ($p < 0.05$, jackknife *t*-test; Figure 3A). By computing the difference between the cross-validated r^2 for the full model and for the two partial models separately, we determined the variance that could be uniquely explained by either pupil size or task engagement in each unit. This produced four categories of state-modulated neurons (Figure 3A): units for which both task and pupil uniquely contributed to the modulation (black); units for which only one variable, task (purple, example unit in Figure 2A) or pupil (green, example unit in Figure 2D), contributed uniquely; and units for which changes in activity could not be uniquely attributed to either task or pupil (dark gray).

A comparison of variance explained across the entire dataset shows consistent improvement by the full model over the null model in both A1 (mean r^2 null: 0.282, full: 0.340, $p = 6.37 \times 10^{-23}$, Wilcoxon signed-rank test) and IC (mean r^2 null: 0.290, full: 0.399, $p = 6.77 \times 10^{-15}$, Figure 3B). Although the proportion of units with any significant state-related modulation was comparable between A1 and IC, only about a quarter of those A1 units showed unique pupil-dependent modulation (15/66), compared to about a half for IC (20/37). The average variance uniquely explained by each state variable also differed between areas. In A1, the variance explained by task (mean r^2 task-unique: 0.017) was significantly greater than variance explained by pupil (mean r^2 pupil-unique: 0.007, $p = 0.00201$, Wilcoxon signed-rank test, Figure 3C). In IC, the variance explained was not significantly different between task and pupil

(mean r^2 task-unique: 0.016, mean r^2 pupil-unique: 0.018, $p = 0.236$). These effects were not significantly different between single- and multiunit data in either area (Figure 3-figure supplement 1). Together, these results suggest that in A1, both the magnitude and prevalence of modulation by task engagement are more pronounced than by pupil size. In IC, modulation equal between the two state variables.

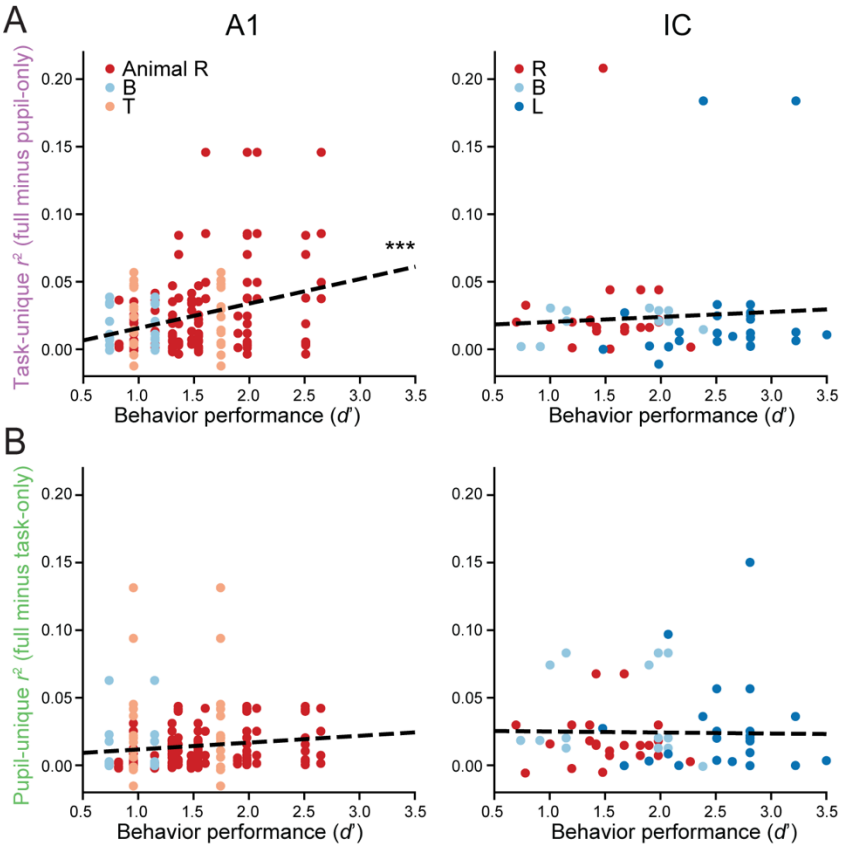


Figure 4: Task and pupil-related modulation of firing rates as a function of behavioral performance in A1 and IC. **A.** Scatter plot of unique variance explained by task plotted against behavioral sensitivity (*d*') for each unit in A1 (left) and IC (right). Each circle represents a unit-behavior pair. Different colors refer to different animal subjects. Points with the same *d*' align on the horizontal axis because they belong to experiments in which multiple units were recorded at the same time using an array. Regression analysis shows a significant correlation between unique variance explained by task and performance in A1 ($r = 0.303$, ***$p = 0.00010$, Pearson's R) but not in IC ($r = 0.069$, $p = 0.580$). **B.** Scatter plot of unique variance explained by pupil plotted against performance for each unit in A1 (left) and IC (right) as in A. Regression analysis shows no significant correlation between unique variance explained by pupil and performance in either A1 ($r = 0.113$, $p = 0.139$) or IC ($r = -0.17$, $p = 0.892$).

The analysis above treated task engagement as a discrete, binary variable, switching between active and passive states. However, behavioral performance, measured by *d*', varied substantially between animals and behavior blocks (Figure 1C). We reasoned that performance represents a more graded

262 measure of task engagement and thus may explain variability in task-related changes across experiments.
263 Indeed, we found a significant correlation between *d*' and the unique variance explained by task in A1 (r^2
264 task-unique; $r = 0.303$, $p = 0.00019$; Figure 4A, left). The same relationship was not observed in IC ($r =$
265 0.069, $p = 0.580$; Figure 4A, right), nor was there a relationship between *d*' and unique variance
266 explained by pupil in either area (r^2 pupil-unique; A1: $r = 0.113$, $p = 0.139$; IC: $r = -0.017$, $p = 0.892$,
267 Figure 4B). Thus, in A1, effects of task engagement cannot be described by a binary variable, even after
268 dissociating effects of pupil. Instead, effects of task engagement are comprised of at least one process that
269 varies continuously between more and less accurate behavior.
270 We also observed that both *d*' and task engagement effects in A1 differed between animals
271 (Figures 4, 4-figure supplement 1). We wondered if the differences in neural modulation could be fully
272 explained by behavioral performance or if they reflected additional between-animal differences. To
273 answer this question, we performed a multiple regression with both *d*' and animal identity as independent
274 variables. This analysis revealed that in A1 *d*' could fully explain the differences in modulation for task
275 engagement (*d*': $F = 16.0$, $p = 0.000093$; animal: $F = 0.66$, $p = 0.52$). Neither *d*' or animal significantly
276 predicted task engagement effects in IC (*d*': $F = 1.11$, $p = 0.29$; animal: $F = 0.22$, $p = 0.80$). A similar
277 analysis of pupil-related effects revealed that r^2 pupil-unique did not depend on *d*', although it did depend
278 on the amount of pupil variability within an experiment (Figure 4-figure supplement 1). Thus, differences
279 in task engagement effects between animals could be explained by differences in the accuracy with which
280 they performed the task.
281
282 **Pupil size accounts for apparent task engagement effects in both A1 and IC**
283 So far, we have shown that neurons in A1 and IC show state-dependent modulation of activity
284 that can be explained variably by task engagement and/or pupil size. The observation that pupil size is
285 correlated with task engagement (Figure 1) suggests previous characterizations of task-related plasticity
286 that did not measure pupil might have attributed changes in neural activity to task engagement when they
287 could, in fact, be better explained by pupil size (Downer et al., 2015; Fritz et al., 2003; Knudsen and

Gentner, 2013; Kuchibhotla et al., 2017; Lee and Middlebrooks, 2011; Otazu et al., 2009; Ryan and Miller, 1977; Yin et al., 2014). Therefore, we asked to what extent pupil-related modulation could explain changes in activity between active and passive blocks that would otherwise be attributed to task engagement alone.

To quantify the magnitude and sign of the modulation by task engagement, we used a modulation index (MI_{AP}), which measured the fraction change in mean predicted firing rate between active and passive conditions (*Eqn. 5*, (Otazu et al., 2009; Schwartz and David, 2018)). In previous work, *MI* was measured directly from differences in the raw firing rate. However, pupil size was correlated with task engagement, and both could contribute to changes in firing rate. To tease apart their respective contributions, we used logic similar to the calculation of unique variance explained (above, Figure 3C). To measure task-related modulation while ignoring possible effects of pupil size, *MI* was computed from predictions of the task only model (MI_{AP} *task-only*). To account for effects of both task engagement and pupil size, *MI* was computed from predictions by the full model (MI_{AP} *full*). Modulation that could be explained by pupil size was computed using the pupil-only model (MI_{AP} *pupil-only*). Then the unique task-related change, MI_{AP} *task-unique*, was computed as the difference between MI_{AP} *full* and MI_{AP} *pupil-only* (*Eqn*. 6). We used a converse approach to quantify modulation uniquely attributable to pupil between large and small pupil states (MI_{LS}, see below).

The model predictions for the example A1 units above (Figure 2) illustrate the method for dissociating contributions of pupil and task engagement to MI_{AP}. The first example shows a strong task engagement effect that cannot be explained by pupil size (MI_{AP} *task-unique* = 0.25, Figure 2C), and the second shows weak task-only modulation that can be completely explained by pupil (MI_{AP} *task-unique* = -0.07, Figure 2F).

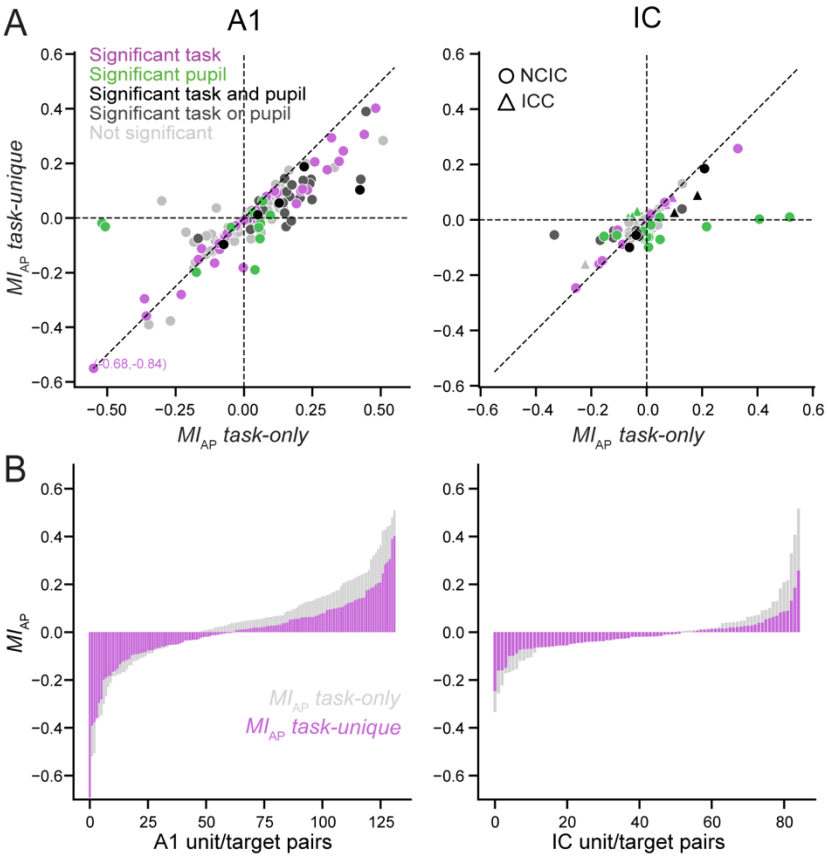


Figure 5: Changes in pupil size account for apparent task engagement effects. **A.** Active versus passive modulation index, MI_{AP} *task-only*, computed from responses predicted by the task only model, in which pupil size is shuffled, plotted against MI_{AP} *task-unique*, in which pupil-dependent modulation is regressed out, for A1 (left, n = 132 unit/target pairs) and IC (right; circles for NCIC, n = 24 unit/target pairs, triangles for ICC units, n = 61). Colors indicate significant unique variance explained by one or two state variables, as in Figure 3 ($p < 0.05$, jackknife t-test). **B.** Overlaid histograms of MI_{AP} *task-only* and MI_{AP} *task-unique*, sorted according to their magnitude for each unit in A1 and IC. Accounting for pupil size reduces the magnitude of MI_{AP} by about 33% in both areas (A1: $p = 6.78 \times 10^{-10}$; IC: $p = 0.041$; Wilcoxon signed-rank test).

