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

Zachary P. Schwartz[1], Brad N. Buran[2], and Stephen V. David[2]
[1]Neuroscience Graduate Program, [2]Oregon Hearing Research Center
Oregon Health and Science University

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

Abstract

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

Significance Statement

Cortical neurons vary in their response to repeated stimuli, and some portion of the variability is due to fluctuations in network state. By simultaneously recording pupil and single-neuron activity in primary auditory cortex of ferrets, we provide new evidence that that network state affects the overall excitability of auditory neurons, but sensory selectivity remains stable. During high-arousal states – indicated by dilated pupil – baseline activity, neuronal gain, and reliability increase, and there is a decrease in acoustic threshold. In addition, we report the occurrence of possible rapid eye movement sleep states, adding to evidence that pupil provides an index of both sleep and physiological arousal.

Introduction

Although the iris primarily functions to regulate the formation of images on the retina, pupil dilation and constriction are also a result of autonomic nervous activity unrelated to visual stimuli (Loewenfeld and Lowenstein, 1999). Growing evidence suggests that these autonomic changes in pupil diameter are an indicator of changes in behavioral state that affect processing in sensory cortex (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015). In mice, pupil dilation is correlated with transitions from stillness to walking (McGinley et al., 2015a; Vinck et al., 2015), which increase firing rates in visual cortex (Niell and Stryker, 2010) and decrease firing rates in auditory cortex (Schneider et al., 2014). Dilated pupil is associated with increases in high-frequency local field potential (LFP), a physiological signature of wakefulness, in hippocampus and sensory cortex (McGinley et al., 2015a). In mouse auditory cortex, intermediate pupil diameter is

also associated with increases in the gain of sound-evoked responses, decreased spontaneous activity, and stable, hyperpolarized subthreshold membrane potential, which have been hypothesized to reflect an "optimal state" for detecting sensory stimuli (McGinley et al., 2015a).

Pupil dilation has also been correlated with changes in sensory selectivity. In visual cortex, orientation tuning becomes more selective when pupil dilates, suggesting that sensory representations are more precise in some behavioral states (Reimer et al., 2014). However, it is not known if pupil-associated state influences the selectivity of neurons in the auditory system. Some population-level associations between pupil size and cortical activity do differ between sensory systems (Shimaoka et al., 2018). Data from mouse auditory cortex – which suggests that pupil size is non-monotonically related to neural activity (McGinley et al., 2015a) – conflicts with data from mouse visual cortex, which show either monotonic increases in neural activity when pupil dilates (Reimer et al., 2014; Vinck et al., 2015), or simultaneous increases and decreases in different neurons recorded at the same time (Stringer et al., 2019). It is not clear whether these observations reflect differences between vision and hearing or different sampling of behavioral states, cell types, or cortical layers.

To explore effects of pupil-associated state on sensory selectivity in auditory cortex, we recorded pupil size and single-unit spiking activity in the primary auditory cortex (A1) of head-restrained, non-anesthetized ferrets during presentation of natural sounds and random tone-pip sequences. We characterized effects of pupil by comparing responses to each distinct sound as pupil size varied. Neurons typically showed an increase in baseline firing rate as well as the gain and reliability of sound-evoked activity when pupil was large, although there was also evidence for non-monotonic effects of pupil-associated state in some neurons. In most neurons, parametric measurements of tuning, including acoustic threshold, spectral bandwidth, and best frequency, remained stable across large changes in pupil size. Across the population, however, we observed a small decrease in acoustic threshold when pupil was dilated. In some recordings, we observed increased saccadic eye movements during periods of sustained pupil constriction. Including the presence of this rapid eye movement (REM) sleep-like state as a separate variable in a regression model of neural variability accounted for some, but not all, of the variability associated with changes in pupil size. The stability of best frequency across large changes in pupil size contrasts with task-related plasticity of receptive fields in ferret auditory cortex (Fritz et al., 2003; David et al., 2012), suggesting a difference between effects of engaging in tasks that require auditory processing and variation within passive behavioral states.

Materials and Methods

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

Neurophysiology

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

After a unit was isolated during tungsten recordings, its auditory responsiveness was tested using tones or narrowband noise bursts. If the site showed no response, the electrode was lowered to search for another site. During depth-array recordings, the probe was lowered into

auditory cortex until auditory-evoked activity was observed across the span of recorded channels. Amplified and digitized signals were stored using open-source data acquisition software (Englitz et al., 2013; Siegle et al., 2017). Ferrets were observed by video monitor during recordings. If a unit became indistinguishable from the noise floor following a body movement, the recording was stopped and the 5 trials preceding the putative isolation loss (28 to 40 seconds of recording time) were excluded from analysis. A1 recording sites were identified by characteristic short latency, frequency selectivity, and tonotopic gradients of single or multiunit activity across multiple penetrations (Bizley et al., 2005; Elgueda et al., 2019).

Spike sorting

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

Pupillometry

During neurophysiological recordings, infrared video of one eye was collected for offline measurement of pupil size. Video was collected using either (1) a commercial CCTV camera fitted with a lens to provide magnification of the eye and a band-stop infrared filter (Heliopan ES 49 RG 780) or (2) an open-source camera (Adafruit TTL Serial Camera) fitted with a lens (M12 Lenses PT-2514BMP 25.0 mm) whose focal length allowed placement of the camera 10 cm from the ferret's eye. To improve contrast, the imaged eye was illuminated by a bank of infrared LEDs. Ambient luminance was provided using a ring light (AmScope LED-144S). At the start of each recording day, the intensity of the ring light was set to a level (~1500 lux measured at the recorded eye) chosen to give a maximum dynamic range of pupil sizes. Light intensity remained fixed across the recording session.

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

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

Blink artifacts were identified by rapid, transient changes in pupil size (McGinley et al., 2015a). The derivative of the pupil trace was taken and bins with derivatives more than 6 standard deviations from the mean were marked. Blinks were identified within these bins by

screening for decreases in pupil size followed by increases. Data during a 1-second period surrounding the blink was then removed from the trace and replaced by a linear interpolation of the pupil size immediately before and after the blink.

When comparing pupil and neural data, a 750 ms offset was applied to pupil trace to account for the lagged relationship between changes in pupil size and neural activity in auditory cortex (McGinley et al., 2015a). Setting the offset at values between 0 and 1.5 s did not affect the mean accuracy of the pupil-based regression models of neural activity described below (data not shown).

To measure changes in eye position, we calculated the Euclidean distance between the center of the ellipse fit to the pupil in each video frame, then multiplied by frame rate to find eye speed. We identified rapid eye movement like (REM-like) states by screening for a combination of tonically constricted pupil and a high rate of saccades. REM-like states were identified by selecting parameter values for each recording (maximum pupil size, maximum pupil standard deviation, minimum saccade speed, minimum saccade rate) that marked abrupt state transitions in the traces of pupil size and eye speed. Gaps of less than 15 seconds between REM episodes were considered artifacts and coded as REM. We restricted our analysis to REM episodes with a duration of at least 30 seconds.

Acoustic stimuli

Stimulus presentation was controlled by custom MATLAB software (code available at https://bitbucket.org/lbhb/baphy). Digital acoustic signals were transformed to analog (National Instruments), amplified (Crown), and delivered through a free-field speaker (Manger).

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

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

During later recording sessions, ferrets were presented with sequences of 10-ms tone pips with varying frequency (4-6 octaves surrounding best frequency, ranging over 32 Hz to 40 kHz) and level (0 to 65 dB SPL, 5 or 10 dB SPL steps). Each pip sequence was 6 seconds long, preceded by a 2-second silent period to allow measurement of spontaneous neural activity. Tone-pip sequences were repeated multiple times in each recording (mean repetitions +/- SEM: 28 +/- 2). During some recordings, tone-pip sequences were interleaved with ferret vocalizations to improve sampling of behavioral states in each data set.

Data analysis: Gain and baseline firing rate

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Data analysis: Non-monotonicity

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

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

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

$$\begin{aligned}
&\text{if } \mathrm{p}_i(\mathrm{t,s}) < \beta_0: && \mathrm{baseline_i(t,s) = [\beta_1 + \beta_2 p_i(t,s)]m_0} \\
&\text{else:} && \mathrm{baseline_i(t,s) = [shift_{base} + \beta_3 p_i(t,s)]m_0} \\
&\text{where:} && \mathrm{shift_{base}} = \beta_1 + \beta_0[\beta_2 - \beta_3] \\
&\text{if } \mathrm{p}_i(\mathrm{t,s}) < \beta_4: && \mathrm{gain_i(t,s) = [\beta_5 + \beta_6 p_i(t,s)]r_0(t,s)} \\
&\text{else:} && \mathrm{gain_i(t,s) = [shift_{gain} + \beta_7 p_i(t,s)]r_0(t,s)} \\
&\text{where:} && \mathrm{shift_{gain}} = \beta_5 + \beta_4[\beta_6 - \beta_7]
\end{aligned}$$

and

$$\mathrm{r_i(t,s) = baseline_i(t,s) + gain_i(t,s) r_0(t,s)}$$

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

To identify non-monotonic neurons, we compared the accuracy of the segmented linear model to an identical model that was constrained to be monotonic. That is, the slope coefficients related to baseline (β_2, β_3) and gain (β_6, β_7) were constrained to have the same sign on both sides

of the breakpoint. If the non-constrained model showed a sign change as well as an improvement in accuracy over the constrained model, we classified the cell as showing a *non-monotonic trend*.

