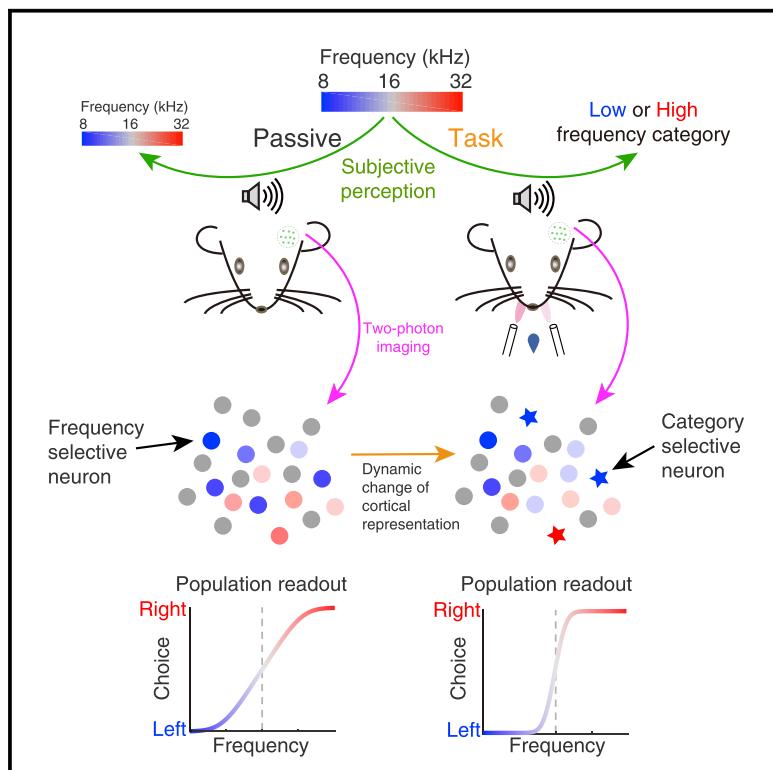


# Sensory-to-Category Transformation via Dynamic Reorganization of Ensemble Structures in Mouse Auditory Cortex

## Graphical Abstract



## Authors

Yu Xin, Lin Zhong, Yuan Zhang,  
Taotao Zhou, Jingwei Pan,  
Ning-long Xu

## Correspondence

xunl@ion.ac.cn

## In Brief

Xin et al. studied the computational mechanism for transforming sensory information into meaningful categories in the auditory cortex using two-photon imaging. Cortical neurons are dynamically recruited to enhance discrimination for stimuli near decision boundaries. Categorical decisions can be accurately decoded from neuronal populations.

## Highlights

- Sensory responses in auditory cortex are modulated during stimulus categorization
- Cortical ensembles show enhanced responses to stimuli near the category boundary
- Auditory cortex neurons accurately predict decisions during categorization behavior



# Sensory-to-Category Transformation via Dynamic Reorganization of Ensemble Structures in Mouse Auditory Cortex

Yu Xin,<sup>1,2</sup> Lin Zhong,<sup>1,2</sup> Yuan Zhang,<sup>1,2</sup> Taotao Zhou,<sup>1</sup> Jingwei Pan,<sup>1</sup> and Ning-long Xu<sup>1,2,3,4,\*</sup>

<sup>1</sup>Institute of Neuroscience, State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology, Shanghai 201210, China

<sup>4</sup>Lead Contact

\*Correspondence: xunl@ion.ac.cn

<https://doi.org/10.1016/j.neuron.2019.06.004>

## SUMMARY

The ability to group physical stimuli into behaviorally relevant categories is fundamental to perception and cognition. Despite a large body of work on stimulus categorization at the behavioral and cognitive levels, little is known about the underlying mechanisms at the neuronal level. Here, combining mouse auditory psychophysical behavior and *in vivo* two-photon imaging from the auditory cortex, we investigate how sensory-to-category transformation is implemented by cortical neurons during a stimulus categorization task. Distinct from responses during passive listening, many neurons exhibited emergent selectivity to stimuli near the category boundary during task performance, reshaping local tuning maps; other neurons became more selective to category membership of stimuli. At the population level, local cortical ensembles robustly encode category information and predict trial-by-trial decisions during task performance. Our data uncover a task-dependent dynamic reorganization of cortical response patterns serving as a neural mechanism for sensory-to-category transformation during perceptual decision-making.

## INTRODUCTION

The ability to classify sensory information into discrete categories is essential for survival and fundamental to cognition (Ashby and Maddox, 2005; Freedman and Assad, 2006; Freedman et al., 2001; Harnad, 1990; Liberman et al., 1957; Seger and Miller, 2010). During categorization, the brain transforms continuous sensory information into discrete classes, a computational process that represents the critical intermediate steps between sensory input and motor output during decision-making (Freedman and Assad, 2011; Grinband et al., 2006). Neural activity correlated with stimulus categories has

been observed in various brain regions of different species (Chang et al., 2010; Freedman and Assad, 2006; Freedman et al., 2001; Prather et al., 2009; Tsunada et al., 2011). However, the patterns of these activities are often directly mapped to the category memberships, reflecting the readout of categorization processes, whereas the neural circuit mechanism for the categorization computations is still unclear (Russ et al., 2007).

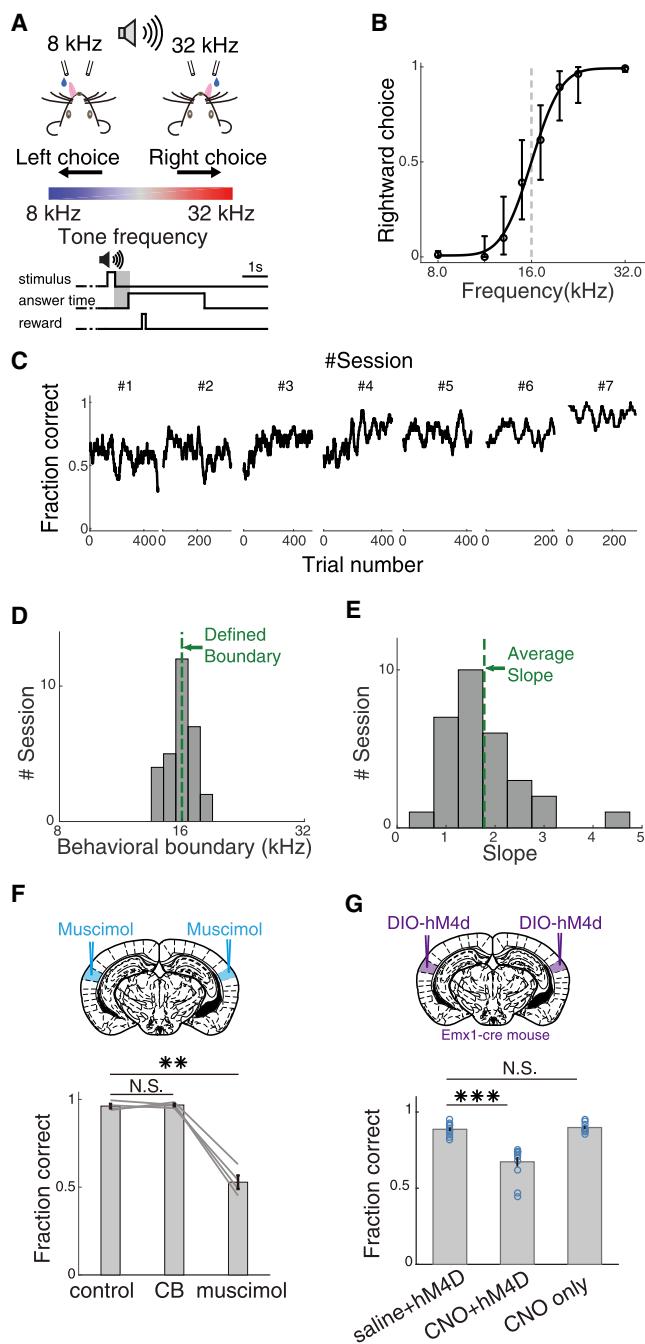
The sensory cortices have been shown to carry out behaviorally relevant computations in addition to sensory feature analysis. For instance, the sensory cortices receive rich top-down feedback inputs (Zhang et al., 2014; Zingg et al., 2014), exhibit choice-related activity (Britten et al., 1996; Newsome et al., 1989; Nienborg and Cumming, 2009; Yang et al., 2016), and are strongly modulated by motor actions (Nelson et al., 2013; Schneider et al., 2014; Xu et al., 2012). The primary auditory cortex has been shown to be modulated by behavioral contexts (Francis et al., 2018; Kato et al., 2015; Kuchibhotla et al., 2017) and involved in more complex computations beyond basic feature extraction (Jaramillo and Zador, 2011; King and Nelken, 2009; Nelken, 2004; Schreiner and Winer, 2007; Tsunada et al., 2011, 2016; Znamenskiy and Zador, 2013). Here, we sought to understand whether and how sensory information is transformed into behaviorally relevant categories in the primary sensory cortex at the single-neuron and population levels. Using mice performing an auditory categorization task and employing *in vivo* two-photon functional imaging from layer 2/3 (L2/3) of the auditory cortex, we found that neuronal ensembles are dynamically reorganized during a categorization task, reflecting implementation of a cortical computation underlying sensory-to-category transformation during perceptual decision-making.

## RESULTS

### Categorical Auditory Decisions in Head-Fixed Mice

To probe how the brain maps continuous sensory information to behaviorally relevant categories, we developed an auditory psychophysics task in head-fixed mice based on a two-alternative forced choice (2AFC) paradigm (Guo et al., 2014; Znamenskiy and Zador, 2013). Head fixation allows reliable and precise stimulus delivery as well as cellular-resolution two-photon imaging during task performance. Mice were first trained to discriminate





**Figure 1. The Auditory Categorization Task in Mice Involves the Auditory Cortex**

- (A) Schematic showing behavioral task configuration. Top: behavioral configuration. Bottom: trial time structure (STAR Methods).
- (B) Example psychometric function of one behavior session (STAR Methods). The dashed black line marks the defined category boundary. Error bars indicate 95% confidence interval.
- (C) Typical learning phase of one mouse.
- (D) Distribution of the psychometric boundary ( $n = 30$  sessions).
- (E) Distribution of the slope values at the boundary ( $n = 30$  sessions).
- (F) Influence of performance after muscimol silencing of bilateral auditory cortex. Top: schematic showing the muscimol injection site. Bottom: task performance. Mice's performance was significantly impaired after muscimol

two exemplar tones (low and high; e.g., 8 and 32 kHz) by directional licking to left and right lick ports (Figure 1A). After reaching a stable performance level (>85% correct; Figure 1C), the intermediate frequencies were introduced (Figure 1B). Psychometric functions were constructed based on the probability of choosing right lick port as a function of tone frequencies (logarithmically scaled in octave), which represent mice's decisions on categorizing tones as high- or low-frequency categories. The hallmark of categorization is a sharp change in choice proportions upon continuously varying stimulus parameters (Freedman and Assad, 2006; Freedman et al., 2001; Liberman et al., 1957; Wytenbach et al., 1996), which indicates the “boundary” between categories. Here, we first examined whether mice show categorical responses to tone frequencies by presenting intermediate frequencies in “probe” trials (STAR Methods). We found that mice exhibited a sharp change in choice proportions near the midpoint of the tested frequency range (Figures 1B, 1D, and 1E; see also Figure S1A), indicating that mice are capable of making categorical judgments on basic acoustic stimuli.