A comparison of MI_{AP} *task-only* and MI_{AP} *task-unique* across the entire dataset showed that a task-only model often overestimated the magnitude of changes due to task engagement (Figure 5A). To measure this effect, we normalized the sign of *MI* for each unit so that the means of MI_{AP} *task-only* and MI_{AP} *task-unique* were always positive. This normalization accounted for the bidirectionality of modulation while avoiding bias that would result from normalizing sign for just a single condition. After sign normalization, accounting for pupil size led to a significant decrease in MI_{AP} magnitude across units in both A1 and IC (A1: $p = 6.78 \times 10^{-10}$; IC: $p = 0.041$; Wilcoxon signed-rank test). Effectively,

accounting for pupil size led to a 33% reduction in the magnitude of MI_{AP} in both A1 and IC (A1: median magnitude MI_{AP} *task-only* = 0.141; MI_{AP} *task-unique* = 0.095; IC: MI_{AP} *task-only* = 0.069; MI_{AP} *task-unique* = 0.046). The effect of removing pupil-related activity on the magnitude of active versus passive modulation was also not different between central and non-central regions of IC (ICC versus NCIC, rank-sum test, statistic = -1.386, p = 0.166, Figure 5-figure supplement 1). Taken together these results show that pupil-indexed arousal accounted for a significant portion of the change in activity between passive and active blocks in A1 and both regions of IC.

Task-related modulation is bidirectional in A1 and mostly suppressive in IC

Previous studies have reported either positive or negative changes in spike rate associated with task engagement (David et al., 2012; Kuchibhotla et al., 2017; Otazu et al., 2009; Slee and David, 2015; Yin et al., 2014). Variability in task-dependent changes in auditory cortex has been explained by differences in neural selectivity for task stimuli (Fritz et al., 2003), task structure (David et al., 2012), behavioral effort (Atiani et al., 2009), and selective attention (Schwartz and David, 2018). Given that we used a positive reinforcement structure similar to previous studies, we expected the average sign of task-related modulation to be positive in A1 (David et al., 2012) and negative in IC (Slee and David, 2015). However, after accounting for effects of pupil size in A1, we found that *MI* was equally likely to be positive or negative (40/69 positive, median MI_{AP} *task-unique* = 0.018; Wilcoxon signed-rank test, statistic = 929.0, p = 0.0959; Figure 5B, left). In IC, results were consistent with previous reports of task-related changes (Slee and David, 2015). Firing rates decreased on average during task engagement (17/49 positive, median MI_{AP} *task-unique* = -0.022; Wilcoxon signed-rank test, statistic = 347.0, p = 0.00827; Figure 5B, right). This decrease was significantly different from the average change in A1 (p = 0.00658, Wilcoxon rank-sum test).

We computed changes in spike rate associated with pupil size by measuring *MI* between large and small pupil conditions, dividing the trials, according to whether pupil size was larger or smaller than the median for the experiment, respectively. Analogous to the calculation for task-unique effects, MI_{LS}

pupil-unique was the difference between MI_{LS} *full* and MI_{LS} *task-only* (*Eqn*. 7). In contrast to task-related MI_{AP}, the average effect of increased pupil size was positive in A1 (55/69 positive, median MI_{LS} *pupil-unique* = 0.038; Wilcoxon signed-rank test, statistic = 929.0, $p = 3.29 \times 10^{-5}$; Figure 6, left). Since task engagement is correlated with a larger pupil, firing rates do tend to increase during behavior in A1. However, this increase can be attributed to a generalized increase in pupil size rather than engagement in the specific task.

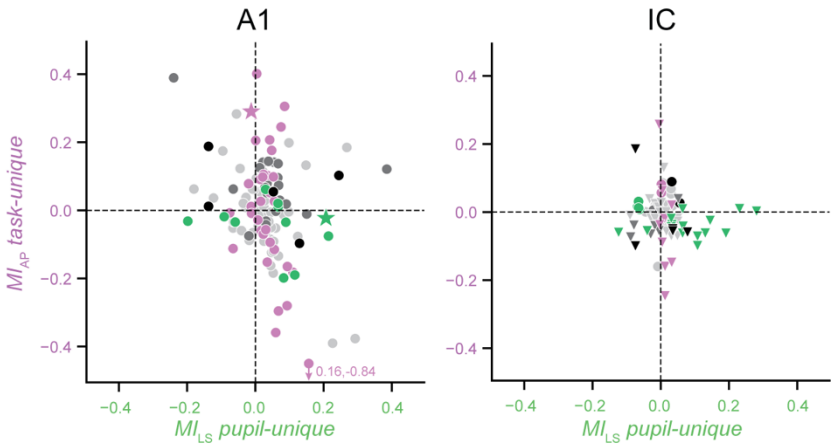


Figure 6: Unique task- and pupil modulation are uncorrelated. Scatter plot compares firing rate modulation attributed to task engagement (MI_{AP} *task-unique*) against modulation attributed to pupil (MI_{LS} *pupil-unique*) for each unit in A1 (left) and IC (right). The two quantities showed a weak negative correlation in A1 (r = -0.281, p = 0.00112, Pearson's R) and were uncorrelated in IC (r = -0.104, p = 0.343). Colors indicate relative model performance as in Figure 3, and stars indicate examples from Figure 2.

In IC, as in A1, firing rates increased for larger pupil size (31/49 positive, median MI_{LS} *pupil-unique* = 0.022; Wilcoxon signed-rank test, statistic = 929.0, p = 0.010; Figure 6, right). Thus, accounting for effects of pupil size unmasked an even stronger overall decrease in spike rate during behavior. After accounting for pupil effects, most state-related effects were similar in ICC and NCIC. However, significant negative task-related modulation was only observed in NCIC (median MI_{AP} *task-unique* ICC: 0.012, p = 0.0869, NCIC: -0.049, p = 0.000731, Figure 5-figure supplement 1).

Some previous work has also reported that the effect of task engagement can depend on the relationship between task-relevant sound features and the receptive field of individual neurons (David et al., 2012; Slee and David, 2015). During a tone detection task, response gain in IC is selectively

379 suppressed for neurons with best frequency near the target tone frequency (Slee and David, 2015). In A1,
380 frequency tuning is suppressed at the target frequency, despite there being no systematic difference in
381 overall response gain between units with best frequency near or far from the target (David et al., 2012).
382 We observed this same pattern of suppression at the target frequency during behavior in the current
383 dataset, even after accounting for pupil size (Figure 6-figure supplement 1). Task engagement effects
384 could also depend on the difficulty or other structural elements of the task (Carcea et al., 2017; Knyazeva
385 et al., 2020). We considered whether state-dependent effects varied with task difficulty but found no
386 differences between pure tone, high SNR tone-in-noise (easy), and low SNR tone-in-noise (hard)
387 conditions (Figure 6-figure supplement 2).
388
389 **Task engagement and pupil-indexed arousal modulate independent neural populations**
390 In some units, the effects of task engagement and pupil size on firing rates were difficult to
391 dissociate by simply looking at raw activity (*e.g.*, Figure 2D). However, the regression analysis could
392 disambiguate their unique contributions and showed that often only one state variable, task or pupil,
393 modulated the activity of individual units. Based on this observation, we wondered if effects of task
394 engagement and pupil size operated via functionally distinct or common pathways. While identifying
395 specific anatomical circuits was outside the scope of this work, we reasoned that if the two modulations
396 were mediated by the same inputs to A1 and IC, a unit modulated by task engagement would also tend to
397 be modulated by pupil size. To test this prediction, we measured Pearson's correlation between MI_{AP} *task-*
398 *unique* and MI_{LS} *pupil-unique* across units (Figure 6, *Eqn*. 7). We found a significant negative correlation
399 between these quantities in A1 (r = -0.281, p = 0.00112) and no correlation in IC (r = 0.104, p = 0.343).
400 The absence of a positive correlation between task- and pupil-related effects suggests that independent
401 populations of neurons in A1 and IC are modulated by these variables. To test directly for dependence
402 between task engagement and pupil size, we performed a permutation analysis, estimating the number of
403 units that would show both task- and pupil-unique effects if the likelihood of these effects was
404 independent ($p < 0.05$, jackknife t-test, counts from Figure 3). The frequency of joint occurrences was no

greater than expected for independent distributions in both areas (A1: $p = 0.41$, IC: $p = 0.59$). This independence is consistent with the hypothesis that separate neural mechanisms mediate effects of task engagement and pupil-indexed arousal.

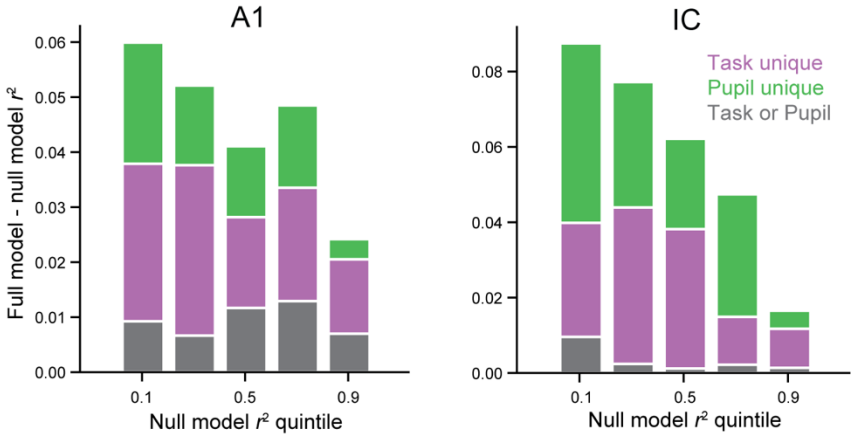


Figure 7: Task- and pupil-related modulation in A1 and IC. Mean variance in firing rate explained by behavioral state variables, broken down by portions attributed to task (purple), attributed to pupil (green), and not uniquely attributable to either state variable (grey) for A1 (left) and IC (right). Data are grouped in quintiles by r^2 for the null model performance, a measure of auditory responsiveness. Variance explained by state was not correlated with null model performance in A1 ($r = -0.134$, $p = 0.123$, Pearson's R), but these values were negatively correlated in IC ($r = -0.323$, $p = 0.00254$). Colors as in Figure 3.