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

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

Data analysis: Frequency and level tuning

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

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

The hierarchical models were implemented in STAN (http://mc-stan.org) and fit using Bayesian inference. In contrast to conventional model fitting, Bayesian analysis allows for simple calculations of credible intervals on derived parameters (e.g., parameters that are mathematical functions of fitted coefficients), is more robust to outliers through the use of Poisson likelihoods for spike count data, and offers simple construction of realistic hierarchical models (Gelman et al., 2003). Each model was fit four times for 1000 samples after a 1000 sample iteration burn-in period. Posterior samples were combined across all chains for inference. Visual inspection of model predictions with observed data were used to verify goodness of fit. Gelman-Rubin were diagnostics are used to assess whether the four fits, each of which started with a random estimate for each parameter, converge onto the same final estimate. A Gelman-Rubin diagnostic less than 1.1 indicates good agreement among the fits (Gelman et al., 2003). For the rate level model, Gelman-Rubin diagnostics for all parameters were below 1.1. For the frequency tuning model, parameters for some cells had Gelman-Rubin diagnostics greater than 1.1; however, visual inspection of the observed data for these cells indicated that there were several reasonable solutions, which was reflected in the predictions. These cells are highlighted in the appropriate plots.

Data analysis: Rate-level functions

To estimate changes in level tuning linked to behavioral state, we first calculated characteristic frequency (CF) before binning data by pupil size. An FRA for all data from the recording was

calculated using data from the stimulus-evoked epoch and smoothed with a two-dimensional box filter with dimensions equivalent to 15 dB and 1.5 octaves. A separate standard error was calculated for each coefficient of the FRA by measuring variance across repetitions of tone pips at that frequency and level. The minimum sound level that evoked a response at least 2 standard errors above the mean spike rate during the 100 ms preceding pip onset, measured across all pips, was calculated for each frequency. CF was defined as the frequency that required the minimum sound level to evoke a response.

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

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

Here, *n* is the number of spikes observed in a time interval, *t*, for cell, *c*, and pupil condition, *p*, on the *i*-th observation of stimulus level, *l*. Since the time interval varied across cells, depending on how long we maintained stable recordings, incorporating this information into our model gave greater weight to data points with longer time intervals.

For levels less than threshold, firing rate was expected to be equal to the baseline firing rate. For stimulus levels greater than threshold, the cell could have an excitatory ($s_c > 0$) or inhibitory ($s_c < 0$) response. Since the FRA was assessed using a cloud of tone pips presented in rapid succession, the cell's firing rate during the 50 ms preceding sound onset (i.e., prestimulus activity) was a better measure of the cell's baseline response for sub-threshold levels than spontaneous activity. To improve our estimate of the baseline rate, spike counts from this period were included when fitting the model.

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

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

1 Since the expected value of the Gamma distribution is α/β, we can compute the posterior for the average spontaneous rate of the population as b_α/b_β when the pupil is small and $b_\alpha/b_\beta \times bg_\mu$ when the pupil is large.

Data analysis: Frequency-tuning curves

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

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

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

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

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

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

Data analysis: REM-like states

To test for neural correlates of REM-like states, we defined a binary variable that indicated the presence or absence of REM ($\mathrm{rem}_i(t, s)$) and added it as a term in the baseline-gain model (see above),

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

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

$$\mathrm{r_i(t,s)} = \beta_0 \mathrm{m_0} + \beta_1 \mathrm{r_0(t,s)} + [\beta_2 \mathrm{m_0} + \beta_3 \mathrm{r_0(t,s)}]\mathrm{rem_i(t,s)}$$

Model fitting and evaluation were completed as for the baseline-gain model.

Experimental design and statistical analyses

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

Code

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

Results

Dilated pupil is correlated with increases in the spontaneous activity, gain, and reliability of auditory neurons

To examine the relationship between behavioral state and neural representation of natural sounds, we simultaneously recorded infrared (IR) video of the eye and single-unit spiking activity in the primary auditory cortex (A1) of head-restrained, non-anesthetized ferrets (Fig 1a, n = 114 neurons from 46 recording sites in 3 ferrets). Pupil size could be extracted from the video, providing a measure of behavioral state (Reimer et al., 2014; McGinley et al., 2015a; Vinck et al., 2015). In order to gather data on the response to the same sounds across a wide range of behavioral states, our stimulus consisted of a small number of sounds (two ferret vocalizations) repeated many times (up to 120 repetitions per recording, mean +/- SD = 77 +/- 40 repetitions). To further increase our sampling of behavioral states, we paused this presentation at set times in the recording to rouse the ferret using auditory stimuli (see Materials and Methods). The variability of neural activity across repetitions of the same vocalization segments exceeded what would be predicted by models of spiking as a Poisson process (Fig. 1b-d, mean Fano factor +/- SEM = 1.5 +/- 0.07). This degree of variability is consistent with previous in vivo studies in cortex (Shadlen and Newsome, 1998; Churchland et al., 2010; Goris et al., 2014).

Although lighting conditions were held static during recordings, we also observed large changes in pupil size across timescales of tens of seconds to minutes (Fig. 2a). Visual inspection of spike rasters suggested that trial-to-trial variability in neural activity sometimes tracked changes in pupil size (Fig. 2b-d).

To quantify the association between pupil size and firing rate, we fit a linear model that allowed baseline spike rate and gain of sound-evoked responses to depend on pupil size (baseline-gain model, see Materials and Methods). In a majority of neurons (n = 75/114, 66%) this model was significantly more accurate at predicting single-trial neural activity than a control model where pupil was shuffled in time (i.e., the prediction was based only on the neuron's peristimulus time histogram [PSTH] response to each vocalization) (Fig. 3a, $p < 0.05$, permutation test). The degree of improvement was not correlated with the prediction accuracy of the control model alone (r = -0.08, p = 0.5, t-test), suggesting that the influence of pupil on trial-to-trial variability did not depend on a neuron's overall auditory responsiveness.

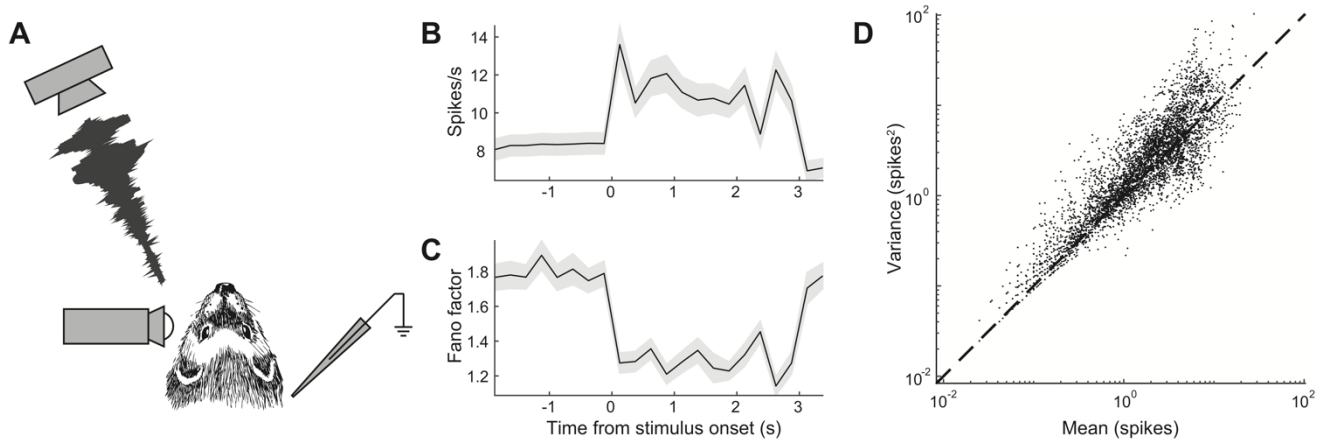

Figure 1: Ferret auditory cortex shows trial-to-trial variability in neural responses to sound. (A) Schematic of experiment, illustrating ferret, free-field speaker, camera recording video of eye, and contralateral extracellular electrode. (B) Mean response to sound across all neurons in the vocalization dataset (n = 114). Shading indicates SEM. (C) Time-varying Fano factor for all neurons in the vocalization dataset. Shading indicates SEM. (D) Variance-to-mean relationship for all neurons in the vocalization dataset. Each data point illustrates the mean and variance of the spike count of one neuron during one 250-ms segment of one vocalization or the surrounding intervals. Dashed line indicates prediction for a Poisson process (slope = 1).

