Cortical efficient coding dynamics shape behavioral performance.

Christopher F. Angeloni[1,2], Wiktor Młynarski[3], Eugenio Piasini[4], Aaron M. Williams[2,4], Katherine C. Wood[2], Linda Garami[2], Ann Hermundstad[5], Maria N. Geffen[2,4]

[1]Psychology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA
[2]Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, PA, USA
[3]Institute of Science and Technology Austria, Klosterneuburg, Austria
[4]Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, PA, USA
[5]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Abstract

The efficient coding hypothesis postulates that neurons shape their response properties to match their dynamic range to the statistics of incoming signals. However, whether and how the dynamics of efficient neuronal adaptation inform behavior has not been directly shown. Here, we trained mice to detect a target presented in background noise shortly after a change in the background contrast. The observed changes in cortical gain and detection behavior followed the predictions of a normative model of efficient cortical sound processing; specifically, target detection and sensitivity to target volume improved in low contrast backgrounds relative to high contrast backgrounds. Additionally, the time course of target detectability adapted asymmetrically depending on contrast, decreasing rapidly after a transition to high contrast, and increasing more slowly after a transition to low contrast. Auditory cortex was required for detection of targets in background noise and cortical neuronal responses exhibited the patterns of target detectability observed during behavior and in the normative model. Furthermore, variability in cortical gain predicted behavioral performance beyond the effect of stimulus-driven gain control. Combined, our results demonstrate that efficient neural codes in auditory cortex directly influence perceptual behavior.

Introduction

As we perceive the world around us, the statistics of the environment can change dramatically. In order to maintain stable percepts, it is crucial for the nervous system to adapt to persistent statistical properties of sensory inputs. The efficient coding hypothesis postulates that the nervous system accomplishes this by matching the limited dynamic range of individual neurons to the statistics of incoming sensory signals[1], allowing them to encode information within many types of environments[2–4]. Neuronal adaptation to environmental statistics has been found in many sensory modalities and species[5–13]. In the auditory system, neurons exhibit contrast gain control, adapting the gain of their response function to match the variability in level (contrast) of the incoming sounds[14–19]. Yet it remains unknown whether and how the dynamics of contrast gain control in the auditory system inform behavior, as a direct link between neuronal adaptation and behavior has not been previously established. The goal of our study was to test the hypothesis that the dynamics of contrast gain adaptation in auditory cortex reflect efficient coding of incoming sounds in a manner that shapes behavioral performance in an auditory task.

The efficient coding hypothesis has been formally implemented through normative models of brain function[3,4,20–23]. These models assess whether and how neuronal adaptation shapes sensory information and simulate how such adaptation might constrain behavior. Models based on efficient codes explain psychophysical biases[24] and shape the rate of information transmission when stimulus statistics change dynamically[21,23]. Psychophysical studies suggest that perception is altered by efficient adaptation to stimulus statistics. In humans, target volume discriminability is greater in low contrast than in high contrast, an effect consistent with gain control observed in primary auditory cortex[19]. Similar relationships between efficient neural encoding and behavioral percepts of sound location have also been found in ferrets[10] and in guinea pigs[25]. However, because neural and behavioral responses in these studies were not simultaneously measured, a direct relationship between gain control and perceptual performance has yet to be assessed. Additionally, recent theoretical work demonstrated that neuronal dynamics reflect efficient coding[21,23], but it is unclear whether and how these dynamics shape behavioral performance.

Our first goal for the study was to build a formal framework based on efficient coding to model the dynamics of contrast gain control and thereby predict how behavioral performance should adapt after a change in contrast. We then derived a novel procedure for estimating moment-to-moment changes in neural gain based on generalized linear models (GLM) and found the dynamics of gain control in auditory cortex matched the predictions of the efficient coding model. Next, to directly test the role of efficient coding in auditory behavior, we trained mice to detect targets in different contrast backgrounds. Contrast-induced changes in behavioral target detection threshold, sensitivity, and background adaptation dynamics followed the normative model predictions. Furthermore, we found that auditory cortex was necessary for target detection in the presence of a background. Building on this finding, we found that the dynamics of cortical encoding of targets were similar to the model predictions and to observed behavioral adaptation, and that population activity in auditory cortex predicted individual variability in task performance. Finally, we estimated cortical gain during the task, finding that variability in neural gain predicted variability in task performance. Combined, our results identify a novel relationship between efficient neuronal coding and acoustic behavior, and they provide a normative framework that can be used to predict the dynamics of behavioral performance in response to changing sensory environments.

Results

A novel target-in-background detection task and normative model for task predictions.

To assess how perceptual performance is impacted by stimulus contrast, we devised a GO/NO-GO task in which mice were trained to detect targets embedded in low and high contrast backgrounds. During each trial, the mouse was presented with dynamic random chords (DRCs) of one contrast, which switched after 3 s to the other contrast. At variable delays after the contrast switch, broad-band target chords were superimposed on the background chords, and mice were trained to lick for a water reward upon hearing the target (henceforth, we refer to high-to-low contrast trials as "low contrast" and low-to-high contrast trials as "high contrast", referring to the contrast where mice detected targets). Target trials were interleaved with background-only trials, during which the mouse was trained to withhold licking, but would receive a 7s timeout for licking after the contrast switch (Figure 1a,b). To assess behavioral sensitivity to targets, we parametrically varied target volume in each contrast and to assess behavioral adaptation, we parametrically varied target timing (Figure 1c). This stimulus design allowed us to quantitatively test whether and how the dynamics of adaptation to background contrast affect behavioral performance.

To predict the optimal time course of contrast gain control and its impact on target detection behavior, we developed a normative model of task performance constrained by efficient neural coding. In this model, we simulated a neuron designed to encode stimuli with minimal error. To efficiently exploit its finite dynamic range, the model neuron estimated the contrast of the recent stimuli, and adjusted the gain of its nonlinearity to minimize the error in estimated contrast (Figure 1d, panels 1-3; *Online Methods*)[21,23]. Adding targets at different levels and times relative to contrast transitions allowed us to probe the sensitivity of the model neuron to targets of varying strength over the time course of adaption (Extended Data Figure 1c,d). When varying target strength and measuring model psychometric performance (*Online Methods*), we found decreased detection thresholds and steeper slopes in low contrast relative to high contrast (Figure 1e). When varying target timing, two factors affected target discriminability: 1) A change in the stimulus distribution after the contrast switch; 2) The effect of gain adaptation on responses to the background (Figure 1f,g; Extended Data Figure 1c,d). These dynamics were well characterized by a single effective timescale, which we quantified by fitting an exponential function to each transition. The normative model presented three primary predictions: When adapted to low contrast, 1) target detection thresholds will be lower and 2) model psychometric functions will have steeper slopes; 3) Discriminability over time will be asymmetric: rapidly decreasing after a switch to high contrast, and slowly increasing after a switch to low contrast (Figure 1h).

Estimated cortical gain dynamics follow normative model predictions.

Previous work on contrast gain control used static models of contrast gain control, measuring steady-state gain after the neuron fully adapted to the new stimulus[14,16,17,19], but see[15,26]. To measure the dynamics of gain control, we developed a Poisson GLM to estimate the gain of neurons in auditory cortex over time following a contrast transition. This model was fit to data recorded from the auditory cortex of a naive mouse (n = 97 neurons) presented with 3 s alternations of low and high contrast DRCs (Figure 2a,b).

The inference model is a Poisson GLM that decomposes the relationship between spiking activity (y_t) and the presented sounds into a stimulus component (x_t), contrast component ($\overline{\sigma}/\sigma_t$), and an interaction between the stimulus and the contrast ($x_t * \overline{\sigma}/\sigma_t$, where $\overline{\sigma}$ is an arbitrary constant, defined as the contrast at which the gain is 1: see *Online Methods*). We calculated a gain control index (w_t) from the fitted model parameters (Figure 2b) which quantified whether gain control estimated from the model was optimal given the background contrast levels (see *Supplementary Information*). For comparison, we also fit previously used linear-nonlinear (LN) models to each neuron[14,16,17,19], one with a static output nonlinearity (static-LN), and one with a contrast-dependent, or gain-controlled (GC), output nonlinearity (GC-LN, Figure 2c; representative neuron: Figure 2d-g). In this neuron, the fits of the GC-LN model and GLM with gain control (GC-GLM) demonstrated contrast gain control, characterized by high gain in low contrast and low gain in high contrast (Figure 2f and g, respectively), suggesting that both models capture similar gain control estimates.

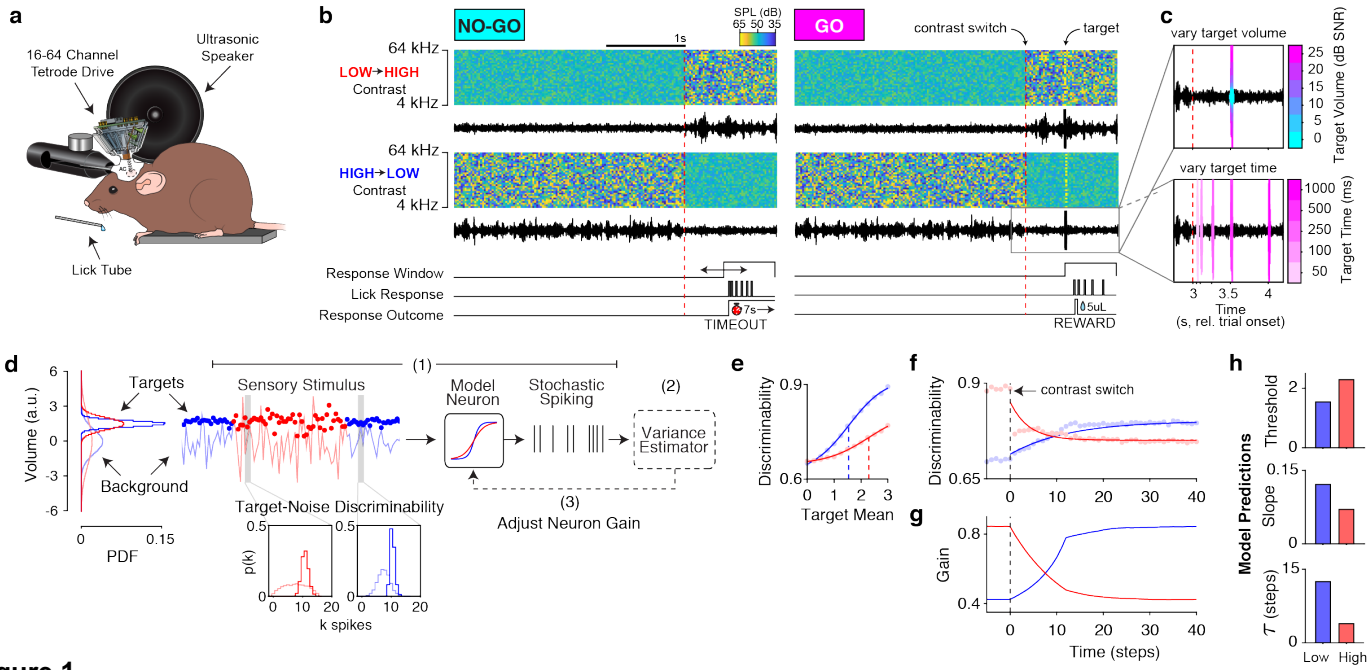


Figure 1.

a, Experimental setup. **b,** GO/NO-GO task design. *Left:* example NO-GO trials. From top to bottom: spectrogram of an example low-to-high contrast trial (color bar indicates volume in dB SPL); waveform for sample spectrogram; example spectrogram for a high-to-low contrast trial; waveform for example spectrogram; response window to determine false alarms; schematic lick responses in the response window; timeout of 7 seconds delivered after the first lick. Vertical red dashed line indicates the contrast switch after 3 seconds. Black scale bar indicates 1s. *Right:* example GO trials. From top to bottom: same as in left panel, except licks in the response window trigger a 5μL reward. In all figures, low to high contrast trials are indicated in red, and high to low contrast trials are indicated in blue. **c,** Target waveforms. *Top:* Overlaid trials where target volume differed. Volume is indicated by the amplitude and color bar. *Bottom:* trials where target timing differed. The red vertical dashed line indicates the contrast switch. **d,** Normative model of the task. *Left inset*: volume distributions for backgrounds (light lines) and targets (dark lines) in low and high contrast. (1) Spike generation process: a model neuron encodes stimuli sampled from a 1-dimensional sensory stimulus stream consisting of a background that transitions between low and high contrast (light lines); target stimuli (solid dots) were used to assess sensitivity. The stimulus response of the neuron is transformed by a sigmoidal function whose output is used to generate stochastic spikes. (2) The observed spike counts are integrated and decoded over a brief time window to estimate the variance of the stimulus. (3) The variance estimate is used to adjust the gain of the model neuron to minimize the expected error in the estimate of stimulus variance at each time step. *Bottom insets*: Sample probability distributions of observing k spikes in response to the background (light lines) or targets (dark lines) 7 time steps after a switch to high (red) or low contrast (blue). **e,** Model psychometric functions at steady state (25 time steps after transition) as a function of contrast and mean target volume. Light dots indicate discriminability of targets from background whereas the solid lines indicate logistic fits to the data. Dashed lines indicate detection thresholds. **f,** Model discriminability as a function of time and contrast. Dashed vertical line indicates the time where the background contrast changes. Light dots denote model discriminability at each time step. Solid lines are exponential fits to the data. Each time course is the discriminability of targets at approximate threshold volume for that contrast (1.5 target mean and 2.25 target mean for low and high contrast respectively). **g,** The average change in gain of the model after each contrast transition. **h,** Model predictions for the effects of contrast on psychometric thresholds, slopes, and adaptation time constants, as estimated by logistic and exponential fits in **e** and **f**.

Qualitatively, the GC-GLM outperformed standard LN models, primarily by capturing the adaptation dynamics after the transition (Figure 2d, middle panel), allowing us to analyze the gain control index as a function of time, w_t (Figure 2d, bottom panel; Figure 2g). To test whether the GC-GLM could better account for the data than standard models, we compared cross-validated correlations of the model predictions with the trial-averaged PSTH for each neuron, finding a significant effect of model type on the correlations (n = 97 neurons; Kruskall-Wallis test: $H(2) = 93.61$, p = 6.70e-21). Post-hoc Wilcoxon Sign-Rank tests showed that the GC-GLM correlation was significantly higher (Median (*Mdn*) = 0.75, Inter-Quartile Range (*IQR*) = 0.24) than the GC-LN model (*Mdn* = 0.54, *IQR* = 0.49, p = 4.41e-6) and the static-LN model (*Mdn* = 0.25, *IQR* = 0.73, p = 9.56e-10). Consistent with previous studies, we also found that the GC-LN model outperformed the static-LN model (p = 3.50e-6, Figure 2h).

We next quantified whether the GC-GLM detected significant gain control in the population. Here, we defined steady-state gain control by calculating the change in w_t between high (w_H) and low contrast (w_L) after the gain has stabilized (1 s after the contrast switch). Based on our definition of w_t, $w_H - w_L = -1$ if gain control is optimal (see *Supplementary Information*). Across all neurons, we found significant gain control (*Mdn*: -0.10, *IQR*: 0.35, Wilcoxon sign-rank test: *rank* = 233, Z = -2.90, p = 0.004; Figure 2i). To further validate the GLM estimates of gain, we compared the GC-GLM gain control indices at steady-state to those of the GC-LN model and found a significant relationship (linear regression: $F(1,95) = 12.20$, p = 7.33e-4, $R^2 = 0.11$; Figure 2j). Together, these results demonstrate that the GC-GLM model better accounts for the neural data by incorporating the dynamics of gain control and conclude that this method captures a similar estimate of steady-state gain control when compared to standard models.

Next, we analyzed the dynamics of gain control by fitting w_t after each contrast switch with an exponential function (Figure 2g). In neurons with gain control ($w_t < 0$ at steady state), the average time course of w_t was asymmetric across contrast transition types, rapidly decreasing after a switch to high contrast, and slowly increasing after a switch to low contrast (n = 45 neurons; Figure 2k). Within this same population, we quantified the timescale of adaptation to each contrast using the time constant (τ) of each exponential fit, finding significantly longer time constants in low contrast (*Mdn* = 0.29, *IQR* = .39) relative to high contrast (*Mdn* = 0.048, *IQR* = 0.094; Wilcoxon sign-rank test: *rank* = 918, Z = 4.52, p = 6.16e-6; Figure 2l). This asymmetry in gain adaptation agreed with the predictions of the normative model (Figure 1g) and with previously described behavior of optimal variance estimators[20].

Mouse behavioral detection is modulated by background contrast.

We next tested whether the asymmetry in gain control observed in cortex was reflected in behavioral sensitivity to targets in background noise. Mice initially trained in a simple version of the GO/NO-GO task where they were required to lick in response to a target and withhold licks on trials without a target (Figure 1b, 3a). Mice learned this task reliably, typically reaching criterion performance of 80% correct within 2-3 weeks in either contrast (Figure 3b). False alarm rates were significantly larger in high contrast than in low contrast (Extended Data Figure 3a), suggesting that detection is more difficult in high contrast, which we discuss next.

By varying the volume of presented targets, we collected psychometric curves for each mouse in each contrast (example mouse performance: Figure 3c; group averages: Figure 3d). Across all mice (n = 25), we found that targets were easier to detect in low contrast, observing significantly lower detection thresholds in low contrast (Mean (*M*) = 7.30, standard deviation (*SD*) = 1.67) compared to high contrast (*M* = 13.20, *SD* = 2.54; paired t-test: $t(23) = -9.11$, p = 4.34e-9, Figure 3e). To quantify the influence of contrast on psychometric slope, we tested a subset of mice with target volumes matched across the contrast conditions. In this cohort (n = 7; Figure 3f), we found significantly lower target thresholds in low contrast (*M* = 6.80, *SD* = 2.73) compared to high contrast (*M* = 14.96, *SD* = 3.51; paired t-test: $t(3) = -3.59$, p = 0.036; Figure 3g) and significantly steeper slopes in low contrast (*M* = 0.051, *SD* = 0.0068) compared to high contrast (*M* = 0.042, *SD* = 0.0064; paired t-test: $t(3) = 3.42$, p = 0.042; Figure 3h). Interestingly, there was no significant change in psychometric slope when combining sessions with different target ranges in each contrast (n = 25; Extended Data Figure 3b). Splitting the data by target range revealed that targets drawn from a narrow range resulted in steeper psychometric slopes than targets drawn from a wide range (Extended Data Figure 3c-f), regardless of the background contrast. Combined, these results demonstrate that background contrast has a substantial impact on detection threshold, and that mice are more sensitive to changes in the volume of targets presented in low contrast.

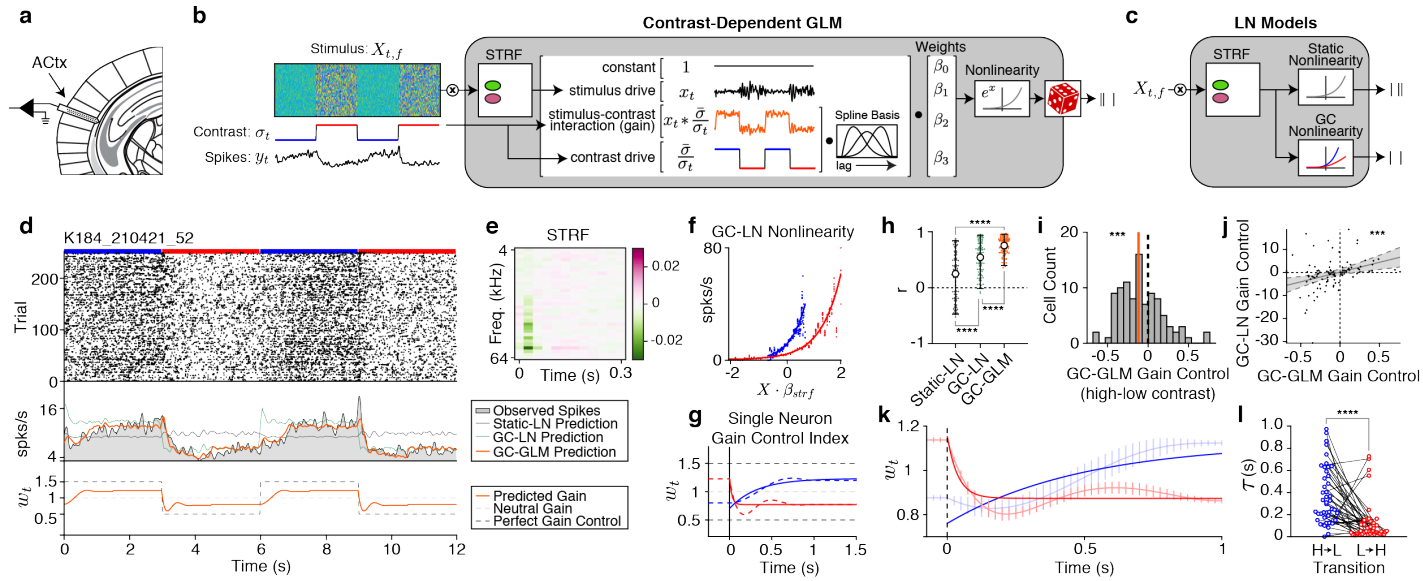


Figure 2.

a, Schematic for acute recordings from auditory cortex. **b,** Schematic of Poisson generalized linear model (GLM) design. *From left to right:* The variables considered by the model are the stimulus spectrogram, the stimulus contrast, and observed spikes. The GLM with gain control (GC-GLM) is fit in two steps: 1) A spectrotemporal receptive field (STRF) is estimated. 2) Weights of stimulus drive (x_t), the multiplicative interaction between the contrast and the stimulus drive ($x_t * \bar{\sigma}/\sigma_t$), and the contrast drive ($\bar{\sigma}/\sigma_t$) are fit using Poisson regression. The history of the contrast predictors were smoothed with a B-spline basis set. **c,** Schematic of linear-nonlinear models. As in **b**, we first fit a STRF which is then passed through either a static exponential nonlinearity (static-LN) or independent nonlinearities fit separately to low and high contrast periods (GC-LN). **d,** Neuronal responses and model fits for a representative neuron. *Top*: a spike raster for the example neuron. Each period of contrast is indicated by the blue (low contrast) and red (high contrast) bars. *Middle*: PSTH of the example cell is plotted in gray. Predictions from the static-LN model are plotted in gray, GC-LN model in green, and GC-GLM model in orange. All traces were smoothed with a 10ms wide Gaussian filter for visualization. *Bottom*: the gain control index, w_t (orange trace). The gain control index of a neuron with no gain control ($w_t = 1$) is plotted as a grey dashed line, and the gain control index of a neuron with optimal gain control ($w_H = 0.5, w_L = 1.5$) is plotted as a black dashed line (see *Online Methods*). **e,** The STRF estimated from this neuron. **f,** The nonlinearities fitted to low (blue) and high (red) contrast in the GC-LN model for the example neuron. Points indicate the mean observed firing rate (ordinate), binned according to observed filter prediction values (abscissa). Solid lines are exponential function fits. **g,** Gain control index, w_t, for the example neuron after each contrast switch (dashed red and blue lines). The solid red and blue lines are fits of an exponential function. Dashed gray and black lines indicate neutral and optimal gain control values as in **d**. **h,** Cross-validated Pearson's correlations between the trial-averaged firing rate trace and the model predictions. Gray, green, and orange dots indicate the correlations for each neuron (n=95) for the static-LN, GC-LN, and GC-GLM models, respectively. Open circles indicate the median correlation, and the error bars indicate 2.5-97.5 percentiles. Results of Wilcoxon Sign-Rank tests are indicated with asterisks. **i,** Distribution of gain control estimated by the GLM for the recorded population (defined as the difference in the gain control index between high and low contrast, measured after the gain has stabilized (ie. after 1s): $w_H - w_L$). Dashed vertical line indicates no gain control, while the solid orange line indicates the median of the distribution. Asterisks indicate the results of a Wilcoxon Sign-Rank test. **j,** Correspondence between gain control estimates from the GC-GLM model (abscissa) and the previously reported GC-LN model (ordinate). Black dots indicate the data for each neuron, while linear model fit and error are indicated by the gray line. Asterisks indicate significance of the linear fit to the data. **k,** Average time course of the gain estimate w_t for neurons with gain control (ie. gain control is less than 0, n = 45). Light red and blue lines indicate the average value of w_t ±SEM over neurons for transitions to high and low contrast, respectively. Solid red and blue lines are exponential fits to the averages after the transition, which is marked by the dashed black line. **l,** Distributions of adaptation time constants of w_t after transitions to low contrast, in blue, and high contrast, in red. Each dot and line indicates a neuron. Asterisks indicate the results of a Wilcoxon Sign-Rank test. In all plots: ns, not significant; †p<0.1; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

To assess behavioral adaptation to the background contrast, we presented targets at threshold volume at variable delays following the contrast transition. We observed behavioral time courses consistent with the normative model and with gain measured in auditory cortex: after a switch to high contrast detection rates decreased quickly over time, but after a switch to low contrast detection rates increased slowly over time (Figure 3i). In high contrast, the first significant drop in performance occurred between the first two time points, while in low contrast the first significant increase in performance occurred between the first and third time points (Figure 3i, Extended Data Table 1). Indeed, behavioral adaptation was significantly faster in high contrast (exponential fit to behavioral performance after contrast transition, *Mdn* = 0.023, *IQR* = 0.082) compared to low contrast (*Mdn* = 0.13, *IQR* = 0.13; Wilcoxon Rank-Sum test (n = 21): *rank* = 547, *Z* = 2.75, *p* = 0.0060; Figure 3j). Taken together, these behavioral results confirm the three predictions from the normative model (Figure 1h): 1) Detection thresholds are lower in low contrast; 2) Psychometric slopes are higher in low contrast; 3) Performance decreases rapidly in high contrast and increases gradually in low contrast.