Relationship between state modulation and auditory responsiveness

While all units were recorded from areas functionally and anatomically characterized as auditory, neurons within a single auditory field can vary substantially in their tendency to respond to sound (Atiani et al., 2014; Gruters and Groh, 2012). We noticed that some units, particularly in IC, did not show strong auditory responses but were modulated during behavior (*e.g.*, Figure 8B). We asked whether the magnitude of the state-dependent modulation in each recorded unit was related to how reliably it responded to repeated presentations of the same auditory stimulus. Auditory responsiveness was quantified by the variance explained by the null model, since it described how accurately the PSTH response to the reference stimuli, averaged across the entire recording, predicted responses on single trials. Larger or smaller values of null model r^2 were associated with units whose auditory responses were

429 more or less reliable across the experiment, respectively. We compared this measure of auditory
430 responsiveness to the change in r^2 between the full and null models, which indicates additional variance
431 explained by the state variables. In IC, but not in A1, units that responded less reliably to sound also
432 showed greater dependence on state (A1: r = -0.134, p = 0.123; IC: r = -0.323, p = 0.00254; Figure 7).
433 These same relationships were observed when auditory responsiveness was defined as the signal-
434 to-noise ratio (SNR) of the spectro-temporal receptive field (STRF) measured during passive listening
435 (Figure 7-figure supplement 1B) (Klein et al., 2000). The results also held true for a larger pool of
436 neurons, which included data collected without pupillometry and analyzed with the task-only model
437 (Figure 7-figure supplement 1C). High prediction accuracy by the null model limits the variance that can
438 be explained by behavioral state (note relatively weak state effect in highest quintile in Figure 7). To
439 explore how this bound might impact the analysis of auditory responsiveness, we performed a similar
440 analysis but measured the correlation between null model performance and the fraction of remaining
441 variance explained by pupil size and task engagement. In this case, the null model performance was
442 positively correlated with the remaining variance explained in A1 (r = 0.301, p = 4.61 x 10^{-4}) and not
443 correlated in IC (r = 0.124, p = 0.258, Figure 7-figure supplement 1A). In this case the overall slope of
444 the relationships changed, but the effects were still relatively larger for IC units with low auditory
445 responsiveness. Thus, across multiple measures of sensory responsiveness, we observed that effects of
446 behavior state in IC were more common in units with weak sensory responses, while in A1, state-
447 dependent modulation was independent of sensory responsiveness.

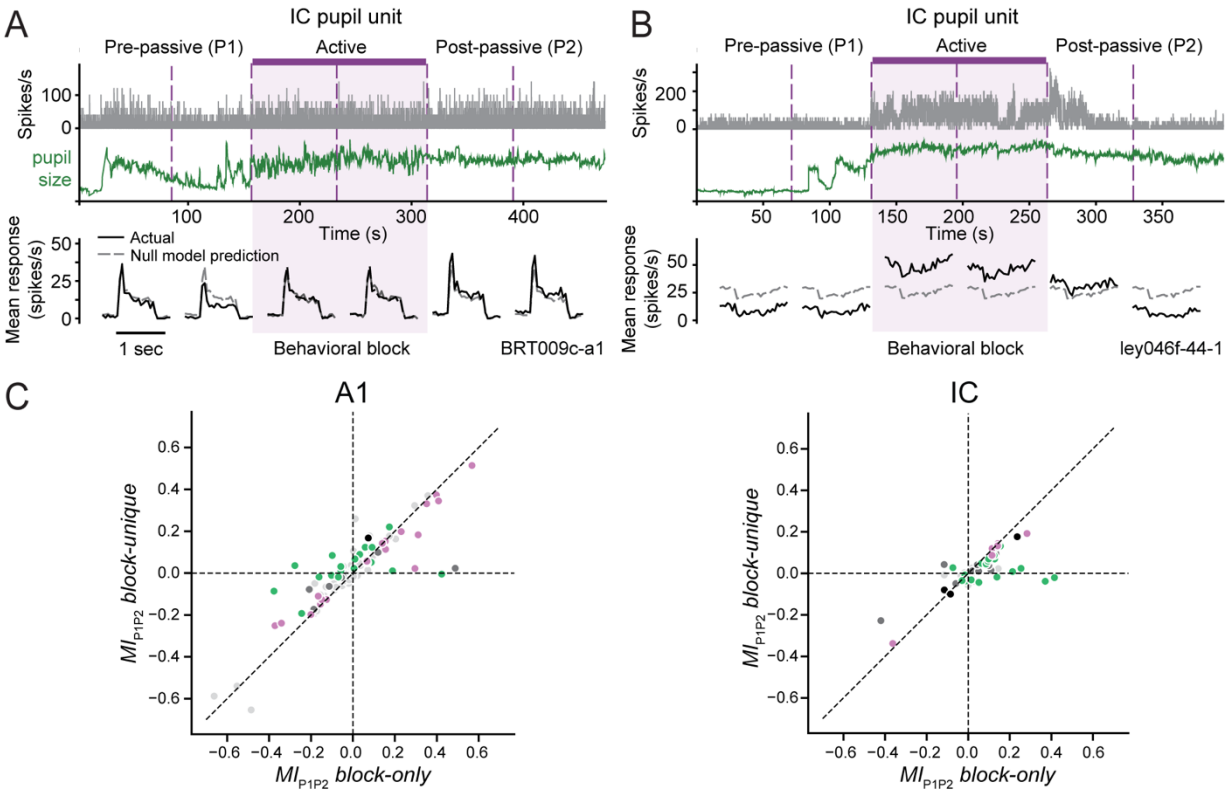


Figure 8: Pupil size explains some persistent task-related modulation. **A.** Top: activity of ICC unit and concurrent pupil size recorded over passive and active blocks, plotted as in Figure 2. Bottom: PSTH response to reference noise averaged over each half block (black) with null model prediction overlaid (dashed gray). The response was enhanced slightly during the post-passive (P2) block relative to pre-passive (P1). Without accounting for effects of pupil size, this suggests a sustained firing rate increase after behavior, MI_{P1P2} *block-only* = 0.25. After accounting for changes in pupil, the task-related effect is reduced, MI_{P1P2} *block-unique* = 0.02. **B.** Data from example NCIC unit, plotted as in A. The PSTH shows a weak, suppressive response to the reference stimuli, but a large increase in mean firing rate that persists into the first half of the post-passive block (MI_{P1P2} *block-only* = 0.37). Again, the persistent change in firing rate can be accounted for by the very large pupil during the first half of P2 (MI_{P1P2} *block-unique* = -0.04). **C.** Scatter plot compares MI_{P1P2} *block-only* and MI_{P1P2} *block-unique* for A1 (left) and IC (right). Colors as in Figure 3. Accounting for pupil reduces *MI* attributed to persistent effects after behavior in both areas (A1: $p = 0.00120$, IC: $p = 9.39 \times 10^{-6}$, Wilcoxon signed-rank test).

Changes in pupil size explain some persistent post-behavior state effects

Previous studies in both A1 (Fritz et al., 2003) and IC (Slee and David, 2015) have reported examples of task-related modulation persisting in passive blocks following active engagement (referred to as post-passive blocks). These effects have been interpreted as persistent task-related plasticity, but they are highly variable and difficult to attribute to any specific aspect of the behavior. We observed these effects in our data as well. In addition to units in A1 (*e.g.*, Figure 2D), some units in IC showed persistent

469 changes in activity during passive blocks following behavior. These changes sometimes lasted through
470 the entire post-passive block (Figure 8A) and other times gradually reverted to pre-passive levels (Figure
471 8B). The timecourse of these changes often tracked fluctuations in pupil size.

472 To measure the presence of persistent task-related modulation—and the extent to which pupil size
473 accounts for it—we compared the activity between passive blocks before (P1) and after (P2) a behavior
474 block. We performed a regression analysis similar to that used for measuring task engagement effects, but
475 treating passive block identity (P1 or P2) as a state variable in addition to pupil size. The modulation
476 index was computed for predictions by a model based only on block identity, ignoring pupil (MI_{P1P2}
477 *block-only*), and for the unique contribution of passive block after accounting for pupil effects (MI_{P1P2}
478 *block-unique*, Figure 8A). We assessed the magnitude of *MI* using the same sign normalization as for the
479 task-unique *MI* analysis above (Figure 5). In both A1 and IC, the magnitude of firing rate modulation
480 between passive blocks was significantly reduced after accounting for effects of pupil (A1: $p = 0.00120$,
481 IC: $p = 9.39 \times 10^{-6}$, Wilcoxon signed-rank test). Accounting for pupil size led to a 20% and a 48%
482 reduction in P1 versus P2 modulation in A1 and IC, respectively (A1: mean signed-normalized MI_{P1P2}
483 *block-only* = 0.140, MI_{P1P2} *block-unique* = 0.100; IC: mean signed-normalized MI_{P1P2} *block-only* = 0.100,
484 MI_{P1P2} *block-unique* = 0.052). This change was not significantly different between animals in A1
485 (F=0.669, p=0.516, one-way ANOVA) or in IC (F=0.446, p=0.643). While there was decrease in post-
486 behavior effects on average, there were distinct groups of units in both A1 and IC that showed changes
487 between P1 and P2, even after accounting for pupil size (compare units falling on the unity line versus x-
488 axis in Figure 8C). Thus, while some post-behavior effects previously reported as persistent plasticity
489 induced by behavior can in fact be explained by fluctuations in pupil size, a small number of units showed
490 persistent post-behavior changes, even after accounting for pupil.

Discussion

This study determined how pupil-indexed arousal contributes to task-related changes in neural activity in primary auditory cortex (A1) and inferior colliculus (IC). Several previous studies have shown that transitioning from passive listening to active behavior leads to sizable changes in neural activity in A1 (Fritz et al., 2003; Niwa et al., 2012; Otazu et al., 2009) and IC (Ryan and Miller, 1977; Slee and David, 2015). The specific changes depend on properties of the task stimuli (Fritz et al., 2003; Jaramillo and Zador, 2011; Lee and Middlebrooks, 2011) and on structural elements of the task, such as reward contingencies (David et al., 2012), task difficulty (Atiani et al., 2009), the degree of engagement (Carcea et al., 2017; Knyazeva et al., 2020), and the focus of selective attention (Hocherman et al., 1976; Lakatos et al., 2013; Schwartz and David, 2018). In the current study, we were able to use pupil size to explain changes in neural activity across both passive and active conditions. Changes in neural activity explained by pupil size are likely to overlap with some of the differences previously reported between task conditions, which may engage different levels of arousal (*e.g.*, (Carcea et al., 2017; Knyazeva et al., 2020; Rodgers and DeWeese, 2014)). At the same time, pupil size is unlikely to explain changes in neural activity specific to a task. By isolating aspects of behavioral state that can be explicitly attributed to changes in pupil, this study moves toward a coherent theory of how a constellation of variables interact to influence sensory coding during behavior.

Aspects of behavioral state, such as arousal, locomotion, and whisking rate, have been shown to change gradually over the course of an experiment, tracking changes in cortical spiking and performance on operant tasks (McGinley et al., 2015; Niell and Stryker, 2010; Poulet and Petersen, 2008). A major obstacle to understanding the relationship between behavioral state and neural activity is that state variables often covary (Musall et al., 2019; Stringer et al., 2019). To dissociate effects of pupil size and task engagement here, we used a stepwise approach in which pupil size and task engagement provided separate regressors (Fritz et al., 2010; Musall et al., 2019). In about half of A1 and IC units, some combination of task engagement and/or pupil size explained changes in spiking activity between blocks of active engagement and passive listening. These behavioral state variables were correlated; that is, pupil

size tended to increase during task engagement. However, these changes affected independent neural populations, suggesting that effects of pupil-indexed arousal and other aspects of task engagement operate via distinct feedback circuits.