The baseline-gain model included separate terms for effects of behavioral state on gain and baseline firing rate, both of which tended to be positive, indicating an increase in firing rate with increasing pupil size (Fig. 3b). Both the baseline and gain terms contributed to model prediction accuracy, compared to models that included only one term (Fig. 3c, Table 1). Adding second-order terms improved median prediction accuracy further (Table 1), and increased the number of neurons that showed a significant fit from 75 to 79.

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

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

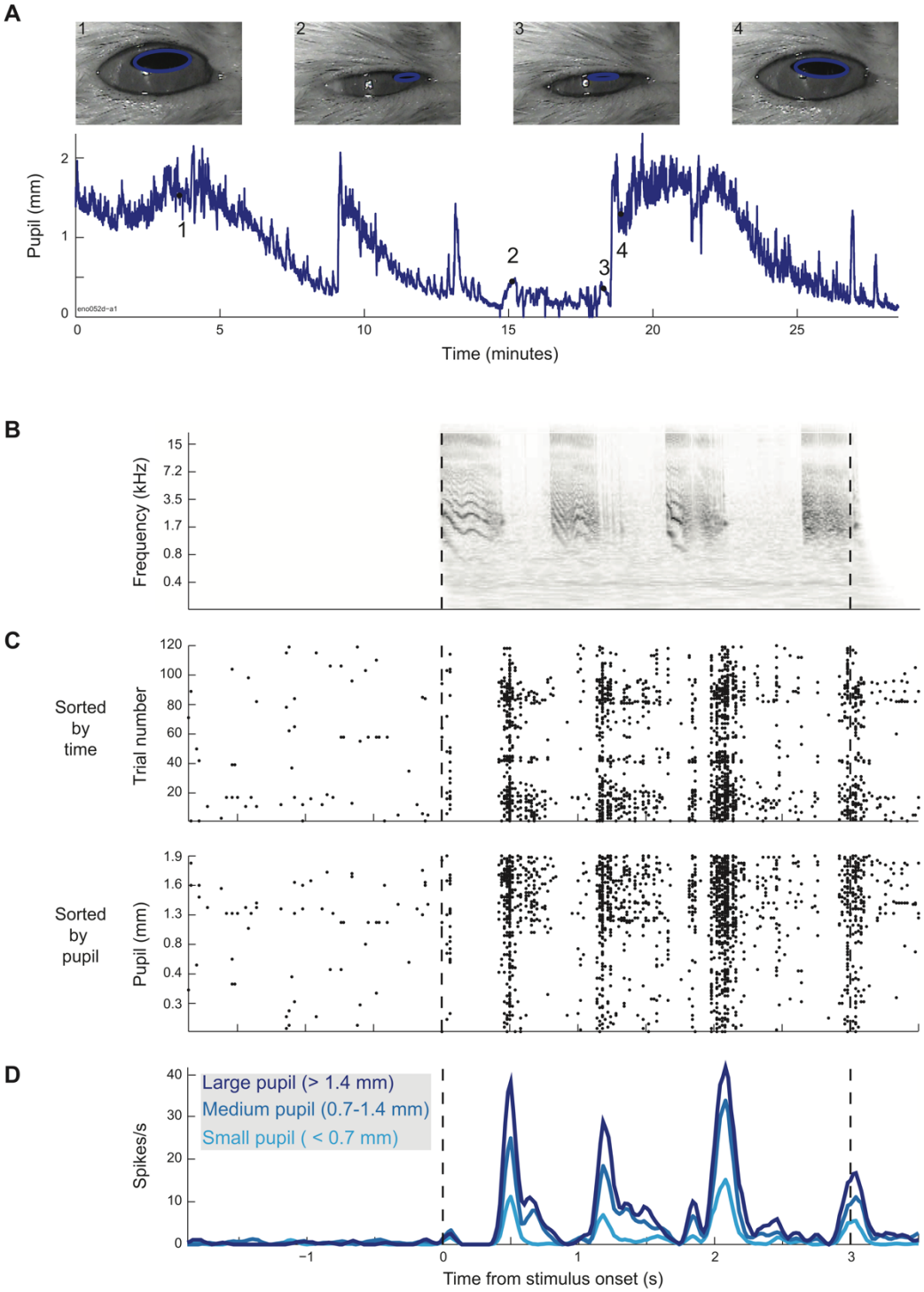

Figure 2: Trial-to-trial variability in neural responses to sound tracks changes in pupil size. (A) Pupil trace recorded across 120 repetitions of one ferret vocalization, with example video frames. (B) Spectrogram of one ferret vocalization. (C) Spike raster of one neuron's response to the vocalization in Fig. 2B recorded during the same experiment as in Fig. 2A, sorted by time (top) and prestimulus pupil size (bottom). (D) Peristimulus time histogram (PSTH) of the same unit, binned by prestimulus pupil size.

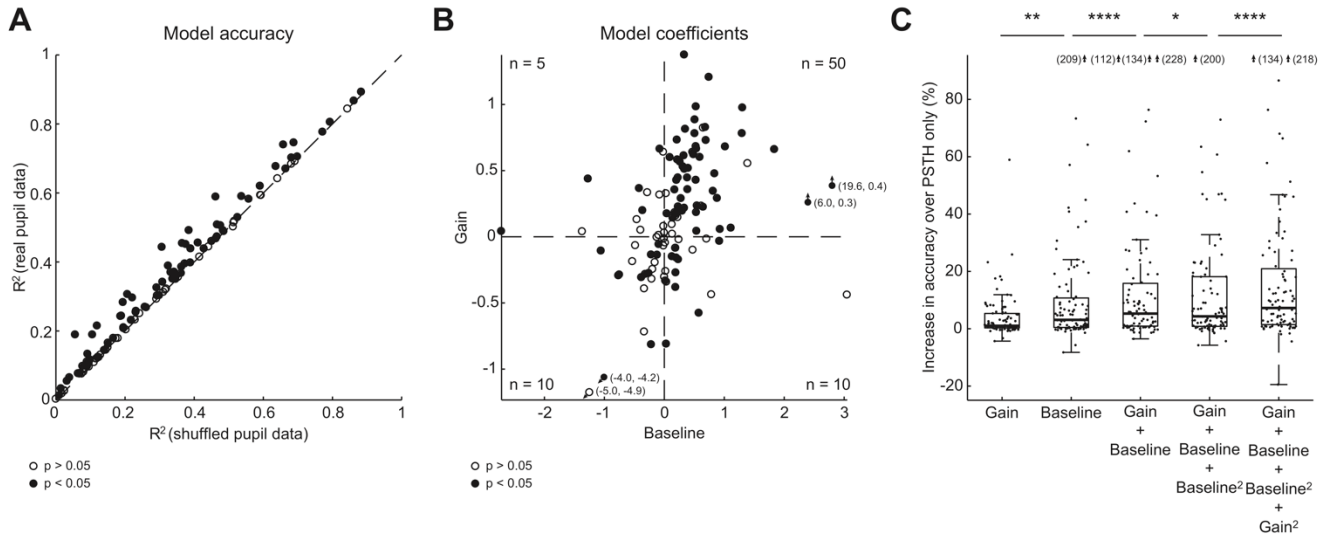

Figure 3: Baseline firing rate and sound-evoked responses of cortical neurons increase when pupil is large. (A) Accuracy of pupil-based baseline-gain regression model for neural responses to ferret vocalizations (n = 114 neurons), plotted against accuracy of a control, PSTH-only model, fit to temporally shuffled pupil data from each cell. Filled dots indicate neurons with a significant improvement over the control model (n = 75/114, p < 0.05, permutation test). (B) Coefficients of baseline-gain model for all data from vocalization recordings (n = 114 neurons). Positive values indicate an increase in baseline spike rate (horizontal axis) or response gain (vertical axis) with larger pupil. Numbers in corners of each quadrant indicate the count of neurons with parameters in that quadrant and a significant improvement in prediction accuracy over the control, pupil-independent model. (C) Improvement in accuracy over PSTH-only model for various models for population of neurons with a significant fit for any model (n = 92/114, *: p < 0.05, **: p < 0.01, ****: p < 0.0001, sign test). Boxplot indicates median, interquartile range, and 1.5 times the interquartile range.

Table 1. Stepwise regression of pupil-based models of neural activity

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

Table 1: Stepwise comparison of accuracy of different pupil-based models of neural activity applied to vocalization dataset. Numbers indicate difference in median accuracy across the population of neurons with significant fit for any model (n = 92/114). Bold text indicates comparisons for which the difference is significant at alpha level 0.05 with Bonferroni correction for multiple comparisons (*: p < 0.05, **: p < 0.01, ***: p < 0.001, ****: p < 0.0001, sign test).