Auditory cortex is necessary for detection in background noise.

Whereas gain control is present in many areas along the auditory pathway, it is strongest in auditory cortex[16,19]. As such, we hypothesized that auditory cortex supports the detection of sounds in the presence of background noise. To test whether auditory cortex is required for task performance, we inactivated auditory cortex using the GABA-A receptor agonist muscimol. We validated that muscimol disrupts cortical coding of target sounds by applying muscimol topically to the cortical surface during passive playback of the behavioral stimuli, finding near complete suppression of target responses (Extended Data Figure 4a-f, *Supplementary Information*).

To test whether inactivation of auditory cortex affects behavioral performance, we repeated the same experiments in behaving mice, administering muscimol or saline bilaterally through chronically implanted cannulae (n = 44 sessions from 4 mice; Figure 4a). We found a profound decrease in the response rates to targets and background in both contrasts (Figure 4b). We quantified these effects on the psychometric curve using a three-way ANOVA with cortical intervention (muscimol or saline), contrast, and target volume as factors. We found significant main effects of cortical intervention ($F(1,307) = 278.63$, $p = 3.83e-44$), contrast ($F(1,307) = 4.39$, $p = 0.037$) and volume ($F(6,307) = 40.90$, $p = 7.54e-36$). Post-hoc tests showed that muscimol application significantly decreased hit rates in both contrasts by 31.45% (95% CI: [27.76, 35.14], $p = 1.060e-10$), whereas an increase in background contrast significantly decreased hit rates in both intervention conditions by 3.95% (95% CI: [2.57, 7.64], $p = 0.036$). Furthermore, we observed significant interactions between target volume and cortical intervention ($F(6,307) = 14.11$, $p = 4.47e-14$), and between target volume and contrast ($F(6,307) = 2.97$, $p = 7.87e-3$), but we did not observe a significant interaction between contrast and cortical intervention, suggesting that muscimol has the same effect in low and high contrast. To quantify the effects of muscimol on psychometric performance, we extracted response rates to the maximum target volume, false alarm rates, thresholds, and slopes of psychometric functions fit to each session, and found that muscimol significantly reduced every measure of psychometric performance, with the exception of behavioral threshold (Figure 4c, Extended Data Table 1). From these results, we can conclude that auditory cortex is necessary for detecting targets in background, regardless of background contrast.

A potential alternative effect of muscimol is a general loss of function that is not specific to hearing target sounds. To control for this, we devised an alternative to the detection in background task where mice detected targets in silence (Figure 4d). To ensure equivalency between the two tasks, we took the highest-volume target trials in the target-in-background task (25dB SNR in high contrast) and removed the background noise during the target detection period (Figure 4e, bottom). Thus, mice were presented with the exact same targets as in the previous task, but without the background DRCs, allowing us to test whether auditory cortex is specifically required for detection in the presence of a background (*Online Methods*).

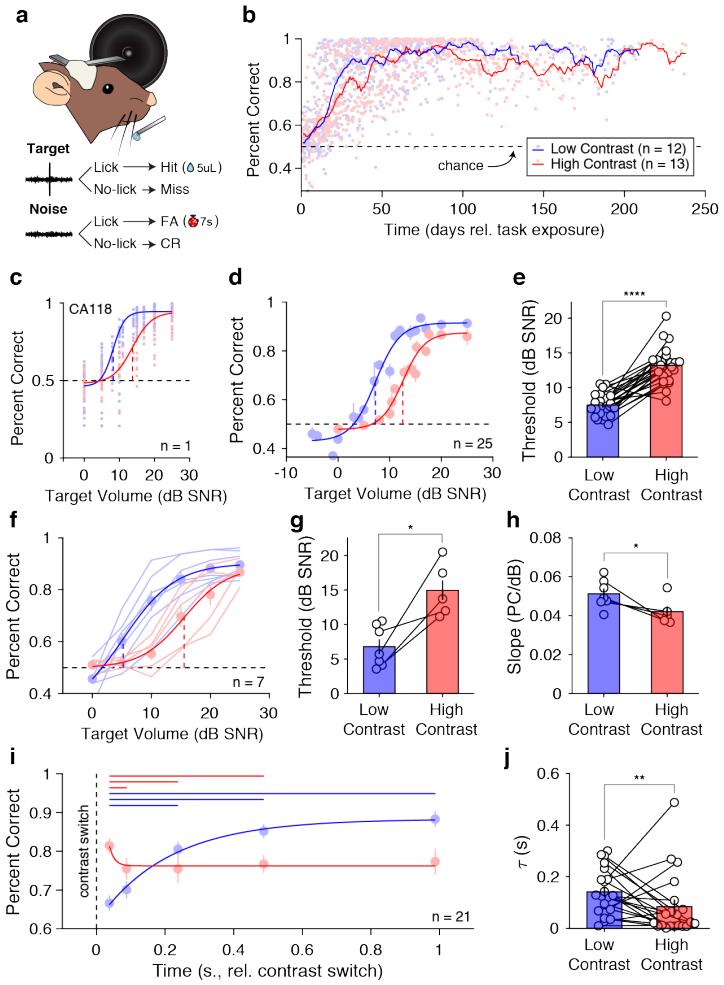


Figure 3.

a, Schematic of GO/NO-GO paradigm. **b,** Behavioral performance the initial training contrast (n=12 mice were first trained in low contrast, n=13 mice were first trained in high contrast). Dots indicate sessions, solid lines are a 7 day moving average. Dashed horizontal line indicates chance performance (percent correct = 0.5). **c,** Psychometric functions in low and high contrast for one mouse (mouse ID indicated in the upper left). Each dot indicates percent correct for a single volume in a single session, while the solid lines indicate average psychometric fits. Colors as in **b**. **d,** Psychometric functions averaged for n=25 mice in each contrast. Dots indicate performance at each target SNR ±SEM over mice, while the solid lines are logistic function fits with thresholds plotted as dashed lines. **e,** Psychometric thresholds per contrast. Each dot and line represents a mouse. Bars indicate the average threshold ±SEM over mice. **f,** Psychometric functions for n=7 mice tested using the same target volumes in each contrast. Dots indicate average performance ±SEM over mice. Lines indicate psychometric fits, with the vertical dashed lines indicating average thresholds. Light lines indicate the psychometric curves of individual mice. **g,** Psychometric thresholds per contrast. Each dot represents a mouse, lines indicate where mice participated in both low and high contrast sessions. Bars indicate the average threshold over mice, while error bars in black indicate threshold ±SEM over mice. **h,** Psychometric slopes per contrast. Presentation as in **g**. **i,** Behavioral performance as a function of contrast and target time relative to the switch in contrast (vertical dashed line) for n=21 mice. Dots indicate average performance ±SEM over mice. Solid curves are exponential function fits. Horizontal lines at the top of the plot indicate significant changes in performance between the first target presentation time and subsequent target presentation times, as assessed by Wilcoxon Sign-Rank tests with false discovery rate correction for multiple comparisons (see Extended Data Table 1). **j,** Average time constant of exponential fits in low and high contrast. Presentation as in **h**. In all plots, asterisks indicate the significance of Wilcoxon Sign-Rank tests: ns, not significant; †p<0.1; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

To assess psychometric performance in this new task, we modulated detection difficulty by attenuating the volume of each target. As observed previously, inactivation of auditory cortex impaired detection in high contrast (Figure 4e, top). However, cortical inactivation had little effect on psychometric performance in silence (Figure 4e, bottom). We quantified these effects on psychometric performance using a three-way ANOVA with cortical intervention (muscimol or saline), task (detection in background or silence), and target volume as factors (n = 26 sessions from 2 mice). We found significant main effects of intervention ($F(1,181) = 62.83$, $p = 3.62e-13$), task ($F(1,181) = 6.82$, $p = 9.86e-3$), and volume ($F(6,181) = 46.16$, $p = 1.69e-32$). Post-hoc tests showed that muscimol significantly reduced hit rates by 20.2% (95% CI: [15.19, 25.17], $p = 1.060e-10$). Hit rates for targets presented in silence were significantly elevated by 6.65% relative to targets presented in background (95% CI: [1.65, 11.64], $p = 0.0090$). Furthermore, we found significant interactions between cortical intervention and task type ($F(1,181) = 6.36$, $p = 0.013$), intervention and volume ($F(6,181) = 3.47$, $p = 2.98e-3$), and volume and task type ($F(6,181) = 8.47$, $p = 5.43e-8$). As before, we parameterized psychometric performance by fitting each session with a psychometric curve, and we extracted the response rates to the maximum target volume, false alarm rates, response rates at threshold volume, and slopes of psychometric functions. During the target-in-background task, we found significant effects of muscimol on the response rates at maximum volume and threshold, a moderate effect on psychometric slope, and no effect on false alarm rate. However, muscimol application had no significant effect on any of these measures in the target-in-silence task (Figure 4f, Extended Data Table 1). Taken together, these results show that while both cortical inactivation and the presence or absence of background noise affected behavioral performance, these effects interacted: muscimol had a larger effect on performance when background noise was present.

Combined, our findings demonstrate that the auditory cortex is specifically required for detection in the presence of background noise, but not in silence. Our next goal was to test whether neuronal activity in AC is predictive of behavioral performance.

Cortical codes predict individual behavioral performance.

To better understand how representations in auditory cortex could give rise to behavior, we chronically recorded from populations of neurons in auditory cortex while mice performed the psychometric task (Figure 5a; n = 12 mice; n = 11 mice participated in low contrast sessions, n = 8 in high contrast sessions).

To quantify the representations of targets and background in the neural population (example responses in Figure 5b,c), we adapted a population vector approach[27] to generate a discriminability metric using population activity (*Online Methods*). This method allowed us to project trial distributions in n-dimensional neural space along a single dimension which separated target and background trials (Figure 5d, left panel). We then estimated the criterion projection value that best predicted whether each trial contained a target or just background[28] (Figure 5d, right).

This population decoding method allowed us to estimate neurometric functions to directly compare to psychometric functions for each mouse (Figure 5e). On average, neurometric and psychometric functions were qualitatively similar (Figure 5f). To test the relationship between contrast and threshold measure we computed average neurometric and psychometric thresholds for each mouse and performed a two-way ANOVA with threshold measure (neurometric or psychometric) and contrast as factors. We found a main effect of contrast ($F(1) = 37.88$, $p = 5.43e-7$), no main effect of threshold measure ($F(1) = 0.060$, $p = 0.81$) and no interaction between measure and contrast ($F(1) = 0.040$, $p = 0.84$), which suggests that psychometric and neurometric thresholds were similarly affected by background contrast. As expected, post-hoc t-tests found no difference between neurometric and psychometric thresholds (0.19, 95% CI: [-1.38, 1.76], $p = 0.81$), and that low contrast significantly decreased thresholds relative to high contrast (-4.77, 95% CI: [-6.34, -3.19], $p = 5.43e-7$).

To quantify the relationship between neurometric and psychometric thresholds, while controlling for the effect of contrast, we fit a mixed-effects model using contrast and neurometric threshold as fixed effects, mouse identity as a random effect and psychometric threshold as the dependent variable (Extended Data Table 1). We tested the significance of each predictor by comparing the full model fit to null models excluding neurometric thresholds or contrast. We found that both neurometric threshold (Likelihood Ratio Test: $\chi^2(1) = 5.89$, $p = 0.015$) and contrast (Likelihood Ratio Test: $\chi^2(1) = 4.70$, $p = 0.030$) significantly improved psychometric threshold predictions (Figure 5g). Taken together, these results demonstrate that population thresholds in auditory cortex are predictive of behavioral thresholds in individual mice, and both psychometric and neurometric thresholds are modulated by contrast as predicted by a normative account of gain control.

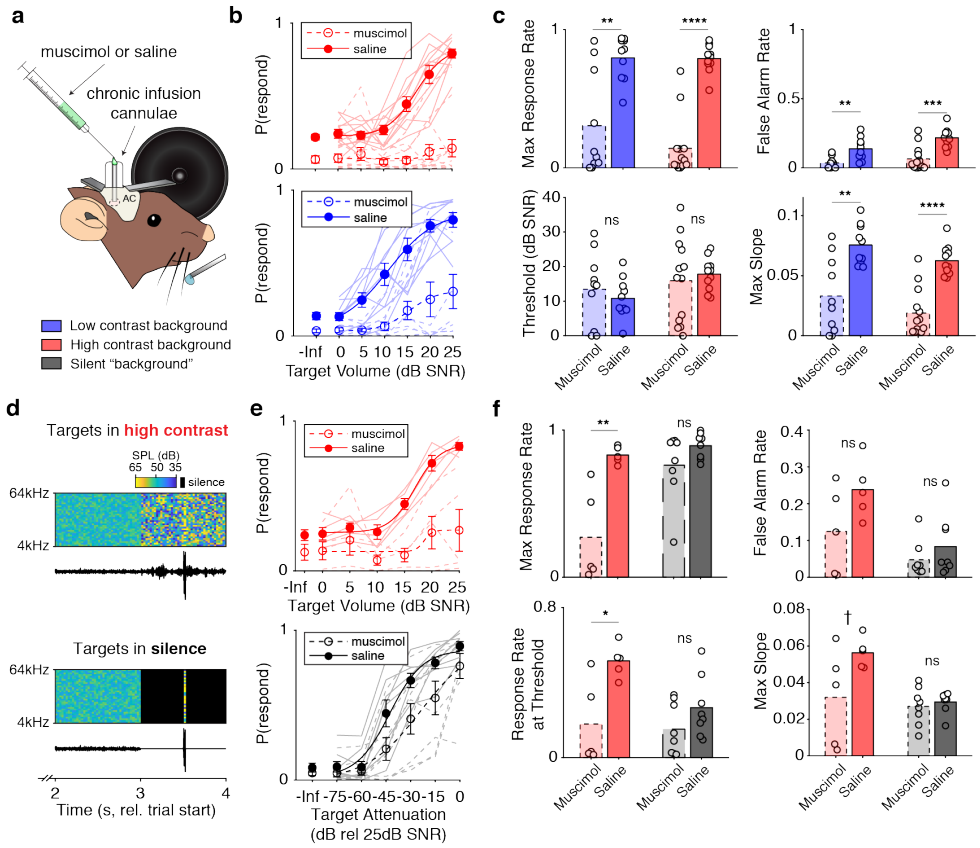


Figure 4.

a, Setup schematic for muscimol application in behaving mice. *Bottom*: legend indicating colors used for each background condition. **b,** Behavioral psychometric functions after muscimol or saline application (n = 43 sessions across 4 mice). *Top*: Performance in high contrast (red). *Bottom:* Performance in low contrast (blue). Dark solid lines and filled circles indicate average performance after saline injection. Dark dashed lines and open circles indicate average performance after muscimol injection. Light solid and dashed lines are psychometric curves from individual sessions. Error bars indicate ±SEM across sessions. **c,** Behavioral performance metrics. Open circles indicate performance in individual sessions. Colored bars indicate average performance across sessions. Opaque bars with solid outlines are averages after saline application, while transparent bars with dashed outlines are averages after muscimol application. **d,** *Top*: Example stimulus spectrogram for the target-in-background detection task with the corresponding waveform below. Color bar indicates sound level (silence is black). *Bottom:* Spectrogram and waveform for the target-in-silence task. **e,** *Top*: psychometric performance with high contrast background (n = 10 sessions across 2 mice), formatting as in top panel of **b**. *Bottom*: psychometric performance (n = 16 sessions from 2 mice) in the target-in-silence task, with target attenuation relative to the highest volume target from the target-in-background task on the abscissa and probability of responding on the ordinate. Black filled circles and dark solid lines indicate average performance after saline injection and psychometric fits to the average. Open circles and dark dashed lines indicate average performance after muscimol injection and psychometric fits to the average. Light gray solid and dashed lines are psychometric curves from individual sessions. Error bars indicate ±SEM across sessions. **f,** Behavioral performance metrics as a function of task type and pharmacological intervention. As in **c**, opaque bars are averages of saline sessions, transparent bars are averages of muscimol sessions. Red bars indicate performance in the detection-in-background task. Gray bars indicate performance in the detection-in-silence task. In all plots: $^{ns}p>0.1$; $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$.

We applied the same statistical analysis to neurometric and psychometric slopes. When analyzing only mice with matched target ranges in low and high contrast, we found a significant main effect of contrast (two-way ANOVA: $F(1) = 5.98$, $p = 0.028$) and slope measure ($F(1) = 10.62$, $p = 0.0057$), but no significant interaction ($F(1) = 2.095$, $p = 0.17$). Post-hoc t-tests found significantly steeper slopes in low contrast compared to high contrast (0.0071, 95% CI: [0.00087, 0.013], $p = 0.028$) and significantly shallower neurometric slopes compared to psychometric slopes (-0.0094, 95% CI: [-0.016, -0.0032], $p = 0.0057$; Extended Data Figure 5a). When including all mice we found significant main effects of slope measure (two-way ANOVA, $F(1) = 5.88$, $p = 0.021$) and contrast ($F(1) = 8.31$, $p = 0.0068$), but no significant interaction between the two ($F(1) = 0.18$, $p = 0.67$). Neurometric slopes were significantly shallower than psychometric slopes (-0.015, 95% CI: [-0.027, -0.0024] PC/dB, $p = 0.021$) and low contrast slopes were significantly shallower than high contrast slopes (-0.018, 95% CI: [-0.030, -0.0052] PC/dB, $p = 0.0068$; Figure 5h). The latter effect may be due to the mixture of target ranges used in the full cohort of mice, which we found to have an impact on psychometric slopes (see *Extended Data Figure* 3b-f). To quantify the relationship between neurometric and psychometric slopes, we applied the same mixed-effects analysis used previously (Extended Data Table 1). For all mice, we found that neurometric slopes (Likelihood Ratio Test: $\chi^2(1) = 9.78$, $p = 0.0018$), but not contrast (Likelihood Ratio Test: $\chi^2(1) = 8.55$, $p = 0.078$) significantly improved psychometric slope predictions (Figure 5h). Overall, these results were consistent with our previous behavioral findings (Figure 3h), demonstrating that, when target volumes are matched, increased contrast reduced neurometric and psychometric slopes and that neurometric slope is predictive of psychometric slope on a mouse-to-mouse basis.

Combined, these results demonstrate that parameters of neurometric and psychometric functions are affected by contrast as predicted by a normative model of gain control. We also find that individual variation in psychometric performance is predicted by population activity in auditory cortex, independently of the effect of contrast.

Dynamics of target detection during adaptation.

We next measured how cortical discriminability evolved as a function of time and contrast in sessions where mice were presented with targets at threshold volume at different offsets relative to the contrast switch. In line with our behavioral results (Figure 3i), we found that in high contrast the first significant drop in cortical discriminability occurred between the first two target times, while in low contrast the first significant drop occurred between the first and third target times (n = 43 recording sessions; Extended Data Table 1; Figure 5i). To quantify the speed of neural adaptation, we fit the average neural discrimination time course for each mouse with an exponential function (n = 8 mice). Consistent with the normative model (Figure 1f-h), gain control dynamics estimated from cortical activity (Figure 2k,l) and behavior (Figure 3i,j), we found asymmetric adaptation in the neural responses, with larger adaptation time constants in low contrast (*Mdn* = 0.14, *IQR* = 0.21) relative to high contrast (*Mdn* = 0.033, *IQR* = 0.16; Wilcoxon sign-rank test (n = 8): *rank* = 28, $p = 0.016$; Figure 5j).

Cortical gain predicts behavioral performance.