Defining meaningful state variables

The correlation between pupil size and task engagement demonstrates the importance of accounting for interactions between state variables when measuring their effects on neural activity. This observation also leads to a more fundamental question: what constitutes a meaningful measure of behavioral state? Concepts like task engagement and arousal make intuitive sense, but they become slippery when one attempts to pin them to distinct physiological processes. The current results illustrate that a commonly used, binary task engagement variable is better described by a combination of pupil size and accuracy of behavioral performance. Other characterizations of task engagement have argued similarly that the degree of engagement can vary, even when relevant acoustic stimuli are held fixed (Atiani et al., 2009; Carcea et al., 2017; Knyazeva et al., 2020; McGinley et al., 2015; Zhou et al., 2014). Larger datasets that track continuous fluctuations in performance are likely to be able to measure fluctuations of engagement effects within behavior blocks. It remains unclear how smoothly internal state can vary. Theories of network attractors suggest that there may in fact be discrete changes in brain state associated with different behavioral contexts (La Camera et al., 2019). Thus, while task engagement is clearly not binary, it could still comprise multiple metastable states.

As a behavioral state variable, arousal also has multiple definitions, including the transition from wakefulness to sleep, the response to emotionally salient stimuli, and a generalized activation of the autonomic nervous system (Satpute et al., 2019). This relatively nonspecific behavioral concept contrasts with pupil size, which has a clear link to neurophysiology (McGinley et al., 2015; Reimer et al., 2016; Zekveld et al., 2018). The detailed circuits that link pupil size to cognitive state are not fully understood, but a connection to behavioral state has been demonstrated repeatedly. Thus, while pupil size may not map trivially onto the concept of arousal, it has a clear link to brain mechanisms.

A clearer understanding of the relationship between behavioral state and sensory processing may be achieved with models that incorporate additional well-defined variables like pupil size, which can be linked to well-defined neurophysiological processes. For example, motor- and reward-related activity are associated with specific neural circuits (Bakin and Weinberger, 1990; Knyazeva et al., 2020; Schneider et al., 2014). The current study excluded epochs that contained licking activity to avoid motor confounds, but motor signals can be included as additional model inputs that reflect aspects of signal detection and decision making (Runyan et al., 2017). These variables may not be directly related to an experimentally imposed task structure, but they capture processes involved in the behavior and can be quantified more concretely than abstract variables based on experimentally imposed task conditions.

Reinterpreting previous studies of auditory behavior

Our findings indicate that changes in firing rate previously attributed to task engagement might instead reflect fluctuations in pupil-indexed arousal, particularly in IC, where pupil-related changes predominate. Changes in pupil size can also explain some modulation that persists after behavior ceases (Fritz et al., 2003; Slee and David, 2015). However, pupil size does not account for all the modulation between active and passive blocks; instead, it appears that both pupil-indexed arousal and task engagement act on independent neural populations to shape sound processing.

Previous studies that measured the effects of task engagement on auditory neuronal activity described different effects on overall excitability between A1 and IC. In A1 the effects of task engagement were shown to be enhancing (David et al., 2012), while in IC these changes were predominantly suppressive (Slee and David, 2015). Separate studies have shown that pupil size is positively correlated with excitability in both A1 and IC (Joshi et al., 2016; Schwartz et al., 2020). Here we also found that increased pupil size enhanced excitability in both areas. However, after accounting for pupil-related changes, task-engagement effects were no longer enhancing in A1, and they were more suppressive in IC. Pupil-related changes did not depend on tuning for the target tone. These non-specific

568 effects of pupil size contrasted with effects of task engagement, which can be selective for task-relevant
569 features (Fritz et al., 2003; Kuchibhotla et al., 2017; Slee and David, 2015).
570 In the current study, significant state-dependent A1 units were equally likely to be modulated by
571 task engagement, by pupil size, or by both. In IC, however, we found that a larger portion of state-
572 dependent changes could be explained by pupil size. Furthermore, in IC but not in A1, units with less
573 reliable auditory responsiveness were those with a stronger state-modulation component. This result is
574 perhaps not surprising given that the majority of IC units in our sample were recorded from non-central
575 regions of the IC (NCIC). These are dorsal and lateral regions of IC that receive feedback from cortex
576 (Winer, 2006), input from neuromodulatory nuclei such as the pedunculopontine and latero-dorsal
577 tegmental nuclei (Motts and Schofield, 2009), as well as multisensory information from somatosensory,
578 visual, and oculomotor centers (Gruters and Groh, 2012). These results suggest that the auditory midbrain
579 is not a passive sound processor. Instead, the network of midbrain nuclei in the IC plays an active role
580 enhancing behaviorally relevant signals. Experiments that sample more units across the different
581 subregions would be needed to confirm the interplay of behavioral and sensory signals across IC.
582 Multiple regression has been used previously to dissociate effects of pupil size and movement on
583 neural activity (Musall et al., 2019; Stringer et al., 2019). This study used this approach to dissociate the
584 effects of state-related variables and differentiate where effects emerge. Several other variables have been
585 shown to impact auditory responses during behavior, including selective attention, temporal expectation,
586 reward value and motor responses (David et al., 2012; Huang et al., 2019; Jaramillo and Zador, 2011;
587 Lakatos et al., 2013; Metzger et al., 2006). The same method can be adopted to explore the contribution
588 of other state variables measured in normal conditions as well as in perturbation experiments to identify
589 the circuits through which such variables operate.
590
591 **The role of neuromodulation in effects of task engagement and pupil-indexed arousal in A1 and IC**
592 The correlation between task engagement and pupil size may help explain a number of results
593 related to neuromodulation in the auditory system. It is well established that neuromodulators can induce

short-term changes in activity and sensory tuning in cortex (Bakin and Weinberger, 1996; Goard and Dan, 2009) and midbrain (Gittelman et al., 2013; Habbicht and Vater, 1996; Hurley and Pollak, 2005). Cholinergic fibers projecting to the auditory cortex from the nucleus basalis play a key role in rapid switching between passive listening and active engagement by modulating the activity of different populations of cortical inhibitory interneurons (Kuchibhotla et al., 2017). In addition, the activity of cholinergic and noradrenergic terminals in auditory cortex was found to be elevated during pupil dilation and reduced during pupil constriction (Reimer et al., 2016). Thus, these neuromodulatory systems have been implicated in effects of both task engagement and pupil-indexed arousal. The correlation between these state variables could be explained by shared cholinergic and noradrenergic signaling.

Materials and Methods

All procedures were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee (protocol IP1561) and conform to National Institutes of Health standards.

Surgical procedure

Animal care and procedures were similar to those described previously for neurophysiological recordings from awake behaving ferrets (Slee and David, 2015). Four young adult male ferrets were obtained from an animal supplier (Marshall Farms, North Rose, NY). After a two-week quarantine, a sterile surgery was performed under isoflurane anesthesia to mount two head-posts for head fixation and to permit access to auditory brain regions for recordings. A UV light-cured composite (Charisma, Heraeus Kulzer International) mount two custom-made stainless-steel head-posts spaced approximately 1 cm apart along the sagittal crest of the skull. The stability of the implant was achieved using 8-10 stainless steel bone screws mounted in the skull (Depuy Synthes, Raynham, MA), which were covered along with the headpost in additional Charisma and acrylic denture material (Co-oral-ite Dental, Diamond Springs. CA). Two 1.2x1.2 cm wells over auditory cortex in the implant were covered with only a thin layer of Charisma to allow access to auditory regions.

Following surgery, animals were treated with prophylactic antibiotics (Baytril 10 mg/kg) and analgesics (Buprenorphin 0.02 mg/kg) under the supervision of University veterinary staff. For the first two weeks the wound was cleaned with antiseptics (Betadine 1:4 in saline and Chlorhexidine 0.2%) and bandaged daily with application of topic antibiotic ointment (Bacitracin). After the wound margin was healed, cleaning and bandaging occurred every 2-3 days through the life of the animals. This method revealed to be very effective in minimizing infections of the wound margin. After recovery (~2 weeks), animals were habituated to a head-fixed posture inside the sound-booth chamber for about 2 weeks prior to the beginning of the training.

628 **Behavioral paradigm and training**

629 The animals were trained by instrumental conditioning to perform a positively reinforced, tone
630 versus noise discrimination task (Ferret L) (Slee and David, 2015) or tone-in-noise detection task (Ferrets
631 R, B, T). During training, animals were provided access to water *ad libitum* on weekends, but were placed
632 on water restriction during the weekdays (Monday through Friday), allowing them to maintain >90% of
633 their baseline body weight long-term. On weekdays, they were given the opportunity to receive liquid
634 reward during behavioral training. Each behavioral trial consisted of a sequence of 2-5 broadband noise
635 reference sounds (TORCs; 30 samples, 5 octaves, 0.75 s duration, 0.7 s inter-stimulus interval) (Klein et
636 al., 2000), followed by a target tone, either alone (tone versus noise task) or overlapping with another
637 TORC (tone-in-noise task). Animals reported the occurrence of the target by licking a water spout (Figure
638 1A). Licks were detected by a piezoelectric sensor glued to the spout. Licks occurring within the target
639 window, 0.1-1.5 s after target onset, were rewarded with 1-3 drops of a 2:1 solution of water and a high-
640 protein, high-calorie dietary supplement (Ensure). False alarms, licks occurring before the target window,
641 and misses, a failure to lick during the target window, resulted in a 5-8 s timeout. Regardless of trial
642 outcome, the next trial began only after animals did not lick for a random period (2.5 ± 0.5 s), reducing
643 the prevalence of accidental, short-latency false alarms. The number of TORCs per trial was distributed
644 randomly with a flat hazard function to prevent behavioral timing strategies (Heffner and Heffner, 1995).
645 Behavioral performance was measured using *d'*, the z-scored difference between hit rate and false alarm
646 rate (Green and Swets, 1966)

647 A behavioral block consisted of 60-100 trials with the same target tone frequency (100 – 20,000
648 Hz) and same distribution of target signal-to-noise ratio (SNR). Animals completed 1-3 behavioral blocks
649 per day, between which target frequency and/or SNR could change. During training, target frequency was
650 chosen at random to span the audiogram of a ferret (Kelly et al., 1986). During electrophysiological
651 recordings, target tone frequency was selected to match the best frequency of the recording site.

652 During the training on the tone-in-noise variant of the task, the TORC sample masking the target
653 varied randomly in each trial to prevent animals from using TORCs' spectro-temporal features to identify

654 targets. At the beginning of training, the tone was presented at +40 dB signal-to-noise ratio (SNR; ratio of
655 peak-to-peak amplitude) relative to the TORCs. This difference was gradually reduced over the course of
656 two-three weeks until the animal consistently performed above chance (three behavioral blocks with
657 performance yielding *d*' > 1, see below) at 0dB SNR.

658 For the tone-in-noise task, the SNR of the tone with respect to the overall level of the TORCs
659 (fixed at 55 or 60dB SPL depending on the animal) varied between +5 and -20dB SNR, in 5dB steps.
660 Each session included five target/noise SNRs. To manipulate task difficulty within each session, the
661 probability of each of the five target/noise SNRs varied, yielding two difficulty conditions: a high SNR
662 ("easy") condition in which 60% of the trials the target/noise SNR was either the highest or the second to
663 the highest SNR within the target/noise SNR distribution; and a low SNR ("hard") condition in which the
664 two lowest target/noise SNRs occurred in 60% of the trials. For example, for ferret B, the distribution
665 after training was completed was kept between 0 and -20dB SNR, so that in the easy condition 0 and -
666 5dB SNR targets would appear 60% of the time, -10dB SNR 20% of the time, and the remaining -15 and
667 -20dB SNR would be presented another 20% of the trials. During electrophysiological tone-in-noise
668 experiments, the tone was embedded in a single TORC sample, which also occurred in the reference
669 period. We confirmed that animals were not biased to respond to this TORC exemplar in the reference
670 phase.