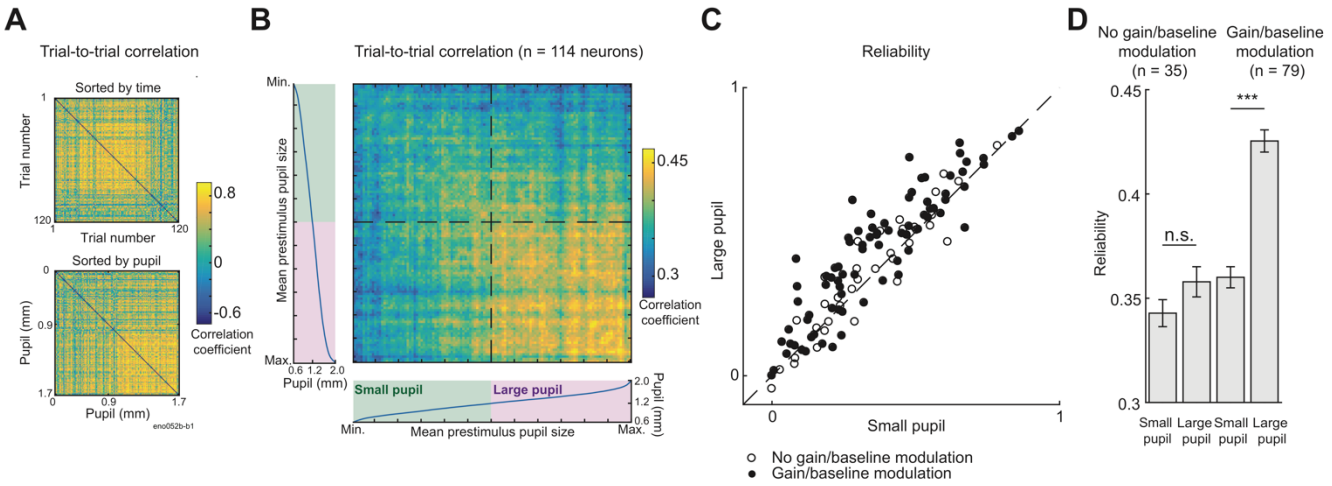

Figure 4: Reliability of neural response to sound increases when pupil is dilated. (A) Correlation between time-varying activity across trials evoked by one ferret vocalization in one neuron, sorted by trial order (top) and mean pupil size before stimulus onset (bottom). (B) Mean trial-to-trial correlation for all neural responses to ferret vocalizations. This heat map was constructed by computing the average correlation matrix sorted by pupil size across all neurons (n = 114). For display, the correlation of each trial with itself was replaced by the mean correlation for the two trials with the most similar pupil size before averaging. Marginal plots indicate the mean pupil size in each trial, across all neurons. Dashed line indicates median pupil size. (C) Comparison of reliability (mean trial-to-trial correlation) for each neuron's response to vocalizations (n = 114) for small versus large pupil. Trials are classified as "large pupil" and "small pupil" based on whether the mean pupil size preceding the vocalization is greater or less than the median pupil size during the recording. Filled circles indicate cells with a significant fit under a second-order baseline-gain regression model (n = 79, p < 0.05, permutation test). (D) Mean reliability (+/- SEM) in each condition for subpopulations of cells that do or do not show effect of pupil-associated state under the regression model (***: p < 0.001, paired t-test).

Previous research indicates that the level of pupil-indexed arousal is associated with changes in the reliability of neurons in auditory cortex, where reliability is defined as the mean cross correlation between spiking responses to the same stimulus across repetitions (Reimer et al., 2014; McGinley et al., 2015a). To test for this effect, we calculated the correlation between the spiking response evoked by each repetition of each vocalization (Fig. 4a). Inter-trial correlations increased with prestimulus pupil size, indicating a more reliable response to sound (Fig 4a-b). To quantify the change in reliability, we split data into large-pupil and small-pupil trials based on whether mean pupil diameter in the silent interval preceding each vocalization was greater or less than the median pupil size during the recording (Fig. 4c). Across the entire population, reliability was significantly greater during large-pupil trials (Fig. 4c, n = 114 neurons, p = 5e-8, paired t-test). This difference in reliability was preserved in the subset of cells that showed modulation of neural activity by pupil-indexed state under our second-order baseline-gain model (Fig. 4d, n = 79 neurons, p = 5e-8, paired t-test), but disappeared when comparing only cells that did not show pupil-associated modulation under this model (Fig. 4d, n = 35 neurons, p = 0.2, paired t-test). These results suggest that pupil-associated state affects both the mean rate at which auditory neurons respond to sound and the variability of their responses, and that these effects occur in the same population of cells.

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

Arousal can have a non-monotonic effect on behavior: both learning (Yerkes and Dodson, 1908; Diamond et al., 2007) and task performance (Aston-Jones and Cohen, 2005; Murphy et al., 2011; van Kempen et al., 2019) are sub-optimal in minimally and maximally aroused states. A previous study found that multiunit spiking activity in mouse auditory cortex showed similarly non-monotonic relation between pupil and firing rate: on average, spontaneous activity was lowest, and gain highest, at intermediate pupil diameter, suggesting that this range of pupil sizes was optimal for detection of auditory signals (McGinley et al., 2015a). We found examples of non-

monotonic pupil-firing rate relationships in some cells (Fig. 5a). To test for their prevalence, we fit data on neural responses to ferret vocalizations with a segmented regression model (Simonsohn, Uri, 2018) that predicted a linear effect of pupil on neural activity but allowed the slope of the fit to take two different values depending on pupil size (Fig. 5b). A difference in the sign of the slope between segments indicated a non-monotonic relationship between pupil size and spiking activity. We then compared the accuracy of this model to a similar model in which slope could vary between segments, but both line segments were constrained to have the same sign (i.e., the fit could not assume the shape of a U or inverted U).

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

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

To examine the effect of changes in behavioral state on stimulus selectivity, we recorded the responses of A1 neurons to tone pips at a variety of frequencies and levels (Fig. 6a-c, 114 neurons from 40 recording sites in 4 ferrets). To test for changes in frequency and level selectivity, we divided the data from each cell into large-pupil and small-pupil bins based on the median pupil size observed during the recording (Fig. 6a-b), and constructed frequency response areas (FRAs) from the data in each bin (Fig. 6c). FRAs showed a variety of patterns typical of the auditory system (Bizley et al., 2005), including a sound-level response threshold and broadening of spectral bandwidth at higher sound levels (Fig. 6c).

Comparing data from the large and small-pupil bins showed a systematic increase in spontaneous activity and the response to tones across all levels and frequencies (Fig. 6b,d-e, , n = 114 neurons, p = 1e-4 [spontaneous rate], p = 0.03 [driven rate], sign test). To confirm that this result did not depend on our division of the data into large and small-pupil bins, and to further examine changes in driven rate, we used linear regression to predict the response to tones based on pupil size and the average FRA of the neuron (see Materials and Methods). As it did for vocalizations, pupil size predicted trial-to-trial variability in a subpopulation of cells (Fig. 6f, n = 57/114 neurons, 50%, $p < 0.05$, permutation test). Most state-modulated neurons showed enhanced gain and baseline firing rate when pupil was large (Fig. 6g), suggesting that pupil-indexed state acted to shift and multiplicatively scale the neuron's auditory tuning curve.