Our results so far provide strong evidence that gain control in the auditory system shapes behavioral performance. To assess the role of cortical gain in behavior, we leveraged the design of the background sounds to estimate spectrotemporal receptive fields (STRFs) and nonlinearities of neurons recorded during task performance. For each neuron, we fit a model with a static nonlinearity (static-LN) or a model with gain control (GC-LN; Figure 6a-d). We then pooled the neurons recorded across sessions, and included only neurons with strong stimulus responses in both contrasts (*Online Methods*). First, we compared the cross-validated performance of the static-LN model versus the GC-LN model, finding higher correlations using the GC-LN model (*Mdn* = 0.82, *IQR* = 0.17) relative to the static-LN model (*Mdn* = 0.67, *IQR* = 0.12; Wilcoxon sign-rank test (n = 2,792 neurons): *rank* = 3.85e5, $Z = -36.74$, $p = 1.88e-295$; Extended Data Figure 5h). We also found significantly higher gain in low contrast (*Mdn* = 0.10, *IQR* = 0.13) than in high contrast (*Mdn* = 0.041, *IQR* = 0.023; Wilcoxon sign-rank test: *rank* = 3.57e6, $Z = 37.92$, $p = 1.070e-314$; Figure 6e, inset). These results demonstrate that LN models can more accurately predict cortical activity when incorporating contrast gain control, and confirm previous reports of robust gain control in mouse auditory cortex[17–19].

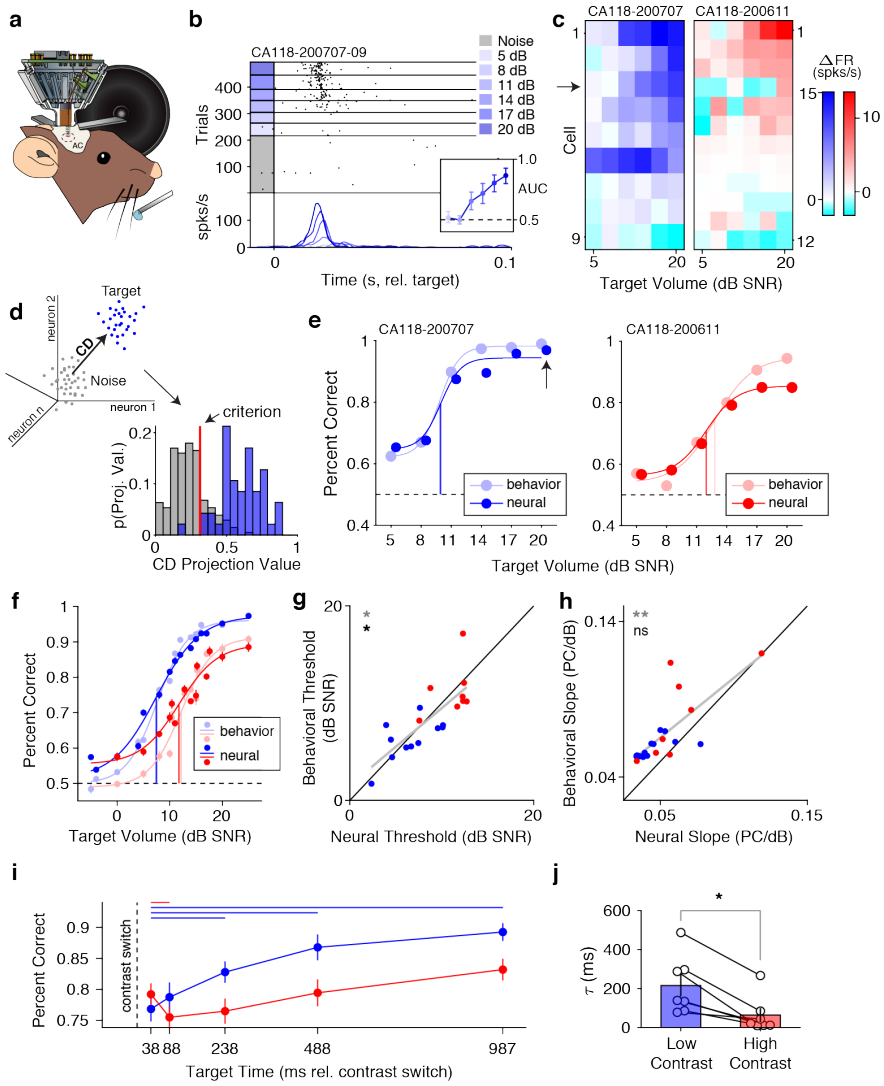


Figure 5.

a, Experimental setup for chronic ACtx recordings from behaving mice. **b,** Example spiking responses to targets and background in low contrast during behavior. *Top*: Spike raster ordered by target volume. *Bottom*: Trial averaged PSTH for each condition (smoothed with a 2 ms wide Gaussian kernel). *Inset:* Area under the ROC curve (AUC) when discriminating background from target responses. The dashed horizontal line indicates chance performance (0.5). Error bars are bootstrapped 95% confidence intervals. **c,** Neurograms of populations of simultaneously recorded neurons during representative low and high contrast sessions. Arrow indicates the neuron in panel **b**. **d,** Discriminating targets from background using population responses. *Left:* schematic of coding direction analysis. In n-dimensional neural space, background trials are represented by a gray point-cloud, while target trials are represented by a blue point-cloud. The coding direction (CD) is the vector defining the average difference between target and noise. *Right:* trial distributions of projections along the CD for one session (low contrast session in **c**). Projection values for 20 dB SNR targets are plotted in blue, projection values for noise trials are plotted in gray. The red line is the criterion optimized using all trials. **e,** Example neurometric and psychometric curves. *Left:* Low contrast curves. Light blue circles and solid lines indicate psychometric performance and logistic fit. Dark blue circles and solid lines indicate neurometric performance from the session plotted in the left panel of **c**. The horizontal dashed line indicates chance performance (percent correct = 0.5). The arrow indicates the neural performance computed from the distributions in **d**. *Right:* High contrast curves for the session plotted in the right panel of **c**. **f,** Average neurometric and psychometric functions for each contrast (n = 19 mice). Formatting as in **e.** Errorbars are ±SEM over mice. **g,** Relationship between behavioral and neural thresholds. Circles represent the average behavioral and neural threshold for each mouse in each contrast. Gray line is the linear best fit, solid black line is unity. Gray asterisk indicates significant relationship between neurometric and psychometric threshold, while black asterisk indicates significant effect of contrast on threshold. **h,** Relationship between behavioral and neural slopes. Appearance as in **g**. **i,** Decoder performance after each contrast transition, as a function of target presentation time. Dashed vertical line indicates the contrast switch. Solid lines and circles indicate the percent correct performance of a target decoder after a switch to low contrast (blue) or high contrast (red) ±SEM over sessions. Horizontal lines indicate significant changes in performance between the first target presentation time and subsequent target presentation times, as assessed by Wilcoxon Sign-rank tests with FDR correction. **j,** Adaptation time constants of exponentials fitted to the average neural decoder performance in each contrast for each mouse. Circles connected with a line indicate data per mouse. Asterisk is the significance of a Wilsoxon Sign-Rank test. In all plots: $^{ns}p>0.1$; $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$.

Based on our previous results, we expected that the amount of gain control in auditory cortex would predict target detectability. When fitting the GC-LN model, we separately estimated neural gain during the adaptation period of the trial and the target period of the trial (defined as the time periods before and after the contrast switch, respectively; Figure 6b). To quantify the effects of contrast and trial period on gain, we performed a two-way ANOVA, with gain as the dependent variable, and contrast, trial period, and their interaction as factors (n = 2,262 neurons, after excluding outliers, see *Online Methods*). As expected, we found a significant main effect of contrast ($F(1,4523) = 431.03$, $p = 1.60e\text{-}91$). Furthermore, there was a significant main effect of trial period ($F(1,4523) = 35.79$, $p = 2.36e\text{-}9$) and a significant interaction between contrast and trial period ($F(1,4523) = 77.91$, $p = 1.51e\text{-}18$). Post-hoc tests revealed that, in low contrast, gain during the target period significantly increased (0.032, 95% CI: [0.024, 0.040], $p = 3.77e\text{-}9$), but did not significantly change in high contrast (0.0062, 95% CI: [-0.017, 0.014], $p = 0.18$; Figure 6e). These findings indicate that neural gain is not only sensitive to stimulus contrast, but also increases during the target period of the trial, specifically in low contrast.

To visualize the gross relationship between gain and psychometric performance, we first averaged the gain of stimulus-responsive neurons during the target period of the trial in each session (n = 168 sessions across 13 mice). We then selected only low contrast sessions and split the data by the median gain in the target period, computing the average psychometric curves for sessions in the bottom versus the top 50th percentile (Figure 6f, inset). We observed that sessions with high gain had steeper slopes and lower thresholds than sessions with low gain (Figure 6f). To quantify the relationship between gain and task performance, we fit a mixed-effects model using contrast and gain during the target period as fixed effects, mouse identity as a random effect and either psychometric slopes or thresholds as the dependent variable. This approach allowed us to separate the neuronal and behavioral impact of contrast gain control from effect of session-to-session fluctuations in gain. We tested whether gain and contrast were significant predictors of behavioral performance by comparing the full model to null models excluding either gain or contrast. We found that the model including gain was a better predictor of behavioral threshold than was the null model (Likelihood Ratio Test: $\chi^2(1) = 5.82$, $p = 0.016$), indicating that thresholds decreased by about 3.046 dB SNR ±1.24 (standard error) for every 10% increase in gain. Using a similar procedure, we found that contrast was also a significant predictor of behavioral threshold (Likelihood Ratio Test: $\chi^2(1) = 5.84$, $p = 0.038$), with the step from low to high contrast inducing a decrease in behavioral thresholds of 3.27 dB SNR ±1.33 (Figure 6g).

We applied the same analysis to test the effects of contrast and gain on psychometric slope (Figure 6f), again finding that gain significantly predicted psychometric slopes (Likelihood Ratio Test: $\chi^2(1) = 6.96$, $p = 0.0083$), such that the psychometric slope increased by 0.16 dB/PC ±0.060 for every 100% increase in gain. However, contrast did not significantly improve the fit of this model (Likelihood Ratio Test: $\chi^2(1) = 2.28$, $p = 0.13$; Figure 6h). This result is not entirely unexpected, given that we observed no effect of contrast on psychometric slopes when comparing across sessions with different target distributions (Extended Data Figure 2b), which is true of the sessions used in this analysis.

Our findings suggest that the relationship between gain and psychometric performance is shaped by two sources: contrast-induced gain control and fluctuations in gain from session to session. To further disentangle the relationship between these two sources of behavioral modulation, we repeated the mixed effects models, this time using gain during the adaptation period as the predictor of interest. We hypothesized that gain in this period should not be predictive of behavioral performance, as there were no targets presented during this portion of the trial. We found that this was the case; we did not observe any predictive relationship between gain estimated in this period and behavioral performance (Extended Data Figure 5i-k; Extended Data Table 1). In summary, we found that cortical gain was modulated by both stimulus contrast and trial period, increasing when contrast is low and when mice were detecting targets. Furthermore, we found that psychometric performance was predicted by both the stimulus contrast and by session-to-session changes in cortical gain during target detection.

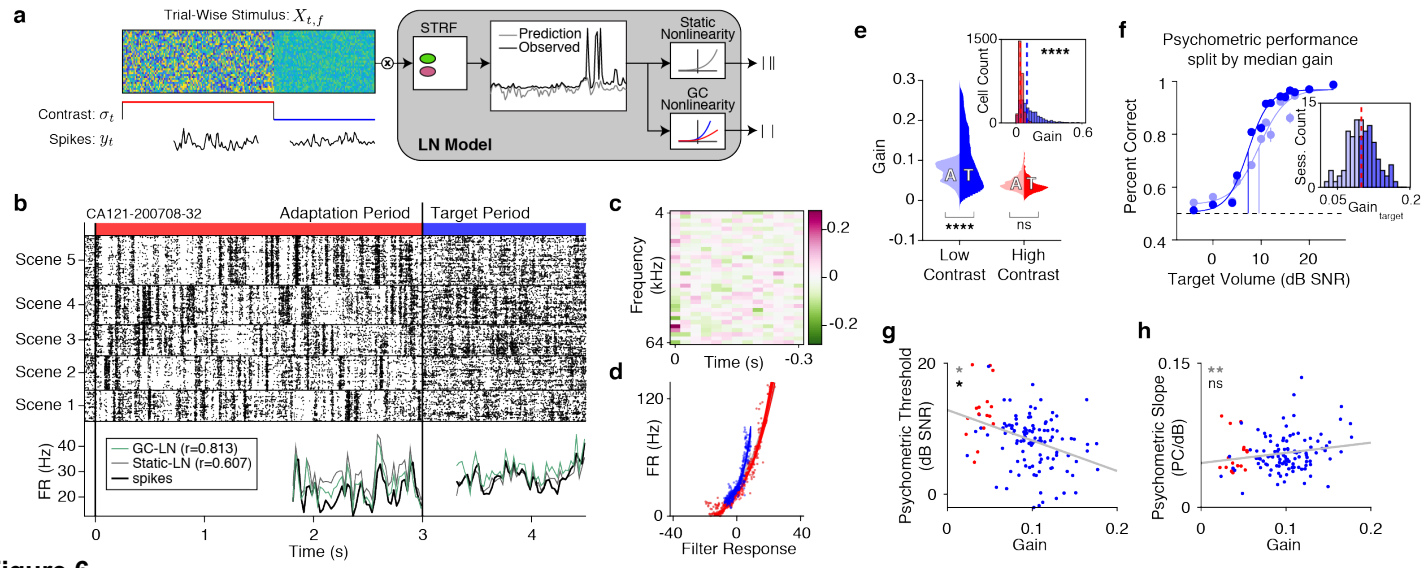


Figure 6.

a, Schematic of the linear nonlinear models fit to behavioral recordings. Spectrograms concatenated across trials were used to estimate a STRF. Nonlinearities were fit for a static and gain-controlled (GC) model. **b,** Example background-locked responses from a well-tuned unit. The top portion of the plot is a spike raster sorted by the background scenes. The contrast of the adaptation and target periods is indicated by the red and blue rectangles on the top of the plot. The bottom portion of the plot is a trial-averaged PSTH of the observed spiking, binned every 25ms (black trace). The colored traces are the predictions of the static-LN model (gray) and GC-LN model (green). Correlations between the model predictions and trial-averaged PSTH are indicated in the legend. **c,** STRF for this example neuron. **d,** Estimated nonlinearities for this example neuron. Points indicate the mean observed firing rate (ordinate), binned according to observed filter prediction values (abscissa). Solid lines indicate exponential function fits to the underlying points across validation folds. **e,** Gain control in auditory cortex during the task. Probability density of gain values across neurons in high and low contrast, separated by adaptation and target periods (labelled by "A" and "T", respectively). Asterisks are the significance of post-hoc tests of target period for a two-way ANOVA testing the effects of target period and contrast on gain. *Inset*: gain distributions for each contrast, across all trial periods. Dashed vertical lines indicate the median of each distribution, asterisks indicate the results of a Wilcoxon Sign-Rank test across contrast. **f,** Average psychometric curves in low contrast split by cortical gain estimated during the target period. Light blue data points indicate average performance in sessions where gain was below the median, dark blue data points indicate average performance in sessions where gain was above the median (±SEM). Solid lines are psychometric fits to the data, with the thresholds plotted vertically from 0.5. *Inset*: Histogram of average target gain over sessions. The dashed red vertical line indicates the median gain, light blue bars indicate sessions below the median, and dark blue bars indicate sessions above the median. **g,** Relationship between gain and behavioral threshold. Circles represent the average gain and behavioral threshold for each session and contrast (blue and red dots indicate low and high contrast target periods, respectively). Gray lines indicate linear best fit. Gray asterisks indicate a significant relationship between gain and psychometric threshold, black asterisks indicates a significant effect of contrast. **h,** Relationship between gain and behavioral slope. Appearance as in **g**. In all plots: $^{ns}p>0.1$; $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$.

Discussion

Our auditory surroundings are characterized by different statistical properties that change over time. Changes in the dynamic range, or contrast, of acoustic inputs poses a challenge to the auditory system, which is composed of neurons with limited dynamic range. The efficient coding hypothesis predicts that as stimulus contrast changes, neurons should adjust their gain in order to match their limited dynamic range to that of the stimulus distribution[1]. Multiple studies have demonstrated that neurons throughout the auditory pathway exhibit such contrast gain control[16,19,29]. Whereas recent work has demonstrated a link between efficient cortical codes and human psychophysical performance[19,25], it is unclear how contrast gain control dynamics are directly related to behavior, as neuronal responses and behavior were not observed simultaneously.

In this study, we directly linked contrast gain control to auditory behavior by combining a theoretical model of efficient coding with simultaneous behavioral psychophysics and recording and manipulation of cortical activity. First, we developed a normative model based on efficient coding[21,23] which predicted that: 1) Detection thresholds of targets should be lower in low contrast than in high contrast; 2) Sensitivity to changes target volume should be greater in low contrast relative to high contrast; and 3) Detection should adapt asymmetrically: increasing slowly after a switch to low contrast, but decreasing rapidly after a switch to high contrast (Figure 1). Then, we used a novel form of Poisson GLM to confirm that gain control dynamics in auditory cortex are indeed asymmetric, as previously observed (Figure 2). To behaviorally test the predictions of the normative model and GLM, we trained mice to detect a target embedded in background DRCs while shifting the contrast of the background between high and low contrast. As predicted by the model, mice had lower detection thresholds and were more sensitive to changes in target volume in low contrast. Behavioral adaptation was also asymmetric, decreasing rapidly after a switch to high contrast, and increasing slowly after a switch to low contrast, in agreement with our model (Figure 3). Furthermore, we found that AC is necessary for this detection-in-background task (Figure 4). When recording in AC, we found that the parameters of neurometric functions were predictive of psychometric functions on a mouse-to-mouse basis, and we also showed that target discriminability adapted asymmetrically, as predicted (Figure 5). Finally, we found that we could predict behavioral performance from cortical gain on a session-to-session basis, independently of the effect of contrast (Figure 6). Taken together, these results support our hypothesis that efficient coding at the neuronal level shapes auditory behavior.

The role of cortex in behavior.

The role of auditory cortex in behavior has been subject of debate. A number of prior studies found that auditory cortex was not required for relatively simple behavioral tasks such as frequency discrimination or detection[30,31]. Rather, many studies found that auditory cortex is primarily involved in more complex behaviors, such those requiring temporal expectation[32], localization[33], or discrimination of more complex sounds[34–36]. Consistent with previous findings[37], we found that AC inactivation selectively impaired detection of targets in a noisy background, but did not impair detection of targets in silence (Figure 4). Furthermore, neuronal activity in AC predicted variability in behavioral performance (Figures 5, 6). This set of results establishes that AC is necessary for the detection of targets in background noise and supports the more general notion that AC is required for more difficult auditory tasks.

While the previous work demonstrates the necessity of auditory cortex in behavioral performance, the brain areas and mechanisms supporting the transformation from stimulus to decision are an active field of study[38,39]. By recording during the task, we were able to leverage behavioral variability to show that task performance covaried with representations of targets within small neural populations (Figure 5), and with cortical gain (Figure 6). There is a large body of literature relating cortical codes to behavioral variability: early studies in the visual system suggested that information from relatively small numbers of neurons was sufficient to match or outperform animal behavior in psychophysical tasks[40–42] and that behavioral choice can be predicted from activity in sensory areas[28,42]. These accounts suggested that variability in bottom-up sensory encoding drives the variability in behavioral output. However, more recent work suggests that variability in sensory areas is driven by top-down influences[43–46], which are modulated by attention and learning[47–50]. A recent study imaging tens of thousands of neurons in the visual cortex supports this notion, finding that cortical representations had higher acuity than behaving mice, yet did not correlate with behavioral performance, suggesting that perceptual discrimination depends on post-sensory brain regions[51].

Our results suggest that bottom-up adaptation to stimulus statistics shapes behavioral output: We observed asymmetric time courses of target discrimination following a change in contrast (Figure 3) that were qualitatively consistent with the predictions of efficient coding (Figure 1), resembled contrast gain adaptation in

auditory cortex in the absence of behavior (Figure 2), and resembled patterns of target-driven activity in auditory cortex during the task (Figure 5). Indeed, there have been other studies demonstrating that individual differences in sensory-guided behaviors are reflected in cortical activity[52,53], are bidirectionally modulated by cortical manipulation[54,55], and can be predicted from tuning properties in auditory cortex[56,57]. While our results cannot rule out top-down input as the causal driver of sensory decisions, they do support the notion that the sensory information upon which decisions are made is shaped by neuronal adaptation, which thereby affects behavioral outcomes.

Roles of gain in the auditory system.

Neurons throughout the auditory system adapt to the statistics of the acoustic environment, including the frequency of stimuli over time[58,59], more complex sound patterns[26,60], and task-related or rewarded stimuli[61–66]. Inspired by the latter studies, we intentionally designed our stimuli using unbiased white-noise backgrounds, which allowed us to fit encoding models to our data. Using these methods, we focused on contrast gain control as a fundamental statistical adaptation that relates to efficient coding[17–19,29]. In this study, we developed a novel application of Poisson GLM that allowed us to quantify the contribution of multiplicative interactions between the stimulus and stimulus contrast to the activity of neurons in auditory cortex. Using the fitted parameters of the model, we were able to accurately estimate neuronal gain as a function of time. This approach allowed us to verify that gain adaptation in auditory cortex is asymmetric (Figure 2), as predicted from the normative theory developed in this work (Figure 1).

Furthermore, we found that behavioral detection of targets adapted asymmetrically (Figure 3), which suggested that the dynamics of contrast gain control influenced task performance. Indeed we found that both stimulus contrast and session-to-session fluctuations in gain predicted psychometric performance (Figure 6). These results suggest two sources of gain modulation in auditory cortex: 1) Bottom-up adaptation to stimulus statistics (ie. contrast gain control), and 2) session-to-session modulation of gain. Previous studies have demonstrated this latter phenomenon, suggesting that top-down gain modulation underlies attention[43,44,67] and the maintenance of optimal behavioral states[68,69]. Our results suggest that automatic of gain control as well as session-to-session fluctuations in gain modulate behavior, and provide a starting point for dissecting the neural mechanisms underlying these two forms of gain modulation.

Cellular mechanisms of gain control.

While this work and other studies observed contrast gain control in the auditory system, the neuronal mechanisms driving gain adaptation at a cellular level remain unclear. In the current study, we have likely recorded from a mixed population of excitatory and inhibitory neurons. Different inhibitory neuronal subtypes exhibit specific roles in adaptation[70,71]. While specific inhibitory neuronal subtypes facilitate divisive or subtractive control of excitatory responses in visual[72,73] and auditory cortex[74,75], the role of these interneurons in contrast gain control has been inconclusive[18]. Furthermore, we were able to separate the behavioral contribution of contrast gain control from stimulus-invariant changes in gain (Figure 6). Whether these two forms of gain control share common neural substrates is unclear. By combining cell-specific optogenetic methods with behavioral tasks, future studies may explore and test the specific role of local circuits and top-down modulation in gain control and behavior.

The missing link between efficient coding and behavior.