671

672 **Sound presentation**

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

speaker, contralaterally to the recording site. Sound level was equalized and calibrated against a standard reference (Brüel & Kjær).

Pupil recording

During experiments, infrared video of one eye was collected for offline measurement of pupil size as an index of arousal (pupil-indexed arousal). Ambient light was maintained at a constant level to prevent light-evoked changes in pupil and to maximize the dynamic range of pupil-related changes (Schwartz et al., 2020). Recordings were collected using a CCD camera (Adafruit TTL Serial Camera 397) fitted with a lens (M12 Lenses PT-2514BMP 25.0 mm) whose focal length allowed placement of the camera 10 cm from the eye. To improve contrast, the imaged eye was illuminated by a bank of infrared LEDs. Ambient luminance was provided using a ring light (AmScope LED-144S). At the start of each recording day, the intensity of the ring light was set to a level (~1500 lux measured at the recorded eye) chosen to give a maximum dynamic range of pupil sizes. Light intensity remained fixed across the recording session. Pupil size was measured from the video signal offline using custom MATLAB software, which is detailed in (Schwartz et al., 2020).

Neurophysiology

Data were recorded during behavior and pupillometry from 129 units in A1 and 66 units in IC. For one supplementary analysis (Figure 7-figure supplement 1), additional data were collected during behavior but without pupillometry (IC: 135 additional units, A1: 125 additional units). After animals demonstrated consistent above-chance performance (target versus reference $d' > 1$, see below), a small craniotomy was opened to access either primary auditory cortex (A1) or inferior colliculus (IC). Extracellular neuronal activity was recorded in non-anesthetized ferrets either using a tetrode (Thomas Recording Inc.), or a linear 64-channel silicon probe (Shobe et al., 2015). The impedance of the tetrode was measured to be 1-2 MOhm, and the 64-channel probe was electroplated to reach a 0.7-MOhm

704 impedance in each channel. The tetrode and probe were inserted using a motorized electrode positioning
705 system (Alpha-Omega).
706 Amplified (tetrodes, AM Systems; arrays, Intan) and digitized (National Instruments)
707 electrophysiological signals were stored using the open-source data acquisition software MANTA
708 (Englitz et al., 2013) or Open Ephys (Black et al., 2017). Recording sites in A1 and IC were initially
709 targeted stereotactically (Moore et al., 1983; Shamma et al., 1993). Sites in A1 were confirmed based on
710 tonotopy—which changes from high to low best frequency as one moves from dorso-medial to ventro-
711 lateral regions—and relatively reliable and simple frequency tuning (Shamma et al., 1993). Recording
712 sites in the IC were classified based on tonotopy and response properties (Aitkin et al., 1975; Moore et al.,
713 1983). Neurons in the central nucleus of the IC (here referred to as ICC) receive input from the auditory
714 brainstem, and have a characteristic short response latency, dorsal-ventral tonotopy, and narrow
715 bandwidth tuning. Conversely, regions around the central nucleus do not receive direct ascending input
716 and present longer response latencies, considerably less sharp tuning, lack consistent tonotopic
717 organization; these areas were grouped together as NCIC (Slee and David, 2015).
718 For tetrode recordings, upon unit isolation, a series of brief (100-ms duration, 200-400ms inter-
719 stimulus interval, 50 dB SPL) tones and/or narrowband noise bursts were used to determine the range of
720 frequencies that evoked the strongest response, the best frequency (BF) of the unit(s). If a unit(s) in the
721 site did not respond to the sound stimuli (an evoked increase or decrease in activity compared to
722 spontaneous activity either during or after sound presentation), the electrode was moved to a new
723 recording depth with small, 5um incremental steps. For the 64-channel recordings, we lowered the probe
724 so that its entire extent (1 mm) spanned the target region (A1 or IC).
725
726 *Spike sorting*
727 Putative spikes were sorted offline by band-pass filtering the raw trace (300–6000 Hz) and
728 extracting events from the continuous signal that deviated ≥4 standard deviations from zero. To separate
729 single units and stable multi units from the electrode signal, we used the Catamaran clustering program

730 (kindly provided by D. Schwarz and L. Carney) (Schwarz et al., 2012) for tetrode recordings, and the
731 software KiloSort (Pachitariu et al., 2016) for array recordings. In both cases, units were defined based on
732 visual inspection of traces and by having less than 1% of inter-spike intervals shorter than 0.75 ms (single
733 units) and less than 2.5% inter-spike intervals shorter than 0.75ms (multi units). Stability of single-unit
734 isolation was verified by examining waveforms and interval histograms. If isolation was lost during an
735 experiment, only activity during the stable period was analyzed.
736
737 **Analysis**
738 Data preprocessing and analysis were performed using custom MATLAB and Python scripts. All
739 analyses were performed on biological replicates, either different behavior sessions, different neurons
740 (single- or multiunit, see above), or different neuron-behavior condition pairs. Neural and pupil activity
741 were binned at 20 samples/s before analysis. A Python library used for the modeling portion of this
742 analysis is available at https://github.com/LBHB/NEMS/. Sample size (number of animals, number of
743 units per condition) was determined based on norms established in previous studies of neural activity in
744 awake, behaving animals (Fritz et al., 2003; Niwa et al., 2012; Otazu et al., 2009).
745
746 *Task performance*
747 Behavioral performance was measured using signal detection theory (Green and Swets, 1966;
748 Stanislaw and Todorov, 1999). Hits or false alarms (*FAs*) occurred when the animal licked the piezo
749 waterspout upon presentation of the target or the reference stimuli, respectively. Misses or correct
750 rejections (*CRs*) occurred when the animal did not lick following presentation of the target or reference.
751 Hit rate (*HR*) was calculated as the proportion of licks that occurred upon presentation of the target during
752 the response window (0.1 to 1.5 s after target onset), $HR = hits / (hits+misses)$. False alarm rate (*FAR*)
753 was calculated as the proportion of reference stimuli that elicited a lick, during the same window
754 following reference onset, $FAR = FAs / (FAs + CRs)$. Sensitivity (*d*'), a measure of the animals' ability to

755 discriminate between target tone and reference noise, was measured by the difference between the z-
756 scored hit rate and the false alarm rate, $d' = HR_z - FAR_z$. No discrimination, that is, equal likelihood of
757 responding to target and reference is indicated by $d' = 0$. Animals were considered trained to the task and
758 ready to undergo electrophysiological recordings when they performed consistently above chance ($d' > 1$)
759 for three consecutive sessions.
760
761 *Auditory tuning properties*
762 Tuning properties of A1 and IC units were characterized using the spectro-temporal receptive
763 field (STRF) estimated by reverse correlation between time-varying neuronal spike rate and the rippled
764 noise used as reference sounds during behavior (Klein et al., 2000). The STRF provides a quick and
765 efficient measure of stimulus frequencies that enhance or suppress neural firing. In the current study, the
766 STRF was used to measure best frequency (BF) as the center of mass of a spectral tuning curve computed
767 from the first principle component of the STRF matrix (Simon et al., 2007). Auditory responsiveness was
768 computed as the signal-to-noise ratio of single trial responses to the reference noise stimuli during passive
769 listening (Klein et al., 2000).
770
771 *State-dependent models*
772 We fit a generalized linear model in which time-varying state variables pupil size, $p(t)$, and task
773 engagement, $b(t)$, were used to re-weight each unit's mean evoked response to each noise stimulus,
774 $r_0(t)$, and spontaneous rate, s_0, to generate a prediction of the single-trial spike rate, $r_{full}(t)$, at each
775 point in time. The model included multiplicative gain parameters (g) and DC offset parameters (d) to
776 capture both types of modulation. We refer to this model as the *full model*:

$$r_{full}(t) = s_0 F_d\left[d_0 + d_p p(t) + d_b b(t)\right] + r_0(t) F_g\left[g_0 + g_p p(t) + g_b b(t)\right] \quad (1).$$

777 To account for nonlinear threshold and saturation of state effects, the summed state signal was passed
778 through a sigmoid nonlinearity, F_d or F_g, before scaling response baseline or gain, respectively (difference

of exponentials) (Thorson et al., 2015). With additional constant terms g_0 and d_0, this model required a total of six free parameters. For comparison, we calculated a state-independent model here referred to as the *null model*, in which the state variable regressors were shuffled in time. Because shuffling removes any possible correlation between state and neural activity, gain and offset parameters are reduced to d_0=g_0=1 and d_p=d_b=g_p=g_b=0, effectively reducing the model to

$$r_{null}(t) = s_0 + r_0(t) \quad (2).$$

In practice, fitting to the shuffled data produces parameter values slightly different from zero, and controls for noise in the regression procedure.

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

$$r_p(t) = s_0 F_d\left[d_0 + d_p p(t)\right] + r_0(t) F_g\left[g_0 + g_p p(t)\right] \quad (3),$$

and the task-only model accounted only for task engagement,

$$r_b(t) = s_0 F_d[d_0 + d_b b(t)] + r_0(t) F_g[g_0 + g_b b(t)] \quad (4).$$

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

By comparing performance of the full model to each partial model, we determined the unique contribution of each state variable to the neuron's activity. We used a 20-fold, nested cross validation procedure to evaluate model performance, which permitted using the full dataset for both model fitting and validation without introducing bias from over-fitting. The model was fit to 95% of the data and used to predict the remaining 5%. Fit and test data were taken from interleaved trials. This procedure was repeated 20 times with non-overlapping test sets, so that the final result was a prediction of the entire response. Model performance was then quantified by the fraction of variance explained, *i.e.*, the squared correlation coefficient, r^2, between the predicted and actual time-varying response. Variance uniquely explained by single state variables was calculated as the difference between r^2 for the full model and for the partial model in which the relevant variable was shuffled in time.

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

Modulation Index

To quantify the modulatory effects of task and pupil size on the firing rate of A1 and IC units, we computed a modulation index (*MI*) (Otazu et al., 2009; Schwartz and David, 2018). The sound-evoked response was computed by averaging firing rate over the 750-ms duration of a stimulus presentation. *MI* was defined as the difference between the mean response to the same stimulus in two conditions, α and β, normalized by the sum,

$$MI_{\alpha\beta} = \frac{r_\alpha - r_\beta}{r_\alpha + r_\beta} \quad (5).$$

MI could be calculated between different behavioral blocks or between state conditions. In the case of task engagement, *MI* was calculated between active and passive conditions, MI_{AP}. For pupil-dependent changes, data from an experiment were divided at the median pupil diameter, and MI_{LS} was computed between large pupil (above median, high arousal) and small pupil (below median, low arousal). To quantify differences between the first and the second passives, MI_{P1P2} was computed between the first and second passive blocks.

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

$$MI_{AP}\ task\ unique = MI_{AP}\ full - MI_{AP}\ pupil\ only \quad (6),$$

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

$$MI_{LS}\ pupil\ unique = MI_{LS}\ full - MI_{LS}\ task\ only \quad (7).$$

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

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

Statistical Tests

Several analyses assessed the changes in the median of repeatedly sampled data, *i.e.*, pupil size or performance measured across multiple behavior blocks, state-dependent modulation measured across multiple units. In this case, significant differences were assessed using a Wilcoxon signed-rank test (Wilcoxon, 1945). State-dependent changes in individual neurons were assessed using a combination of nested cross validation and a jackknife *t*-test (see above, (Efron and Tibshirani, 1986)).

844 **References Cited**

845 Aitkin LM, Webster WR, Veale JL, Crosby DC. 1975. Inferior colliculus. I. Comparison of response
846 properties of neurons in central, pericentral, and external nuclei of adult cat. *J Neurophysiol*
847 **38**:1196–1207.