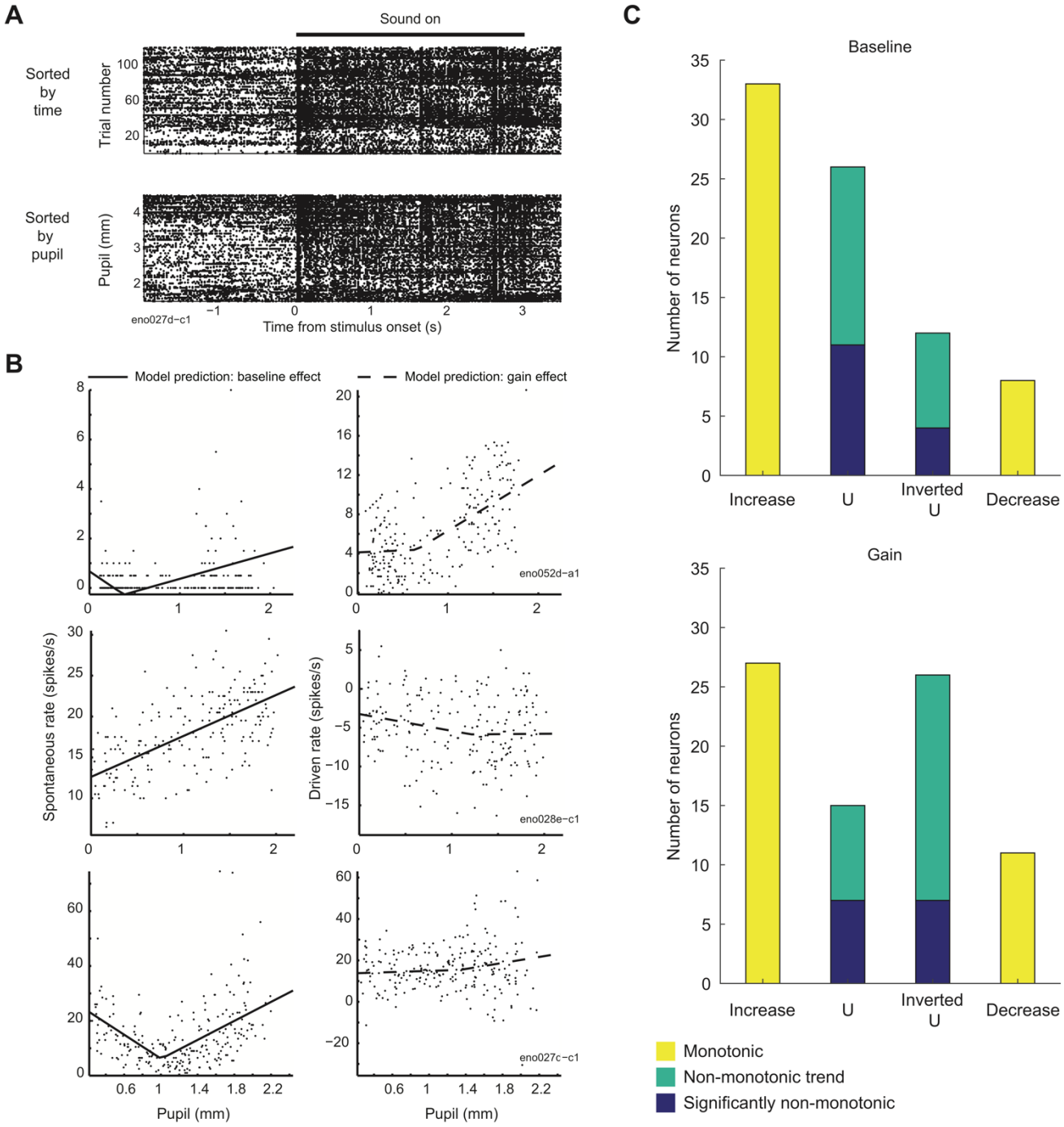

Figure 5: Non-monotonic effects of pupil-related state. (A) Spike raster of one neuron's response to single ferret vocalization sorted by time (top) and pupil size preceding the stimulus (bottom). A non-monotonic relationship is evident between pupil size and spontaneous firing rate. (B) Examples of segmented regression model fits to spontaneous and driven activity for three neurons. Each point represents spike rate for one neuron on one trial (spontaneous activity: 2 s, driven activity: 3 s). The bottom row displays data from the neuron in panel A. (C) Results of test for non-monotonicity using segmented regression model, counting the number of neurons with significant or trends toward non-monotonic changes in baseline or gain.

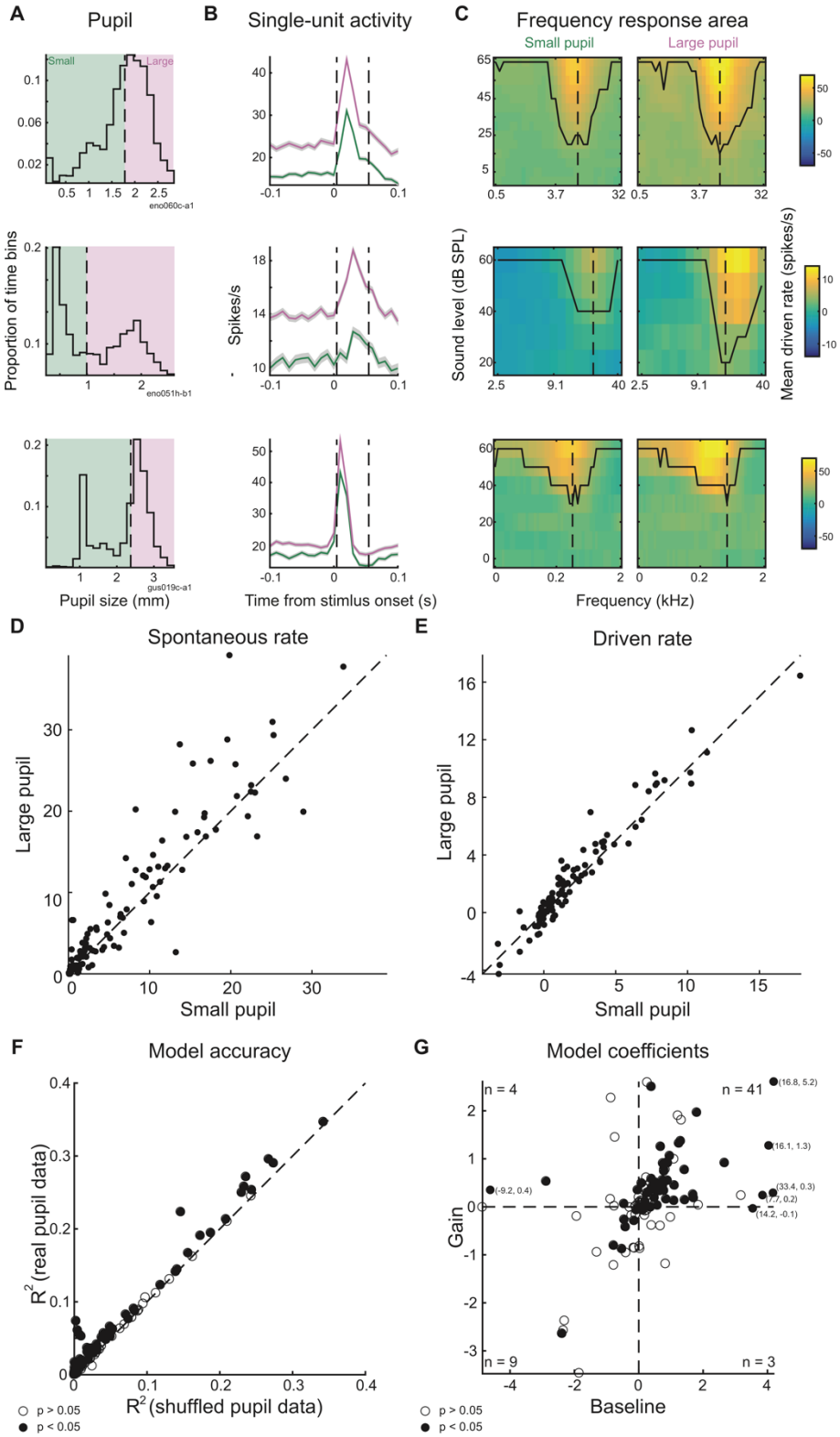

Figure 6: Spontaneous activity and evoked response to tones increases when pupil is dilated. (A) Distribution of pupil size during three recordings of neural responses to tone pips, indicating division into small-pupil and large-pupil bins based on median pupil size during the recording. (B) Mean peristimulus time histograms (PSTH)s of neural responses for the three example cells, computed separately for data in the smal-pupil (green) and large-pupil (purple) bins and averaged across all tone pip levels and frequencies. Shading indicates SEM. Dashed lines indicate window for measurement of response to sound. (C) Frequency-response areas for the three examples cells, computed separately for small and large pupil bins. Contour indicates lowest level that shows an auditory response at each frequency. Dashed line indicates characteristic frequency. (D-E) Spontaneous and driven rate for all neurons' responses to tone pips (n = 114 neurons). (F) Single-trial prediction accuracy of linear baseline-gain model and control, pupil-independent model for all neurons' response to tone pips. Filled dots indicate neurons with a significant improvement for the baseline-gain model (n = 57/114, p < 0.05, permutation test). (G) Coefficients of baseline-gain model fit for all data from tone-pip recordings. Numbers indicate count of neurons with significant improvement for the pupil-dependent model in each quadrant.