The involvement of cortical areas in a given perceptual task often depends on task conditions and training stages (Gimenez et al., 2015; Hong et al., 2018; Jaramillo and Zador, 2011; Kuchibhotla et al., 2017; Otchy et al., 2015; Talwar et al., 2001). Here, we first examined whether the primary auditory cortex was involved in our behavioral task. We performed reversible inactivation by bilateral injection of muscimol (a GABA<sub>A</sub> receptor agonist) in the primary auditory cortex and found that mice's discrimination for the exemplar tones was significantly impaired (Figure 1F). This result is consistent with previous studies that also used muscimol silencing of bilateral auditory cortex to assess the requirement of auditory cortex in task performance (Jaramillo and Zador, 2011; Kuchibhotla et al., 2017). In addition, we expressed hM4D, a designer receptor exclusively activated by designer drug (DREADD) (Armbruster et al., 2007), in bilateral primary auditory cortex of Emx1-IRES-Cre mice using adeno-associated virus (AAV). Intraperitoneal injection of clozapine-N-oxide (CNO) significantly impaired task performance in hM4D expressing mice, but not in control mice (Figure 1G). These results suggest that the mouse primary auditory cortex may contribute to our frequency categorization task.

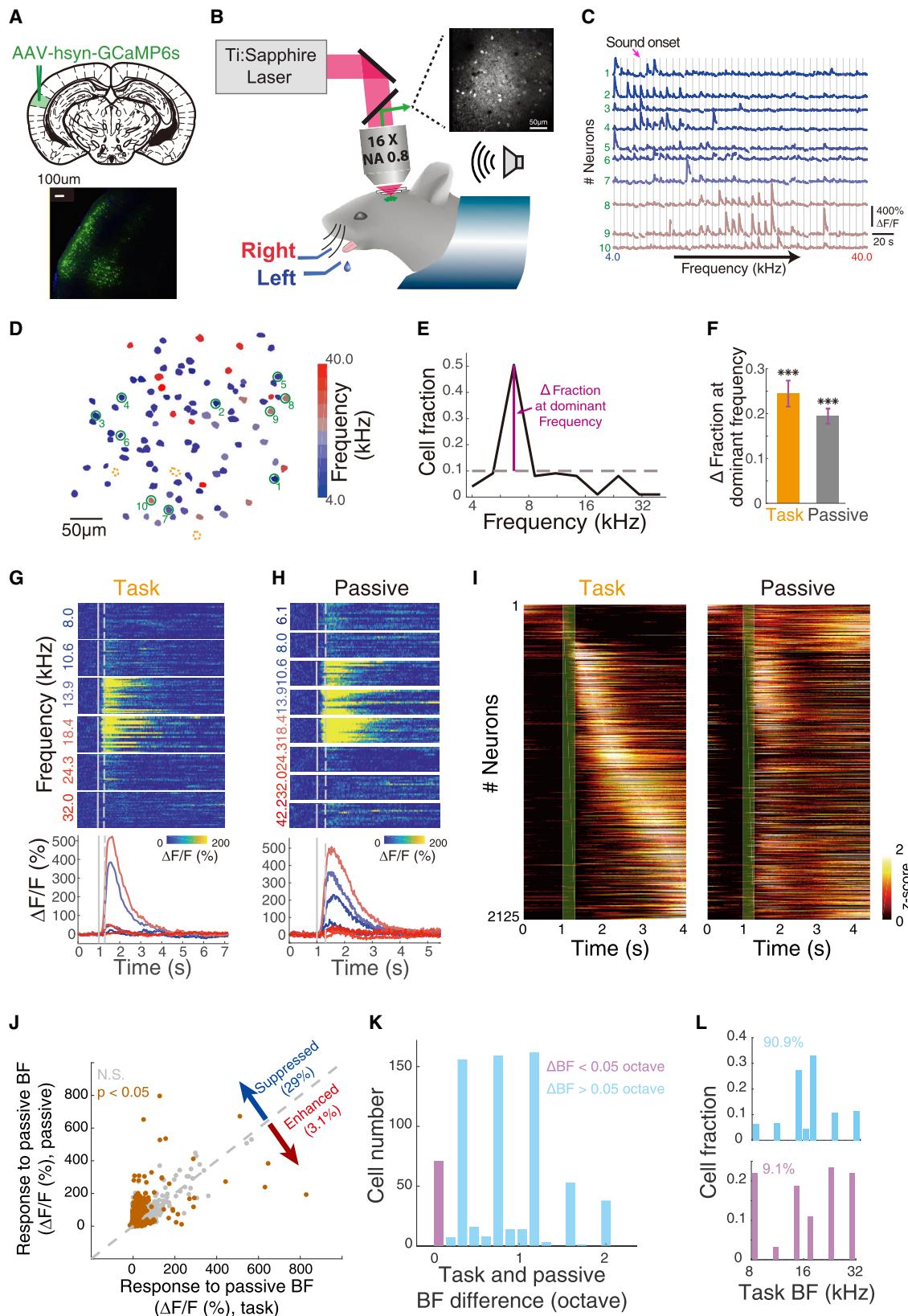
### Two-Photon Calcium Imaging in Awake Mice Reveals Spatial Organization of Local Cortical Populations

To examine how neuronal activity in auditory cortex may contribute to stimulus categorization, we performed *in vivo* two-photon calcium imaging from L2/3 neuronal populations. We expressed a genetically encoded calcium indicator,

injection ( $0.53 \pm 0.077$ ) compared to control mice ( $0.96 \pm 0.025$ ) and mice injected with cortex buffer ( $0.97 \pm 0.014$ ). Two-sample t test;  $n = 4$  animals for each condition. Error bars indicate SEM. N.S.,  $p > 0.05$ ; \*\*,  $p < 0.01$ .

(G) Similar to (F), but with chemogenetic inactivation of the auditory cortex. Performance was significantly impaired in mice expressing hM4D in bilateral auditory cortex and CNO injection ( $0.67 \pm 0.034$ ) compared with saline injection ( $0.89 \pm 0.035$ ), but not for CNO injection only in control mice ( $0.90 \pm 0.033$ ). Two-sample t test;  $n = 12$  sessions for each condition. Error bars indicate SEM. N.S.,  $p > 0.05$ ; \*\*,  $p < 0.001$ .

See also Figure S1.



(legend on next page)

GCaMP6s (Chen et al., 2013), in mouse auditory cortex using AAV vector and implanted a chronic imaging window (Figures 2A and 2B; STAR Methods). Reliable single-neuron responses with a high signal-to-noise ratio were evoked by pure tone stimuli during task performance and passive tone stimulation (Figures 2C, 2G, and 2H; STAR Methods). We first examined the frequency selectivity of individual neurons in response to pure tones of various frequencies randomly delivered under a passive awake condition (Figures 2C and S2C). In some cases, nearby neurons show distinct frequency selectivity (Figures 2D), consistent with previous imaging results in anesthetized mice (Bandyopadhyay et al., 2010; Rothschild et al., 2010). However, for local populations, we also found that the best frequency (BF) of individual neurons within the same imaging field ( $\sim 250 \times 250 \mu\text{m}$ ) are not randomly distributed, with a higher proportion of neurons showing similar frequency preference (Figures 2D and 2E). We defined the frequency preferred by majority of neurons in the local imaging field as the dominant frequency and found that during both task and passive stimulation, the proportion of neurons preferring the dominant frequency was significantly higher than random distribution (Figures 2E and 2F). These results indicate that the frequency representations in local population of the L2/3 of auditory cortex of awake mice are spatially organized, with clustering ensemble structures, consistent with a previous study in unanesthetized mice (Issa et al., 2014).

### Changes in Frequency Selectivity of Auditory Cortex Neurons during Task Performance

Previous studies comparing sound evoked responses in the auditory cortex between passive stimulation and behavior conditions reported general enhancement or suppression of neuronal responses by behavior (David et al., 2012; Francis et al., 2018; Fritz et al., 2003; Kato et al., 2015; Kuchibhotla et al., 2017; Otazu et al., 2009). It is yet unclear whether behavioral tasks may specifically modulate stimulus selectivity of auditory cortex neurons in favor of behavioral needs. Our two-choice auditory categorization task provides a unique opportunity to examine category-related modulation. Figure 2I shows Z-scored mean  $\text{Ca}^{2+}$

signals of all imaged neurons sorted according to the peak time in behavioral trials. A significant proportion of neurons show short-latency tone evoked responses aligned to the stimulus time (Figures 2G and 2H). We first compared responses to the BF determined using passive stimulation ( $\text{BF}_{\text{pass}}$ ) between task performance and passive stimulation. We observed significant difference in responses to the  $\text{BF}_{\text{pass}}$  in 32% of neurons, among which majority of neurons (29%) show suppressed response during the task while only a small proportion of neurons (3%) show enhanced responses (Figure 2J).

We further asked whether the auditory cortical neurons show a change in frequency selectivity during task compared to passive stimulation. We found that among neurons showing suppressed responses to the  $\text{BF}_{\text{pass}}$ , a majority (90.9%) showed different frequency preference during the task (Figure 2K). Remarkably, for these neurons, the frequency preferences during the task were not uniformly distributed across the tested frequencies, with more neurons preferring the frequencies near the category boundary (Figure 2L). A smaller fraction (9.1%) of neurons showing suppressed responses to  $\text{BF}_{\text{pass}}$  showed similar frequency preference during task and passive stimulation. However, these neurons exhibited near-uniform distribution of frequency preference during the task (Figure 2L). These results indicate that rather than a broad-range and nonselective suppression of responses, task-dependent modulation changed the stimulus selectivity of auditory cortex neurons.

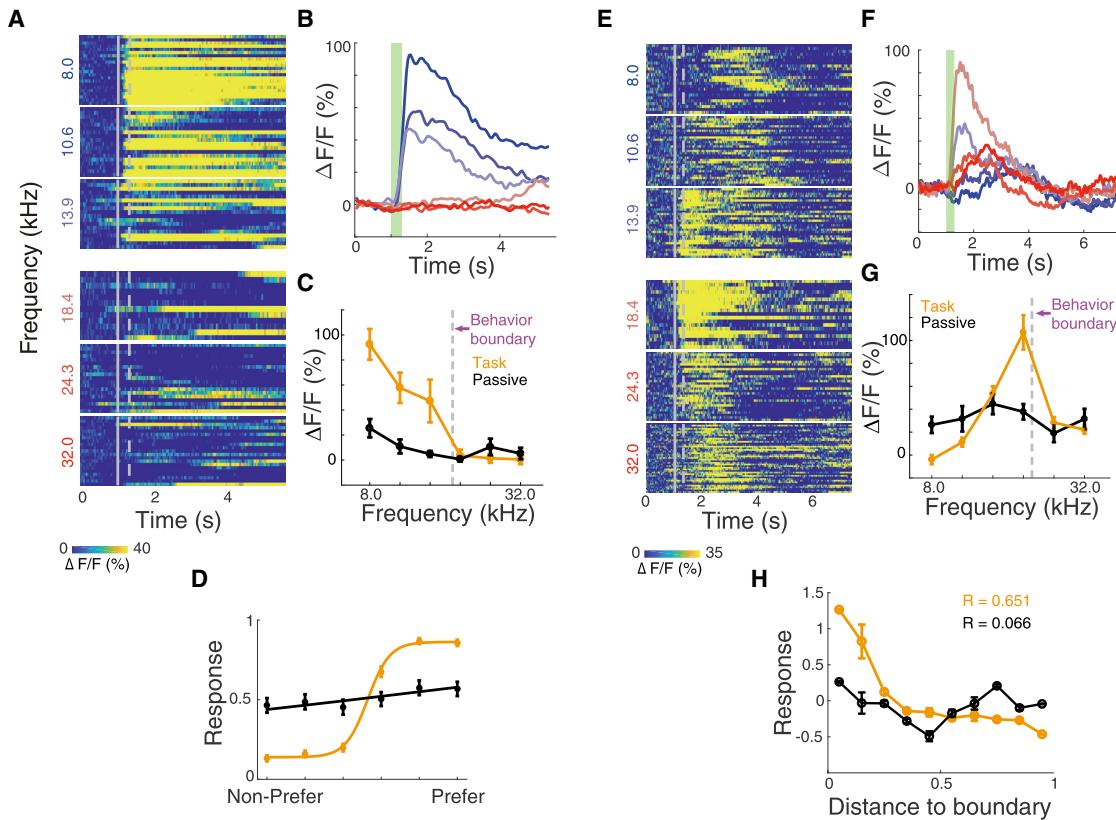
### Categorization Task Reshapes Cortical Neuronal Response Profiles