848 Atiani S, David SV, Elgueda D, Locastro M, Radtke-Schuller S, Shamma SA, Fritz JB. 2014. Emergent
849 selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* **82**:486–499.
850 doi:10.1016/j.neuron.2014.02.029

851 Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA. 2009. Task difficulty and performance induce
852 diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*
853 **61**:467–480.

854 Bagur S, Averseng M, Elgueda D, David S, Fritz J, Yin P, Shamma S, Boubenec Y, Ostojic S. 2018.
855 Go/No-Go task engagement enhances population representation of target stimuli in primary
856 auditory cortex. *Nat Commun* **9**:2529. doi:10.1038/s41467-018-04839-9

857 Bakin JS, Weinberger NM. 1996. Induction of a physiological memory in the cerebral cortex by
858 stimulation of the nucleus basalis. *Proc Natl Acad Sci U A* **93**:11219–11224.

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

861 Beaton R, Miller JM. 1975. Single cell activity in the auditory cortex of the unanesthetized, behaving
862 monkey: Correlation with stimulus controlled behavior. *Brain Res* **100**:543–62.
863 doi:10.1016/0006-8993(75)90157-2

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

866 Bizley JK, Cohen YE. 2013. The what, where and how of auditory-object perception. *Nat Rev Neurosci*
867 **14**:693–707. doi:10.1038/nrn3565

Bizley JK, Walker KMM, Nodal FR, King AJ, Schnupp JWH. 2013. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr Biol* **23**:620–5. doi:10.1016/j.cub.2013.03.003

Black C, Voigts J, Agrawal U, Ladow M, Santoyo J, Moore C, Jones S. 2017. Open Ephys electroencephalography (Open Ephys + EEG): a modular, low-cost, open-source solution to human neural recording. *J Neural Eng* **14**:035002. doi:10.1088/1741-2552/aa651f

Brosch M, Selezneva E, Scheich H. 2005. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci* **25**:6797–6806.

Carcea I, Insanally MN, Froemke RC. 2017. Dynamics of auditory cortical activity during behavioural engagement and auditory perception. *Nat Commun* **8**:14412. doi:10.1038/ncomms14412

David SV, Fritz JB, Shamma SA. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci USA* **109**:2150–55. doi:10.1073/pnas.1117717109

Downer JD, Niwa M, Sutter ML. 2015. Task engagement selectively modulates neural correlations in primary auditory cortex. *J Neurosci* **35**:7565–7574. doi:10.1523/JNEUROSCI.4094-14.2015

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* **1**:54–77.

Eliades SJ, Wang X. 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* **453**:1102–6. doi:10.1038/nature06910

Englitz B, David SV, Sorenson MD, Shamma SA. 2013. MANTA-an open-source, high density electrophysiology recording suite for MATLAB. *Front Neural Circuits* **7**:69. doi:10.3389/fncir.2013.00069

Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA, Elhilali M. 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci* **13**:1011–9. doi:10.1038/nn.2598

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

894 Fritz JB, Elhilali M, Shamma SA. 2005. Differential dynamic plasticity of A1 receptive fields during
895 multiple spectral tasks. *J Neurosci* **25**:7623–7635.
896 Fritz JB, Shamma SA, Elhilali M, Klein DJ. 2003. Rapid task-related plasticity of spectrotemporal
897 receptive fields in primary auditory cortex. *Nat Neurosci* **6**:1216–1223.
898 Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, Huang ZJ, Stryker MP. 2014. A
899 cortical circuit for gain control by behavioral state. *Cell* **156**:1139–1152.
900 doi:10.1016/j.cell.2014.01.050
901 Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. 2010. Pupil diameter tracks changes in control state
902 predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci*
903 **10**:252–69. doi:10.3758/CABN.10.2.252
904 Gittelman JX, Perkel DJ, Portfors CV. 2013. Dopamine modulates auditory responses in the inferior
905 colliculus in a heterogeneous manner. *J Assoc Res Otolaryngol* **14**:719–729.
906 Goard M, Dan Y. 2009. Basal forebrain activation enhances cortical coding of natural scenes. *Nat*
907 *Neurosci* **12**:1444–1449.
908 Green DM, Swets J. 1966. Signal detection theory and psychophysics. New York: Wiley.
909 Gruters KG, Groh JM. 2012. Sounds and beyond: multisensory and other non-auditory signals in the
910 inferior colliculus. *Front Neural Circuits* **6**:96. doi:10.3389/fncir.2012.00096
911 Habbicht H, Vater M. 1996. A microiontophoretic study of acetylcholine effects in the inferior colliculus
912 of horseshoe bats: implications for a modulatory role. *Brain Res* **724**:169–179.
913 Heffner HE, Heffner RS. 1995. Conditioned Avoidance In: Klump GM, Dooling RJ, Fay RR, Stebbins
914 WC, editors. Methods in Comparative Psychoacoustics. Basel: Birkhauser Verlag. pp. 79–93.
915 Hocherman S, Benson DA, Goldstein Jr. MH, Heffner HE, Hienz RD. 1976. Evoked unit activity in
916 auditory cortex of monkeys performing a selective attention task. *Brain Res* **117**:51–68.
917 Huang Y, Heil P, Brosch M. 2019. Associations between sounds and actions in early auditory cortex of
918 nonhuman primates. *Elife* **8**:e43281.

Hurley LM, Pollak GD. 2005. Serotonin shifts first-spike latencies of inferior colliculus neurons. *J Neurosci* **25**:7876–7886.

Jaramillo S, Zador AM. 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat Neurosci* **14**:246–51. doi:10.1038/nn.2688

Joshi S, Li Y, Kalwani RM, Gold JI. 2016. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron* **89**:221–34. doi:10.1016/j.neuron.2015.11.028

Kahneman D, Beatty J. 1966. Pupil diameter and load on memory. *Science* **154**:1583–5.

Kelly JB, Kavanagh GL, Dalton JC. 1986. Hearing in the ferret (Mustela putorius): thresholds for pure tone detection. *Hear Res* **24**:269–75.

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

Knudsen DP, Gentner TQ. 2013. Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. *J Neurophysiol* **109**:1690–1703. doi:10.1152/jn.00461.2012

Knyazeva S, Selezneva E, Gorkin A, Ohl FW, Brosch M. 2020. Representation of Auditory Task Components and of Their Relationships in Primate Auditory Cortex. *Front Neurosci* **14**. doi:10.3389/fnins.2020.00306

Kuchibhotla KV, Gill JV, Lindsay GW, Papadoyannis ES, Field RE, Sten TAHH, Miller KD, Froemke RC. 2017. Parallel processing by cortical inhibition enables context-dependent behavior. *Nat Neurosci* **20**:62–71. doi:10.1038/nn.4436

La Camera G, Fontanini A, Mazzucato L. 2019. Cortical computations via metastable activity. *Curr Opin Neurobiol*, Computational Neuroscience **58**:37–45. doi:10.1016/j.conb.2019.06.007

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE. 2013. The spectrotemporal filter mechanism of auditory selective attention. *Neuron* **77**:750–61. doi:10.1016/j.neuron.2012.11.034

Lee C-CC, Middlebrooks JC. 2011. Auditory cortex spatial sensitivity sharpens during task performance. *Nat Neurosci* **14**:108–14. doi:10.1038/nn.2713

Lin P-A, Asinof SK, Edwards NJ, Isaacson JS. 2019. Arousal regulates frequency tuning in primary auditory cortex. *Proc Natl Acad Sci* **116**:25304–25310. doi:10.1073/pnas.1911383116

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

Metzger RR, Greene NT, Porter KK, Groh JM. 2006. Effects of reward and behavioral context on neural activity in the primate inferior colliculus. *J Neurosci* **26**:7468–7476.

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

Motts SD, Schofield BR. 2009. Sources of cholinergic input to the inferior colliculus. *Neuroscience* **160**:103–14. doi:10.1016/j.neuroscience.2009.02.036

Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK. 2019. Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci* **22**:1677–1686. doi:10.1038/s41593-019-0502-4

Niell CM, Stryker MP. 2010. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* **65**:472–479. doi:10.1016/j.neuron.2010.01.033

Niwa M, Johnson JS, O'Connor KN, Sutter ML. 2012. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J Neurosci* **32**:9323–34. doi:10.1523/JNEUROSCI.5832-11.2012

Otazu GH, Tai L-HH, Yang Y, Zador AM. 2009. Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* **12**:646–54. doi:10.1038/nn.2306

Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD. 2016. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. doi:10.1101/061481

Poulet JFA, Petersen CCH. 2008. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**:881–5. doi:10.1038/nature07150

Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tolias AS. 2016. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat Commun* **7**:13289. doi:10.1038/ncomms13289

Ringach DL. 2009. Spontaneous and driven cortical activity: implications for computation. *Curr Opin Neurobiol* **19**:439–444. doi:10.1016/j.conb.2009.07.005

Rodgers CCC, DeWeese MRR. 2014. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron* **82**:1157–70. doi:10.1016/j.neuron.2014.04.031

Runyan CA, Piasini E, Panzeri S, Harvey CD. 2017. Distinct timescales of population coding across cortex. *Nature*. doi:10.1038/nature23020

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

Satpute AB, Kragel PA, Barrett LF, Wager TD, Bianciardi M. 2019. Deconstructing arousal into wakeful, autonomic and affective varieties. *Neurosci Lett*, Functional Neuroimaging of the Emotional Brain **693**:19–28. doi:10.1016/j.neulet.2018.01.042

Schneider DM, Nelson A, Mooney R. 2014. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* **513**:189–94. doi:10.1038/nature13724

Schriver BJ, Bagdasarov S, Wang Q. 2018. Pupil-linked arousal modulates behavior in rats performing a whisker deflection direction discrimination task. *J Neurophysiol* **120**:1655–1670. doi:10.1152/jn.00290.2018

Schwartz ZP, Buran BN, David SV. 2020. Pupil-associated states modulate excitability but not stimulus selectivity in primary auditory cortex. *J Neurophysiol* **123**:191–208. doi:10.1152/jn.00595.2019

Schwartz ZP, David SV. 2018. Focal suppression of distractor sounds by selective attention in auditory cortex. *Cereb Cortex N Y N 1991* **28**:323–339. doi:10.1093/cercor/bhx288

Schwarz DM, Zilany MSA, Skevington M, Huang NJ, Flynn BC, Carney LH. 2012. Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. *J Neurosci Methods* **206**:120–31. doi:10.1016/j.jneumeth.2012.02.013

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

Shobe JL, Claar LD, Parhami S, Bakhurin KI, Masmanidis SC. 2015. Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *J Neurophysiol* **114**:2043–52. doi:10.1152/jn.00464.2015

Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA. 2007. Temporal symmetry in primary auditory cortex: implications for cortical connectivity. *Neural Comput* **19**:583–638. doi:10.1162/neco.2007.19.3.583

Slee SJ, David SV. 2015. Rapid task-related plasticity of spectro-temporal receptive fields in the auditory midbrain. *J Neurosci* **35**:13090–102. doi:10.1523/JNEUROSCI.1671-15.2015

Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput J Psychon Soc Inc* **31**:137–49.