Combined, our results develop a framework and provide support for the role of efficient neuronal coding in behavior. The efficient coding hypothesis has emerged as one of the leading principles in computational neuroscience that has shaped our understanding of neuronal coding, architecture and evolution[1,22,76–78]. Prior research found that human behavior follows principles of efficiency[19,24]. Our work now provides a framework for linking the principles of neuronal coding with behavioral performance. Additionally, we have introduced a novel application of Poisson GLM designed to detect multiplicative interactions between presented stimuli and other variables. While in this study we focused on the multiplicative effect of contrast, this approach could in theory be applied to any other time-varying signal that modulates neuronal gain, such as movement[79,80], arousal[68,69], or targeted experimental interventions[72–75]. In summary, we expect the theoretical frameworks and modelling methods applied here to have broad utility in the study of neuronal adaptation, a fundamental function of the nervous system.

Online Methods

Animals.

All experiments were performed in adult male (n = 19) and female (n = 19) C57BL/6 (Stock No. 000664) or B6.CAST-*Cdh23*$^{Ahl+}$ (Stock No. 002756) mice (The Jackson Laboratory; age 12-15 weeks; weight 20-30g). Some of the mice used in these experiments were crossed with other cell-type specific -cre lines, as detailed in Extended Data Table 2. All mice were housed with, at most, five mice per cage, at 28°C on a 12-h light:dark cycle with food provided ad libitum, and a restricted water schedule (see *Water restriction*). All experiments were performed during the animals' dark cycle. All experimental procedures were in accordance with NIH guidelines and approved by the Institutional Animal Care and Use Committee at the University of Pennsylvania.

Surgery.

Mice were anesthetized under isoflurane (1-3%). Prior to implantation, all mice were administered subcutaneous doses of buprenorphine (Buprenex, 0.05-0.1 mg/kg) for analgesia, dexamethasone (0.2 mg/kg) to reduce brain swelling, and bupivicane (2 mg/kg) for local anesthesia. In mice implanted with microdrives, two ground screws attached to ground wires were implanted in the left frontal lobe and right cerebellum, with an additional skull screw implanted over the left cerebellum to provide additional support. A small craniotomy was performed over the target stereotactic coordinates relative to bregma, -2.6mm anterior, -4.3mm lateral. Either custom 16-channel microdrives, 32-, or 64-channel shuttle drives (cite) holding tetrode bundles of polyimide-coated nichrome wires were chronically implanted over auditory cortex, and tetrodes were lowered 800um below the pial surface. The exposed tetrodes were covered with GelFoam (Pfizer) or sterile silicone lubricant and sealed with Kwik-Cast (World Precision Instruments). The plastic body of the microdrive and a custom stainless-steel headplate were secured to the skull using dental cement (C&B Metabond) and acrylic (Lang Dental). Mice undergoing only behavioral experiments were implanted with two skull screws in the cerebellum, and a headplate was mounted on the skull as previously described. An antibiotic (Baytril, 5mg/kg) and analgesic (Meloxicam, 5mg/kg) were administered daily (for 3 days) during recovery.

Water restriction.

Following surgical recovery (3 days post-operation), each mouse's weight was monitored for three additional days to establish a baseline weight. Over the next seven days, mice were water deprived, beginning with a daily ration of 120uL/g and gradually decreasing their ration to 40-50uL/g. During the task, if mice did not receive their full ration, the remainder of their ration was provided in their home cage. Mouse weight relative to baseline was monitored during all stages of water restriction. Additional health signs were used to determine a health score and subsequent treatment plan if a mouse lost more than 20% of baseline weight, as described by previously published methods[81] and approved by the Institutional Animal Care and Use Committee at the University of Pennsylvania.

Behavioral apparatus.

During the GO/NO-GO task, the mouse was head-fixed in a custom-built, acoustically isolated chamber. A capacitive touch sensor (AT42QT1010, SparkFun) soldered to a lick spout monitored lick activity. Water rewards were dispensed from a gravity fed reservoir, controlled by a solenoid valve (161T011, Neptune Research) calibrated to deliver approximately 4-5uL of water per reward[82]. Low-level task logic – such as lick detection, reward and timeout delivery, and task timing intervals – was directly controlled by an Arduino Uno microprocessor running custom, low-latency software routines. High-level task logic, such as trial randomization, stimulus buffering and presentation, and online data collection and analysis were controlled by custom MATLAB (Mathworks) software communicating with the Arduino over a USB serial port. Acoustic waveforms were generated in MATLAB and converted to analog signals via a soundcard (Lynx E44, Lynx Studio Technology, Inc.) or a National Instruments card (NI PCIe-6353) and delivered through an ultrasonic transducer (MCPCT-G5100-4139, Multicomp). The transducer was calibrated to have a flat frequency response between 3 kHz and 80 kHz using a 1/4-inch condenser microphone (Brüel & Kjær) positioned at the expected location of the mouse's ear, as described previously[83,84]. During electrophysiological recording sessions, licks were detected using an optical interrupt sensor (EE-SX771, Omron Automation), to prevent lick-related electrical artifacts introduced by contact with a capacitive sensor.

Behavioral timeline.

Each mouse underwent four stages in the behavioral task: 1) water restriction and habituation, 2) behavioral training, 3) psychometric testing, and, 4) offset testing. During the induction of water restriction, mice were habituated to head-fixation in the behavioral chambers and received water through the lick spout, getting a drop of water for licks separated by more than 2 s. After the mouse began to receive its entire ration by licking in the booth, behavioral training was initiated (typically after 1 week). Each mouse was initially trained and tested in one contrast condition (see *Stimuli*), with the initial training condition counterbalanced across mice. Behavioral performance was monitored during training, and mice were considered trained after completing at least three consecutive sessions with over 80% percent correct. After completing training, behavioral thresholds were measured during at least three sessions in which psychometric stimuli were presented (see *Stimuli*). After estimating the behavioral threshold for each mouse, offset stimulus sets were generated using threshold-level targets. After completion of at least three sessions in the offset task, each mouse was then retrained on the remaining contrast condition. Upon reaching the training criterion of 80% in the new contrast condition, mice were then tested in the psychometric and offset tasks as previously described. For mice in electrophysiological experiments, this sequence of training and testing was continued until the recording site yielded less than three units, or until the mouse stopped performing in the task.

Stimuli.

All stimuli were created in MATLAB and sampled at 192 kHz or 200 kHz and 32-bit resolution. A set of dynamic random chords (DRCs) were created with different contrasts, similarly to those described in previous studies[17,19,29]. To construct a DRC, amplitude modulated pure tones were generated at multiple frequencies and then superimposed to create a chord. In some experiments, 34 frequencies were sampled between 4 and ~40kHz in 1/10 octave steps, in the remaining experiments, 33 frequencies were sampled between 4 and 64kHz in 1/8 octave steps. The amplitude envelope of each tone was generated as follows: every 25 ms, amplitudes for each frequency were sampled from a uniform distribution with a mean of 50 dB and a width of ±5 dB in low contrast or ±15 dB in high contrast. Between each 20 ms chord, the amplitude envelope of each frequency band was linearly ramped over 5 ms to the amplitude value for the next chord, such that the total duration of each chord and its ramp was 25 ms. To synthesize the stimuli, amplitude envelopes were multiplied by a sine wave of their respective frequencies, and summed to produce the final waveform. Each time a set of DRCs was generated, 5 unique random number generator seeds were used to restrict the background noise to 5 distinct scenes (see raster in Figure 6 for an example of spike-locking to the repeated scenes).

In all stages of behavioral training and testing, stimuli created for each trial consisted of a DRC background containing a change in contrast, and the presence or lack of a target at a delay after the change in contrast. Each trial began with 3 seconds of DRC background from one contrast, followed by a switch to the other contrast. Targets consisted of a fixed chord composed of 17 frequencies pseudo-randomly sampled from the frequencies contained in the DRC background, such that the target frequencies were uniformly distributed across the frequency range of the background. To add targets to the background noise, the target amplitude at each target frequency was simply added to a single chord in the amplitude envelope of the background, and ramped as described previously: this procedure ensured that target timing was perfectly aligned to changes in the background noise, removing asynchronous timing cues that could be used to detect the target. Target amplitudes are described in values of signal-to-noise ratio (SNR) relative to the average level of the background noise (ie. a 50 dB target embedded in 50 dB background would have an SNR of 0 dB). See Extended Data Table 3 for SNRs used for each mouse. In all trials, targets were embedded after a change in the background contrast, with a delay and volume dependent on the current training or testing stage.

Efficient coding model.

We simulated a model neuron that encodes incoming stimuli via an adapting neural nonlinearity. Stimuli were drawn from a Gaussian distribution whose mean μ was fixed over time but whose standard deviation σ_t could switch over time between a low and a high value ($\sigma_t = \sigma^L$ and $\sigma_t = \sigma^H$, respectively). At each time t, a stimulus s_t was drawn from the distribution $p(s_t|\sigma_t) = \mathcal{N}(s_t; \mu, \sigma_t^2)$, transformed via a saturating nonlinearity of the form $1/\left(1 + e^{-k(s_t - s_0)}\right)$, distorted by Gaussian noise with variance σ_n^2, and finally discretized into N discrete levels to generate a response r_t. This discrete response was linearly decoded to extract an estimate $\hat{s}_t$ of the current stimulus: $\hat{s}_t = p_1 r_t + p_0$. The recent history of L stimulus estimates was used to update an estimate $\hat{\sigma}_t$ of the underlying standard deviation: $\hat{\sigma}_t = \mathrm{std}(\hat{s}_{t-L+1}{:}\,\hat{s}_t)$. The estimate $\hat{\sigma}_t$ was then used to select the parameters of the encoder (k, s_0) and the decoder (p_1, p_0) on the next timestep. The encoding and decoding parameters were chosen to minimize the expected error in decoding stimuli given the neuron's current estimate of the underlying standard deviation: $k, s_0, p_1, p_0 = \mathrm{argmin}\langle(\hat{s}_t - s_t)^2\rangle_{p(s_t|\hat{\sigma}_t)}$[21,23].

The parameters of the encoder and decoder were adapted based on a background stimulus with a mean μ_B that was fixed over time and a standard deviation σ_B that switched between low and high values σ_B^L and σ_B^H, respectively. We used this adapting nonlinearity to determine how well this model neuron could discriminate target stimuli from this background. Target stimuli were sampled from a Gaussian distribution with a fixed mean μ_T and with a variance σ_T that was scaled in proportion to the variance of the background ($\sigma_T^L = f\sigma_B^L$ and $\sigma_T^H = f\sigma_B^H$, respectively). At each timestep, we computed the Bhattacharyya coefficient (BC) of the response distributions produced by background versus target stimuli: $BC = \sum\sqrt{p(r_B)p(r_T)}$. We used $1 - BC$ as our measure of discriminability.

We simulated the behavior of this model using a background "probe" stimulus whose standard deviation switched every T timesteps. We simulated N_c cycles of this probe stimulus, where each cycle consisted of T timesteps in the low state, followed by T timesteps in the high state. This yielded timeseries of the gain k and offset s_0 of the adapting nonlinearity, as well as distributions of the neural response to the background and target stimuli at each timepoint following a switch in standard deviation. We averaged the gain and offset across cycles to obtain the average properties of the encoder at each timepoint following a switch. We used the distribution of responses to target and background stimuli, measured across cycles, to compute the discriminability at each timepoint following a switch. All simulations were performed with the following values: $T = 50$, $N_c = 1{,}000$, $\mu_B = 0$, $\mu_T = 0$ to 3 in 0.25 steps, $\sigma_B^L = 1$, $\sigma_B^H = 3$, $f = 0.25$, $\sigma_n^2 = 0.01$, $N = 15$, $L = 12$. For Figure 1g, model discriminability in each contrast was fit with a logistic function to estimate the sensitivity and threshold of the model. To approximate the stimulus conditions used in the offset task, the target thresholds for each contrast were then used to select target volumes to plot discriminability over time ($\mu_T^L = 1.50$, $\mu_T^H = 2.25$; Figure 1f).

Behavioral task.

We employed a GO/NO-GO task to measure the detectability of targets in background. In this task, each trial consisted of a noise background with a contrast shift, along with the presence or absence of a target after the change in contrast. Mice were trained to lick when they detected a target (hit), or to withhold licking in the absence of a target (correct reject). This behavior was reinforced by providing a 4-5 uL water reward when the mouse licked correctly (hit), and by initiating a 7-10 s timeout when the mouse licked in the absence of a target (false alarm). Any licks detected during the timeout period resulted in the timeout being reset. In a subset of mice, we introduced an additional trial abort period coincident with the first part of the contrast background, before the contrast switch. Any licks detected in this abort period resulted in the trial being repeated after a 7-10 s timeout, until the mouse withheld from licking during this period. In this task, misses and correct rejects were not rewarded or punished. Trials were separated by a minimum 1.5s inter-trial-interval (ITI). To discourage spontaneous licking, licks were monitored during this period, and if any licks occurred the ITI timer was reset.

To prevent mice from predicting target time, we varied the timing of the target relative to the contrast shift. This required a method for estimating hit rates and false alarm rates at different times during each trial, and to reward and punish the animal during these times in an unbiased manner. To approach this issue, we considered licks only during a 1 s response window after a target presentation (eg. if a target was presented 500 ms post-contrast-switch, the response window persisted from 500 to 1500 ms post-contrast-switch). To apply this method to background-only trials, in which no targets were presented, we considered background trials to be target trials containing infinitely small target amplitudes. For each background trial, we assigned a response window with equiprobable delay matched to the target conditions and considered only licks within those "target" response windows. Thus, over the course of a session, we randomly sampled lick probabilities in background trials during the same temporal windows as those considered during target trials. Using this scheme, we treated target and background-only trials identically, and estimated hit rates and false alarm rates over time in an unbiased manner.

Each mouse performed three stages in the behavioral task: training, psychometric testing, and offset testing. During the training task, trials consisted of two types, background-only trials or target trials presented with equal probability. To facilitate learning, we selected target SNRs at the highest end of the range described previously: in low contrast training sessions, targets were 16 dB SNR, and in high contrast training sessions, targets were 20 dB SNR. To prevent response bias as a function of target timing, we randomly varied the target delay between 250, 500, 750 and 1000 ms after the contrast change in each trial. During the psychometric testing task, there were 7 trial types consisting of background-only trials and target trials spanning six different SNRs (Extended Data Table 3). Based on behavioral piloting, we presented high SNR trials with a greater probability, to ensure that mice were consistently rewarded during the task. In low and high contrast psychometric sessions, the probability of a background trial was 0.4, the probability of the four lowest target SNRs was 0.05 each, and the probability of the two highest target SNRs was 0.2 each. As in training, target timing was varied randomly

between 250, 500, 750 and 1000 ms after the contrast change in each trial. After completing at least three sessions of the psychometric task, stimuli were generated for the offset testing task. This task consisted of 15 unique trial types: 3 target volumes (background trials, threshold target trials, and high SNR target trials), and 5 target delays relative to the contrast change (25, 75, 225, 475, 975 ms delay). Threshold target amplitudes were determined individually for each mouse by fitting performance averaged over several sessions with a psychometric function, and extracting the volume at which the slope of the psychometric curve was steepest. Based on behavioral piloting, background trials, threshold target trials, and high SNR target trials were presented with probabilities of 0.4, 0.2, and 0.4, respectively. Target delay on each trial was selected with equal probability. In all behavioral stages, trial order was pseudorandomly generated, such that there were no more than three target or background trials in a row.

A subset of mice (n = 2), were presented targets in silence (Figure 4). To generate this stimulus set without changing the basic structure of the task or stimuli, we simply took the spectrograms of all stimuli containing 25 dB SNR targets from the low-to-high contrast stimulus sessions, and set the stimulus power flanking each target to zero. This manipulation was only performed in the target period, and the low contrast adaptation period of the trials remained the same. Thus, the targets and adaptation periods were identical to those presented in the target-in-background task. To vary the difficulty of the task, the volume of the target was attenuated using the following values: -75, -60, -45, -30, -15, and 0 dB attenuation relative to the 25 dB SNR target. Mice were previously trained in the target-in-background task prior to performing the target in silence task. Before psychometrically varying the target attenuation, mice were trained in the new task to criterion performance. Mice generalized very rapidly to the new task, reaching 97% and 94% training accuracy on the first day of exposure to targets in silence (mice CA124 and CA125, respectively).

Chronic muscimol application.

A separate cohort of mice (n = 4) were bilaterally implanted with 26 GA guide cannulae (PlasticsOne, C315GMN-SPC mini, cut 5 mm below pedestal) in auditory cortex. The surgery was performed as described previously with the following modifications. After the skull was leveled using a stereotax, two small craniotomies were performed -2.6 mm anterior, ±4.3 mm lateral from bregma, over auditory cortex. The guide cannulae and dummy infusion cannulae (PlasticsOne, C315DCMN-SPC mini, cut to fit 5 mm C315GMN with a 0.5 mm projection depth) were sterilized in an autoclave. The dummy cannulae were partially screwed into the guide cannulae and placed in a stereotaxic clamp. After zeroing the tip of the guide cannula to the brain surface, the cannula was lowered to 500 µm below the cortical surface. This depth was chosen because the infusion cannulae (PlasticsOne, C315LIMN-SPC mini) project 500 µm from the end of the guide cannulae when completely inserted, leading to a final depth of 1000 µm – the location of auditory cortex. The dummy cannulae were then fully inserted and this procedure was repeated for the next cortical hemisphere.

Prior to injecting, two injection syringes (Hamilton Syringe, 10µL Gaslight #1701) and tubing (C313CT tubing 023x050 PE50) were backfilled with mineral oil. Sterilized infusion cannulae were then attached to each syringe and ~500nL of muscimol (diluted with 1x PBS to .25 mg/mL; Sigma Aldrich, M1523) or 0.9% sterile saline was drawn up into the injection cannulae using a dual injector (Harvard Apparatus, Pump 11 Pico Plus Elite). The mouse was then headfixed and the dummy cannulae were removed and sterilized. The loaded infusion cannulae were then screwed all the way into the guide cannulae and 400 nL of muscimol or saline was infused bilaterally at a rate of 250 nL/minute. The infusion cannulae were then replaced with the dummy cannulae and the mouse rested in its home cage for 30-45 minutes before beginning the behavioral session.

Acute electrophysiological recordings.

For acute recordings used to fit the GC-GLM model (Figure 2), neuronal signals were recorded from n = 1 awake, untrained mouse. Prior to the recording session, the mouse was anesthetized and a headpost and ground pin were implanted on the skull (see *Surgery*). On the day of the recording, the mouse was briefly anesthetized with 3% isoflurane and a small craniotomy was performed over auditory cortex using a dental drill or scalpel (~1 mm x 1 mm craniotomy centered approximately 1.25 mm anterior to the lambdoid suture along caudal end of the squamosal suture). A 32 channel silicon probe (Neuronexus) was then positioned perpendicularly to the cortical surface and lowered at a rate of 1-2 µm/s to a final depth of 800-1200 µm. As the probe was lowered, trains of brief noise bursts were repeated, and if stimulus locked responses to the noise bursts were observed, the probe was determined to be in auditory cortex. The probe was then allowed to settle for up to 30 minutes before starting the recording. Neuronal signals were amplified and digitized with an Intan headstage (RHD 32ch) and recorded by an openEphys acquisition board[85] at a rate of 30 kHz.

For this experiment, the mouse was presented with 3 s DRCs alternating between low and high contrast (uniform distribution with a mean of 50 dB and a width of ±5 dB in low contrast or ±15 dB in high contrast at a chord rate of 25 ms, as described in *Stimuli*). In order to accurately fit the GLM in an unbiased manner, these stimuli were highly random, composed of 100 unique chord patterns for each contrast (Extended Data Figure 2i,j**)**. For each of the two recording sites, 5 repeats of this stimulus set were played.

Behavioral electrophysiological recordings.

Neural signals were acquired from awake, behaving mice as they performed the psychometric and offset testing tasks described previously. Chronically implanted, 16-, 32-, or 64-channel microdrives[85,86] were connected to one or two 32 channel Intan amplifier headstages. Amplified signals were recorded at 30 kHz using an openEphys acquisition board via an SPI cable, where the signals were digitized.

For all recordings, broadband signals were filtered between 500 and 6000 Hz, offset corrected, and re-referenced to the median across all active channels. The preprocessed data was then sorted using KiloSort[87] or KiloSort2 and the resulting clustering was manually corrected in phy2 according to community-developed guidelines. The resulting units were labelled as single units if they exhibited a clear refractory period and did not need to be split. Splitting assessments were made through manual examination of principle component features for the two best channels of a cluster. If two noticeable clusters in feature space were evident in a unit, the unit was either manually split, or classified as a multiunit.

Generalized linear model.

To justify the form of GLM used here, we discuss a how a model neuron could implement gain control in the simplest terms, and then structure our inference model to extract the parameters of this model neuron. We will assume that the activity of the model neuron is driven by three sources: 1) stimulus drive, 2) stimulus contrast, and 3) the multiplicative interaction between the two, which we use to define the gain (for a formal definition of this forward model and the inference model, see *Supplementary Information*).

As discussed previously, the stimulus used in our experiments is composed of many frequencies that change in loudness in discrete time steps:

$$X_{t,f} \sim \mathcal{U}(\mu, \sigma_t)$$

where $X_{t,f}$ is the stimulus spectrogram that varies as a function of time t and frequency f. Each time and frequency bin of X is sampled from a uniform distribution defined by an average value μ and contrast σ_t.

We assume that the hypothetical neuron responds selectively at some frequency and time lag, defined by a filter, or STRF $\beta_{h,f}$ with history h and frequency f components. Given β, we can define the stimulus drive x_t as

$$x_t = X_t \beta \tag{1}$$

where at each time t, X_t is a row vector of size F frequencies times H lags (ie. the "unrolled" lagged stimulus spectrogram) and β is the STRF unrolled to a single column vector of the same size.

In the spirit of efficient coding theory, and as shown in previous work, we assume that the gain g of the neuron should be inversely proportional to the contrast, such that $g(\sigma) \propto 1/\sigma$ (ie. when contrast is low gain should be high, and vice-versa). We also define "neutral" gain to be the average of the gain of the neuron in low and high contrast. Putting these two features together, we can summarize the gain of the neuron as

$$g(\sigma) = \frac{\bar{\sigma}}{\sigma_t} \tag{2}$$

where $\bar{\sigma}$ is the harmonic mean of the contrast in the low and high conditions (see *Supplementary Information*). In the case of a 3-fold change in contrast, this function constrains the gain of the neuron between 0.5 and 1.5, with a neutral value of 1. As mentioned previously, we consider gain to be the multiplicative interaction between the stimulus drive and the contrast, such that the contribution of gain control to the response of the neuron is related to $x_t \cdot \frac{\bar{\sigma}}{\sigma_t}$.