To further test whether the behavioral modulation of neuronal responses may reshape sensory representations in a task-relevant manner, we compared response profiles of individual neurons in the auditory cortex between task performance and passive stimulation conditions. We found that a substantial proportion (23.2%) of the imaged neurons exhibited categorization-related modulations. First, some neurons became strongly selective to category membership. As shown in Figures 3A and 3B, the neuron exhibited strong responses to tone stimuli belonging to the

### Figure 2. Two-Photon Imaging of Auditory Cortex L2/3 Neurons

- (A) Schematic showing the injection site of the virus (top) and GCaMP6s expression in auditory cortex (bottom).
- (B) Schematic showing *in vivo* two-photon imaging of auditory cortex during auditory-guided 2AFC task. The mouse was head-fixed with body constrained in a plastic tube. Top right: an example image of L2/3 neurons in auditory cortex.
- (C) Calcium signals of 10 example neurons. Vertical gray lines represent stimulus onset time. Trials were sorted by stimulus frequency (horizontal direction). Neurons were sorted by their best frequency (BF; vertical axis), which is indicated by the color map in (D).
- (D) BFs of individual neurons from an example imaging field. The dashed orange circles indicate neurons with no significant response. Green circles and numbers indicate neurons shown in (C).
- (E) Histogram of the BF values in (D). Dashed gray line indicates the distribution of BF values of a pseudo-randomly (evenly) distributed representation.
- (F) Bar plot of the difference between the dominant frequency fraction and the evenly distributed fraction (purple line in E; task,  $0.24 \pm 0.029$ ; passive,  $0.19 \pm 0.017$ ; Student's t test,  $n = 21$ , \*\*\*,  $p < 0.001$ ). Error bars indicate SEM.
- (G and H) Response profile of example neurons for task (G) and passive conditions (H). Top: color-coded calcium signals (%  $\Delta F/F$ ). Trials were aligned to sound onset time (gray solid lines, gray dashed lines, sound offset time). Bottom: color-encoded mean calcium trace for each frequency.
- (I) Temporal dynamics of all imaged neurons during task (left) and passive conditions (right). Each row was the across-trial averaged activity of one neuron, sorted by peak time under the task condition. Green shading indicates sound period.
- (J) Single-neuron response value to  $\text{BF}_{\text{pass}}$  (STAR Method) between task and passive conditions. Orange dots depict neurons showing significant difference between task and passive conditions (two-sample t test).
- (K) Frequency difference between  $\text{BF}_{\text{pass}}$  and  $\text{BF}_{\text{task}}$  (in octaves); only neurons showing suppressed  $\text{BF}_{\text{pass}}$  responses in (J) were used.
- (L) Histogram of  $\text{BF}_{\text{task}}$  for the population of neurons in (K). Top:  $\text{BF}_{\text{task}}$  for neurons having different BFs between task and passive conditions. Bottom:  $\text{BF}_{\text{task}}$  for neurons having the same BF in two conditions.

See also Figure S2.



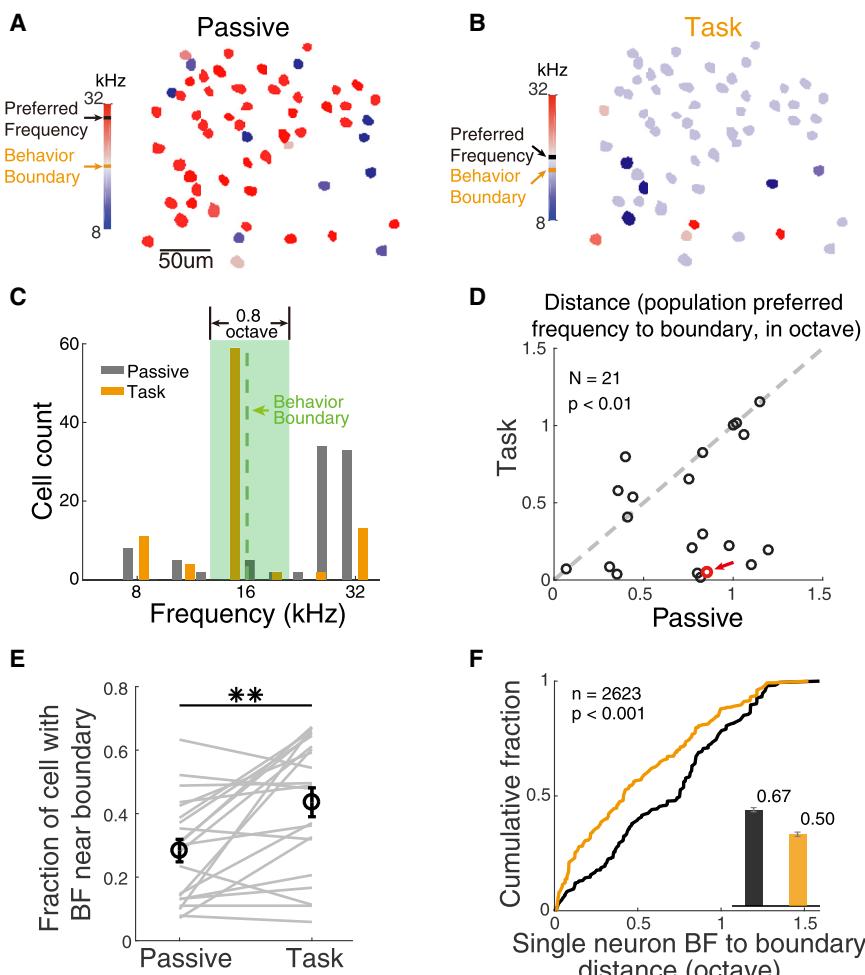
**Figure 3. Single-Neuron Activity Is Modulated by a Categorization Task**

(A) Color raster plot of calcium signals for one example category-selective neuron in the task session (same arrangement as in Figure 2H).  
 (B) Averaged calcium traces for the neuron in (A). Color encodes tone frequency. Green shading indicates the sound period.  
 (C) Responses amplitude for the neuron in (A) as a function of tone frequencies for both task (orange) and passive (black) sessions. Error bars indicate SEM.  
 (D) Averaged response of all category-selective neurons as a function of frequency ordered by preferred versus non-preferred categories. Single-neuron activity was normalized before average. Error bars indicate SEM.  
 (E) Color plot as in (A) showing calcium signals of an example neuron selective to near-boundary stimuli.  
 (F) Averaged calcium traces for (E).  
 (G) Responses amplitude for neuron in (E).  
 (H) Averaged responses (Z-scored) of all neurons showing preference to near-boundary frequency, plotted as a function of distance to behavior boundary. Circles and error bars indicate the mean and SEM of response values within each bin.

See also Figure S3.

low-frequency category but showed almost no responses to tone stimuli belonging to the high-frequency category. Such category preference, however, was not present during passive stimulation with the same set of tone stimuli (Figure 3C). We identified these neurons by comparing sigmoidal fits for responses from task and passive conditions (Figures S3A and S3B; 9.8% of all imaged neurons; **STAR Methods**). Averaging the responses during task performance across these neurons as a function of tone frequencies ordered by an individual neuron's preferred category gave rise to a sigmoidal population response curve, exhibiting strong selectivity to category membership resembling behavioral categorization function (Figures 3D). However, during passive stimulation, these neurons show diverse frequency selectivity, and averaging across neurons gave rise to an almost flat population response curve lacking category selectivity (Figures 3D). Therefore, the strong selectivity to stimulus category membership in these neurons arose from task-related modulation.

Second, we found that many neurons showed enhanced responses to the tone frequencies near category boundaries during task performance compared to passive stimulation (Figure 2L). As shown in Figures 3E–3G, the neuron exhibited stronger responses to stimuli near the category boundary during task performance than during passive stimulation. The enhanced neuronal responses to the near-boundary stimuli may either reflect a task-demanded enhancement in differentiating stimuli near category boundaries (Bonnasse-Gahot and Nadal, 2008; Raizada and Poldrack, 2007) or coincide with an enhanced sensory response by gain modulation. The former would suggest that the response enhancement for near-boundary stimuli was independent of the sensory selectivity during passive stimulation, while the latter would suggest an enhanced response without changes in stimulus selectivity. To distinguish these possibilities, we identified neurons with enhanced responses to near-boundary stimuli by first using a Gaussian fit to an individual



**Figure 4. Dynamic Reorganization of Frequency Representation in Local Ensembles during Categorization Task**

(A and B) Color-coded BF distribution for passive (A) and task (B) condition of an example imaging field.

(C) Histogram of the BF distributions in (A) and (B). The light green shade indicates the frequency range considered as the near-boundary range.

(D) Comparison of the distance between preferred frequency and behavior boundary between passive and task conditions (in octaves; passive,  $0.74 \pm 0.071$ ; task,  $0.44 \pm 0.084$ ; paired t test). Red arrow indicates the example field in (A)–(C).

(E) The fraction of neurons with near-boundary BFs compared between passive and task conditions. Each gray line represents one imaging field. Black circles and error bars represent the mean and SEM (passive,  $0.28 \pm 0.036$ ; task,  $0.44 \pm 0.046$ ; paired t test,  $n = 21$ ,  $p = 0.008$ ).

(F) Comparison of single-neuron BF to behavior boundary distance between task (brown) and passive (black) conditions. The inserted bar plot is the averaged distance, and error bars indicate 95% confidence interval.  $p$  value was calculated using the Wilcoxon rank-sum test.

See also Figures S4.

neuron's tuning curve during the task and then selected neurons with preferred frequency near the category boundary (within 0.2 octaves from the inflection point of the psychometric function; see Figures S3C and S3D; STAR Methods). When averaging the responses across the population (13.4% of all imaged neurons), we found a consistent enhancement for responses near the category boundary, which was not present for responses during passive stimulation (Figure 3H). Thus, the task-dependent selective enhancement of neuronal responses near the category boundary likely reflects an increased discrimination demand for near-boundary stimuli during stimulus categorization.