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**. doi:10.1126/science.aav7893

Thorson IL, Liénard J, David SV. 2015. The essential complexity of auditory receptive fields. *PLoS Comput Biol* **11**:e1004628. doi:10.1371/journal.pcbi.1004628

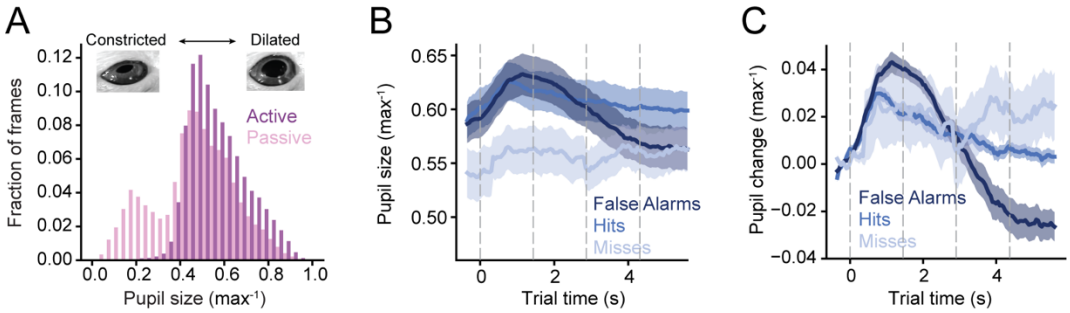
Tsunada J, Liu ASK, Gold JI, Cohen YE. 2015. Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat Neurosci* **19**:135–142. doi:10.1038/nn.4195

Vinck M, Batista-Brito R, Knoblich U, Cardin JA. 2015. Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron* **86**:740–54. doi:10.1016/j.neuron.2015.03.028

Wilcoxon F. 1945. Individual Comparisons by Ranking Methods. *Biom Bull* **1**:80–83. doi:10.2307/3001968

1021 Wimmer RD, Schmitt LI, Davidson TJ, Nakajima M, Deisseroth K, Halassa MM. 2015. Thalamic control
1022 of sensory selection in divided attention. *Nature* **526**:705–9. doi:10.1038/nature15398
1023 Winer JA. 2006. Decoding the auditory corticofugal systems. *Hear Res* **212**:1–8.
1024 Winn MB, Edwards JR, Litovsky RY. 2015. The Impact of Auditory Spectral Resolution on Listening
1025 Effort Revealed by Pupil Dilation. *Ear Hear* **36**:e153-65. doi:10.1097/AUD.0000000000000145
1026 Yin P, Fritz JB, Shamma SA. 2010. Do ferrets perceive relative pitch? *J Acoust Soc Am* **127**:1673–1680.
1027 Yin PP, Fritz JB, Shamma SA. 2014. Rapid spectrotemporal plasticity in primary auditory cortex during
1028 behavior. *J Neurosci* **34**:4396–408. doi:10.1523/JNEUROSCI.2799-13.2014
1029 Zekveld AA, Koelewijn T, Kramer SE. 2018. The Pupil Dilation Response to Auditory Stimuli: Current
1030 State of Knowledge. SAGE Publications Inc. doi:10.1177/2331216518777174
1031 Zhou M, Liang F, Xiong XR, Li L, Li H, Xiao Z, Tao HW, Zhang LI. 2014. Scaling down of balanced
1032 excitation and inhibition by active behavioral states in auditory cortex. *Nat Neurosci* **17**:841–50.
1033 doi:10.1038/nn.3701
1034

1035 **Supplementary Figures**



1036

1037 Figure 1-figure supplement 1: Pupil size correlates with task engagement and performance. **A.** Histogram
1038 of pupil size during passive and active blocks, averaged across all animals and physiological recording
1039 sessions. Pupil size is normalized to maximum per session. **B.** Mean timecourse of pupil size during
1040 behavior, grouped according to trial outcome (hit, miss, or false alarm). Mean pupil size was not
1041 significantly different for false alarm and hit trials ($p = 0.12$, Wilcoxon signed-rank test), but it was
1042 smaller during miss trials ($p = 0.0017$, Wilcoxon signed-rank test). Shading indicates SEM. **C.** Trial
1043 timecourse of average pupil change during active behavioral blocks and grouped according to
1044 performance, plotted as in B.

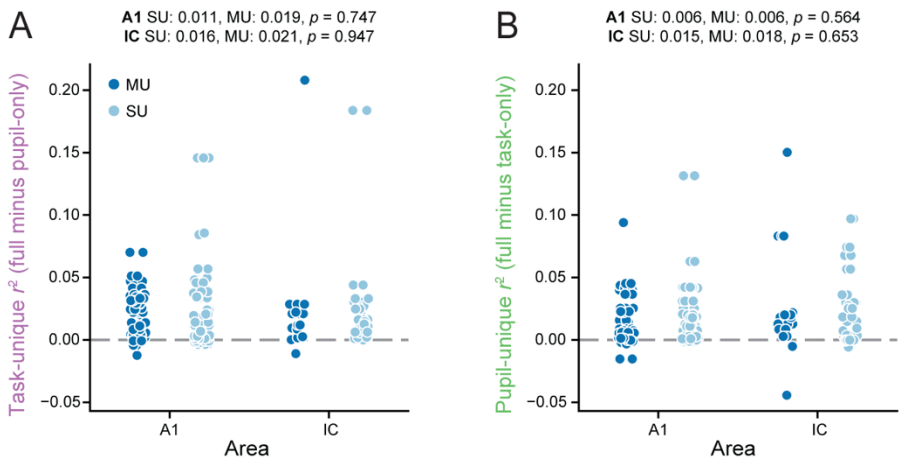


Figure 3-figure supplement 1: Unique effects of task and pupil on neural activity did not differ across single units (SU) and multi units (MU). **A.** Unique variance explained by task separated for units categorized as MU (dark blue circles) and SU (light blue circles) in A1 and IC. Medians for the different categories as well as *p* values extracted from a rank sum test run between SU and MU groups in the two areas are reported. **B.** Data as in A for unique variance explained by pupil.

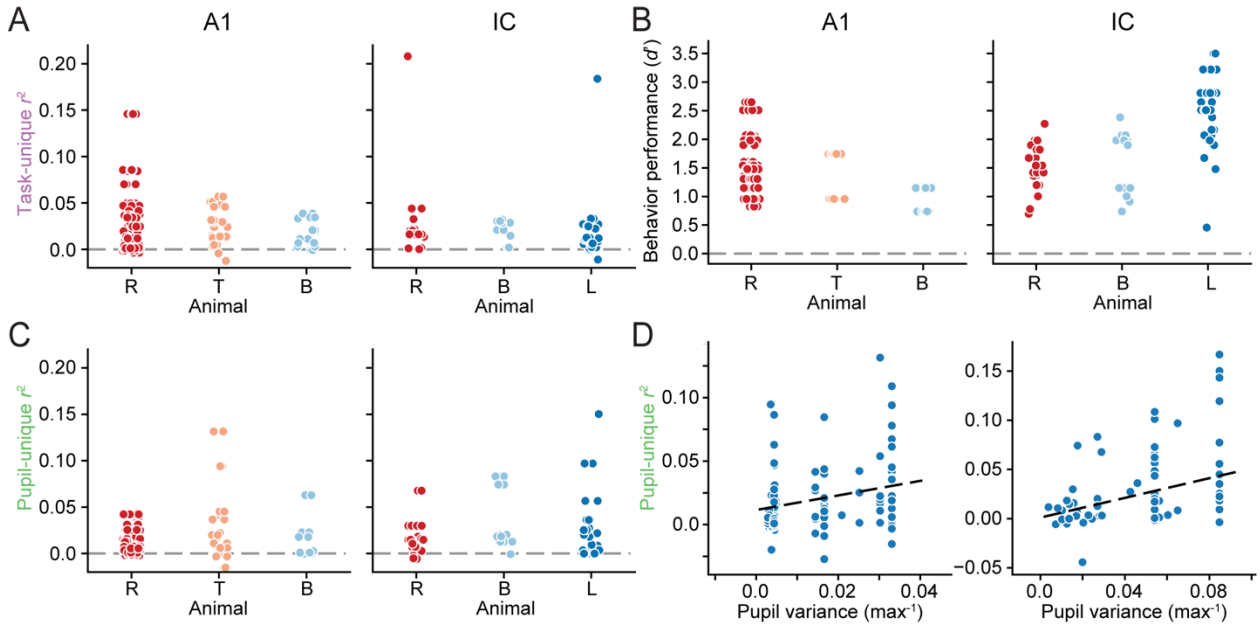


Figure 4-figure supplement 1: Variability of neural behavior state effects and task performance across animals. **A.** Unique variance explained by task engagement (r^2 task-unique) in A1 (left) and IC (right) for each unit, grouped by animal. Mean variance explained did not differ significantly between animals in either area (A1 animal R: 0.025, T: 0.023, B: 0.017, $F = 0.606$, $p = 0.547$; IC animal R: 0.028, B: 0.021, L: 0.022, $F = 0.221$, $p = 0.803$, ANOVA). **B.** Task performance (*d*') for each behavior block, grouped by animal and recordings in A1 (left) and IC (right). Mean *d*' differed between animals in both groups of experiments (A1 animal R: 1.55, T: 1.35, B: 0.93, $F = 15.8$, $p < 0.0001$; IC animal R: 1.52, B: 1.61, L: 2.48, $F = 23.9$, $p < 0.0001$, ANOVA). Interactions between animal and *d*' on neural behavior state effects are reported in Figure 4 in the main text. **C.** Unique variance explained by changes in pupil size (r^2 pupil-unique), plotted as in A. Mean variance explained differed between animals in A1, but not in IC (A1 animal R: 0.011, T: 0.030, B: 0.012, $F = 10.9$, $p < 0.0001$; IC animal R: 0.018, B: 0.038, L: 0.022, $F = 1.82$, $p = 0.170$, ANOVA). **D.** Scatter plots show r^2 pupil-unique as a function of variance of pupil size during each experiment. This relationship is significant in both A1 (left, $r = 0.341$, $p \leq 10^{-6}$) and IC (right, $r = 0.409$, $p = 0.000007$, Pearson's *R*). Dashed line shows regression fit. Multivariate regression shows that r^2 pupil-unique depends significantly on both animal and pupil variance but not on *d*' in A1 (animal: $F = 11.5$, $p = 0.000021$; pupil variance: $F = 11.4$, $p = 0.00092$; *d*': $F = 0.145$, $p = 0.703$, ANOVA). In IC,

r^2 pupil-unique depends significantly only on pupil variance (animal: $F = 2.23$, $p = 0.116$; pupil variance: $F = 16.2$, $p = 0.00016$; d': $F = 0.032$, $p = 0.86$, ANOVA).

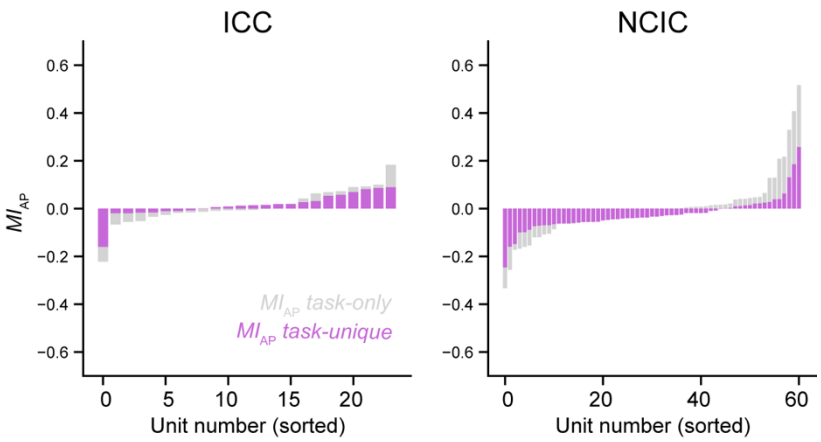


Figure 5-figure supplement 1: Task-related changes in central versus external inferior colliculus.