To summarize, we considered a hypothetical neuron driven by the stimulus according to a STRF β and by the interaction between the stimulus drive and the contrast $x_t \cdot \frac{\bar{\sigma}}{\sigma_t}$. To infer the relative weights of each of

these components of the neural response, we defined a Poisson GLM with an intercept term and the following predictors:

$$x_t, \qquad \frac{\overline{\sigma}}{\sigma_t}, \qquad x_t \cdot \frac{\overline{\sigma}}{\sigma_t} \tag{3}$$

In other words, the model is composed of a stimulus predictor x_t, a contrast predictor $\overline{\sigma}/\sigma_t$, and their interaction. Therefore, the GLM models the firing rate λ at time t as a Poisson distribution with the following mean:

$$\lambda_t = exp\left[\beta_0 + x_t\beta_1 + x_t \cdot \frac{\overline{\sigma}}{\sigma_t}\beta_2 + \frac{\overline{\sigma}}{\sigma_t}\beta_3\right] \tag{4}$$

where $\beta_0 \ldots \beta_3$ are the parameters to be inferred. Based on our behavioral data (Figure 3) and the predictions of the efficient coding model (Figure 1), we expected the influence of contrast on neural gain to be asymmetric and smooth. To enable the GLM to capture both of these qualities, we first defined the contrast predictors from a set of cubic B-spline temporal basis functions, then defined separate contrast predictors for transitions to low and high contrast. Incorporating these changes, we can redefine equation 4 above as

$$ln\lambda_t = \beta_0 + x_t\beta_1 + x_t \circ C_t'\beta_2 + C_t'\beta_3 \tag{5}$$

where $\circ$ denotes element-by-element "broadcasting" multiplication and C_t' is a matrix of contrast predictors $\overline{\sigma}/\sigma_t$ convolved with a set of basis functions and separated by contrast transitions (see *Supplementary Information*). For the sake of clarity, note that in the expression above, β_0 is a number, x is a column vector of length T, β_1 is a number, C' is a T-by-$2B$ matrix, and β_2 and β_3 are column vectors of length $2B$, where B is the number of splines.

So far, we outlined a hypothetical neuron which implements gain control, and a GLM with which we can approximate the behavior of this neuron. Next, we describe how to use the fitted parameters to quantify the gain of the neuron. Conceptually, an increase or decrease in the gain of the neuron is analogous to more or less sensitivity to small changes in the stimulus. Based on this intuition, we focused on how the response of the neuron (as modelled by a fitted GLM) is expected to change between conditions where the gain is expected to contribute (ie. in the presence of gain control) and where it is not (ie. in the absence of gain control, where gain is "neutral"). Following this logic, we derived a definition for gain w_t as the ratio between the sensitivity of the fitted model with changes in contrast, compared to the sensitivity of the same model when the contrast is at a reference value, which we defined previously as $w_t = 1$:

$$w_t = \frac{\beta_1 + C_t'\beta_2}{\beta_1 + C^0\beta_2} \tag{6}$$

where w_t is the estimated gain at time t, and C^0 is a reference contrast design matrix identical to C_t' except that all non-zero elements are set to 1 (see *Supplementary Information* for full derivation of w_t).

To fit the model, we implemented a two-step procedure. In the first step, the STRF β of the neuron was estimated according to the model

$$ln\,\lambda_t = \alpha + X_t\beta \tag{7}$$

For the second step, we calculated the stimulus drive as described in equation 1, and then fit equation 5 to the data for each neuron using glmfit in MATLAB. This entire fitting procedure was 10-fold cross-validated with folds stratified across trials of each contrast. In the first step, we fit the STRF β with F frequency bins according to the stimulus spectrogram (F = 33 or 34, see *Stimuli*) and a history window of 300 ms (H = 12). When fitting the full model, we defined the contrast design matrix C_t' to capture 1000 ms of contrast history around each transition (H' = 40), convolved with a set of B-spline temporal basis functions[88] (here, we used B-splines with a degree of 3 and 3 equally-spaced knots, constrained to go smoothly to zero at the longest lag, which implied that B = 4).

To validate the model, we first simulated neurons according to the forward model outlined above (Extended Data Figure 2a) while varying the amount of gain control and the temporal trajectory of gain in different simulation runs. We found that the GLM accurately predicted the STRF shape, spike rates and gain trajectories across a variety of simulation parameters (Extended Data Figure 2c, e-h). For a detailed description and discussion of the simulation results, see *Supplementary Information* and Extended Data Table 4.

Behavioral and neural detection performance.

To calculate performance in the target-in-background detection task we adopted commonly used signal detection theory methods[40,89] to estimate the ability of an ideal observer to discriminate between two sensory distributions: in our case, a distribution for target trials and a distribution for background trials. When analyzing behavior, we computed the percent correct performance of an ideal observer[90] as a function of the probability of hits and false alarms:

$$pc = z^{-1}\left(\frac{z(H) - z(FA)}{\sqrt{2}}\right) \tag{8}$$

where z^{-1} is cumulative probability of the normal distribution ($normcdf$ in MATLAB), z is the inverse of the normal distribution (ie. the z-score, $norminv$ in MATLAB), H is the hit rate, and FA is the false alarm rate. For psychophysical performance, hit rates and false alarm rates near 0 and 1 were adjusted using the log-linear rule[91], to reduce biases in performance estimation caused by low numbers of trials.

To calculate neural performance in the same reference frame as the behavior, we employed similar ideal observer techniques. First, neuronal responses (either spike rates or single units, or population projection values), were averaged in a 100ms window post target onset (for background trials, this window was randomly chosen on each trial to coincide with target presentation times on target trials). Then, using the distributions of responses during target and background trials, we computed receiver-operating-characteristic curves and took the area under the curve (AUC) as the percent correct of an ideal observer discriminating between the target and background distributions. To determine whether the AUC value for a given set of trial distributions was significantly different from chance, we performed a bootstrap procedure where we sampled from all the trials with replacement 500 times and recomputed AUC for each sample. If the 95% confidence intervals for this bootstrapped distribution did not include chance (.5), we defined that AUC value as significant. For population analyses which generated single-trial predictions, neural hit and false alarm rates were transformed to percent correct as described above.

To characterize performance, psychometric curves were fit with a logistic function:

$$y = \gamma + (1 - \gamma - \lambda) * \frac{1}{1 + e^{\alpha - \beta x}} \tag{9}$$

where α is the x-offset of the function, β determined the sensitivity of the function, γ determined the guess rate (lower bound), λ determined the lapse rate (upper bound) and x was stimulus volume. α/β determined the threshold of this function, defined as the volume corresponding to the steepest part of the curve. This function was fit to behavioral or neural performance using constrained gradient descent ($fmincon$ in MATLAB) initialized with a 10x10 grid-search of parameters α and β.

To characterize adaptation time constants, adaptation curves were fit with an exponential function

$$y = a + b * e^{-\frac{t}{\tau}} \tag{10}$$

where a determined the y-offset of the function, b was a multiplicative scaling factor, and τ was the time constant of the exponential in units of time t. This function was fit to behavioral or neural responses using constrained gradient descent initialized with a 10x10x10 grid search across all three parameters.

Population response metrics.

On sessions where three or more neurons were simultaneously recorded, we used a population vector technique[27] to estimate the ability of neural populations to discriminate targets from background. First, spike rates in each trial were averaged in a 100ms window post-target onset. Then, using a leave-one-out procedure, we computed a trial averaged population vector for target trials, v_T, and a separate average population vector for background trials, v_B. We then estimated the coding direction in high dimensional neural space that best separated the target and background responses: $CD = v_T - v_B$. The held out trial was then projected along this dimension, by taking the population response vector on that trial v_{trial} and projecting it along the estimated coding direction using the dot product: $projection\ value = v_{trial} * CD$. This procedure was repeated holding out each trial, and estimating the coding direction from the remaining trials. For psychometric testing sessions, the target responses from the two loudest target volumes were used to estimate coding direction, and in offset testing

sessions the target responses from the high SNR target trials were used. After computing projections for every trial, the resulting matrix was normalized between 0 and 1.

Population classifier.

Based on previously described methods[28], we used a criterion-based decision rule to estimate how a hypothetical down-stream neuron may read out the neural activity of a population of neurons. As before, trial distributions of neural responses to targets or background were created from the average activity in a 100ms window post-target. Then, we sampled 100 criterion values between the minimum and maximum response, and for each criterion estimated the proportion of correct trials under two decision rules: 1) report target present if the response is greater than the criterion, or, 2) report target present if the response is less than the criterion. By assessing these two decision rules, neurons that were suppressed by target presence were treated equally to neurons that were enhanced by target presence. Finally, we chose the criterion and decision rule that yielded the highest proportion of correct trials, and computed neural hit rates and false alarm rates for each target level, and background-only trials. These hit rates and false alarm rates were then transformed to percent correct according to Equation 8.

Linear-nonlinear model.

First, we selected only neurons in the dataset which had reliable responses to stimulus repeats. To determine response reliability, we computed a noise ratio (NR) for each neuron, which describes the amount of variability in the response due to noise versus the amount of variability in the response driven by the stimulus[92,93]. Values approaching 0 indicate increasingly reliable responses to the stimulus, so for the remaining analyses, we included neurons with NR < 100.

The linear nonlinear model was composed of a spectrotemporal receptive field (STRF) and a set of rectifying nonlinearities. The STRF β was fit using normalized reverse correlation

$$\beta = [XX^T]^{-1}X\lambda \tag{11}$$

where X is the stimulus design matrix X_t defined in equation 1 and λ is the spike count in each 25 ms bin of the DRC stimulus. When defining X, we used a history window of 300 ms (H = 12) and frequency bins corresponding to the frequencies composing the dynamic random chords (see *Stimuli*). After fitting the STRF, we fit the nonlinearities of the neuron. This two-step fitting procedure was repeated using 10 fold cross-validation, as described below.

For each fold, we selected 90% of the trials for training, leaving the remaining 10% to be held out for testing. Within each trial, we excluded neuronal responses around transitions from silence, or transitions in contrast, to prevent the model from overfitting strong transients in the neural response. Additionally, we excluded neural responses within a 50ms window after target presentation, to prevent overfitting of target responses. Given these exclusion criteria, we calculated the duration of stimulus sampled in the target period for each trial, and, for each trial, sampled the same duration of stimulus within the adaptation period. This procedure ensured that the model was fit to the same amount of high and low contrast stimulation per trial, to minimize overfitting to one contrast condition. Then, a stimulus design matrix X was defined using these stimulus periods, and the STRF was fit using equation 11. During an initial pilot experiment, we tested whether STRF properties were affected by stimulus contrast, and found STRFs to be largely stable when estimated separately for each contrast (*Supplementary Information* and Extended Data Figure 5b-g). Therefore, we used both periods of contrast to estimate β.

Using the STRF fit to the training data, we computed the linear drive x_t by convolving the STRF with the lagged spectrogram of the training stimulus (equation 1). For the GC-LN model we separated the linear predictions into low and high contrast periods, while for the static-LN model all matched time points were used. We generated a histogram of the linear prediction values (50 bins), and for each bin, computed the mean spike rate of the neuron when the linear prediction fell within those bin edges (Figure 6d, scatter points). The resulting set of linear prediction values and average spike rates were fit with an exponential function:

$$y = a + be^{c(x-d)} \tag{12}$$

where a determined the minimum firing rate, b was a multiplicative scaling factor, c determined the gain of the exponent, and d determined the x-offset, or firing threshold of the neuron. This function was fit to each cell using constrained gradient descent ($fmincon$ in MATLAB), using a 10x10 grid search for parameters b and c. The gain

for each neuron was defined as c. This entire process was repeated for each cross-validation fold, and the final parameter estimates for the STRF and nonlinearities were taken as the average over the 10 runs.

To determine the relationship between neuronal gain and behavioral performance, we computed the average neural gain across all noise responsive neurons (NR < 100) in each session for the adaptation and target periods in the trial. We then compared the session-averaged gain values to the fitted thresholds and slopes of the psychometric curve across sessions using the mixed-effects linear models outlined in the main text.

Inclusion criteria.

Unless otherwise noted, behavioral sessions in which the false alarm rate exceeded 50% were discarded from analysis. One mouse (ID: CA122) had consistently high false alarm rates in the high contrast condition, so we excluded high contrast sessions from this mouse from all analyses. For Figures 5 and 6, we removed neurons with low spike rates (<1Hz) and noise-like or inverted (ie. upward inflected) spike waveforms. To determine waveform quality, we computed the width of each waveform at half of the minimum value (FWHM) and its correlation with the average waveform over all neurons. Neurons whose waveforms had outlier FWHM values ($isoutlier$ in MATLAB), negative correlations, or were not significantly correlated with the average (Bonferoni corrected p > 5.85e-6) were removed from further analysis. For Figure 5g-i, sessions with stable population decoding performance were included (defined as sessions where more than half of the target volumes or times elicited significant population AUC values, as determined by the bootstrap procedure described previously). For Figure 6e-h, only neurons with noise ratios less than 100 were included in all analyses.

1. Barlow, H. B. Possible Principles Underlying the Transformations of Sensory Messages. in *Sensory Communication* **6**, 216–234 (2013).
2. Brenner, N., Bialek, W. & De Ruyter Van Steveninck, R. Adaptive rescaling maximizes information transmission. *Neuron* **26**, 695–702 (2000).
3. Bharioke, A. & Chklovskii, D. B. Automatic Adaptation to Fast Input Changes in a Time-Invariant Neural Circuit. *PLoS Comput Biol* **11**, 1004315 (2015).
4. Borst, A. & Theunissen, F. E. Information theory and neural coding. *Nature Neuroscience* **2**, 947–957 (1999).
5. Baccus, S. A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36**, 909–919 (2002).
6. Dean, I., Harper, N. S. & McAlpine, D. Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* **8**, 1684–1689 (2005).
7. Lesica, N. A. *et al.* Adaptation to Stimulus Contrast and Correlations during Natural Visual Stimulation. *Neuron* **55**, 479–491 (2007).
8. Gutnisky, D. A. & Dragoi, V. Adaptive coding of visual information in neural populations. *Nature* **452**, 220–224 (2008).
9. Wen, B., Wang, G. I., Dean, I. & Delgutte, B. Dynamic range adaptation to sound level statistics in the auditory nerve. *J. Neurosci.* **29**, 13797–13808 (2009).
10. Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L. & King, A. J. Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space. *Neuron* **66**, 937–948 (2010).
11. Wen, B., Wang, G. I., Dean, I. & Delgutte, B. Time course of dynamic range adaptation in the auditory nerve. *J. Neurophysiol.* **108**, 69–82 (2012).
12. Clarke, S. E., Longtin, A. & Maler, L. Contrast coding in the electrosensory system: Parallels with visual computation. *Nature Reviews Neuroscience* **16**, 733–744 (2015).
13. Clemens, J., Ozeri-Engelhard, N. & Murthy, M. Fast intensity adaptation enhances the encoding of sound in Drosophila. *Nat. Commun.* **9**, 1–15 (2018).
14. Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H. & King, A. J. Contrast Gain Control in Auditory Cortex. *Neuron* **70**, 1178–1191 (2011).
15. Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H. & King, A. J. Spectrotemporal contrast kernels for neurons in primary auditory cortex. *J. Neurosci.* **32**, 11271–11284 (2012).
16. Rabinowitz, N. C., Willmore, B. D. B., King, A. J. & Schnupp, J. W. H. Constructing Noise-Invariant Representations of Sound in the Auditory Pathway. *PLoS Biol.* **11**, e1001710 (2013).
17. Cooke, J. E., King, A. J., Willmore, B. D. B. & Schnupp, J. W. H. Contrast gain control in mouse auditory cortex. *J. Neurophysiol.* **120**, 1872–1884 (2018).
18. Cooke, J. E. *et al.* Contrast gain control occurs independently of both parvalbumin-positive interneuron activity and shunting inhibition in auditory cortex. *J. Neurophysiol.* **123**, 1536–1551 (2020).
19. Lohse, M., Bajo, V. M., King, A. J. & Willmore, B. D. B. Neural circuits underlying auditory contrast gain control and their perceptual implications. *Nat. Commun.* **11**, 1–13 (2020).
20. DeWeese, M. & Zador, A. Asymmetric Dynamics in Optimal Variance Adaptation. *Neural Comput.* **10**, 1179–1202 (1998).
21. Młynarski, W. F. & Hermundstad, A. M. Adaptive coding for dynamic sensory inference. *Elife* **7**, (2018).
22. Młynarski, W., Hledík, M., Sokolowski, T. R. & Tkačik, G. Statistical analysis and optimality of neural systems. *Neuron* **109**, 1227-1241.e5 (2021).
23. Młynarski, W. F. & Hermundstad, A. M. Efficient and adaptive sensory codes. *Nat. Neurosci.* **24**, 998–1009 (2021).
24. Wei, X.-X. & Stocker, A. A. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nat. Neurosci.* **18**, 1509–1517 (2015).
25. Maier, J. K. *et al.* Adaptive coding is constrained to midline locations in a spatial listening task. *J. Neurophysiol.* **108**, 1856–1868 (2012).
26. Pennington, J. R. & David, S. V. Complementary effects of adaptation and gain control on sound encoding in primary auditory cortex. *eNeuro* **7**, 1–17 (2020).
27. Li, N., Daie, K., Svoboda, K. & Druckmann, S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532**, 459–464 (2016).
28. Christison-Lagay, K. L., Bennur, S. & Cohen, Y. E. Contribution of spiking activity in the primary auditory cortex to detection in noise. *J. Neurophysiol.* **118**, 3118–3131 (2017).
29. Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H. & King, A. J. Contrast Gain Control in Auditory

Cortex. *Neuron* **70**, 1178–1191 (2011).
30. Talwar, S. K., Musial, P. G. & Gerstein, G. L. Role of mammalian auditory cortex in the perception of elementary sound properties. *J. Neurophysiol.* **85**, 2350–2358 (2001).
31. Gimenez, T. L., Lorenc, M. & Jaramillo, S. Adaptive categorization of sound frequency does not require the auditory cortex in rats. *J. Neurophysiol.* **114**, 1137–1145 (2015).
32. Jaramillo, S. & Zador, A. M. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* **14**, 246–253 (2011).
33. Wood, K. C., Town, S. M., Atilgan, H., Jones, G. P. & Bizley, J. K. Acute inactivation of primary auditory cortex causes a sound localisation deficit in ferrets. *PLoS One* **12**, (2017).
34. Kato, H. K., Gillet, S. N. & Isaacson, J. S. Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance. *Neuron* **88**, 1027–1039 (2015).
35. Ceballo, S., Piwkowska, Z. & Bourg, J. Targeted Cortical Manipulation of Auditory Perception In Brief. *Neuron* **104**, 1168-1179.e5 (2019).
36. Li, Z. *et al.* Corticostriatal control of defense behavior in mice induced by auditory looming cues. *Nat. Commun.* **12**, 1–13 (2021).
37. Town, S., Wood, K. & Bizley, J. Signal processing in auditory cortex underlies degraded speech sound discrimination in noise. *bioRxiv* 833558 (2019). doi:10.1101/833558
38. Musall, S., Urai, A. E., Sussillo, D. & Churchland, A. K. Harnessing behavioral diversity to understand neural computations for cognition. *Current Opinion in Neurobiology* **58**, 229–238 (2019).
39. Shadlen, M. N. & Kiani, R. Decision making as a window on cognition. *Neuron* **80**, 791–806 (2013).
40. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
41. Britten, K. H. *et al.* The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
42. Shadlen, M. N., Britten, K. H., Newsome, W. T. & Movshon, J. A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
43. Nienborg, H. & Cumming, B. G. Decision-related activity in sensory neurons reflects more than a neurons causal effect. *Nature* **459**, 89–92 (2009).
44. Cumming, B. G. & Nienborg, H. Feedforward and feedback sources of choice probability in neural population responses. *Current Opinion in Neurobiology* **37**, 126–132 (2016).
45. Tsunada, J., Liu, A. S. K., Gold, J. I. & Cohen, Y. E. Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat. Neurosci.* **19**, 135–142 (2015).
46. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
47. Cohen, M. R. & Newsome, W. T. Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron* **60**, 162–173 (2008).
48. Cohen, M. R. & Newsome, W. T. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* **29**, 6635–6648 (2009).
49. Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J. & Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science (80-.).* **359**, 463–465 (2018).
50. Downer, J. D., Niwa, M. & Sutter, M. L. Task Engagement Selectively Modulates Neural Correlations in Primary Auditory Cortex. *J. Neurosci.* **35**, 7565–7574 (2015).
51. Stringer, C., Michaelos, M., Tsyboulski, D., Lindo, S. E. & Pachitariu, M. High-precision coding in visual cortex. *Cell* (2021). doi:10.1016/j.cell.2021.03.042
52. Hires, S. A., Gutnisky, D. A., Yu, J., O'Connor, D. H. & Svoboda, K. Low-noise encoding of active touch by layer 4 in the somatosensory cortex. *Elife* **4**, (2015).
53. Hobbs, J. A., Towal, R. B. & Hartmann, M. J. Z. Spatiotemporal patterns of contact across the rat vibrissal array during exploratory behavior. *Front. Behav. Neurosci.* **9**, 356 (2016).
54. Aizenberg, M. & Geffen, M. N. Bidirectional effects of aversive learning on perceptual acuity are mediated by the sensory cortex. *Nat. Neurosci.* **16**, 994–996 (2013).
55. Aizenberg, M., Mwilambwe-Tshilobo, L., Briguglio, J. J., Natan, R. G. & Geffen, M. N. Bidirectional Regulation of Innate and Learned Behaviors That Rely on Frequency Discrimination by Cortical Inhibitory Neurons. *PLOS Biol.* **13**, e1002308 (2015).
56. Briguglio, J. J., Aizenberg, M., Balasubramanian, V. & Geffen, M. N. Cortical neural activity predicts sensory acuity under optogenetic manipulation. *J. Neurosci.* **38**, 2094–2105 (2018).