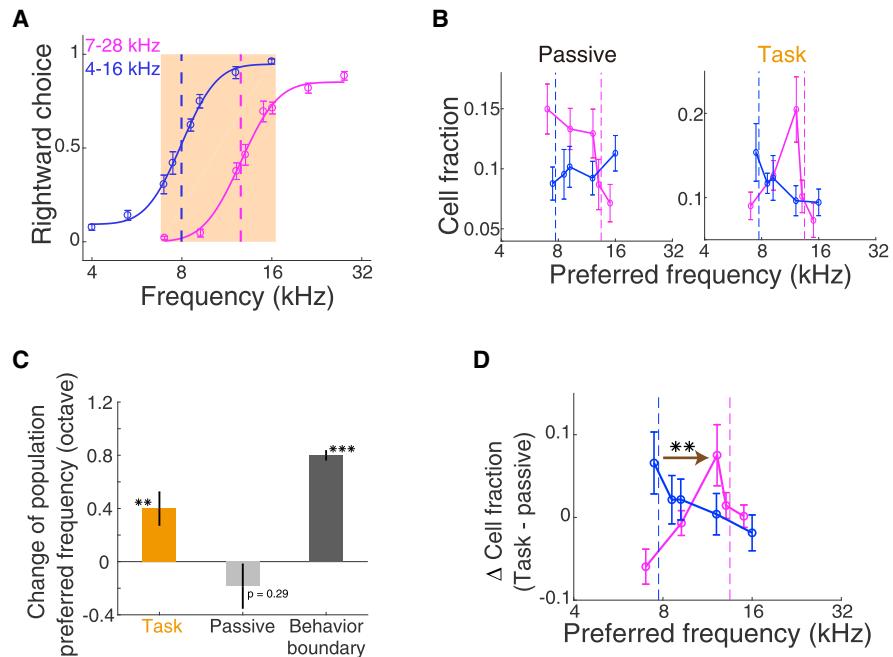
#### Enhanced Ensemble Representation of Near-Boundary Stimuli via Task-Dependent Neuronal Recruitment

The increased responses to the near-boundary stimuli may reflect an increased demand for neuronal resources to enhance discrimination near the decision boundary during categorization (Bonassie-Gahot and Nadal, 2008; Freedman and Assad, 2006; Freedman et al., 2001; Guenther et al., 2004; Liberman et al., 1957; Raizada and Poldrack, 2007). We thus asked whether the categorization task was accompanied by a reorganization of local ensemble structures to meet behavioral demands. We found that during task performance, the same population of neu-

rons exhibited a different population preference than during passive stimulation, such that the local dominant frequency was closer to the category boundary (Figures 4B and 4C). Overall, for each imaging field, the distance from the population dominant frequency during the task to the behavioral category boundary is significantly smaller than that for the dominant frequency during passive stimulation (Figure 4D). In each imaging field, a significantly greater fraction of neurons show a preference for near-boundary frequencies during the task than during passive stimulation (Figure 4E). For individual neurons across the entire dataset, the distance from the BF to the behavioral boundary was also significantly smaller during task performance than during passive stimulation (Figures 4F). Thus, during the categorization task, the stimuli near the category boundary became more highly represented, reflecting task-related modulation to meet behavioral demands.

#### Task-Related Modulation of Stimulus Selectivity Followed Learned Category Boundary Changes

It has been shown that mice can learn to categorize tones based on different category boundaries (Jaramillo and Zador, 2014). We wonder whether the category-related modulation of neuronal activity we observed in the auditory cortex would also change when the animal learns to perform stimulus categorization based on new category boundaries. We trained mice to first perform stimulus categorization on tones in a lower frequency range (4–16 kHz). After mice reached performance criterion, they were then trained to categorize tones in a higher frequency range



**Figure 5. Task-Related Modulation of Stimulus Selectivity following Changes in Category Boundary**

(A) Summarized psychometric function for two frequency-categorization tasks using different categorization boundaries.

(B) Neuron's preferred frequency distribution within the overlapped frequency range of the two task contexts.

(C) Summary of changes in population preferred frequencies across sessions ( $n = 12$ ) when category-boundary was changed from the low to the high value. Population preferred frequency during the task, but not passive, session was significantly shifted to the same direction as the training boundary change. Student's t test,  $n = 12$ . \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ .

(D) Summary of the neural fraction difference between task and passive conditions compared between two frequency ranges. Student's t test,  $n = 12$ . \*\*,  $p < 0.01$ . All error bars indicate SEM. See also Figures S5.

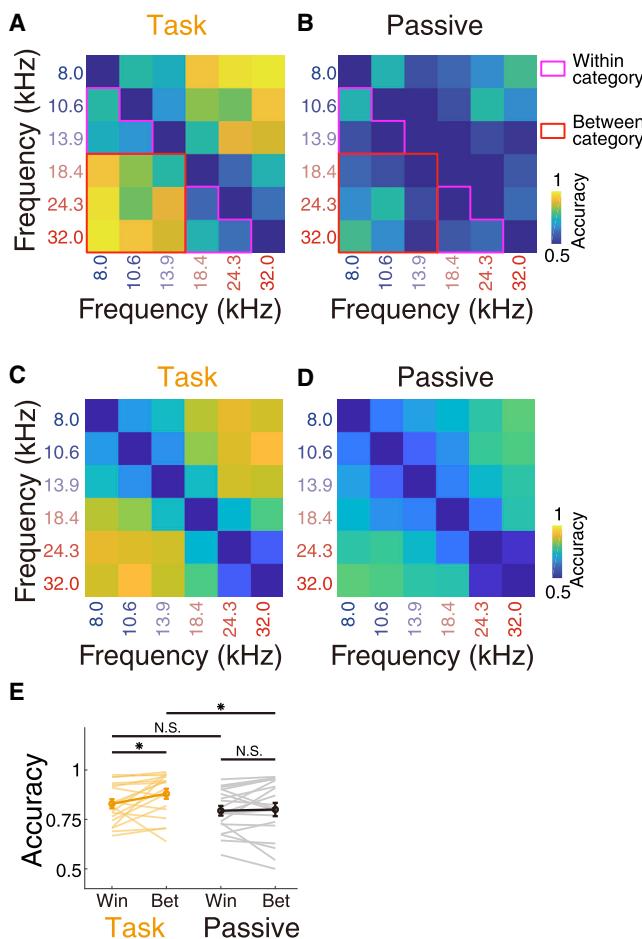
(7–28 kHz). Mice were capable of performing stimulus categorization under both task contexts, changing category boundaries at a lower frequency in the first task context to a higher frequency in the second task context (Figures 5A and 5C). This result also excludes the possibility that the steep slope at the category boundary we observed may be due to a coincidence that mice could have higher sensory discrimination near the category boundary without task training.

*In vivo* chronic two-photon imaging allowed us to track the same population of neurons in the auditory cortex before and after mice learned to perform stimulus categorization under the two task contexts (two frequency ranges). Similar to what we found for the frequency range of 8–32 kHz, under both of the new task contexts, the neuron population best frequencies were closer to the category boundary under the task condition than under the passive condition (Figure S5). For categorization in the first task context (4–16 kHz), the population preference during task performance was closer to the lower frequency category boundary, while for categorization in the second task context (7–28 kHz), the population preference was shifted toward the higher frequency boundary (Figures 5B–5D). In contrast, under passive stimulation, although the population preferred frequencies showed some degree of variability between the two frequency ranges (Figure 5B), presumably due to differential training histories, the population preference to the category boundary and its changes following the task

context change was not present (Figures 5B and 5C). These results indicate that the category-related modulation of neuronal activity co-varies with the learning-induced changes in the behavioral category boundary under different task contexts.

#### Categorical Frequency Discrimination by Simultaneously Imaged Neuronal Populations

While individual neurons in the auditory cortex exhibit category-related modulation during the categorization task, we wonder whether categorical information is present at the population level of simultaneously imaged neurons. We used a linear classifier (Hung et al., 2005) to decode stimulus information from the population activity by training the classifier to discriminate pairs of tone frequencies (STAR Methods). We then constructed population discrimination matrices to visualize the degree of discrimination by simultaneously imaged neurons for pairs of tone frequencies belonging to the same or different categories. Interestingly, the discrimination matrix constructed from population responses during task shows a distinctive pattern, with higher discrimination accuracy for stimuli belonging to different categories than for stimuli belonging to the same categories (Figures 6A and 6C), indicating that the population coding for the difference in tone stimuli also exhibits categorical structure. In contrast, this categorical discrimination pattern was largely diminished for responses from the same population of neurons during passive stimulation (Figures 6B and 6D). Across the



**Figure 6. Population Discrimination of Tone Stimulus**

(A and B) Population discrimination accuracy of paired-tone frequencies from an example field for task (A) and passive (B) conditions.  
 (C and D) Average population discrimination accuracy for all sessions for task (C) and passive (D) conditions ( $n = 21$  sessions).  
 (E) Comparison of population discrimination accuracy for different types of tone pairs (STAR Methods). Win, within category; Bet, between category. Task: within category,  $0.83 \pm 0.023$ ; between category,  $0.88 \pm 0.025$ . Passive: within category,  $0.79 \pm 0.025$ ; between category,  $0.80 \pm 0.034$  (paired t test;  $n = 21$  sessions, N.S.,  $p > 0.05$ ; \*,  $p < 0.05$ ). Error bars indicate SEM.

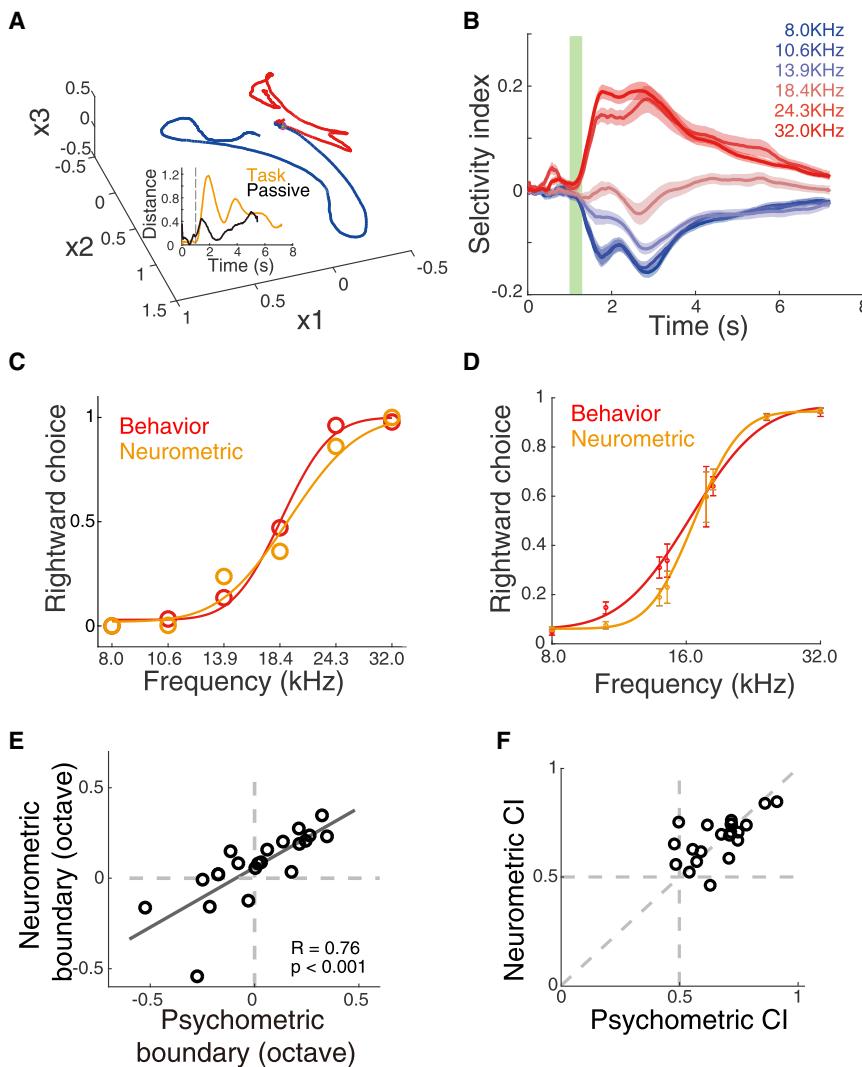
grouped data over all imaging fields, we compared the population discrimination accuracies for between-category stimulus pairs versus within-category stimulus pairs (Figure 6E; STAR Methods). We found that during task performance, the population discrimination accuracy for between-category tones is significantly higher than that for within-category tones (Figure 6E). In contrast, the discrimination accuracy from population activity during passive stimulation is not significantly different for between- and within-category stimulus pairs (Figure 6E). Thus, cortical neuron populations exhibit categorical stimulus discrimination during the categorization task, but not during passive stimulation, suggesting a task-dependent modulation imposing a category structure to population sensory representations.