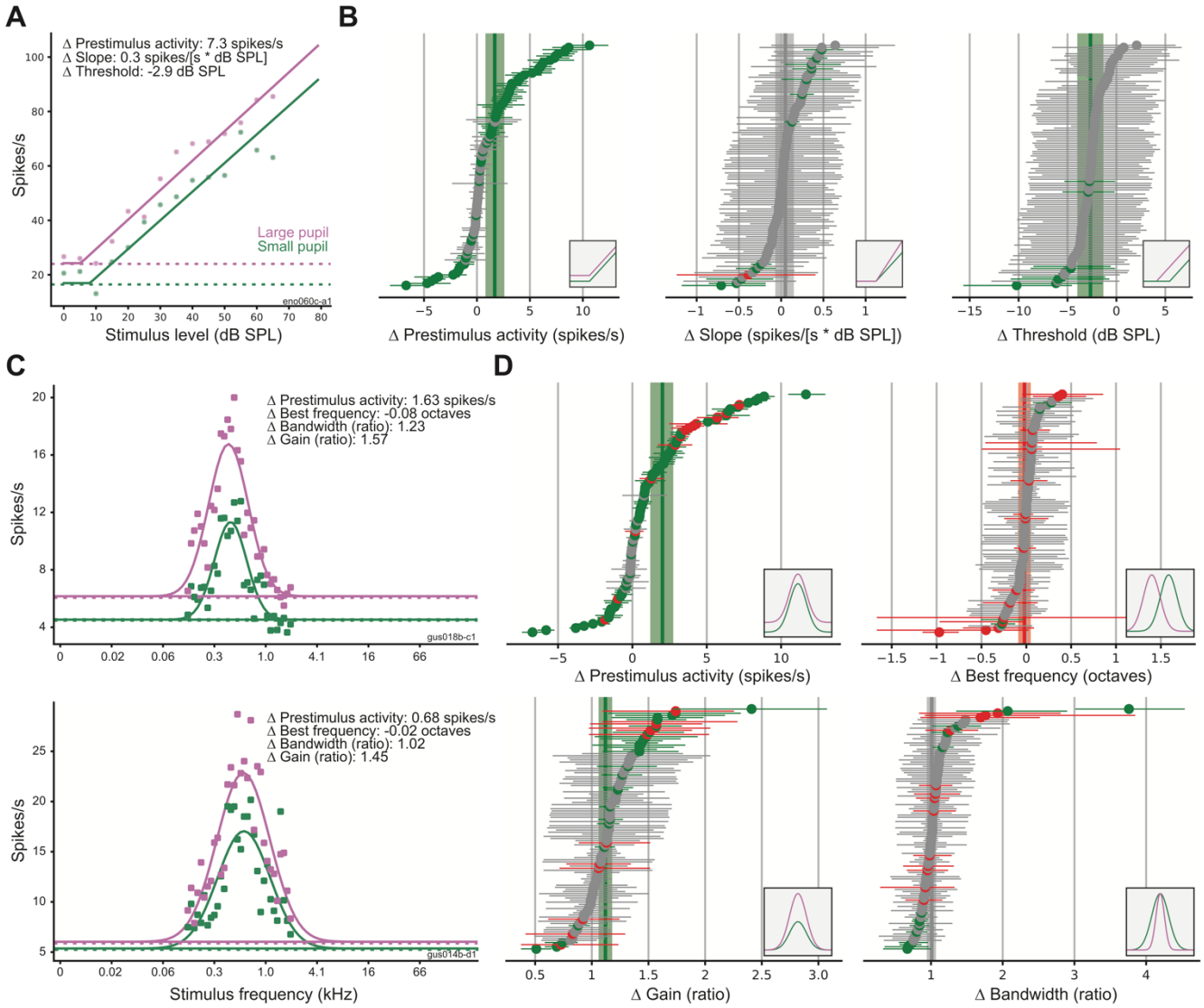

Figure 7: Effects of pupil-related state on frequency and level tuning. (A) Example of hinge function fit to neural response at characteristic frequency for large-pupil (purple) and small-pupil (green) conditions. Dashed lines indicate prestimulus firing rate in each condition. (B) Change in each parameter of rate-level function between pupil conditions (large pupil – small pupil) for all neurons (n = 114). Inset cartoons illustrate effect of changing each parameter fit. Red dot indicates a neuron whose fit failed to converge. Green dots indicate neurons with a significant change in the tuning parameter (90% credible interval outside 0). Gray dots indicate neurons with changes that were not significant at the single-cell level. Vertical line and shading shows mean and 90% credible interval for the population. (C) Examples of Gaussian functions fit to neural responses at best level for large-pupil (purple) and small-pupil conditions (green). (D) Change in each parameter of frequency-tuning curve across pupil conditions (gain and bandwidth: large pupil/small pupil, baseline and best frequency: large pupil – small pupil) for all neurons (n = 114). Insets and color coding as in panel B.

Does the stimulus selectivity of auditory neurons change depending on pupil-indexed state? To test for changes in level tuning, we again divided data at the median pupil value in each recording, then fit a hinge function (see Methods and Materials) to the characteristic-frequency rate-level function in each pupil condition (Fig. 7a-b). Across the population, there was a significant decrease in threshold when pupil was large (mean change: -2.65 dB, 90% credible interval: -4.00 to -1.33 dB), but no change in the slope of the rate-level function (mean change: 0.05 [spikes/s]/dB, 90% credible interval: -0.07 to 0.15 [spikes/s]/dB SPL). Consistent with previous analyses, we also observed an increase in the baseline firing rate of the cells when pupil was large (mean change: 1.68 spikes/s, 90% credible interval: 0.80 to 2.59 spikes/s). At the single-cell level, most neurons showed a significant change in baseline firing rate (n = 68/114), but few showed significant changes in slope (n = 12/114) or threshold (n = 8/114). The low number of cells showing significant changes in threshold may be due to the small effect size (-2.65 dB average across the population) and the uncertainty in model fits for individual cells.

To test for changes in frequency tuning, we fit a Gaussian function to the best-level frequency tuning curve (FTC) in each pupil condition (Fig. 7c-d). To isolate changes in frequency tuning (i.e., bandwidth) from nonspecific changes in baseline firing rate or gain, our model included multiplicative gain term and additive offset terms. Consistent with data from previous analysis, the gain and baseline parameters of the FTC showed a systematic increase across the population when pupil was large (gain: mean large/small ratio = 1.12, 90% credible interval = 1.06 to 1.18, baseline: mean change = 2.02 spikes/s, 90% credible interval = 1.20 to 2.75). There was no change in the spectral bandwidth of the FTC when pupil was large (mean large/small ratio = 1.01, 90% credible interval = 0.94 to 1.07). There was also no change in mean best frequency across pupil conditions (mean change = -0.02 octaves, 90% credible interval = -0.09 to 0.05). The number of neurons that showed a significant change in each parameter was greater for gain (n = 27/114) and baseline firing rate (n = 78/114) than it was for bandwidth (n = 14/114) or best frequency (n = 4/114). Thus, while A1 neurons did sometimes show pupil-dependent changes in response threshold or other tuning parameters, these changes were relatively small compared to the changes in baseline activity and response gain.

REM-like states account for additional neural variability

Although we did not explicitly seek to study sleep state, during some recordings we observed periods of constricted pupil accompanied by an increase in saccade-like eye movements (Fig. 8a-b, Movies 1 and 2). The short duration of the episodes (mean +/- SEM = 2.0 +/- 0.2 minutes, n = 49 episodes) was consistent with previous reports of rapid eye-movement (REM) sleep in ferrets (Marks and Shaffery, 1996 [mean: 5.6 minutes]; Jha et al., 2006 [mean: 3.7 minutes]).

Given that neural responses to sound in auditory cortex are preserved during natural sleep states, but sometimes suppressed or enhanced (Issa and Wang, 2008), we wondered if these episodes might show changes in firing rate distinct from those associated with changes in pupil size. We therefore fit linear regression models that included putative REM state, pupil size, or both REM state and pupil size as predictors of neural activity (Fig. 8c-d and Table 2). REM state significantly predicted neural activity in 81% of neurons recorded in experiments that included one or more REM bouts (n = 38/47 neurons). Across the population of neurons with a significant fit for any tested model (n = 44), REM state predicted less variability than pupil, which in turn predicted less variability than both pupil and REM state (Fig. 8d and Table 2, median change in accuracy over