Histograms of MI_{AP} *task-only* and MI_{AP} *task-unique* sorted according to their magnitude for each unit in ICC (left) and NCIC (right). The median magnitude of MI_{AP} *task-unique* was reduced by 43% relative to MI_{AP} *task-only* in ICC (Wilcoxon signed-rank test: statistic = 57.0, p = 0.00788). Units in NCIC also showed a trend toward reduction, but the decrease was not significant (median decrease 31%, statistic = 865.0, p = 0.0563). Among ICC units with significant unique task modulation, n = 9/13 showed increased activity (MI_{AP} *task-unique* > 0) during behavior, and the median was not significantly different from zero (median 0.012, Wilcoxon signed rank test, statistic = 21.0, p = 0.08). Among NCIC units, only n = 8/36 were enhanced during behavior, and median was significantly less than zero (median -0.049, statistic = 118.0, p = 0.000731).

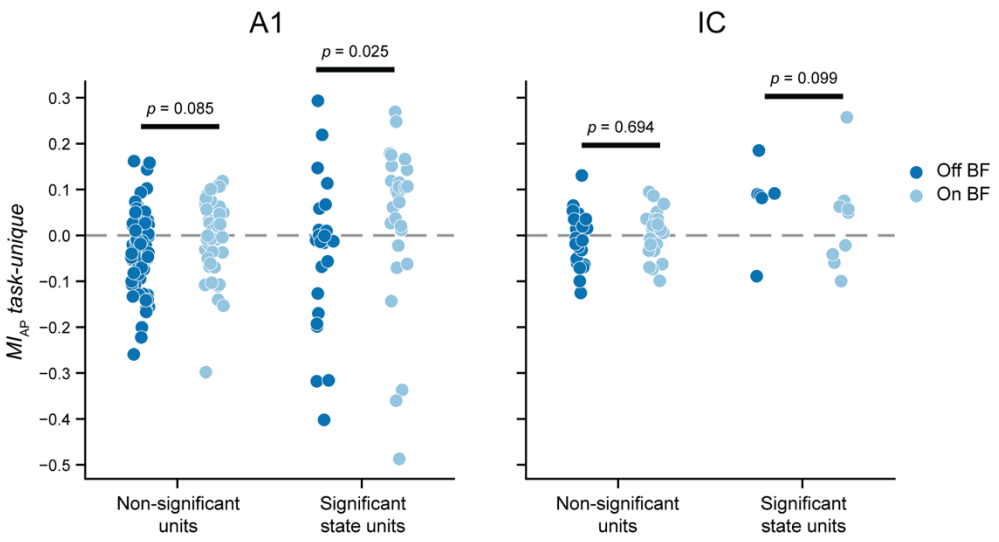


Figure 6-figure supplement 1: Frequency tuning versus task-related modulation of neural activity. Strip plots compare task-unique changes in activity (MI_{AP} *task-unique*) between units with best frequency similar to the target tone (on-BF, < 0.5 octave difference, blue) and different from the target tone (off-BF, > 0.5 octave difference, orange) in A1 (left) and IC (right). Data are further broken down by units with no task-unique changes (left) and significant changes ($p < 0.05$, permutation test, right). In A1, median MI_{AP} *task-unique* is significantly greater for modulated on-BF units (on-BF: median 0.082, $n = 28$; off-BF: -0.012, $n = 24$; $p = 0.025$, rank-sum test) but not for unmodulated units (on-BF: -0.002, $n = 44$; off-BF: -0.019, $n = 77$; $p = 0.085$). In IC, there are no differences between on- and off-BF units in the modulated (on-BF: median 0.05, $n = 9$; off-BF: 0.088, $n = 6$; $p = 0.099$, Wilcoxon signed-rank test) or unmodulated groups (on-BF: 0.006, $n = 25$; off-BF: 0.005, $n = 27$; $p = 0.694$). The relatively small number of significantly modulated IC units makes it difficult to interpret the absence of a significant on- versus off-BF difference, although it confirms that units with task-unique modulation are relatively rare in IC.

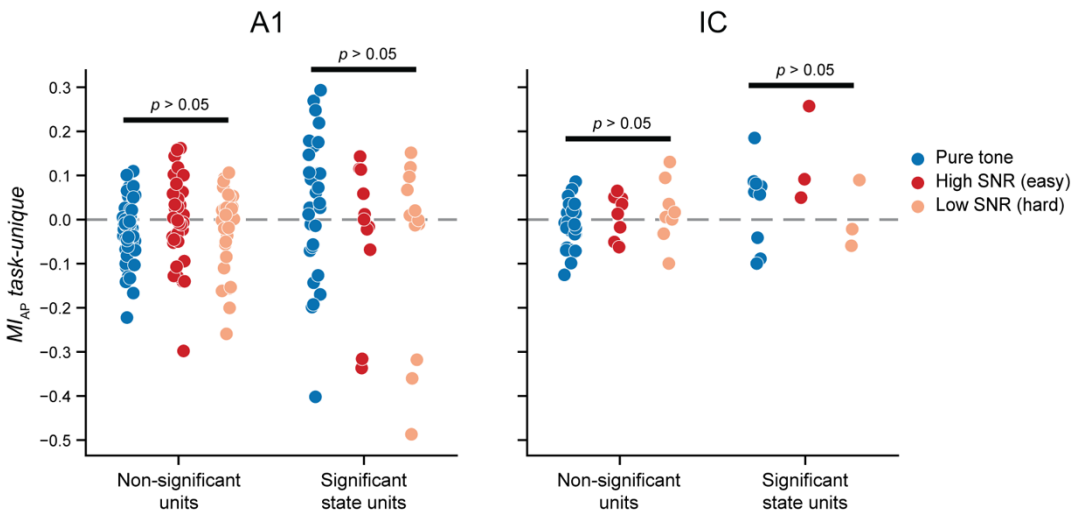


Figure 6-figure supplement 2: Task difficulty versus task-related modulation of neural activity. Swarm plots compare task-unique changes in activity (MI_{AP} *task-unique*) broken down by task difficulty, area (A1 or IC) and significance of task-unique modulation, plotted as in Figure 6-figure supplement 1. We found no difference in MI_{AP} *task-unique* between any pair of difficulty conditions in any group ($p > 0.05$, rank-sum test).

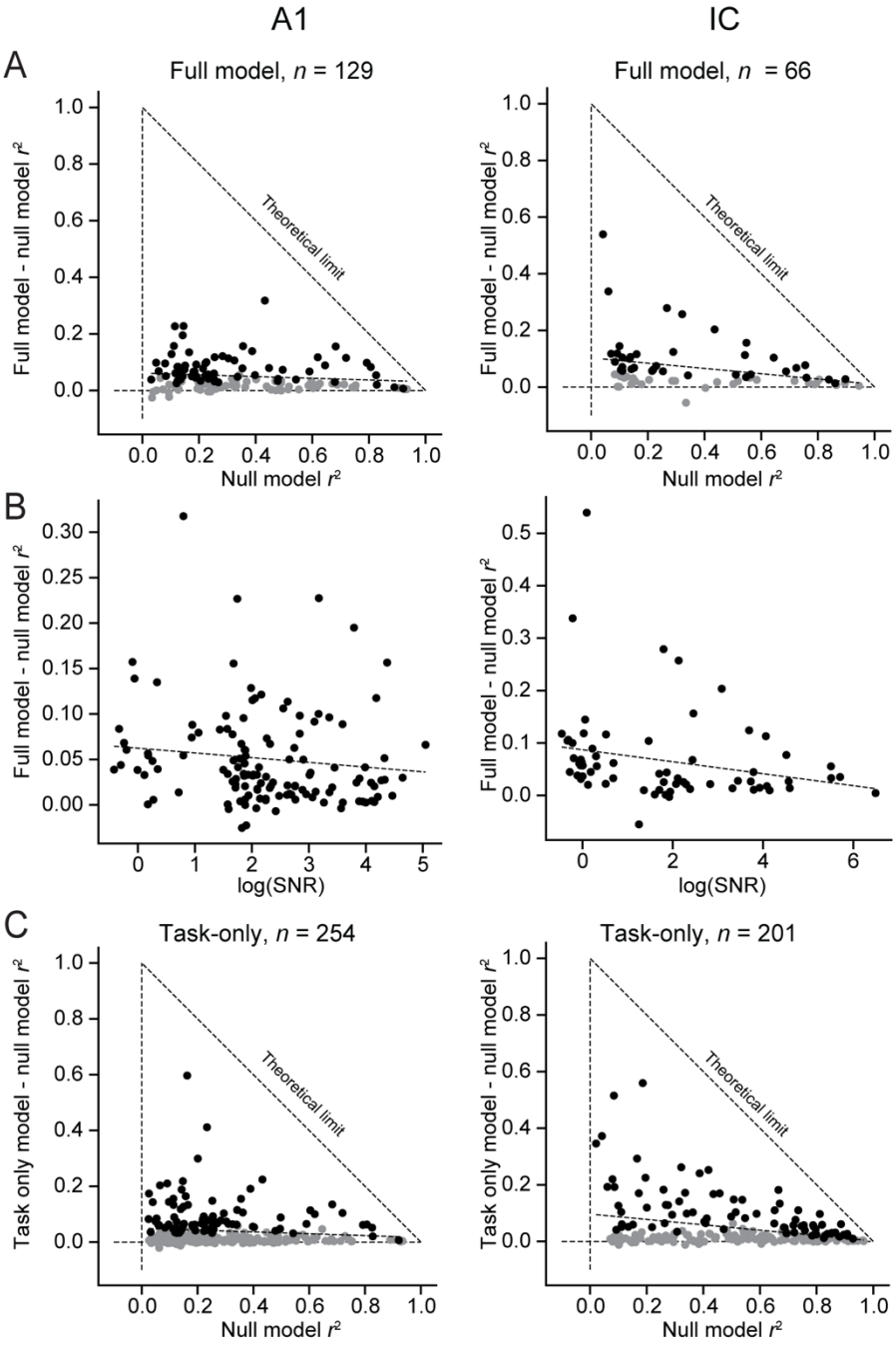


Figure 7-figure supplement 1: Auditory responsiveness predicts state-dependent modulation in IC. **A.** Scatter plots compare variance explained by the null model, a measure of auditory responsiveness for each neuron, against additional variance explained by the full model, which incorporates modulation by task engagement and pupil size. In A1 (left), the additional variance explained by the full model is uncorrelated with null model performance (n = 129, r = -0.134, p = 0.125, Pearson's R). In IC, the change

is negatively correlated (IC: $n = 66$, $r = -0.323$ $p = 0.00254$), indicating a larger behavior effect in units with weaker auditory responses. Dashed diagonal line shows the theoretical limit on additional variance explained by the full model. If the additional variance explained is normalized by this limit, the data are positively correlated in A1 ($r = 0.301$, $p = 0.000461$) and not correlated in IC ($r = 0.124$, $p = 0.258$). **B.** Scatter plots compare log signal-to-noise ratio (SNR) of response evoked by reference noise-evoked, an alternative measure of auditory responsiveness, against additional variance explained by the full model. Again, there is no significant correlation in A1 ($r = -0.121$, $p = 0.165$), but the correlation is negative in IC ($r = -0.252$, $p = 0.0202$). **C.** Variance explained by the null model versus additional variance explained by a behavior-dependent model for larger populations of neurons. Some data did not include pupillometry and the model accounted only for task engagement. Again, there was no correlation in A1 ($n = 254$, $r = -0.106$, $p = 0.0910$), but there is a negative correlation in IC ($n = 201$, $r=-0.317$, $p = 2.78 \times 10^{-6}$).