### Population Representation of Categorical Perceptual Decisions

Neural correlates of perceptual decisions based on activity of individual neurons have been widely studied in sensory regions by constructing individual neurons' neurometric functions to link neuronal activity to perceptual judgments (Liu et al., 2013; Newsome et al., 1989; Parker and Newsome, 1998). However, the link between simultaneously recorded population activity patterns in sensory cortex and the categorical perceptual judgments is yet to be demonstrated. To examine this, we first analyzed the population dynamics as a trajectory in a state space of neuronal population activity (Briggman et al., 2005; Harvey et al., 2012). The averaged trajectories across trials of a given stimulus (visualized using factor analysis for dimensionality reduction) depict the population representation for this stimulus evolving over the trial time (Figure 7A). We constructed a population selectivity index based on the Euclidean distance from the trajectories of individual trials to the mean trajectories of all trials of the opposite stimulus category as a measure of population discrimination (Figure 7B; STAR Methods). The trajectories for the tone stimuli of high- and low-frequency categories start to separate soon after stimulus onset, indicating sensory-driven changes in population dynamics. The population trajectories exhibit a fast separation following stimulus onset and a lasting separation throughout the trial time (Figures 7A and 7B), reflecting early sensory responses and ongoing feedback activity during the conscious perceptual process (Dehaene and Naccache, 2001; Dehaene and Changeux, 2011; Gaillard et al., 2009). This lasting separation was unlikely due to the delayed calcium signaling, since it was not present for the passive stimulation condition (Figure 7A). In addition, population decoding using an estimated spike rate from calcium signals (STAR Methods), which reduced the decay time of calcium fluorescence signals, also revealed a lasting effect in decoding accuracy for activity during the task, but not during passive stimulation (Figure S6).

To quantify the relation between population dynamics and categorical decisions, we examined the peak values of the selectivity index for each stimulus as a function of tone frequency and compared it with psychometric function (STAR Methods). We found that the population neurometric function largely captured the characteristics of the behavioral psychometric function (Figures 7C and 7D). Across all the imaging sessions, the category boundaries of the population neurometric functions are strongly correlated with the behavioral category boundaries (Figure 7E). In addition, we constructed a categorization index (CI) based on the differences between averaged proportion of choices on the two categories (STAR Methods). A CI value  $>0.5$  would indicate significant categorization. We found that the CI values based on neuronal population activity during task performance are comparable to those computed from behavioral choices (Figure 7F) and are on average  $>0.5$  in both cases. Taken together, these results support a population representation of stimulus categorization in the auditory cortex that matches behavioral performance.

In addition to using model-free population dynamics to examine population representation of stimulus categorization, we also used a population decoding approach (as shown in Figure 6) to examine the stimulus categorization by cortical



**Figure 7. Population Dynamics for Stimulus Categorization**

(A) Mean trajectories of the 8-kHz (blue) or 32-kHz (red) tone stimuli trials in state space. Inset: Euclidean distance between the two trajectories as a function of time, for task and passive conditions. Dashed gray line indicates sound onset time.

(B) Mean traces of selectivity index for each frequency as a function of time for one session. Green stripe represents sound period. Shadings indicate SEM.

(C) Neurometric and behavioral function for the example session in (B).

(D) Averaged neurometric and psychometric functions across all sessions ( $n = 21$ ). Error bars indicate SEM.

(E) Correlation of neurometric boundary and psychometric boundary. Black line represents a linear fit of all data.  $p$  value was calculated using Student's  $t$  test.

(F) Comparison of neurometric and psychometric CI (behavior,  $0.67 \pm 0.026$ ; task,  $0.68 \pm 0.022$ ; paired  $t$  test,  $p = 0.61$ ; STAR Methods).

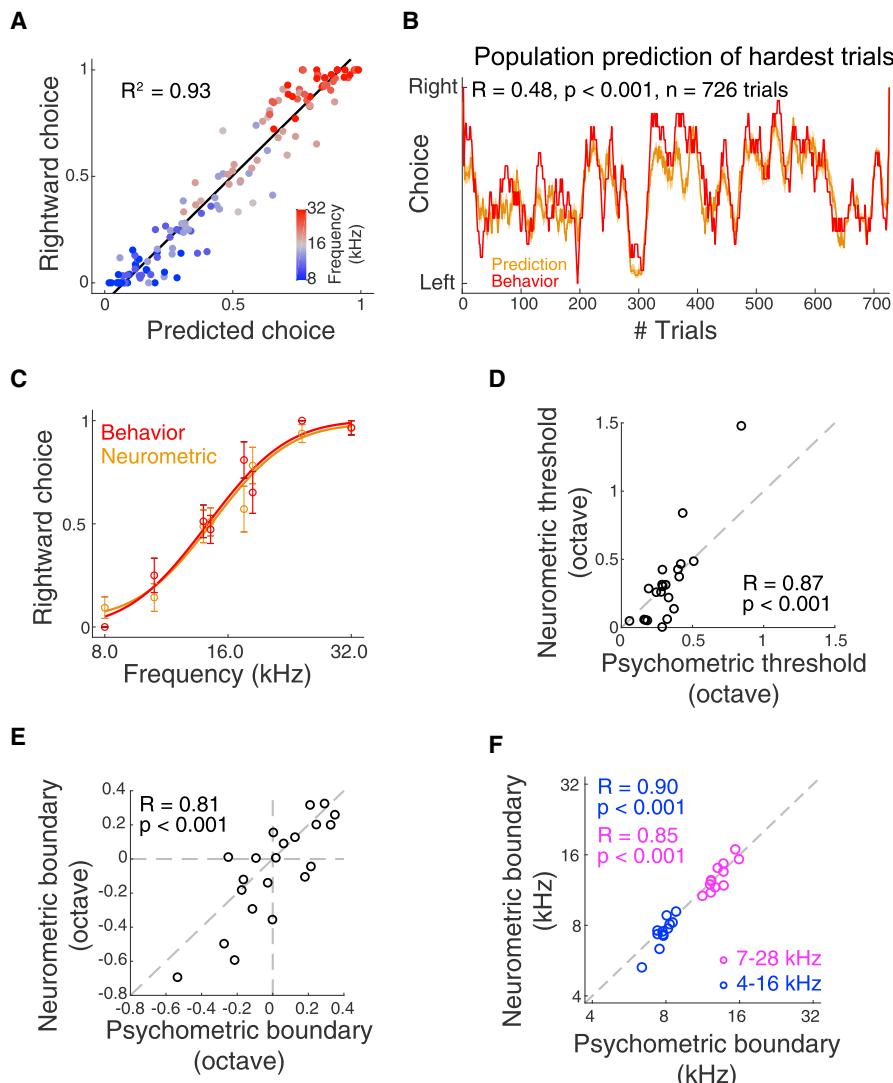
See also Figure S6.

population activity. We found that the probability of rightward choices upon different tone frequencies was strongly predicted by the classifier based on the population activity in the same sessions (Figure 8A). Furthermore, we examined whether population activity could predict the trial-by-trial varying choices (STAR Methods). We found that the predicted choices for the most difficult stimuli in each session were strongly correlated with the behavioral choices across individual trials (Figure 8B), indicating that the population activity can indeed predict trial-by-trial behavioral choices. To confirm that temporally the  $\text{Ca}^{2+}$  signals can indicate the contribution of population activity to perceptual choices, we used deconvolution on the calcium signals to recover spike rates and performed population decoding over the time course of each trial. We found that significant decoding occurred before the end of the 300-ms tone stimulus (Figure S6), suggesting that the population activity indicated by  $\text{Ca}^{2+}$  signals can indeed contribute to decisions.

As shown in Figure 7, we also constructed the neurometric function using the trial-by-trial decoded choices from simultaneously imaged population neurons. We found that the

population neurometric function also well captured the behavioral psychometric function (Figure 8C), with significant session-by-session correlation of the threshold (Figure 8D) and boundary (Figure 8E) between the neurometric function and psychometric function. Furthermore, for the experiments where we changed the category boundary (Figure 5), we also found that the neurometric boundaries based on population decoding in different frequency ranges were strongly correlated with the psychometric boundaries in the corresponding frequency ranges (Figure 8F). Thus, the behavioral choices during stimulus categorization can be accurately predicted based on simultaneously imaged population activity.

To confirm the robustness of the population decoding results, instead of directly decoding trial-by-trial choices, we used the linear classifier to find a hyperplane in the high-dimensional population activity space that best separates the left and right choices (Figure S7; STAR Methods). The average distance from the points in the activity space, which represent the population responses to a particular stimulus, to the discriminant hyperplane represents the probability of the population activity to report a stimulus as high- or low-frequency categories to an ideal observer (Kiani et al., 2014). We define this distance as the population decision variable (PDV). We found that the neurometric function constructed using PDVs also captured characteristics of the psychometric function from the same imaging session (Figure S7). Taken together, the activity of simultaneously imaged neuronal populations in the auditory cortex robustly encodes sufficient information to predict the stimulus categorization at the behavioral level.



**Figure 8. Population Decoding for Perceptual Decisions**

(A) Scatterplot showing behavioral rightward choice against the rightward choices predicted by population decoding (STAR Methods). Line represents a linear fit of all data.

(B) Prediction of single-trial choice using population activity. The binary choices were moving averaged for visualization (span, 20 trials).

(C) Neurometric function based on trial-by-trial population decoding and psychometric function averaged across sessions ( $n = 21$ ). Error bars indicate SEM.

(D) Correlation of threshold value for neurometric function and psychometric function. Student's t test;  $n = 21$  sessions.

(E) Correlation of the neurometric boundary and psychometric boundary. Student's t test;  $n = 21$ .

(F) Correlation of the neurometric boundary and psychometric boundary for 4–16 kHz sessions ( $n = 12$ ) and 7–28 kHz sessions ( $n = 12$ ). See also Figure S7.