57. Wood, K. C., Angeloni, C. F., Oxman, K., Clopath, C. & Geffen, M. N. Neuronal activity in sensory cortex predicts the specificity of learning. *bioRxiv* 2020.06.02.128702 (2020). doi:10.1101/2020.06.02.128702
58. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* **6**, 391–398 (2003).
59. Natan, R. G., Carruthers, I. M., Mwilambwe-Tshilobo, L. & Geffen, M. N. Gain Control in the Auditory Cortex Evoked by Changing Temporal Correlation of Sounds. *Cereb. Cortex* **27**, 2385–2402 (2017).
60. Espejo, M. L., Schwartz, Z. P. & David, S. V. Spectral tuning of adaptation supports coding of sensory context in auditory cortex. *PLoS Comput. Biol.* **15**, e1007430 (2019).
61. Fritz, J., Shamma, S., Elhilali, M. & Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **6**, 1216–1223 (2003).
62. Mesgarani, N., Fritz, J. & Shamma, S. A computational model of rapid task-related plasticity of auditory cortical receptive fields. *J. Comput. Neurosci.* **28**, 19–27 (2010).
63. David, S. V., Fritz, J. B. & Shamma, S. A. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2144–2149 (2012).
64. Yin, P., Fritz, J. B. & Shamma, S. A. Rapid spectrotemporal plasticity in primary auditory cortex during behavior. *J. Neurosci.* **34**, 4396–408 (2014).
65. Niwa, M., Johnson, J. S., O'Connor, K. N. & Sutter, M. L. Active Engagement Improves Primary Auditory Cortical Neurons' Ability to Discriminate Temporal Modulation. *J. Neurosci.* **32**, 9323–9334 (2012).
66. Fritz, J. B., Elhilali, M. & Shamma, S. A. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* **98**, 2337–46 (2007).
67. Reynolds, J. H. & Heeger, D. J. The Normalization Model of Attention. *Neuron* **61**, 168–185 (2009).
68. McGinley, M. J., David, S. V. & McCormick, D. A. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* **87**, 179–192 (2015).
69. Reimer, J. *et al.* Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* **7**, 1–7 (2016).
70. Natan, R. G. *et al.* Complementary control of sensory adaptation by two types of cortical interneurons. *Elife* **4**, (2015).
71. Natan, R. G., Rao, W. & Geffen, M. N. Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation. *Cell Rep.* **21**, 878–890 (2017).
72. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-Expressing Interneurons Linearly Transform Cortical Responses to Visual Stimuli. *Neuron* **73**, 159–170 (2012).
73. Wilson, N. R., Runyan, C. A., Wang, F. L. & Sur, M. Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature* **488**, 343–348 (2012).
74. Seybold, B. a, Phillips, E. a K., Schreiner, C. E. & Hasenstaub, A. R. Inhibitory Actions Unified by Network Integration. *Neuron* **87**, 1181–1192 (2015).
75. Phillips, E. A. K. & Hasenstaub, A. R. Asymmetric effects of activating and inactivating cortical interneurons. *Elife* **5**, e18383 (2016).
76. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
77. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annual Review of Neuroscience* **24**, 1193–1216 (2001).
78. Simoncelli, E. P. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology* **13**, 144–149 (2003).
79. Schneider, D. M., Nelson, A. & Mooney, R. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* **513**, 189–94 (2014).
80. Schneider, D. M., Sundararajan, J. & Mooney, richard. A cortical filter that learns to suppress the acoustic consequences of movement. *Nature* (2018). doi:10.1038/s41586-018-0520-5
81. Guo, Z. V. *et al.* Procedures for behavioral experiments in head-fixed mice. *PLoS One* **9**, (2014).
82. Isett, B. R., Feasel, S. H., Lane, M. A. & Feldman, D. E. Slip-Based Coding of Local Shape and Texture in Mouse S1. *Neuron* **97**, 418-433.e5 (2018).
83. Carruthers, I. M., Natan, R. G. & Geffen, M. N. Encoding of ultrasonic vocalizations in the auditory cortex. *J Neurophysiol* **109**, 1912–1927 (2013).
84. Carruthers, I. M. *et al.* Emergence of invariant representation of vocalizations in the auditory cortex. *J. Neurophysiol.* jn.00095.2015 (2015). doi:10.1152/jn.00095.2015
85. Voigts, J. *et al.* An easy-to-assemble, robust, and lightweight drive implant for chronic tetrode recordings in freely moving animals. *J. Neural Eng.* **17**, 26044 (2020).
86. Voigts, J., Siegle, J., Pritchett, D. L. & Moore, C. I. The flexDrive: An ultra-light implant for optical control

and highly parallel chronic recording of neuronal ensembles in freely moving mice. *Front. Syst. Neurosci.* **7**, 8 (2013).
87. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv. Neural Inf. Process. Syst.* 4455–4463 (2016).
88. Eilers, P. H. C. & Marx, B. D. Flexible smoothing with B-splines and penalties. *Stat. Sci.* **11**, 89–102 (1996).
89. Stanislaw, H. & Todorov, N. Calculation of signal detection theory measures. *Behav. Res. Methods, Instruments, Comput.* **31**, 137–149 (1999).
90. Rocchi, F. & Ramachandran, R. Neuronal adaptation to sound statistics in the inferior colliculus of behaving macaques does not reduce the effectiveness of the masking noise. *J. Neurophysiol.* **120**, 2819–2833 (2018).
91. Hautus, M. J. Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behav. Res. Methods, Instruments, Comput.* **27**, 46–51 (1995).
92. Sahani, M. & Linden, J. F. How linear are auditory cortical responses? *Adv. Neural Inf. Process. Syst.* 109–116 (2003). doi:10.1124/dmd.105.005157.concerning
93. Sahani, M. & Linden, J. F. Evidence optimization techniques for estimating stimulus-response functions. in *Advances in Neural Information Processing Systems* (2003).
94. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

Supplementary Information

Supplemental Experimental Procedures

Acute electrophysiological recordings with muscimol or saline.

Neural signals were recorded from n = 2 awake, untrained mice. Prior to the recording session, each mouse was anesthetized and a headpost and ground pin were implanted on the skull (see *Surgery* in the main text). On the day of the recording, the mouse was briefly anesthetized with 3% isoflurane and a small craniotomy was performed over auditory cortex using a dental drill or scalpel (~1mm x 1mm craniotomy centered approximately 1.25mm anterior to the lambdoid suture along caudal end of the squamosal suture). A 32-channel silicon probe (Neuronexus) was then positioned perpendicularly to the cortical surface and lowered at a rate of 1-2μm/s to a final depth of 800-1200μm. As the probe was lowered, trains of brief noise bursts were repeated, and if stimulus locked responses to the noise bursts were observed, the probe was determined to be in auditory cortex. The probe was then allowed to settle for up to 30 minutes before starting the recording.

For the muscimol and saline recordings (Extended Data Figure 3), a durotomy was performed over the injection site and baseline neural responses to the behavioral stimuli were recorded. Then, 2.5μL of .25mg/mL muscimol or 0.9% sterile saline solution was topically applied to the surface of auditory cortex and allowed 30 minutes to penetrate the tissue. The same stimuli were then recorded again after the elapsed time. In these recordings, the same targets and DRC background presented during behavior were presented. Neural signals from n = 2 mice (1 mouse for muscimol application, 1 mouse for saline application) were amplified and digitized using a Cheetah Digital LYNX system (Neuralynx) at a rate of 32kHz.

Acute electrophysiological recordings for Extended Data Figure 5b-g

Neural signals were recorded from n = 9 awake, untrained mice of several -cre strains (somatostatin-cre, n = 5; parvalbumin-cre, n = 2; VGAT-cre, n = 2). These mice were implanted with a headplate and groundpin, as described in *Surgery*. Additionally, each mouse was bilaterally injected with 700 μL of Flex-ChR2 during the initial surgery in auditory cortex, then bilaterally implanted with opto-cannulae which projected 500 um below the brain surface above auditory cortex. During the recordings, mice were presented with dynamic random chord stimuli (DRC) which changed contrast every 3 s. At each time step, the chords were randomly drawn from a uniform distribution with a center of 50 dB SPL and a spread of either 7.5 dB SPL or 15 dB SPL in low and high contrast respectively. Each chord was presented for 4 ms with a 1 ms linear ramp between each chord. Chords were composed of 25 frequencies between 1 and 64 kHz, spaced 0.25 octaves apart. On a subset of trials, 470 nm LED or laser light was continuously shone or pulsed at 25 Hz through the opto-cannulae for the duration of the 3 s of contrast period (power measured at the fiber tip ~2-5 mW). For the purposes of this study, we discarded all trials with light presentation.

Supplemental Results

Muscimol application disrupts cortical encoding of targets.

In n = 2 awake, naïve mice, we first recorded baseline responses to the stimuli used in the psychometric task, then topically applied muscimol or saline, waited 30 minutes, and recorded stimulus responses again. After muscimol application, there was a marked decrease in neural responses to targets compared to the baseline recordings (Extended Data Figure 1b, left). Notably, in our saline control, we observed little to no change in neural responses after saline application (Extended Data Figure 4b, right). We next compared how contrast, volume and muscimol or saline application changed the responses during the pre- and post-application periods, finding that muscimol significantly reduced the firing rates between pre- and post-application periods, while saline significantly increased firing rates (Extended Data Figure 4c,d, Extended Data Table 1). We speculate that the small increase in firing rate between pre- and post-saline application was due to changes in recording quality or due to neural drift over the ~1 hour recording session, and note that the effect size of saline pre-post application is very small ($\eta^2 = 0.0046$) compared to the effect size of muscimol ($\eta^2 = 0.38$). We then used a three-way ANOVA to compare the effects of muscimol, contrast, and target volume on target responses in the saline and muscimol recording sessions. We found a significant main effect of muscimol ($F(1) = 322.65$, $p = 4.88e\text{-}67$) and volume ($F(6) = 15.48$, $p = 1.98e\text{-}17$), but no main effect of contrast ($F(1) = 0.39$, $p = 0.53$), indicating nearly complete suppression of responses to both targets and background in high and low contrast (Extended Data Figure 4e,f). These results confirmed that muscimol effectively disrupts the cortical coding of our behavioral stimuli.

Muscimol application does not prevent licking.

An additional alternative effect of muscimol is a general loss of the ability to lick. To assess this, we monitored the lick probability of the mice throughout the trial duration, and found that muscimol specifically reduced licking responses during the period where targets were presented (Wilcoxon rank-sum test: T = 337, z = -4.23, p = 2.34e-5; Extended Data Figure 4g, right panel of Extended Data Figure 4h). Mice also tended to lick immediately after the trial onset (Extended Data Figure 4i, green trace), but we found that the lick rates under muscimol and saline conditions were identical during this period (Wilcoxon rank-sum test: T = 528, z = 0.23, p = 0.81; Extended Data Figure 4h, left panel). These results suggest that muscimol does not impair the mouse's ability to lick in general, but results in a specific deficit in licking in response to targets.

STRF are stable across contrasts.

Based on a pilot study of neuronal data acutely recorded from auditory cortex, we tested whether STRF properties were affected by stimulus contrast. We recorded spiking activity in response to DRCs that changed contrast every 3 seconds. Out of the 700 units identified from n = 9 mice, we selected the subset of neurons with noise ratios (NR) below 100 for further analysis (n = 129). For each neuron, we computed the spectrotemporal receptive field (STRF) using a spike triggered average in each contrast (Extended Data Figure 5b), then computed 100 "random" STRFs by shuffling the stimulus in time within each contrast. For each shuffle, we computed the correlation of the true low contrast STRF with the shuffled high contrast STRFs to generate a null distribution of low-high contrast STRF correlations. We then compared the true correlations of the low and high contrast STRF with this null distribution, defining them as significantly correlated if the true correlation fell outside the 99th percentile of the null distribution. We found that nearly all of the low and high contrast STRFs were significantly correlated (124/129 neurons, 96%), suggesting that contrast doesn't change the overall structure of the STRF (Extended Data Figure 5d).

To further quantify these results, we tested whether more concrete STRF properties such as best frequency (BF), lag, and max value were affected by contrast. First, we de-noised each STRF by determining the significance of each pixel. To do this, we compared the value of each pixel to the distribution of shuffled values for that pixel, and retained only pixels greater than three standard deviations of the shuffled value. Based on the de-noised STRFs, we computed frequency and temporal components by averaging over each STRF dimension (Extended Data Figure 5c). We then estimated the BF and lag as the max of these components, and determined the max STRF value by finding the max value over all pixels. Next, we compared each measure across STRFs from low and high contrast. We found that the maximum pixel value was significantly greater in high contrast (Median (*Mdn*) = 1.33, inter-quartile range (*IQR*) = 1.28) than in low contrast (*Mdn* = 0.56, *IQR* = 0.62; Wilcoxon signed-rank test: z = -9.78, *rank* = 0, p = 1.39e-22; Extended Data Figure 5e, e). On the other hand, we found a non-significant trend towards lower BFs in low contrast (*Mdn* = 19.03 kHz, *IQR* = 35.74 kHz) compared to high contrast (*Mdn* = 22.63 kHz, *IQR* = 47.09 kHz; Wilcoxon signed-rank test: z = 1.78, *rank* = 1761, p = 0.076; Extended Data Figure 5f), and no significant change in lag (Wilcoxon signed-rank test: z = -0.93, *rank* = 1776, p = 0.35; Extended Data Figure 5g). Taken together, these results demonstrate that the frequency and temporal modulation of sound responses are consistent across contrasts, supporting previously published findings.

Generalized linear model of contrast gain control dynamics

A primary goal of the current study was to estimate the influence of stimulus contrast on neural gain dynamics, for instance, after a switch from one contrast to another. To approach this problem, we first define a model neuron with dynamic gain control.

Forward model

To best approximate the stimuli used in our experiments, we define the stimulus environment of our model as an F-dimensional signal that evolves in discrete time steps:

$$X_{t,f} \sim \mathcal{N}(\mu, \sigma_t),$$

where $X_{t,f}$ is a stimulus spectrogram that varies as a function of time t and frequency f. Each time and frequency bin of X is sampled from a normal distribution defined by an average value μ and contrast σ_t at time t.

To approximate the behavior of real neurons, we define a model neuron that has a two-dimensional linear filter (representing the STRF of the neuron):

$$\beta_{h,f} = \mathcal{N}(m,\, C;\; h, f),$$

where stimulus filter $\beta_{h,f}$ is defined as a two-dimensional gaussian distribution evaluated at lag h and frequency f. The filter location in frequency-history space is defined by its mean m and covariance matrix C. The stimulus drive of the neuron at each time step, x_t, is then computed as the convolution of the stimulus matrix and the linear filter:

$$x_t = \mathrm{X_t}\beta \tag{13}$$

where X_t at each time t is a row vector of length $F \cdot H$ (ie. the unrolled stimulus spectrogram lagged by H lags) and β is the filter, unrolled as a column vector of the same length.

The model neuron has a firing rate that depends only on the stimulus drive x_t and the contrast σ_t at time t. We then assume that the number of spikes y_t emitted by the neuron at each time step follow a Poisson distribution:

$$y_t \sim Poisson(\lambda_t)$$

where λ_t is the firing rate at time t, given by

$$\lambda_t = exp[a + g(\sigma_t)b(x_t - c)] \tag{14}$$

where g is a gain control function, and a, b, and c are parameters of the model. The parameter a represents the baseline response of the neuron, b is a scaling factor of the stimulus drive, and c represents the operating point of the gain. We remove the obvious degeneracy in the definition of *g* and *b* (only their product matters) by requiring that *g* be adimensional and such that

$$\frac{1}{2}[\mathrm{g}(\sigma_\mathrm{H}) + g(\sigma_L)] \;=\; 1 \tag{15}$$

where σ_H and σ_L are the high and low contrast values. This constraint forces the neutral value of the gain, $g \;=\; 1$ to be the midpoint between gain in the high and low contrast conditions.

Optimal gain control

In the spirit of the efficient coding principle, we derived a form for $g(\sigma)$ that will guarantee that, under certain conditions, the dynamic range of the neuron will be approximately conserved under changes in contrast. To do this, we define the dynamic range as

$$R(\sigma) = \lambda(\mu + \sigma) - \lambda(\mu - \sigma) \tag{16}$$

which can be rewritten using equation 2 as

$$R(\sigma) = e^a\big[exp\big(g(\sigma)b(\mu + \sigma - c)\big) - exp\big(g(\sigma)b(\mu - \sigma - c)\big)\big]. \tag{17}$$

If the argument of the exponentials is not too large, we can linearize this expression to obtain

$$R(\sigma) \simeq 2e^a b\sigma \;\cdot g(\sigma) \tag{18}$$

and that R is approximately independent of σ provided that $g(\sigma) \propto 1/\sigma$. So, for our model, we set

$$g(\sigma) = \frac{\overline{\sigma}}{\sigma_t} \tag{19}$$

where $\overline{\sigma}$ is the harmonic mean of σ_H and σ_L:

$$\bar{\sigma} := \left[\frac{1}{2}\left(\frac{1}{\sigma_H} + \frac{1}{\sigma_L}\right)\right]^{-1} = 2\,\frac{\sigma_H \sigma_L}{\sigma_H + \sigma_L} \tag{20}$$

Finally, to validate that our fitting methods are sensitive to real world neurons, which do not necessarily adjust their gain to account for changes in contrast according to the model just described, we consider an interpolation scheme that smoothly transforms a model with positive gain control to a similar model without gain control, or with "anti" gain control. To do this, we redefine g as follows:

$$g(\sigma) \rightarrow \xi g(\sigma) + (1 - \xi), \qquad -1 \leq 0 \leq 1 \tag{21}$$

so that by changing ξ we can control whether gain control is optimal ($\xi = 1$), non-existant ($\xi = 0$), or "anti" ($\xi = -1$).

Putting everything together, the final expression for the firing rate of the forward model is

$$\lambda_t = exp\left[a + b\left(\xi\frac{\bar{\sigma}}{\sigma_t} + (1 - \xi)\right)(x_t - c)\right] \tag{10}$$

Generalized linear model

The forward model developed in the previous section provides a simple approximation of the relationship between the stimulus, stimulus contrast and neuronal responses. We also note that the form of the forward model lends itself to estimation using a Poisson GLM, provided that the predictors are chosen appropriately. As such, we define the inference model as a Poisson GLM with an intercept term and the following predictors:

$$(x_t - \mu), \qquad \frac{\bar{\sigma}}{\sigma_t}, \qquad (x_t - \mu)\frac{\bar{\sigma}}{\sigma_t}$$

In other words, the model is composed of a stimulus predictor $(x_t - \mu)$, a contrast predictor $(\bar{\sigma}/\sigma_t)$, and their interaction. Therefore, the GLM models the data at time t as a Poisson distribution with the following mean:

$$\lambda_t = exp\left[\beta_0 + \beta_1(x_t - \mu) + \beta_2\frac{\bar{\sigma}}{\sigma_t}(x_t - \mu) + \beta_3\frac{\bar{\sigma}}{\sigma_t}\right] \tag{22}$$

where $\beta_0 \ldots \beta_3$ are the parameters to be inferred, and, as defined previously, x_t is the stimulus drive of the neuron determined by its STRF.

Model fitting

To fit the model, we took a two-step approach. First we found the best-fit filter (STRF) for the neuron. Then, we fit the full GLM to determine how the linear drive determined by the STRF is modulated by contrast. In the first step, the linear drive is obtained by fitting the model

$$ln\,\lambda_t = \alpha + X_t\beta \tag{23}$$

where X_t is a design matrix defined as a function of frequency bins f and history lags h, and β is the fitted STRF. Stimulus drive x_t is then computed as in equation 1.

We then define the full model according to equation 11,

$$ln\,\lambda_t = \beta_0 + \beta_1 x_t + \sum_{i=1}^{B}\beta_{2i}\,x_t \cdot (b_i * c)(t) + \sum_{i=1}^{B}\beta_{3i}(b_i * c)(t) \tag{24}$$

where $c(t) = \bar{\sigma}/\sigma_t$ and $\{b_i\}_{i=1}^{B}$ is a set of cubic B-spline temporal basis functions. By defining a matrix C as follows

$$C_{ti} = (b_i * c)(t) \tag{25}$$

we can rewrite equation 13 in a more compact form:

$$\ln \lambda = \beta_0 + x\beta_1 + x \circ C\beta_2 + C\beta_3 \tag{26}$$

where $\circ$ denotes element-by-element "broadcasting" multiplication.

To fit asymmetric changes in firing rate after transitions to low or high contrast, we took the simple approach of defining separate sets of contrast predictors for each transition type. This amounted to modifying C by masking transitions to high contrast or transitions to low contrast with zeros, such that the model fit a window H' of 40 time bins around each contrast transition. To do so, we created a new matrix C' by duplicating C column-wise. Then, we define the first B columns as predictors for the transition to low contrast by masking a 1 second period around each transition to high contrast with zeros. This same procedure was repeated for the remaining columns in C', instead masking out the transition to low contrast. Substituting this into equation 15, we obtain

$$ln\,\lambda = \beta_0 + x\beta_1 + x \circ C'^{\beta_2} + C'^{\beta_3} \tag{27}$$

For the sake of clarity, note that in the expression above, β_0 is a number, x is a column vector of length T, β_1 is a number, C' is a T-by-$2B$ matrix, and β_2 and β_3 are column vectors of length $2B$.

Defining gain

We have outlined a forward model for simulating neural activity according to efficient coding of stimulus contrast, and described an inference model (a Poisson GLM) for estimating the influence of the stimulus, stimulus contrast, and their interaction. In this section, we describe how to use the fitted parameters to quantify the amount of gain control in the neuron.

Conceptually, an increase or decrease in the gain of a system is analogous to more or less sensitivity to small changes in the stimulus, dependent what is modulating the gain (in our case, the recent history of the contrast). Based on this intuition, we focus on how the response of the neuron (as modeled by a fitted GLM) is expected to change between conditions where the gain is expected to contribute (i.e. in the presence of gain control) and where it is not (ie. in the absence of gain control, where gain is "neutral").

To do this, we start by considering the gradient of the link function (the log rate) at time t with respect to X_t:

$$\begin{aligned}\eta_t = \nabla_{x_t} ln\lambda_t &= \nabla_{x_t}[\beta_0 + x_t\beta_1 + x_t C_t\beta_2 + C_t\beta_3] \\ &= \nabla_{x_t}[\beta_0 + (X_t\beta)\beta_1 + (X_t\beta)C_t\beta_2 + C_t\beta_3] \\ &= \beta(\beta_1 + C_t\beta_2)\end{aligned} \tag{28}$$

We can immediately read equation 17 as "the STRF of the model is modulated by a factor of $\beta_1 + C_t\beta_2$ at time t", and define the gain based on this intuition, but we will take a slightly longer and more formal route to get to the same result.

The gradient η is a vector with the same dimensionality of β_1 and $C_t\beta_2$, and it encapsulates all information about the sensitivity of the link function to small changes in X_t at a given time. Because X_t is not a scalar (it has $H \cdot F$ components), these changes can happen along many dimensions, and the sensitivity can be different in different directions. We can define the gain based on the sensitivity to changes in a specific direction v (assuming for concreteness that $||v|| = 1$, although this is not necessary for the derivation below). If $X_t = r \cdot v$, where r is some scalar, then

$$\frac{dln\lambda_t}{dr} = \langle \eta_t, v \rangle \tag{29}$$

by definition of the gradient. We can then define the gain w along direction v as the ratio between the sensitivity of the log rate to changes along v and the sensitivity one would have if the contrast C_t was at some reference value C^0 where we define $w = 1$ by construction. If we do so, we obtain

$$\begin{aligned}w_t &= \left(\frac{dln\lambda_t}{dr}\right)\left(\frac{dln\lambda(C = C^0)}{dr}\right)^{-1} \\ &= \frac{\langle \beta, v \rangle(\beta_1 + C_t\beta_2)}{\langle \beta, v \rangle(\beta_1 + C^0\beta_2)} \\ &= \frac{\beta_1 + C_t\beta_2}{\beta_1 + C^0\beta_2}\end{aligned} \tag{30}$$

Note that this definition does not depend on the initial choice of v, or even on the specifics of the choice of basis functions used to define C. In conclusion, by reasoning about the sensitivity of the response of the fitted GLM, we define a value w_t which captures the relationship between the true gain g and the stimulus contrast c_t.