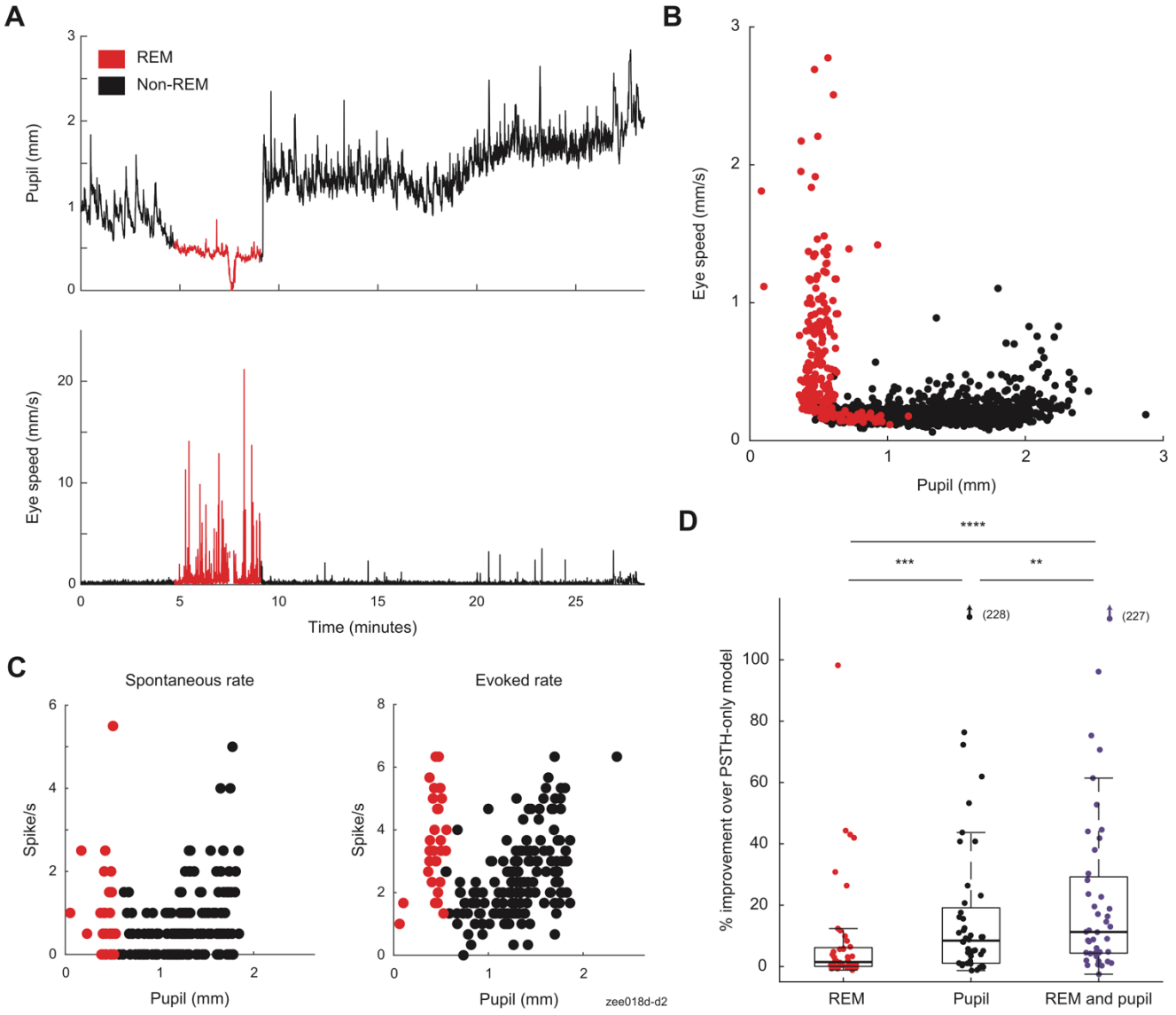

Figure 8: Rapid eye movement (REM)-like states account for additional variability in neural responses. (A) Pupil size and eye-speed traces from one recording. (B) Pupil size and eye speed data from one ferret (13 recordings). Each point illustrates average pupil size and eye speed during one trial (5.5 s). Color indicates whether a trial was classified as REM (red) or non-REM (black), using criteria based on eye movement and pupil size data (see Methods). (C) Spontaneous (left) and sound-evoked neural data (right) from the recording shown in A. Each point indicates spike rate before or during one vocalization (prestimulus period = 2 s, stimulus period = 3 s). (D) Stepwise regression results for models including REM state, pupil size or both REM state and pupil size as predictors of spike rate (n = 44 neurons with significant fit for any model and recorded during experiments that include REM episodes, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$, sign test). Boxplot indicates median, interquartile range, and 1.5 times the interquartile range.

Table 2. Stepwise comparison of pupil- and/or REM-based models of neural activity

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

Table 2: Stepwise comparison of accuracy of pupil and/or REM-state based models of neural activity (see Methods) applied to all vocalization data including REM episodes (n = 47 neurons). Text indicates median accuracy for population of neurons with significant fit for any model (n = 44/47 neurons). Bold text indicates that the difference is significant at alpha level 0.05 with Bonferroni correction for multiple comparisons (*: p < 0.05, **: p < 0.01, ***: p < 0.001, ****: p < 0.0001, sign test).

PSTH-only model: 1.5% [REM], 8.4% [pupil], 11% [pupil and REM]). There was no correlation between the duration of REM during the recording and the improvement in accuracy of the REM-based model (r = 0.2, p = 0.16, t-test), suggesting that REM effects vary across neurons.

In some neurons, visual inspection of spiking activity and pupil state indicated that REM-like episodes were associated with a non-monotonic change in firing rates (Fig. 8c). To quantify this effect, we compared trials recorded during REM episodes to non-REM trials falling in an equivalent range of pupil sizes (Fig. 9a-b). Across the population, there was no significant difference in spontaneous rate, sound-evoked rate, or reliability between these two conditions (Fig. 9c, p = 0.16 [spontaneous rate], p = 0.38 [evoked rate], p = 0.17 [reliability], sign test, n = 47 neurons with recordings that included REM episodes,). However, a subpopulation of neurons did show a difference in spontaneous or sound-evoke rates at the single-cell level (Fig. 9c, n = 7/47 [spontaneous rate], n = 15/47 [evoked rate], p < 0.05, unpaired t-test). To further examine the effect of REM state, we removed data from trials recorded during REM episodes from the dataset and refit the segmented regression model initially used to test for non-monotonic effects. Fitting to data that excluded REM state reduced the number of neurons that showed non-monotonic effects in both baseline and gain (Fig. 9d), indicating that REM accounted for non-monotonic effects in some neurons (n = 7/42 [change in non-monotonicity of baseline effect], n = 6/42 [change in non-monotonicity of gain effect], n = 42 neurons with recordings that included REM episodes and a significant fit for second-order baseline-gain model).

Discussion

Our results show that pupil size, an index of behavioral state, is correlated with changes in the gain and baseline firing rate of neurons in the primary auditory cortex of non-anesthetized ferrets. Non-monotonic effects of changes in pupil state were observed in some neurons, but the majority of effects were monotonic and showed a positive correlation between pupil size and spike rate. Across our population of recorded neurons, pupil size was also correlated with increases in the reliability of responses to sound, decreases in acoustic threshold, and no change in spectral bandwidth or best frequency. Together, these observations suggest that pupil size tracks the gross level of activity evoked by auditory stimuli: sounds become more salient to the rest of the brain. While relatively small, the changes in threshold and reliability may also support a more sensitive or precise representation of sound in high-arousal states.

Pupil-associated changes in sensory tuning

In mouse visual cortex, single neurons' orientation selectivity increases when pupil is dilating rather than constricting: the neurons' response to their preferred direction, but not the