## DISCUSSION

Decision-making, even in its simpler form of perceptual decisions, involves an aggregation of distinct subprocesses, including sensory detection, categorization, motor planning, and outcome evaluation. Among these, categorization is an essential step for making perceptual judgments linking sensory input and motor output. Despite a large body of work on categorization in perceptual and cognitive processes (Ashby and Maddox, 2005; Harnad, 1990; Zentall et al., 2008), few studies have

examined how neuronal circuits compute to transform high-dimensional and continuous physical stimuli into discrete categories. In the current study, we used a two-choice auditory psychophysics task to evaluate mice's ability to classify equal-spaced tone frequencies into high and low categories, and we used *in vivo* two-photon imaging to simultaneously record large populations of neurons in the auditory cortex. We found that in addition to conventional general modulation of neuronal responses by choice, reward, or motor planning (Francis et al., 2018; Kato et al., 2015; Kuchibhotla et al., 2017; Niell and

Stryker, 2010; Schneider et al., 2014; Yang et al., 2016), a stimulus categorization task dynamically reshapes both single-neuron response profiles and local cortical ensemble structures in a stimulus-specific manner, transforming sensory coding into category representations. First, we found some neurons show stronger selectivity to category membership compared to passive stimulation, similar to those observed in higher-order cortical areas (Freedman and Assad, 2006; Freedman et al., 2001), which may reflect a readout of categorization computation (Figure 3D). Second, a greater proportion of neurons show enhanced responses to stimuli near category boundaries during task performance, which may reflect a demand for increased discriminability for more ambiguous stimuli near boundaries during categorization (Figure 3H). At the population level, local cortical ensembles were dynamically reorganized to form augmented representations for near-boundary stimuli (Figure 4), which followed the flexible changes in behavioral level categorization criteria (Figure 5). The category-related modulation was also manifested in population discrimination of individual stimuli (Figure 6). Finally, the population dynamics of simultaneously imaged neurons (Figure 7) as well as the readout of population coding by an ideal observer (Figure 8) accurately captured behavioral categorical decisions, indicating that neuronal populations in the auditory cortex encode sufficient information for stimulus categorization.

It has been controversial whether the cerebral cortex is causally involved in various perceptual behaviors (Gimenez et al., 2015; Hong et al., 2018; O’Shea et al., 2015; Talwar et al., 2001). Due to the highly redundant brain circuits that are prone to plasticity changes, the requirement for a particular cortical region is likely to be strongly dependent on experimental conditions such as the animal species, type of tasks, the stimuli being used, training history, and methods of manipulation. Recent studies on neuronal activity in the auditory cortex showed that pharmacological or optogenetic inactivation impaired performance on auditory tasks (Jaramillo and Zador, 2011; Kato et al., 2015; Kuchibhotla et al., 2017; Talwar et al., 2001), while permanent lesions of the auditory thalamus, but not the auditory cortex, impaired an auditory discrimination task (Gimenez et al., 2015). It is therefore important to first assess whether the region of auditory cortex under examination contributes to the frequency categorization task. We used two independent methods of inactivation to address this question and found a consistent effect of impairment on task performance (Figures 1F and 1G).

There are various types of stimulus categorization with different levels of abstraction, some of which can be accompanied by strong perceptual effects known as categorical perception (Goldstone and Hendrickson, 2010; Harnad, 1990; Russ et al., 2007), while others can be more flexible and abstract (Ashby and Maddox, 1998; Rosch, 1978; Russ et al., 2007; Wutz et al., 2018; Zenall et al., 2002). Our behavioral task captures the basic feature of stimulus categorization, which is the ability to group stimuli or sensory items based on behavioral needs or task rules. The animals were able to perform categorization based on a short period of training with exemplar stimuli (Figures 1 and S1) and exhibited considerable flexibility by learning to categorize tone stimuli using different category boundaries (Figure 5). The neural representations exhibited dynamic reorganization that mirrored behavioral

flexibility. Thus, our results may have general implications for neural mechanisms of stimulus categorization.

Activity related to perceptual decisions in various sensory cortices has been widely reported in past decades (Britten et al., 1996; Celebrini and Newsome, 1994; Cohen and Newsome, 2009; Liu et al., 2013; Nienborg and Cumming, 2009; Nienborg et al., 2012; Romo et al., 2002; Tsunada et al., 2016; Yang et al., 2016). Although these studies provide important links between sensory cortex activity and choice behavior, a mechanistic understanding of how choice-related activity may contribute to specific subprocesses of decision-making (e.g., categorization) remains absent. A classical pooling model framework has been used to account for choice-related activity in the sensory cortex (Nienborg et al., 2012; Shadlen et al., 1996). It is, however, unclear whether and how task-related top-down information may play a role in this framework (Cumming and Nienborg, 2016; Nienborg et al., 2012). Our data show that the categorization task reshapes response profiles of both single neurons and neuronal ensembles, reflecting strong top-down influences. When such top-down modulation is incorporated into a conventional pooling model via stimulus-specific amplification (Raizada and Poldrack, 2007), the pooling sum from the two oppositely tuned populations may produce a steeper neurometric slope mirroring the behavioral level stimulus categorization. This is demonstrated by a simple model incorporating selective weighting by the response function based on the response selectivity to near-boundary stimulus into the pooling model (Figures S8). A similar idea has also been demonstrated previously (Jazayeri and Movshon, 2007). Thus, incorporating top-down modulation via selective amplification into conventional pooling model may represent a circuit level framework for understanding the computational mechanisms of stimulus categorization.

In summary, by combining an auditory psychophysics task in mice with *in vivo* two-photon imaging, we demonstrate that neuronal ensembles in the auditory cortex are markedly modulated by task demands, exhibiting dynamic reorganization of population response profiles that may serve as a mechanism for sensory-to-category transformation during perceptual decision-making. An important future direction would be to investigate the local and long-range circuit mechanisms for such task-dependent modulation.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Surgery and virus injection
  - Behavioral Apparatus
  - Animal behavior
  - Reversible silencing of the auditory cortex
  - Two-photon calcium imaging
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Imaging data analysis

- Neuronal response classification
- Factor analysis
- Population decoding
- Modeling selective amplification
- Statistical Analysis

## ● DATA AND CODE AVAILABILITY

### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.neuron.2019.06.004>.

### ACKNOWLEDGMENTS

We thank J. Anthony Movshon and Bruce Cumming for advice on population coding; Roian Egnor for advice on the auditory behavioral apparatus; Aaron Kerlin for communication on behavioral control software; and Yang Dan, Matthew McGinley, Yong Gu, Liping Wang, and Chengyu Li for valuable discussions on data analysis. This work was supported by the National Natural Science Foundation of China (grant 31571081); the National Key R&D Program of China (grant 2017YFA0103900/2017YFA0103901); the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB32010000); Key Research Program of Frontier Sciences, Chinese Academy of Sciences (grant QYZDB-SSW-SMC045); and the Youth Thousand Talents Program (to N.-I.X.).

### AUTHOR CONTRIBUTIONS

Y.X. and N.-I.X. conceived the project and designed the experiments. Y.X. performed all the experiments and data analysis. L.Z. performed the muscimol inactivation experiments. Y.Z. performed the DREADDs inactivation experiments. T.Z. contributed to behavioral experiments. L.Z., J.P., and N.-I.X. developed the hardware and software of the behavior and imaging system. Y.X. and N.-I.X. wrote the manuscript.

### DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: May 22, 2018

Revised: March 8, 2019

Accepted: June 10, 2019

Published: July 8, 2019

### REFERENCES

- Armbruster, B.N., Li, X., Pausch, M.H., Herlitze, S., and Roth, B.L. (2007). Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl. Acad. Sci. USA* **104**, 5163–5168.
- Ashby, F.G., and Maddox, W.T. (1998). Stimulus categorization. In *Measurement, Judgment and Decision Making*, M.H. Birnbaum, ed. (Academic Press), pp. 251–301.
- Ashby, F.G., and Maddox, W.T. (2005). Human category learning. *Annu. Rev. Psychol.* **56**, 149–178.
- Bandyopadhyay, S., Shamma, S.A., and Kanold, P.O. (2010). Dichotomy of functional organization in the mouse auditory cortex. *Nat. Neurosci.* **13**, 361–368.
- Bonnasse-Gahot, L., and Nadal, J.-P. (2008). Neural coding of categories: information efficiency and optimal population codes. *J. Comput. Neurosci.* **25**, 169–187.
- Briggman, K.L., Abarbanel, H.D.I., and Kristan, W.B., Jr. (2005). Optical imaging of neuronal populations during decision-making. *Science* **307**, 896–901.
- Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13**, 87–100.
- Carandini, M., and Churchland, A.K. (2013). Probing perceptual decisions in rodents. *Nat. Neurosci.* **16**, 824–831.
- Celebrini, S., and Newsome, W.T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J. Neurosci.* **14**, 4109–4124.
- Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., and Knight, R.T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* **13**, 1428–1432.
- Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300.
- Cohen, M.R., and Newsome, W.T. (2009). Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *J. Neurosci.* **29**, 6635–6648.
- Cumming, B.G., and Nienborg, H. (2016). Feedforward and feedback sources of choice probability in neural population responses. *Curr. Opin. Neurobiol.* **37**, 126–132.
- David, S.V., Fritz, J.B., and Shamma, S.A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc. Natl. Acad. Sci. USA* **109**, 2144–2149.
- Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227.
- Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37.
- Francis, N.A., Winkowski, D.E., Sheikhattar, A., Armengol, K., Babadi, B., and Kanold, P.O. (2018). Small networks encode decision-making in primary auditory cortex. *Neuron* **97**, 885–897.e6.
- Freedman, D.J., and Assad, J.A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**, 85–88.
- Freedman, D.J., and Assad, J.A. (2011). A proposed common neural mechanism for categorization and perceptual decisions. *Nat. Neurosci.* **14**, 143–146.
- Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316.
- Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* **6**, 1216–1223.
- Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., Cohen, L., and Naccache, L. (2009). Converging intracranial markers of conscious access. *PLoS Biol.* **7**, e61.
- Gimenez, T.L., Lorenc, M., and Jaramillo, S. (2015). Adaptive categorization of sound frequency does not require the auditory cortex in rats. *J. Neurophysiol.* **114**, 1137–1145.
- Goldstone, R.L., and Hendrickson, A.T. (2010). Categorical perception. Wiley Interdiscip. Rev. Cogn. Sci. **1**, 69–78.
- Grinband, J., Hirsch, J., and Ferrera, V.P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron* **49**, 757–763.
- Guenther, F.H., Nieto-Castañon, A., Ghosh, S.S., and Tourville, J.A. (2004). Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear. Res.* **47**, 46–57.
- Guo, Z.V., Hires, S.A., Li, N., O'Connor, D.H., Komiyama, T., Ophir, E., Huber, D., Bonardi, C., Morandell, K., Gutinsky, D., et al. (2014). Procedures for behavioral experiments in head-fixed mice. *PLoS ONE* **9**, e88678.
- Harnad, S.R. (1990). *Categorical Perception: The Groundwork of Cognition* (Cambridge University Press).
- Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68.
- Heffner, H.E., and Heffner, R.S. (2007). Hearing ranges of laboratory animals. *J. Am. Assoc. Lab. Anim. Sci.* **46**, 20–22.
- Hong, Y.K., Lacefield, C.O., Rodgers, C.C., and Bruno, R.M. (2018). Sensation, movement and learning in the absence of barrel cortex. *Nature* **561**, 542–546.

- Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866.
- Issa, J.B., Haeffele, B.D., Agarwal, A., Bergles, D.E., Young, E.D., and Yue, D.T. (2014). Multiscale optical Ca<sup>2+</sup> imaging of tonal organization in mouse auditory cortex. *Neuron* 83, 944–959.
- Jaramillo, S., and Zador, A.M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* 14, 246–251.
- Jaramillo, S., and Zador, A.M. (2014). Mice and rats achieve similar levels of performance in an adaptive decision-making task. *Front. Syst. Neurosci.* 8, 173.
- Jazayeri, M., and Movshon, J.A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature* 446, 912–915.
- Kato, H.K., Gillet, S.N., and Isaacson, J.S. (2015). Flexible sensory representations in auditory cortex driven by behavioral relevance. *Neuron* 88, 1027–1039.
- Kiani, R., Cueva, C.J., Reppas, J.B., and Newsome, W.T. (2014). Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Curr. Biol.* 24, 1542–1547.
- King, A.J., and Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat. Neurosci.* 12, 698–701.
- Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., and Froemke, R.C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. *Nat. Neurosci.* 20, 62–71.
- Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368.
- Liu, S., Gu, Y., DeAngelis, G.C., and Angelaki, D.E. (2013). Choice-related activity and correlated noise in subcortical vestibular neurons. *Nat. Neurosci.* 16, 89–97.
- Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480.
- Nelson, A., Schneider, D.M., Takatoh, J., Sakurai, K., Wang, F., and Mooney, R. (2013). A circuit for motor cortical modulation of auditory cortical activity. *J. Neurosci.* 33, 14342–14353.
- Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. *Nature* 341, 52–54.
- Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65, 472–479.
- Nienborg, H., and Cumming, B.G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* 459, 89–92.
- Nienborg, H., Cohen, M.R., and Cumming, B.G. (2012). Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu. Rev. Neurosci.* 35, 463–483.
- Otazu, G.H., Tai, L.-H., Yang, Y., and Zador, A.M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* 12, 646–654.
- Otchy, T.M., Wolff, S.B.E., Rhee, J.Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S.M.H., and Ölveczky, B.P. (2015). Acute off-target effects of neural circuit manipulations. *Nature* 528, 358–363, advance online publication.
- Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* 21, 227–277.
- Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* 2, 13.
- Prather, J.F., Nowicki, S., Anderson, R.C., Peters, S., and Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* 12, 221–228.
- Raizada, R.D.S., and Poldrack, R.A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56, 726–740.
- Romo, R., Hernández, A., Zainos, A., Lemus, L., and Brody, C.D. (2002). Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat. Neurosci.* 5, 1217–1225.
- Rosch, E. (1978). Principles of categorization. In *Cognition and Categorization*, E. Rosch and B.B. Lloyd, eds. (L. Erlbaum Associates), pp. 28–49.
- Rothschild, G., Nelken, I., and Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nat. Neurosci.* 13, 353–360.
- Russ, B.E., Lee, Y.-S., and Cohen, Y.E. (2007). Neural and behavioral correlates of auditory categorization. *Hear. Res.* 229, 204–212.
- Schneider, D.M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* 513, 189–194.
- Schreiner, C.E., and Winer, J.A. (2007). Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* 56, 356–365.
- Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. *Annu. Rev. Neurosci.* 33, 203–219.
- Shadlen, M.N., Britten, K.H., Newsome, W.T., and Movshon, J.A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* 16, 1486–1510.
- Talwar, S.K., Musial, P.G., and Gerstein, G.L. (2001). Role of mammalian auditory cortex in the perception of elementary sound properties. *J. Neurophysiol.* 85, 2350–2358.
- Tsunada, J., Lee, J.H., and Cohen, Y.E. (2011). Representation of speech categories in the primate auditory cortex. *J. Neurophysiol.* 105, 2634–2646.
- Tsunada, J., Liu, A.S.K., Gold, J.I., and Cohen, Y.E. (2016). Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat. Neurosci.* 19, 135–142.
- Uchida, N., and Mainen, Z.F. (2003). Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.* 6, 1224–1229.
- Vogelstein, J.T., Packer, A.M., Machado, T.A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* 104, 3691–3704.
- Wichmann, F.A., and Hill, N.J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313.
- Wutz, A., Loonis, R., Roy, J.E., Donoghue, J.A., and Miller, E.K. (2018). Different levels of category abstraction by different dynamics in different pre-frontal areas. *Neuron* 97, 716–726.e8.
- Wytenbach, R.A., May, M.L., and Hoy, R.R. (1996). Categorical perception of sound frequency by crickets. *Science* 273, 1542–1544.
- Xu, N.L., Harnett, M.T., Williams, S.R., Huber, D., O'Connor, D.H., Svoboda, K., and Magee, J.C. (2012). Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* 492, 247–251.
- Yang, H., Kwon, S.E., Severson, K.S., and O'Connor, D.H. (2016). Origins of choice-related activity in mouse somatosensory cortex. *Nat. Neurosci.* 19, 127–134.
- Zentall, T.R., Galizio, M., and Critchfield, T.S. (2002). Categorization, concept learning, and behavior analysis: an introduction. *J. Exp. Anal. Behav.* 78, 237–248.
- Zentall, T.R., Wasserman, E.A., Lazareva, O.F., Thompson, R.K.R., and Rattermann, M.J. (2008). Concept learning in animals. *Comp. Cogn. Behav. Rev.* 3, 13–45.
- Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665.
- Zingg, B., Hintiryan, H., Gou, L., Song, M.Y., Bay, M., Bienkowski, M.S., Foster, N.N., Yamashita, S., Bowman, I., Toga, A.W., and Dong, H.W. (2014). Neural networks of the mouse neocortex. *Cell* 156, 1096–1111.
- Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* 497, 482–485.

**STAR★METHODS****KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
AAV2/9-hSyn-Gcamp6s	Taitool Bioscience (Shanghai)	Cat# S0225
AAV2/9-hSyn-DIO-hM4D(Gi)-mCherry	Taitool Bioscience (Shanghai)	Cat# S0193
Chemicals, Peptides, and Recombinant Proteins		
Muscimol	Sigma	Cat# M1523-10MG
Vybrant Dil Cell-Labeling Solution	ThermoFisher scientific	Cat# V22885
Clozapine N-oxide	Sigma	Cat# C0832-5MG
Deposited Data		
All data	The current study	<a href="https://doi.org/10.17632/bvx47zg5mt.1">https://doi.org/10.17632/bvx47zg5mt.1</a>
Experimental Models: Cell Lines		
Mouse: C57BL/6J	SLAC laboratory	N/A
Mouse: Emx1-IRES-Cre	Jackson Laboratory	Cat# 005628
Software and Algorithms		
MATLAB	MathWorks	<a href="https://www.mathworks.com/products/matlab.html?stid=hp_products_matlab">https://www.mathworks.com/products/matlab.html?stid=hp_products_matlab</a> ; RRID: SCR_001622
Deconvolution algorithm	<a href="#">Vogelstein et al., 2010</a>	<a href="https://github.com/jovo/oopsi">https://github.com/jovo/oopsi</a>
Scanimage	Vidrio Technologies	<a href="http://scanimage.vidriotechnologies.com/display/SIH/Scanimage">http://scanimage.vidriotechnologies.com/display/SIH/Scanimage</a>

**LEAD CONTACT AND MATERIALS AVAILABILITY**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, N.L.X. ([xunl@ion.ac.cn](mailto:xunl@ion.ac.cn)). This study did not generate new unique reagents.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

All experimental aspects were carried out in compliance with the procedures approved by the Animal Care and Use Committee of the Institute of Neuroscience, Chinese Academy of Sciences. Data were mainly acquired from male C57BL/6J (SLAC), some Emx1-IRES-Cre mice (Jax number: 005628) were also used for chemogenetic inactivation experiments. Animal age was 8–10 weeks at the start of behavioral training and 10–14 weeks during imaging. Mice had no previous history of any other experiments. On days not trained, mice received 1 mL of water. On training days, mice were tested in experimental sessions lasting 1 to 2 hours where they received all their water (0.5 to 1 ml). Each mouse's weight was measured daily to ensure that it was not below 80% of the mouse's pre-water restriction weight.

**METHOD DETAILS****Surgery and virus injection**

During surgery, mice were anaesthetized with isoflurane (1~2%). For chronic imaging window implantation, a craniotomy (~2 mm in diameter) was made over the left auditory cortex, with the dura left intact. AAV-hSyn-GCaMP6s virus was slowly injected (~50–150 nL per site, 3 - 4 injection sites per animal). The injection system comprised of a pulled glass pipette (25–30  $\mu$ m O.D.; Drummond Scientific, Wiretrol II Capillary Microdispenser) back-filled with mineral oil. A fitted plunger was inserted into the pipette and advanced to displace the contents using a hydraulic manipulator (Narashige, MO-10). Retraction of the plunger was used to load the pipette with virus solution. The injection pipette was positioned with a Sutter MP-225 manipulator. After injection, the craniotomy was covered with a double-layered glass coverslip, sealed in place with dental cement (Jet Repair Acrylic, Lang Dental Manufacturing). The double-layered glass comprised a 200- $\mu$ m-thick glass coverslip (diameter, ~2 mm) attached to a larger glass coverslip (diameter, 5 mm, 0.22 mm thickness) using ultraviolet cured optical adhesives. A titanium head-post with an opening on the left side was attached to

the skull with cyanoacrylate glue and dental cement to permit head fixation and two-photon imaging over the cranial window. Mice were allowed to recover for at least 7 days before water restriction.

After all imaging sessions were finished, the imaging location was verified by Dil injection (Figure S2B). The chronic window was removed and the dura was kept intact, the imaging field was estimated by the vasculature location. Then the Dil (10–20 nl) was injected at the center of all imaging fields (2–3 fields for each animal). After surgery the animal was perfused immediately using 4% paraformaldehyde and frozen sliced (~40 µm) to confirm the two-photon imaging position.

### Behavioral Apparatus

Experiments were conducted inside custom-designed and fabricated double-walled sound-attenuating boxes. Mice were head-fixed with a pair of clamps and thumbscrews (Guo et al., 2014; Xu et al., 2012). The mouse body was contained in an acrylic ‘body tube’ (1 inch i.d.), with the mouse head extending out and the front paws gripping the tube edge after head-fixation. The holder and body tube in turn were attached to a custom-designed tube holder mounted to a caddy which was mounted to the behavior box with table clamps after head-fixation. Water reward was provided by two custom-made metal lickports placed in front of the mouse mouth (bilateral sides alongside animal mouth). The lickports were connected to a capacitive-sensing board that sense the contact of the tongue during licking. The amount of water delivered was controlled by valves open time calibrated at least once a week. The fluid pressure was similar for each behavioral session.

The mouse auditory decision behavior was controlled by a custom-developed low-cost, high-precision real-time control system, the PX-Behavior System. The system includes a custom-designed tone-generating module (TGM) to generate sound waveforms of arbitrarily high frequencies with high fidelity, and an Arduino microcontroller implementing a real-time virtual state machine framework for programming the behavioral protocols, stimulus delivery, and measurement of behavioral events. The Arduino board communicates with the TGM through a custom-designed cache board without interrupting the flow of the behavioral protocols. The TGM sends the specified sound waveforms to an amplifier and speakers to produce acoustic stimuli. Behavioral data were logged via serial port using custom-written software in python. Behavioral trials were synchronized with two-photon image acquisition by digital outputs from the PX-Behavior System to the ScanImage system.

Sounds were delivered through electrostatic speakers (ES1, Tucker-Davis Technologies) placed on the right side of the mice. The sound system was calibrated using a free-field microphone (Type 4939, Brüel and Kjær) over 3–60 kHz and showed a smooth spectrum ( $\pm 5$  dB SPL). Measurements were performed with the behavioral box closed and the microphone positioned at the location and orientation of the mouse ear-position in the presence of the mouse mounting system. The microphone was connected to a preamplifier (Type 2670, Brüel and Kjær), and signals were digitized with a National Instruments acquisition card (NI 9201) at 500,000 samples per second for further analysis, 5 ms cosine ramps are applied to the rise and fall of all tones.