Simulations

To validate our inference model, we simulated neural activity according to the generative model defined in the *Forward Model* section (Extended Data Figure 2a). We were interested in capturing several dimensions upon which the generative model could vary, namely, the amount of gain control in the simulated neurons ξ, and the dynamics of the gain function g.

To parametrically control the evolution of gain over time, we simulated different temporal trajectories of gain control, by modifying $g(\sigma_t)$ as follows

$$g(\sigma_c, \tau_c)_t = \mathrm{g}(\sigma_{c-1}) + (\mathrm{g}(\sigma_c) - \mathrm{g}(\sigma_{c-1})) \cdot exp(\tau_c, t) \tag{31}$$

where the gain g after a switch to contrast σ_c transitions from the gain in the previous contrast $g(\sigma_{c-1})$ to the gain in the current contrast $g(\sigma_c)$ according to an exponential function with time constant τ_c. Note that τ_c could vary between the two contrasts to simulate asymmetric dynamics.

For each neuron, we first generated a STRF and linear drive according to equation 1 (Extended Data Figure 2b,d). For different sets of simulated neurons, we parametrically varied the amount of gain control ξ between -1 and 1, and varied the gain time courses to simulate three types of gain adaptation dynamics: 1) Slow transitions to low contrast with fast transitions to high contrast, 2) Fast, symmetric transitions to each contrast, 3) Fast transitions to low contrast and slow transitions to high contrast (Extended Data Figure 2f).

We simulated 100 neurons for each combination of ξ and τ, with other simulation parameters held constant (Extended Data Table 4). Extended Data Figure 2e plots the average firing rates and overlaid model fits for three sets of simulations with optimal gain control ($\xi = 1$) while varying τ. Importantly, the model flexibly captured the gain dynamics in the three simulated adaptation conditions, with the gain estimate w_t following the true gain trajectory (Extended Data Figure 2f). For additional values of ξ, the model accurately predicted the firing rates (Extended Data Figure 2g) and gain trajectories (Extended Data Figure 2h). We observed that some combinations of ξ and τ elicited large firing rate transients, particularly in the cases where simulated gain slowly adapted after a switch to high contrast (bottom panels in Extended Data Figure 2e, f, g, h). This behavior is expected, as gain remains relatively high for a longer period after the switch, causing large fluctuations in firing rate as the stimulus drive during high contrast is increased. These large firing rate transients seemed to reduce the accuracy of gain estimate w, but we observed that the predicted time courses still captured the overall asymmetries present in the underlying model.

During our behavioral recordings, we used a limited number of background noise scenes (n = 5) to reduce the overall size of the stimulus set. However, it became clear that our model required a larger sample of stimulus space to accurately estimate gain. To demonstrate this, we plotted the simulation results when neurons were exposed to 100 unique background scenes (Extended Data Figure 2i) compared to simulations where neurons were only exposed to 5 unique background scenes, as in our behavioral recordings (Extended Data Figure 2j). We observed that with 100 scenes, estimates of w were very close to the true gain values, but were consistently underestimated in the case of 5 background scenes, even in the case of perfect gain control. Therefore, when analyzing our behavioral recordings, we used a standard linear-nonlinear model to estimate neural gain (Figure 5), as we previously found that gain estimates from the GLM were highly correlated with gain estimated from the LN model (Figure 2i).

Extended Data Table 1: Statistical Comparisons.

Comparison	Figure	Center	Spread	N	Test	Statistic	Effect Size	p-value
Behavior percent correct, low contrast: time 1 vs. time 2	2g	T1: 0.68 T2: 0.70 (median)	T1: 0.10 T2: 0.15 (IQR)	21 mice	Two-tailed Wilcoxon sign-rank test (FDR corrected[94] for multiple comparisons)	Z = -1.93 Rank: 60	$Z/\sqrt{n}$ = -0.42	0.054
Behavior percent correct, low contrast: time 1 vs. time 3		T1: 0.68 T3: 0.82 (median)	T1: 0.10 T3: 0.092 (IQR)			Z = -4.01 Rank: 0	$Z/\sqrt{n}$ = -0.88	5.96e-5
Behavior percent correct, low contrast: time 1 vs. time 4		T1: 0.68 T4: 0.87 (median)	T1: 0.10 T4: 0.190 (IQR)			Z = -4.01 Rank: 0	$Z/\sqrt{n}$ = -0.88	5.96e-5
Behavior percent correct, low contrast: time 1 vs. time 5		T1: 0.68 T5: 0.91 (median)	T1: 0.10 T5: 0.11 (IQR)			Z = -4.01 Rank: 0	$Z/\sqrt{n}$ = -0.88	5.96e-5
Behavior percent correct, high contrast: time 1 vs. time 2		T1: 0.82 T2: 0.77 (median)	T1: 0.083 T2: 0.19 (IQR)			Z = 2.84 Rank: 181	$Z/\sqrt{n}$ = 0.62	0.0045
Behavior percent correct, high contrast: time 1 vs. time 3		T1: 0.82 T3: 0.77 (median)	T1: 0.083 T3: 0.14 (IQR)			Z = 2.17 Rank: 163	$Z/\sqrt{n}$ = 0.47	0.030
Behavior percent correct, high contrast: time 1 vs. time 4		T1: 0.82 T4: 0.78 (median)	T1: 0.083 T4: 0.16 (IQR)			Z = 3.36 Rank: 195	$Z/\sqrt{n}$ = 0.73	7.80e-4
Behavior percent correct, high contrast: time 1 vs. time 5		T1: 0.82 T5: 0.79 (median)	T1: 0.083 T5: 0.12 (IQR)			Z = 1.94 Rank: 157	$Z/\sqrt{n}$ = 0.42	0.052
ANOVA for effects of pre-post muscimol application, contrast, and volume on firing rate in ACtx	S4c	n/a	n/a	42 neurons	three-way ANOVA	$F_{pre\text{-}post}(1)$ = 812.54 $F_{contrast}(1)$ = 22.64 $F_{volume}(6)$ = 21.70	η^2 = 0.38 η^2 = 0.011 η^2 = 0.061	4.48e-136 2.19e06 2.77e-24
ANOVA for effects of pre-post saline application, contrast, and volume on firing rate in ACtx	S4d	n/a	n/a	104 neurons	three-way ANOVA	$F_{pre\text{-}post}(1)$ = 15.40 $F_{contrast}(1)$ = 0.43 $F_{volume}(6)$ = 76.067	η^2 = 0.0046 η^2 = 1.29e-4 η^2 = 0.14	8.89-5 0.51 1.76e-88
Percent correct max dB SNR, low contrast: muscimol vs. saline	3c	Musc.: 0.10 Saline: 0.85 (median)	Musc.: 0.67 Saline: 0.27 (IQR)	10 musc. sessions, 10 saline sessions (4 mice)	Two-tailed Wilcoxon rank-sum test	Z = -2.76 Rank: 68	$Z/\sqrt{n}$ = -0.62	0.0058
Threshold (dB SNR), low contrast: muscimol vs. saline		Musc.: 14.78 Saline: 9.66 (median)	Musc.: 18.46 Saline: 6.88 (IQR)			Z = 0.72 Rank: 115	$Z/\sqrt{n}$ = 0.16	0.47
FA rate, low contrast: muscimol vs. saline		Musc.: 0.026 Saline: 0.132 (median)	Musc.: 0.10 Saline: 0.85 (IQR)			Z = -2.91 Rank: 66	$Z/\sqrt{n}$ = -0.65	0.0036
Max slope (PC/dB), low contrast: muscimol vs. saline		Musc.: 0.026 Saline: 0.072 (median)	Musc.: 0.056 Saline: 0.030 (IQR)			Z: -2.68 Rank: 69	$Z/\sqrt{n}$ = -0.60	0.0073
Percent correct max dB SNR, high contrast: muscimol vs. saline		Musc.: 0.06 Saline: 0.80 (median)	Musc.: 0.10 Saline: 0.85 (IQR)	13 musc. sessions, 10 saline sessions (4 mice)		Z = -4.06 Rank: 92	$Z/\sqrt{n}$ = -0.83	4.96e-5
Threshold (dB SNR), high contrast: muscimol vs. saline		Musc.: 16.77 Saline: 18.80 (median)	Musc.: 21.33 Saline: 5.89 (IQR)			Z = -0.35 Rank: 156	$Z/\sqrt{n}$ = -0.071	0.73
FA rate, low contrast: muscimol vs. saline		Musc.: 0.027 Saline: 0.213 (median)	Musc.: 0.10 Saline: 0.85 (IQR)			Z = -3.19 Rank: 107	$Z/\sqrt{n}$ = -0.65	0.0014
Max slope (PC/dB), high contrast: muscimol vs. saline		Musc.: 0.012 Saline: 0.058	Musc.: 0.024 Saline: 0.018			Z = -3.77 Rank: 97	$Z/\sqrt{n}$ = -0.77	1.66e-4

		(median)	(IQR)					
Percent correct max dB SNR, target in high contrast : muscimol vs. saline	3f	Musc.: 0.07 Saline: 0.82 (median)	Musc.: 0.51 Saline: 0.095 (IQR)	5 musc. sessions, 5 saline sessions (2 mice)	Two-tailed Wilcoxon rank-sum test	Z = nan Rank: 15	$Z/\sqrt{n}$ = nan	0.0079
Percent correct at threshold, target in high contrast: muscimol vs. saline		Musc.: 0.03 Saline: 0.53 (median)	Musc.: 0.35 Saline: 0.11 (IQR)			Z = nan Rank: 17	$Z/\sqrt{n}$ = nan	0.032
FA rate, target in high contrast : muscimol vs. saline		Musc.: 0.12 Saline: 0.23 (median)	Musc.: 0.22 Saline: 0.11 (IQR)			Z = nan Rank: 21	$Z/\sqrt{n}$ = nan	0.22
Max slope (PC/dB), target in high contrast : muscimol vs. saline		Musc.: 0.038 Saline: 0.057 (median)	Musc.: 0.046 Saline: 0.012 (IQR)			Z = nan Rank: 19	$Z/\sqrt{n}$ = nan	0.095
Percent correct max dB SNR, target in silence: muscimol vs. saline		Musc.: 0.85 Saline: 0.92 (median)	Musc.: 0.23 Saline: 0.15 (IQR)	8 musc. sessions, 8 saline sessions (2 mice)		Z = nan Rank: 53	$Z/\sqrt{n}$ = nan	0.13
Percent correct at threshold, target in silence : muscimol vs. saline		Musc.: 0.11 Saline: 0.22 (median)	Musc.: 0.28 Saline: 0.22 (IQR)			Z = nan Rank: 55	$Z/\sqrt{n}$ = nan	0.20
FA rate, target in silence : muscimol vs. saline		Musc.: 0.029 Saline: 0.041 (median)	Musc.: 0.038 Saline: 0.11 (IQR)			Z = nan Rank: 60	$Z/\sqrt{n}$ = nan	0.44
Max slope (PC/dB), target in silence : muscimol vs. saline		Musc.: 0.028 Saline: 0.031 (median)	Musc.: 0.015 Saline: 0.0048 (IQR)			Z = nan Rank: 63	$Z/\sqrt{n}$ = nan	0.65
Mixed-effects model: `behavioral_threshold ~ neural_threshold + contrast + (contrast-1|mouse)`	5g	**Model Coefficients** Estimate ± standard error [tstat(df), p-value]		19 mice	Likelihood ratio test against model without neural threshold: `beh_thresh ~ contrast + (contrast-1|mouse)`	$\chi^2(1) = 5.89$		0.015
		Intercept: 0.23±1.21 t(16) = 0.19, p = 0.85 Neural threshold: 0.48±0.18 t(16) = 2.63, p = 0.018 Contrast: 2.83±1.14 t(16)= 2.48, p = 0.025			Likelihood ratio test against model without contrast: `beh_thresh ~ neur_thresh + (contrast-1|mouse)`	$\chi^2(1) = 4.68$		0.030
Mixed-effects model: `behavioral_slope ~ neural_slope + contrast + (contrast-1|mouse)`	5h	Intercept: 0.023±0.0069 t(16) = 3.34, p = 0.0042 Neural slope: 0.58±0.16 t(16) = 3.58, p = 0.0025 Contrast: 0.0096±0.0052 t(16)= 1.85, p = 0.082			Likelihood ratio test against model without neural slope: `beh_thresh ~ contrast + (contrast-1|mouse)`	$\chi^2(1) = 9.78$		0.0018
					Likelihood ratio test against model without contrast: `beh_slope ~ neur_slope + (contrast-1|mouse)`	$\chi^2(1) = 3.10$		0.078

Neural percent correct, low contrast: time 1 vs. time 2	5i	T1: 0.79 T2: 0.83 (median)	T1: 0.15 T2: 0.22 (IQR)	43 sessions	Two-tailed Wilcoxon sign-rank test (FDR corrected[94] for multiple comparisons)	Z = -1.12 Rank: 418	$Z/\sqrt{n}$ = -0.17	0.26
Neural percent correct, low contrast: time 1 vs. time 3		T1: 0.79 T3: 0.85 (median)	T1: 0.15 T3: 0.15 (IQR)			Z = -3.61 Rank: 198	$Z/\sqrt{n}$ = -0.56	0.00031
Neural percent correct, low contrast: time 1 vs. time 4		T1: 0.79 T4: 0.92 (median)	T1: 0.15 T4: 0.20 (IQR)			Z = -4.68 Rank: 103	$Z/\sqrt{n}$ = -0.72	2.89e-6
Neural percent correct, low contrast: time 1 vs. time 5		T1: 0.79 T5: 0.91 (median)	T1: 0.15 T5: 0.16 (IQR)			Z = -5.34 Rank: 31	$Z/\sqrt{n}$ = -0.82	9.44e-8
Neural percent correct, high contrast: time 1 vs. time 2		T1: 0.78 T2: 0.74 (median)	T1: 0.15 T2: 0.12 (IQR)			Z = 2.62 Rank: 690	$Z/\sqrt{n}$ = 0.40	0.0088
Neural percent correct, high contrast: time 1 vs. time 3		T1: 0.78 T3: 0.76 (median)	T1: 0.15 T3: 0.13 (IQR)			Z = 1.45 Rank: 593	$Z/\sqrt{n}$ = 0.22	0.15
Neural percent correct, high contrast: time 1 vs. time 4		T1: 0.78 T4: 0.83 (median)	T1: 0.15 T4: 0.20 (IQR)			Z = -0.24 Rank: 453	$Z/\sqrt{n}$ = -0.037	0.81
Neural percent correct, high contrast: time 1 vs. time 5		T1: 0.78 T5: 0.83 (median)	T1: 0.15 T5: 0.16 (IQR)			Z = -2.00 Rank: 307	$Z/\sqrt{n}$ = -0.31	0.045
Mixed-effects model: `threshold ~ gain_target + contrast + (contrast-1|mouse)`	6g	**Model Coefficients** Estimate ± standard error [tstat(df), p-value]		168 sessions	Likelihood ratio test against model without gain: `threshold ~ contrast + (contrast-1|mouse)`	$\chi^2(1)$ = 5.82		0.016
		Intercept: 10.97±1.33 t(120) = 8.27, p = 2.059e-13 Target gain: -30.46±12.45 t(120) = -2.45, p = 0.016 Contrast: 3.27±1.55 t(120)= 2.10, p = 0.038			Likelihood ratio test against model without contrast: `threshold ~ gain_target + (contrast-1|mouse)`	$\chi^2(1)$ = 3.71		0.054
Mixed-effects model: `slope ~ gain_target + contrast + (contrast-1|mouse)`	6h	Intercept: 0.039±0.0064 t(120) = 6.23, p = 7.14e-9 Target gain: 0.16±0.060 t(120) = 2.67, p = 0.0085 Contrast: 0.0094±0.062 t(120)= 1.52, p = 0.13		168 sessions	Likelihood ratio test against model without gain: `slope ~ contrast + (contrast-1|mouse)`	$\chi^2(1)$ = 6.96		0.0083
					Likelihood ratio test against model without contrast: `slope ~ gain_target + (contrast-1|mouse)`	$\chi^2(1)$ = 2.28		0.13
Mixed-effects model: `thresh ~ gain_adapt + contrast + (contrast-1|mouse)`	S5m	Intercept: 5.33±1.64 t(120) = 3.26, p = 0.0015 Adaptation gain: 56.66±35.62 t(120) = 1.59, p = 0.11 Contrast: 2.77±1.92 t(120)= 1.44, p = 0.15		168 sessions	Likelihood ratio test against model without gain: `thresh ~ contrast + (contrast-1|mouse)`	$\chi^2(1)$ = 2.51		0.11
					Likelihood ratio test against model without contrast: `thresh ~ gain_adapt + (contrast-1|mouse)`	$\chi^2(1)$ = 2.020		0.16

Mixed-effects model: `slope ~ gain_adapt + contrast + (contrast-1|mouse)`	S5n	Intercept: 0.062±0.0078 t(120) = 7.98, p = 9.63e-13 Adaptation gain: -0.14±0.17 t(120) = -0.80, p = 0.43 Contrast: 0.0049±0.0084 t(120)= 0.58, p = 0.57	168 sessions	Likelihood ratio test against model without gain: `slope ~ contrast + (contrast-1|mouse)`	$\chi^2(1) = 0.64$		0.43
				Likelihood ratio test against model without contrast: `slope ~ gain_adapt + (contrast-1|mouse)`	$\chi^2(1) = 0.33$		0.57

Extended Data Table 2: Mouse strains and genders.

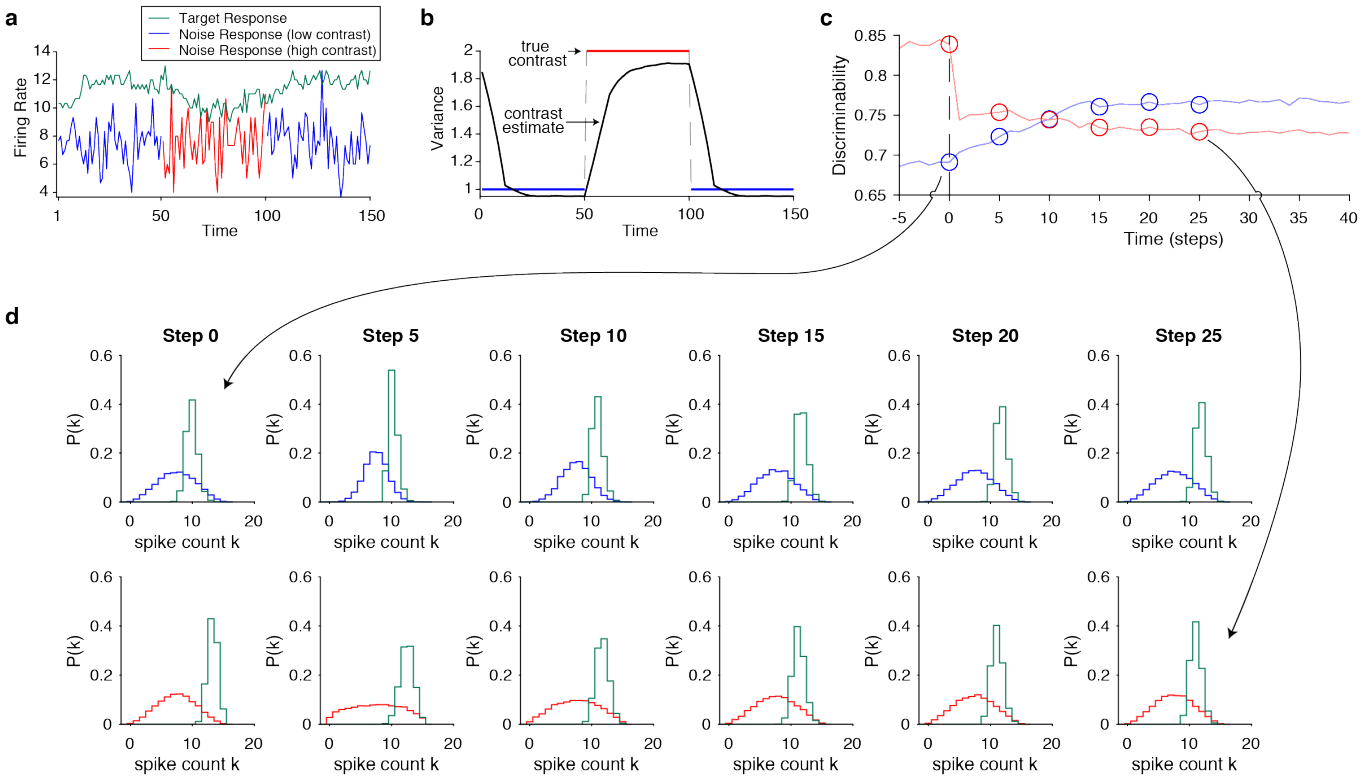
Experiment	Figures	Strain	N [female, male]
Acute ACtx recordings	Figure 2	CDH23	1 [M]
Behavior (no microdrive)	Figure 3	C57BL/6 x CamKII-cre	1 [F], 4 [M]
		C57BL/6 x PV-cre	1 [F]
		CDH23 x SOM-cre	1 [F], 1 [M]
Behavior (microdrive)	Figure 3, Figure 5, Figure 6	CDH23	4 [F], 4 [M]
		C57BL/6 x PV-cre	1 [F]
		C57BL/6 x SOM-cre	1 [F]
		CDH23 x SOM-cre	1 [F], 2 [M]
		CDH23 x CamKII-cre	1 [F]
Muscimol (behavior)	Figure 4	CDH23	2 [F], 2 [M]
Muscimol (acute recording)	Supplemental Figure 4	CDH23 x CamKII-cre	1 [M]
		CDH23	1 [M]
Acute ACtx recordings	Supplemental Figure 5	CDH23 x SOM-cre	3 [F], 2 [M]
		CDH23 x PV-cre	1 [F], 1 [M]
		CDH23 x VGAT	2 [F]
		Total:	19 [F], 19 [M]

Extended Data Table 3: Target SNRs used during psychometric testing.