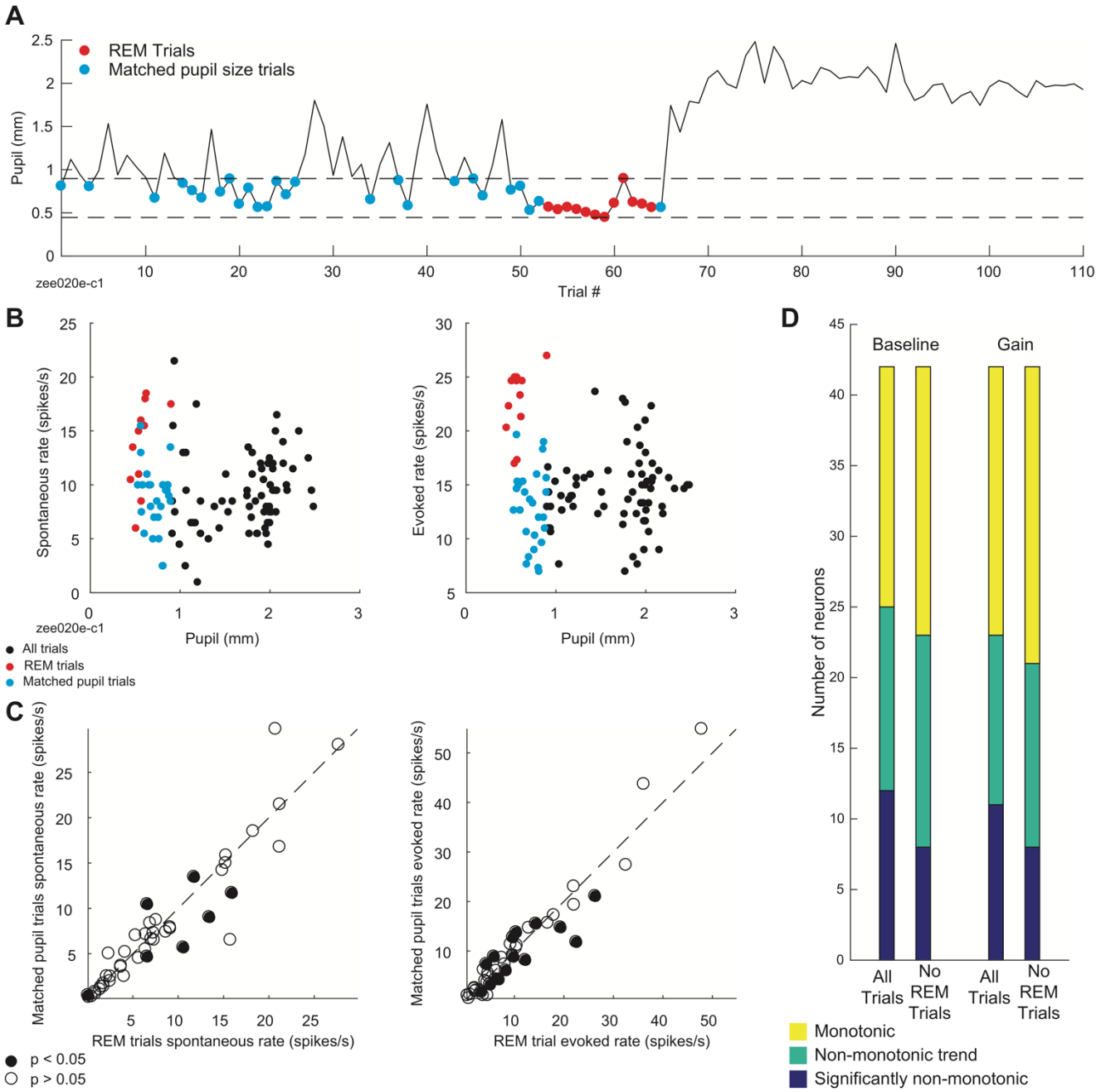

Figure 9: REM state accounts for non-monotonicity in some neurons. (A) Mean prestimulus pupil size for responses to one ferret vocalization, highlighting REM trials (red) and trials with matched prestimulus pupil size (blue). (B) Spontaneous (i.e. prestimulus) and stimulus-evoked neural activity from the recording in A, illustrating greater spike rate in REM trials as compared to trials with matched pupil size. (C) Average spontaneous (left) and sound-evoked spike rate (right) during REM trials versus trials with matching pupil size (n = 47 neurons recorded in experiments including REM episodes, filled circles: $p < 0.05$, unpaired t-test). (D) Results of test for non-monotonicity using segmented regression model fit to all data or data with REM trials removed (n = 42 neurons recorded in experiments including REM episodes, and showing significant fit to second-order baseline-gain regression model).

orthogonal direction, increases (Reimer et al., 2014). Although the impact of this change in orientation selectivity on tuning bandwidth has not been quantified, it appears to involve a narrowing of orientation tuning in the mean response across the population (Reimer et al., 2014). Our observation of constant tuning bandwidth in A1 may therefore indicate differences between how pupil-indexed states affect sensory selectivity in visual and auditory cortex.

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

Comparison to previous pupillometry studies

We used absolute pupil size as a measurement of behavioral state. Our work therefore complements human pupillometry's traditional focus on small, rapidly-decaying pupil dilations that coincide with behavioral events (Kahneman and Beatty, 1966; Beatty, 1982a). The amplitude of these task-evoked dilations depends on a variety of cognitive variables (Zekveld et al., 2018). Although the variations in pupil size we observed provide a measurement of physiological arousal, they do not allow us to infer changes in a particular cognitive state, given that they occurred outside a controlled behavior. However, mechanisms like those we observed may underlie correlations between absolute pupil size and the efficiency of responses to sensory stimuli in some tasks (Beatty, 1982b; Murphy et al., 2011; van Kempen et al., 2019).

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

Several factors could explain this inconsistency between studies. The difference could reflect sampling of cortical layers. The earlier study targeted layers 4/5. We did not target any particular layer, but our method of recording auditory cells as the electrode advanced through cortex may have biased our sample towards more superficial layers. It is also possible that, compared to the earlier study, we tended to sample a different range of behavioral states, either due to differences in experimental technique (the earlier study recorded head-restrained animals on a treadmill, while we did not use a treadmill) or differences in the behavioral patterns of ferrets and mice. Ferrets were not able to run and rarely showed substantial motor activity during the recordings. Thus, they may not have achieved the very high arousal state observed during bouts of running and other motor activity in mice.

Pupil as an index of noradrenergic tone

Activity in neuromodulatory centers, particularly the noradrenergic and cholinergic systems, has been proposed as a mechanism underlying correlations between pupil size and the level of neural activity in sensory cortex (McGinley et al., 2015b; Reimer et al., 2016). Evidence from multiple labs, species, and experimental techniques indicates that central release of noradrenaline from locus coeruleus is causally involved in pupil dilation and that noradrenaline release in sensory cortex accompanies pupil dilation (Aston-Jones and Cohen, 2005; Murphy et al., 2014; Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017; Liu et al., 2017; Lovett-Barron et al., 2017; Larsen et al., 2018). Despite the substantial evidence linking pupil size to noradrenergic tone, the gain increases we observed during pupil dilation do not directly match in vivo measurements of the effects of noradrenaline in auditory cortex. Iontophoresis of noradrenaline increases the signal-to-noise ratio of some auditory cortex neurons via a decrease in spontaneous, but not sound-evoked, activity (Foote et al., 1975; Manunta and Edeline, 1997, 1999). The gain increases we observe persist after subtracting mean spontaneous activity and are present when analyzing data from the evoked period alone. It is possible that iontophoresis does not replicate the spatial distribution or temporal dynamics of noradrenaline release in non-anesthetized animals. It is also possible that the gain increases we observed were a result of multiple modulatory systems acting on auditory cortex, both directly and indirectly. Activity in the cholinergic and dopaminergic systems are also correlated with changes in pupil size (Reimer et al., 2016; de Gee et al., 2017; Larsen et al., 2018). Given that pupil size is correlated with activity in multiple neuromodulatory systems, our failure to directly replicate results from spatially-localized, single-neuromodulator experiments is unsurprising.

Involvement of other cortical and subcortical brain regions

Some of the trial-to-trial variability we observed in auditory cortex may also be due to feedback from motor cortex (Schneider et al., 2014). Pupil size varies even when mice are still, and explains more variability in neural activity than locomotion (McGinley et al., 2015a). However, in non-anesthetized mice motor activities such as running, whisking, and licking a water reward are associated with pupil dilation (Lee and Margolis, 2016; Stringer et al., 2019). Detailed tracking of face and body movements explains more variability than pupil alone in mouse visual cortex (Stringer et al., 2019) and dorsal cortex (Musall et al., 2018). We did not track detailed motor behavior in the current study, and thus the extent to which motor activity predicts trial-to-trial variability in our data is not known. However, our results add to the evidence that pupil is one of several variables related to movement that can be used to predict cortical activity.

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

Sleep states during head-restrained recordings

Our observation of rapid eye movement (REM)-like sleep states complicates the suggestion that the pupil size of head-restrained animals reveals a continuum of awake arousal states analogous to, but distinct from, sub-states of sleep (McGinley et al., 2015b). Our data is consistent with recent results in head-restrained mice that showed fluctuations between awake, non-REM, and REM sleep states associated with changes in pupil size, with constricted pupil associated with REM (Yuzgec et al., 2018). Thus, a full characterization of arousal states must incorporate both pupillometry and sleep state.

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

Acknowledgments

This work was supported by grants from the National Institute on Deafness and Other Communication Disorders (R01 DC0495, R21 DC016969, F31 DC016204) and the ARCS Foundation Oregon Chapter. The authors would like to thank Daniela Saderi, Luke A. Shaheen, and Sean Slee for training, advice, and assistance with neurophysiological recording, Ulysses Duckler for animal care, and Matthew J. McGinley, Avinash Bala, and Robert Peterka for technical assistance with pupillometry. The authors declare no competing financial interest.

Supplementary Material

Movie 1. Real-time example of pupil size and eye movement dynamics during passive, non-REM-like state.

Movie 2. Real-time example of pupil size and eye movement dynamics during REM-like state.

References

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

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

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

Bizley JK, Nodal FR, Nelken I, King AJ (2005) Functional organization of ferret auditory cortex. Cereb Cortex N Y N 1991 15:1637–1653.

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

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

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

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

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

Elgueda D, Duque D, Radtke-Schuller S, Yin P, David SV, Shamma SA, Fritz JB (2019) State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. Nat Neurosci Available at: https://doi.org/10.1038/s41593-018-0317-8.

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.

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

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

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

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

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

Goris RLT, Movshon JA, Simoncelli EP (2014) Partitioning neuronal variability. Nat Neurosci 17:858–865.

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

Jha SK, Coleman T, Frank MG (2006) Sleep and sleep regulation in the ferret (mustela putorius furo). Behav Brain Res 172:106–113.

Joshi S, Li Y, Kalwani RM, Gold JI (2016) Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron 89:221–234.

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

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

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

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

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

Loewenfeld IE, Lowenstein O (1999) The Pupil: Anatomy, Physiology, and Clinical Applications. Oxford: Butterworth-Heinemann.

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

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

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

Marks GA, Shaffery JP (1996) A preliminary study of sleep in the ferret, mustela putorius furo: a carnivore with an extremely high proportion of REM sleep. Sleep 19:83–93.

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

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

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

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

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

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

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

Musall S, Kaufman MT, Gluf S, Churchland AK (2018) Movement-related activity dominates cortex during sensory-guided decision making. bioRxiv:308288.

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

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

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

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

Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD (2016) Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv Available at: http://biorxiv.org/content/early/2016/06/30/061481.abstract.

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

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

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

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

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

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

Simonsohn, Uri (2018) Two-lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. Adv Methods Pract Psychol Sci Forthcoming Available at: https://ssrn.com/abstract=3021690.

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

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

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

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

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

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