Target Volumes [range]	[n]: Mouse IDs	n Sessions (total)	n High-Low Contrast Sessions	n Low-High Contrast Sessions
0, 5, 10, 15, 20, 25 dB SNR [25]	[12]: CA102, CA104, CA106, CA107, CA118, CA119, CA121, CA122, CA123, CA124, CA125, CA126	214	111	103
-5, 0, 5, 10, 15, 20 dB SNR [25]	[8]: CA102, CA104, CA106, CA107, CA118, CA119, CA121, CA122	31	31	0
0, 4, 8, 12, 16, 20 dB SNR [20]	[1]: CA046	1	0	1
5, 8, 11, 14, 17, 20 dB SNR [15]	[4]: CA118, CA119, CA121, CA122	68	52	16
8, 10.4, 12.8, 15.2, 17.6, 20 dB SNR [12]	[15]: CA046, CA047, CA048, CA049, CA051, CA052, CA055, CA061, CA070, CA072, CA073, CA074, CA075, CA104, CA107	111	0	111
-4, 0, 4, 8, 12, 16 dB SNR [20]	[11]: CA051, CA052, CA055, CA061, CA070, CA072, CA073, CA074, CA075, CA102, CA106	91	91	0
-5, -1, 3, 7, 11, 15 dB SNR [20]	[5]: CA046, CA047, CA048, CA049, CA051	19	19	0
-75, -60, -45, -30, -15, 0 dB attenuation rel. 25dB SNR	[2]: CA124, CA125	20	n/a	n/a

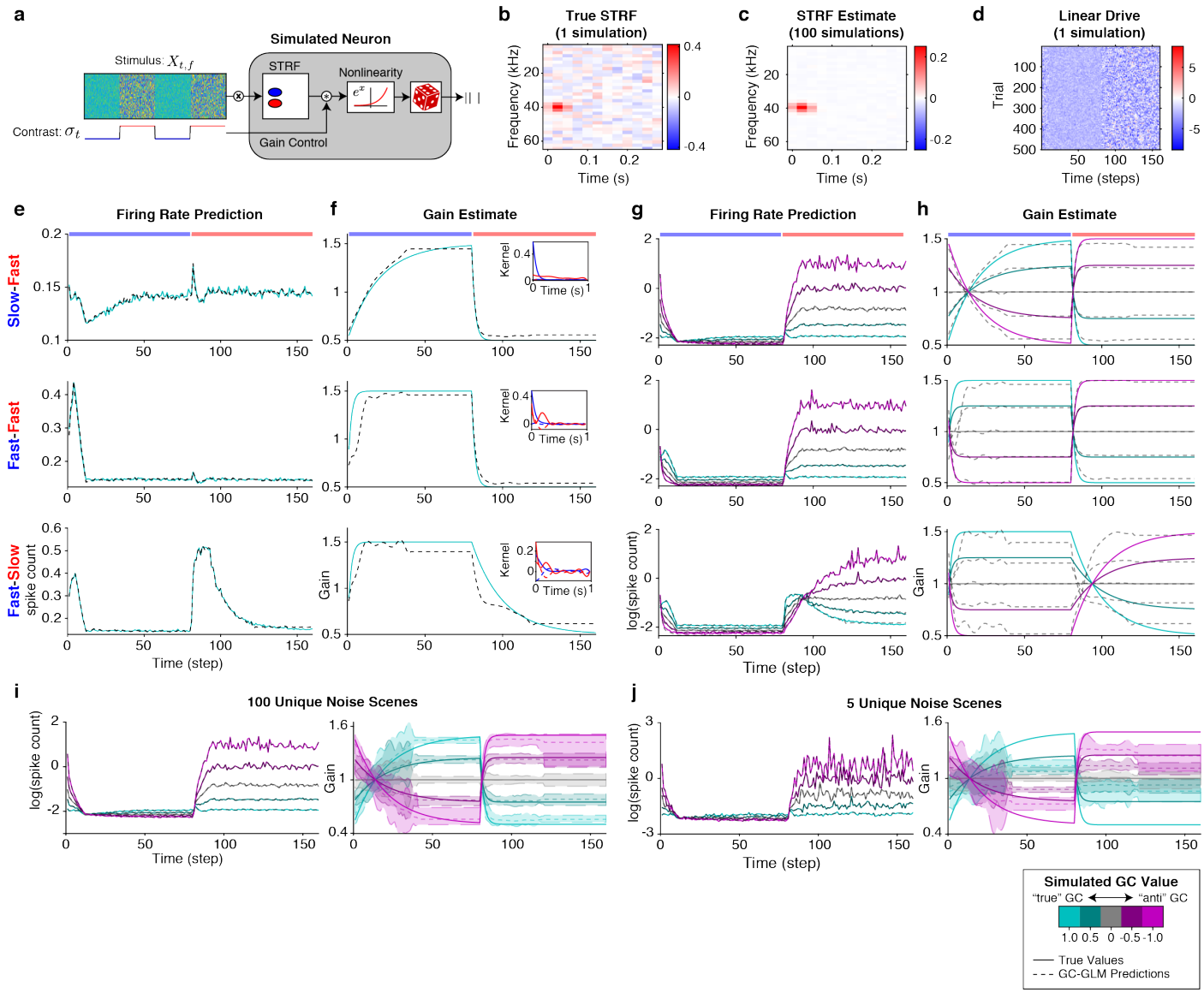
Extended Data Table 4: GLM Simulation Parameters

Parameter	Value
μ	30
σ_L, σ_H	$[1,3]$
β centroid m (frequency bin f, history bin h)	$[20,2]$
β covariance matrix C	$\begin{bmatrix} 0.8 & 0.1 \\ 0.1 & 0.5 \end{bmatrix}$
β dimensions $(F \cdot H)$	$[33, 12]$
Baseline rate a	0.1
Stimulus scaling b	1
Gain operating point c	μ
Gain control ξ	$[-1.0, -0.5, 0, 0.5, 1.0]$
Adaptation time constants $[\tau_L\ \tau_H]$	$\begin{bmatrix} Slow - Fast: 0.05 & 0.5 \\ Fast - Fast: 0.5 & 0.5 \\ Fast - Slow: 0.5 & 0.05 \end{bmatrix}$
Simulated background scenes	$100\ or\ 5$
Contrast history H'	40
B-spline degree, knots	$[3, 7]$



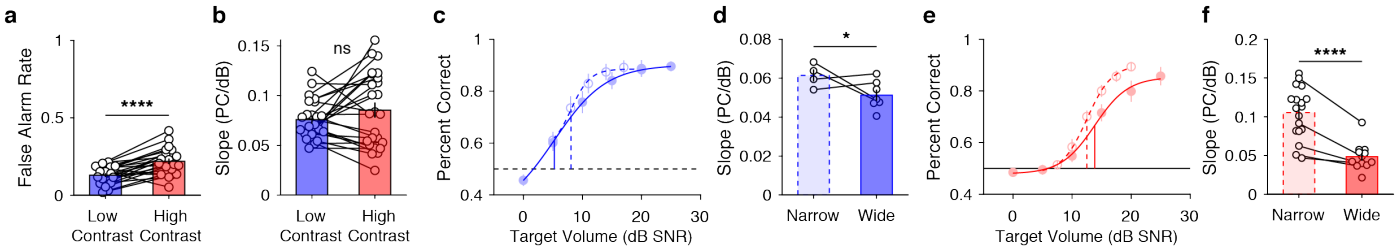
Extended Data Figure 1 (related to Figure 1). Normative model responses, predictions, and example response distributions.

a, The firing rate of the simulated neuron as a function of time. Traces shaded in blue or red indicate the firing rate to periods of low or high contrast background noise, respectively. The green trace indicates the model response to overlaid targets. **b,** The true contrast (labelled as variance) of the stimulus (blue, red, and dashed gray lines) along with the average model estimate of the contrast (solid black line) over time. **c,** Discriminability as a function of time and contrast, with the trace color indicating the contrast after the switch. The dashed vertical line indicates the time of the contrast switch. Open circles indicate time samples used to plot the distributions in **d**. **d,** Target (green) and background (blue or red) distributions as a function of time and contrast. The top row includes responses to targets and background in low contrast. Each column denotes a different time step relative to the change in contrast, as indicated by the column title. The bottom row is the same, but for high contrast. Arrows between **c** and **d** indicate distributions which yielded the indicated value of discriminability in the trace.



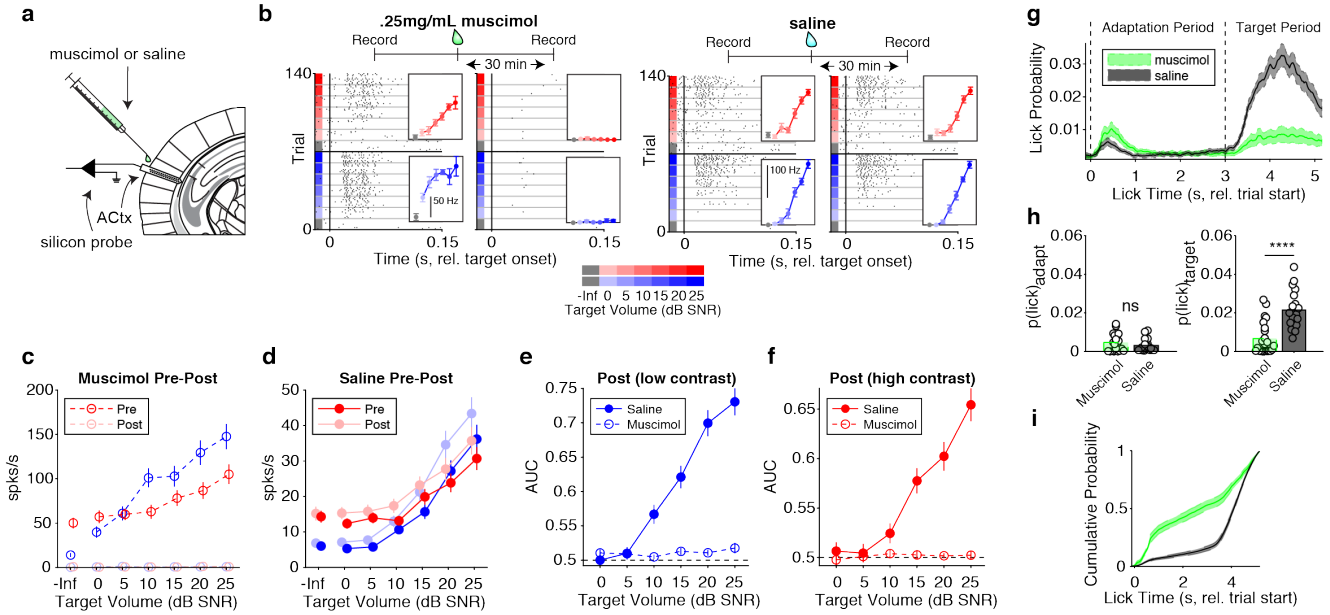
Extended Data Figure 2 (related to Figure 2). Simulation results to validate the GC-GLM.

a, Schematic of simulated neurons in the forward model. Each neuron received broadband noise inputs which changed contrast every 2s ($X_{t,f}$). A STRF modelled by a 2D-gaussian function with added noise filtered the stimulus to generate a linear response. This filter response was then modulated by a gain control function, which controlled the amount and time-course of gain control based on the stimulus contrast. This gain modulated output was then exponentiated and stochastic spikes were generated using a Poisson process. **b,** Example STRF from one simulated neuron. Colorbar indicates STRF magnitude. **c,** Model estimate of the STRF averaged across 100 simulated neurons. **d,** Example linear drive for one simulated neuron over 500 trials (ie. the filter response of the STRF convolved with the stimulus). **e,** Each panel plots the average firing rates of 100 simulated neurons (solid teal lines) and corresponding GC-GLM fits (dashed black lines) when simulating perfect gain control (GC = 1.0). Each row corresponds to 100 simulations of different gain time courses, with the top row depicting a slow transition to low contrast, with a fast transition to high contrast. The middle row plots simulations were both transitions were fast. The bottom row plots simulations where the transition to low contrast was fast, with a slow transition to high contrast. The corresponding rows of panels **f**, **g**, and **h**, are the results of simulations with the same gain time courses. **f,** Average gain time-course of the simulated neurons (solid teal lines) and the corresponding GC-GLM estimate of the gain, w, averaged over 100 simulations (black dashed lines). Insets of each panel depict the contrast kernels (dashed lines) and gain kernels (solid lines) estimated for each contrast. Blue lines indicate kernels after a switch to low contrast and red lines indicate kernels after a switch to high contrast. **g,** Average log firing rate for simulations with different gain time-courses and different degrees of gain control (GC value; the legend in the lower right indicates the color-GC value mapping). Each plotted line indicates the average firing rate/prediction for 100 simulations. **h,** Average gain time-course of all simulations (solid colored lines) and the average estimates of w (dashed gray lines). **i,** Simulations with 100 unique stimulus scenes, repeated 5 times each. Left panel plots the average firing rates and model fits. Right panel plots the true gain time-course (solid lines) and the average model gain estimate, w (dashed lines). The shaded areas indicate 2.5 and 97.5 percentiles of the gain estimates. **j,** Simulations with 5 unique stimulus scenes, repeated 100 times each. Formatting as in **i**. For panels **e-j**, the GC value colors and line formatting are indicated in the legend on the bottom right.



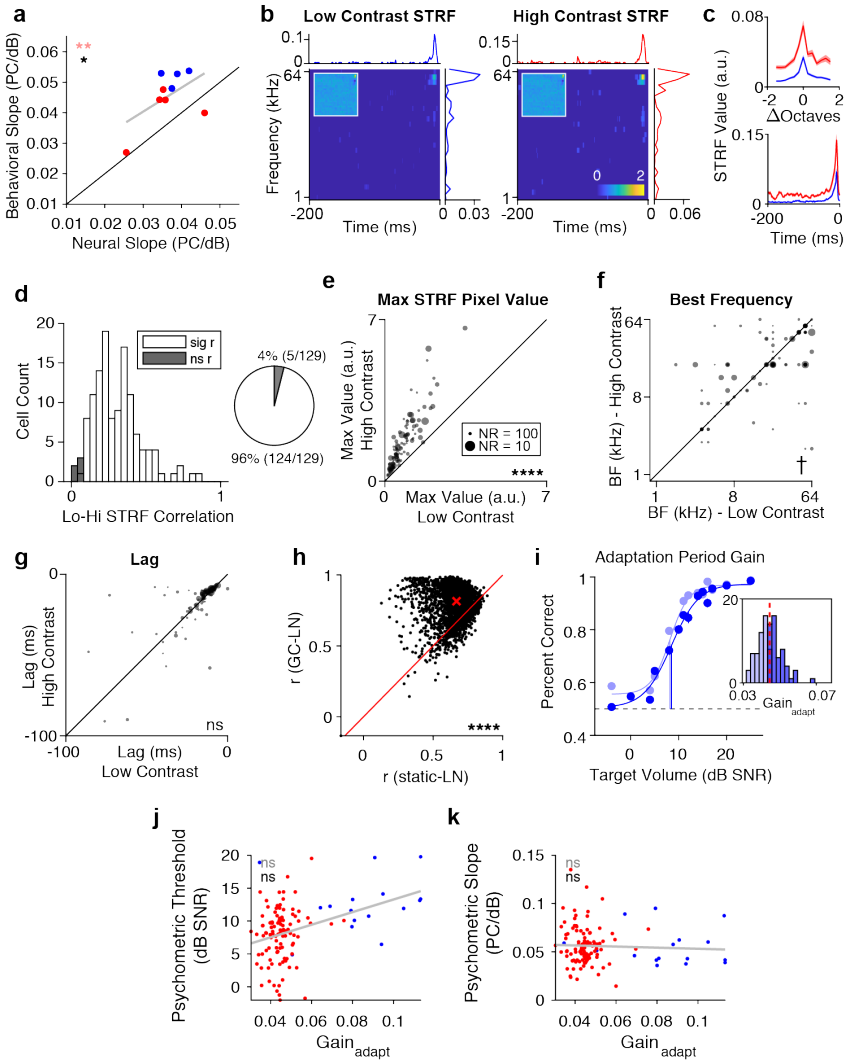
Extended Data Figure 3 (related to Figure 3). Behavioral slopes are affected by the target volume range.

a, The effect of contrast on the false alarm rates in psychometric sessions (n = 25 mice). Each dot and line represent a mouse, the blue and red bars indicate the mean false alarm rate for low and high contrast ±SEM. Results of a paired t-test ($t(23) = -6.16$, $p = 2.75e-6$) across contrast revealed a significantly higher false alarm rate in high contrast (Mean (M) = 0.22, standard deviation (std) = 0.080) compared to low contrast ($M = 0.13$, $std = 0.054$). **b,** Comparison of psychometric slopes across all mice (n = 25). Formatting as in **a**. Results of a paired t-test ($t(23) = -1.51$, $p = 0.14$) across contrast revealed no significant difference between the slopes. **c,** Average psychometric curves and percent correct for mice presented with a narrow range of targets (range = 15 dB SNR; Extended Data Table 3, row 4; dashed lines and open dots), and those presented with a wide range of targets (range = 25 dB SNR; Extended Data Table 3, row 1; solid lines and filled dots) in low contrast. Error bars indicate ±SEM. **d,** Psychometric slope for each mouse when low contrast targets were from narrow or wide target distributions. Each bar indicates the mean for each condition ±SEM. Results of an unpaired t-test ($t(9) = 2.34$, $p = 0.044$) indicated significantly larger slopes in response to narrow target distributions ($M = 0.061$, $std = 0.0060$) compared to wide target distributions ($M = 0.051$, $std = 0.0073$). **e**, Average psychometric curves and percent correct for mice presented with a narrow range of targets (average of ranges = 12 and 15 dB SNR; Extended Data Table 3, rows 4 and 5; dashed lines and open dots) or wide range of targets (range = 25 dB SNR; Extended Data Table 3, row 1; solid lines and filled dots) in high contrast. **f,** Psychometric slope for each mouse when high contrast targets came from narrow or wide distributions. Formatting as in **d**. Results of an unpaired t-test ($t(28) = 5.49$, $p = 7.29e-6$) indicated a significantly larger slopes in response to narrow target distributions ($M = 0.11$, $std = 0.033$) compared to wide target distributions ($M = 0.049$, $std = 0.017$) in high contrast. In all plots: $^{ns}p>0.1$; $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$.



Extended Data Figure 4 (related to Figure 4). Confirmation of cortical inactivation with muscimol.

a, Setup schematic for acute muscimol recordings in ACtx. **b,** Example spike rasters from two different neurons pre- and post-muscimol or saline application. On top of the raster is the timeline for each recording. Rasters are sorted by contrast and target volume, with color indicating low or high contrast backgrounds, color shade indicating target volume, and gray indicating background-only trials (-Inf). *Left panel:* spike raster of a representative neuron recorded prior to muscimol application, followed by the raster for the same neuron 30 minutes after muscimol application. *Insets:* Mean firing rate for each condition. Shade indicates target volume and the scale bar indicates the firing rate. Error bars are ±SEM across trials. *Right panel:* Example neuron before and after application of saline. Formatting as in left panels. **c,** Firing rates before and after muscimol application as a function of target volume and contrast. Dark dashed lines indicate spike rates recorded pre-muscimol application and light dashed lines indicate the responses post-application. **d,** Firing rates before and after saline application. As in **c**, dark lines are responses recorded prior to saline application and light lines indicate responses recorded after saline application. In **c** and **d**, blue and red plots indicate responses during low contrast and high contrast, respectively, and the circles not connected by a line and labelled "-Inf" are responses to background alone. **e,** Area under the ROC curve (AUC) averaged across neurons after drug application in muscimol and saline recording sessions in low contrast. Filled circles and solid lines are responses after saline was applied while open circles and dashed lines are responses after muscimol was applied. Error bars indicate ±SEM across neurons. **f,** Same as **e**, but for high contrast. **g,** Lick probability over time during muscimol or saline sessions. Dashed vertical lines indicate trial onset (0 s) and the contrast switch (3 s). Green traces are muscimol sessions and black traces are saline sessions. The shading around each trace indicates ±SEM across sessions. **h,** *Left:* comparison of lick probability during the adaptation period. *Right:* comparison of lick probability during the target period. Each circle indicates a session and color is as in **g**. **i,** Cumulative probability of licking throughout the trial, normalized within muscimol or saline conditions to sum to 1. Colors as in **g**, **h**. Shading indicates ±SEM across sessions. In all plots: $^{ns}p>0.1$; $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$.



Extended Data Figure 5 (related to Figures 5 and 6). STRFs are unaffected by contrast, and the relationship between gain during adaptation periods in the trial and behavior.

a, Neural against behavioral psychometric slopes for n = 6 mice presented with matched target volumes in high and low contrast. Formatting as in Figure 5g, except pink asterisk indicates a significant effect of slope measure and black asterisk indicates a significant effect of contrast determined using a two-way ANOVA. **b,** Example STRFs from one neuron estimated from each contrast period. *Left*: Low contrast STRF. The main plot depicts thresholded STRF values as a function of time and frequency. Inset is the original STRF, which has the same axes. Above the main plot is the temporal average across columns of the STRF, and to the right is the frequency average across rows. *Right*: High contrast STRF. Color bar indicates the color-mapping for both of the thresholded STRF plots. **c,** Average centered frequency (*top*) and temporal (*bottom*) STRF components for low and high contrast (red and blue traces, respectively) ±SEM across neurons. **d,** Histogram of correlations between low and high contrast STRFs for neurons with noise ratios (NR) below 100 (n = 129 neurons). Shaded bars indicate correlations that were not significantly different from chance, while unshaded bars indicate significant correlations, as determined by a permutation test. *Inset*: Proportions of the correlations in the population found not-significant (gray) and significant (white). **e,** Maximum STRF value across all pixels for low and high contrast, plotted for each neuron. Solid line indicates unity. The size of each circle indicates the NR of each neuron, with larger dots for smaller NR (see legend). Significance markers indicate the results of a Wilcoxon sign-rank test. **f,** Best frequency for each neuron in low and high contrast. Formatting as in **e**. **g,** Lag of the maximum STRF response for each neuron in low and high contrast. Formatting as in **e** and **f**. **h,** Correlation coefficients between the prediction of a linear-nonlinear model using STRFs estimated from the model without gain control (static-LN) versus a model with gain control (GC-LN). Each dot indicates a neuron. The red solid line indicates unity. The red "x" indicates the median correlation in each contrast. Asterisks indicate the significance of a Wilcoxon Sign-Rank test. **i,** Psychometric performance in low contrast, averaged based on a median split of average cortical gain during the adaptation period of the trial. Light dots and lines indicate the session average and psychometric fit to sessions in the bottom 50th percentile of gain, while dark dots and lines indicate the same values for sessions in the top 50th percentile of gain. Errorbars indicate ±SEM across sessions. *Inset*: distribution of average gain in each session estimated

from the adaptation period. The red dashed line indicates the median of the distribution, and the histogram bars are shaded according to whether they fall above (light blue) or below (dark blue) the median. **j,** Session-wise relationship between average gain in the adaptation period and psychometric threshold. Each dot indicates the gain and threshold for a single session, and its color indicates the contrast of the adaptation period. The gray line is the best linear fit to the data. The text in the lower right indicates the results of Likelihood Ratio Tests for models including gain as a predictor (in gray) or contrast as a predictor (in red). Full statistical results in Extended Data Table 1. Grey and black "ns" indicate that gain in the adaptation period and contrast, respectively, did not significantly predict psychometric slopes. **k,** Same as in **j**, but plotting psychometric slope as a function of gain. In all plots: $^{ns}p>0.1$; $^{\dagger}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$, $^{****}p<0.0001$.