### Animal behavior

After one week’s recovery following surgery, mice were started with water restriction procedure. Each mouse received 1 mL water per day and the body weight was monitored. After ~7 days of water restriction, behavior training was started. Water consumption was calculated using the body weight change before and after each training session. If mice consumed less than 0.5 mL water, additional water supplement was provided. Mice were allowed to perform the task until sated.

The behavioral task is based on the auditory-guided two-alternative-forced-choice task (Guo et al., 2014; Uchida and Mainen, 2003; Znamenskiy and Zador, 2013). During the initial training stage, mice were trained to discriminate two easy tone stimuli that were 2 octaves apart (e.g., 8 kHz and 32 kHz). Tone frequencies were chosen according to the typical hearing range of laboratory mouse (Heffner and Heffner, 2007). The initiation of each trial is not explicitly cued, and the animal needs to wait for the tone stimulus to occur. Following the inter-trial interval of the previous trial, a 0.5–1 s random delay was imposed before tone stimuli (duration, 300 ms) to ensure that the onset of tone stimuli cannot be predicted by animal. Mice were required to respond within a 3 s answer period after a 500 ms post-stimulus-offset delay (Shading in Figure 1A) following the stimulus offset, by licking left or right lickport placed in front of the animal. During the post-stimulus delay, licking will not produce any consequences. Correct answers were defined as licking the left lickport in response to the lower frequency tone (e.g., 8 kHz), or licking the right lickport in response to the higher frequency tone (e.g., 32 kHz). Correct responses lead to the water valve open to dispense a small amount of water reward (~6ul). Error responses lead to a 2~6 s time-out punishment, during which licking to the wrong side would reinitiate the time-out period. If mice made no response lick within the 3 s response window, the trial was defined as a ‘miss’ trial, leading to the initiation of inter-trial-interval (ITI). Once mice reached criteria of > 85% correct performance, the second training stage was started for testing categorization performance. In this stage, 6–8 different frequencies (from 8 to 32 kHz, separated by linearly spaced intervals in octave) were delivered in randomly interleaved trials. For example, since the difference between 8 and 32 kHz in octave is 2, for 6 different tones, the difference in octave between neighboring tones is 0.4, therefore the 6 tone frequencies are 8000 Hz, 10556 Hz, 13929 Hz, 18379 Hz, 24252 Hz, and 32000 Hz. In some behavioral sessions, 8 stimuli were used, with 2 additional more difficult tones located 0.1 octave from the defined boundary on either side (8000 Hz, 10556 Hz, 13929 Hz, 14929 kHz, 17148 kHz, 18379 Hz, 24252 Hz, and 32000 Hz). The sound intensity was 70 or 75 dB SPL for all frequencies. Reward contingency was based on the mid frequency (16 kHz) as the boundary, with correct answers defined as licking leftward when tone frequencies were lower than the boundary and licking rightward when the tone frequencies were higher than the boundary. Imaging experiments were started after 2–5 days of training in the second training stage.

In a separate group of animals, we tested mice's internal categorization boundary after the initial training with the 8 and 32 kHz exemplar tones, the intermediate frequencies were delivered in 'probe' trials, constituting a minor fraction of total trials (< 30%), and were randomly rewarded with no punishment (Figures 1 and S1).

Psychometric function was obtained by fitting the behavioral data using a 4-parameter sigmoidal (logistic) function (Carandini and Churchland, 2013; Wichmann and Hill, 2001)

$$y(x) = g + (1 - g - l) * 0.5 * (1 + \text{erf}((x - u)/\sqrt{2 * v^2}))$$

where  $y(x)$  is the probability that animal would make a right choice, and  $x$  is the tone frequency (in octave). Parameters to be fitted are:  $g$  the guess rate,  $l$  the lapse rate,  $u$  the subject bias (boundary), and  $v$  the discrimination sensitivity (threshold).  $\text{erf}()$  represents error function. Using this function, parameter  $u$  indicates the psychometric boundary, and parameter  $v$  indicates the psychometric threshold.

For experiments with shifted category boundaries, mice were first trained with 8 kHz and 32 kHz to learn 2AFC task. The animals were then trained with 4 kHz and 16 kHz tones to perform categorization task in 4 – 16 kHz frequency range with reward contingencies based on the boundary at 8 kHz. The same animals were then further trained to perform categorization task in the 7 – 28 kHz frequency range based on the boundary at 14 kHz. Imaging experiments were carried out after animal performed the categorization task with > 75% correct.

### Reversible silencing of the auditory cortex

Two parallel approaches were used to reversibly silence the auditory cortex to test its involvement in our auditory decision task. First, we used the GABA (c-aminobutyric acid) agonist muscimol hydrobromide (Sigma-Aldrich) dissolved in saline (5 mg/ml). 60nl of the solution was injected slowly (10nl per min) using a thin glass pipette to bilateral auditory cortex (4.8 mm lateral to midline and 2.6 mm anterior to Bregma, depth ~500 um under the pia). As control, the same dose of muscimol was injected to primary visual cortex bilaterally in different sessions in the same animals (Figure S1). The animals were left to recover for 1 hour following muscimol injection before the start of behavioral sessions. Second, we used a chemogenetic approach based on DREADDs to reversibly silence excitatory neurons in the auditory cortex. AAV2/9-hSyn-DIO-hM4D(Gi)-mCherry ( $1.72 \times 10^{12}$  genomic copies per milliliter) was injected to bilateral auditory cortex of Emx1- IRES-Cre mice, followed by 3 weeks virus expression. Clozapine-N-Oxide (CNO) was dissolved in saline (0.9% NaCl solution) to a stocking solution of 20mg/mL stored at  $-20^{\circ}\text{C}$ , and then diluted to a working concentration of 0.2 mg/mL each day before experiments. At the day of experiments, saline or CNO (2 mg/kg b.w.) was administered intraperitoneally (i.p.) to the mice expressing hM4D(Gi) in the auditory cortex 30~40 min before behavioral session. For CNO control experiments, animals without hM4D(Gi) virus injection received the same dose of CNO treatment before behavior session.

### Two-photon calcium imaging

Calcium imaging was performed using a custom built two-photon microscope (<https://wiki.janelia.org/wiki/display/shareddesigns/MIMMS>). To eliminate potential influence from ambient noise of the two-photon imaging system and the noise of the laser scanning system on the auditory guided behavior and auditory evoked activity, the entire microscope was enclosed in a custom-designed and fabricated, double-walled sound attenuation box (internal noise level < 30 dB SPL with the two-photon imaging system running). The noise produced by the resonant scanner were attenuated to < 30 dB SPL using an optical window sealing the output opening of the resonant scanner module. GCaMP6s was excited using a Ti-Sapphire laser (Chameleon Ultra II, Coherent) tuned to 925 nm. Images were acquired using a 16x 0.8 NA objective (Nikon), and the GCaMP6s fluorescence was isolated using a bandpass filter (525/50, Semrock), and detected using GaAsP photomultiplier tubes (10770PB-40, Hamamatsu). Horizontal scanning was accomplished using a resonant galvanometer (Thorlabs; 16 kHz line rate, bidirectional). The average power for imaging was ~70 mW, measured at the entrance pupil of the objective. The field of view was  $\sim 300 \times 300$  imaged at high frame rate (~55 Hz for  $256 \times 256$  pixels, or ~28 Hz for  $512 \times 512$  pixels). The acquisition system was controlled using ScanImage (<https://scanimage.org>) (Pologruto et al., 2003). For each mouse the optical axis was adjusted (45-50 degree from vertical) to be perpendicular to the imaging window in the auditory cortex. Different fields-of-view in the same mouse were imaged on different sessions. At the initiation of each trial, ScanImage acquisition was triggered by a starting signal from the PX-Behavior System and a fixed number of frames (500 frames at 55 Hz frame rate, or 280 frames at 28 Hz frame rate) covering the entire behavioral trial were recorded. After completion of one behavioral session, we imaged a passive stimulation session for the same imaging field. During passive session, lick ports were moved away from the animal and the same set of sound stimulus were delivered in pseudo-randomly interleaved trials (Figure S2A).

For chronic imaging experiments, to image the same population of neurons across sessions/days, the supporting platform position was fixed and the optical axis was adjusted to the same angle for each animal across days. The same field of views were tracked by vascular features and the spatial location of landmarks (bright and stable structures of somas and neuropils) in the field of view. The presence of each individual neuron within the same field of view across days was visually checked. Only neurons showing clear and similar cell body morphology across sessions were included for analysis.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Imaging data analysis

Since the acquisition rate of image frames was high, we do not find detectable intra-frame movement. We thus performed frame by frame registration to correct brain motion. All imaging frames from each imaging/behavior session were aligned to a target image frame using a cross-correlation-based registration algorithm (discrete Fourier transformation, DFT, algorithm). The target image was obtained by mean projection of visually identified frames ( $> 30$  frames) with few motion artifacts. To extract fluorescence signals from individual neurons, regions of interest (ROIs) were drawn manually based on neuronal shape using a custom-written GUI software in MATLAB. The pixels within each ROI in a frame were averaged as the fluorescence intensity of a neuron at that frame time. The fluorescence intensity of each neuron was then extracted over all the frames as the fluorescence time series ( $F$ ). For each trial, averaged fluorescence value before sound onset was used as  $F_0$  for that trial, and  $\Delta F/F_0$  was calculated as  $(F - F_0)/F_0 \times 100\%$ . The  $\Delta F/F_0$  traces were used for all following analysis. Miss trials were not included for all analysis.

To focus on behavioral modulation of sound evoked responses, we only used the mean  $\text{Ca}^{2+}$  signals within a 1 s time window following stimulus onset to avoid the confounding effects from uncontrolled behavioral variables in late trial epochs (Figure 2G and 2H). The sound evoked responses were calculated for each frequency, and then single neuron best frequency (BF) was defined as the frequency that could elicit maximum response. During task and passive condition, single neuron could have different frequency preference. So here defined BF<sub>pass</sub> as the BF measured under passive condition, and BF<sub>task</sub> as the BF measured under task condition, respectively. The population preferred frequency is the dominant BF (Figure 2E) for specific imaging sessions.

To estimate the spike rates from the calcium signals, we used the Fast Nonnegative Deconvolution algorithm (Vogelstein et al., 2010) to deconvolve the fluorescence traces. Estimated firing rates for the initial four frames were manually set to zeros to avoid artificial fluctuations (Figure S6).

### Neuronal response classification

To classify single neuron response type, we fit single neuron activity to a sigmoidal function (Figure S3A) as:

$$SR(x) = b_1 + \frac{b_2}{1 + e^{-\frac{(x - b_3)}{b_4}}}$$

and a Gaussian function (Figure S3C)

$$GR(x) = c_1 e^{-\frac{(x - c_2)^2}{2*c_3^2}} + c_4$$

To evaluate the goodness of fitting, we calculated an error function as:

$$E = \sum_x (R_{Real}(x) - R_{Fit}(x))^2 / \sum_x (R_{Real}(x))^2$$

in which  $R_{Real}(x)$  and  $R_{Fit}(x)$  was the real response value and fit value to frequency  $x$ . We set a threshold of 0.1, such that when the fitting error was below the threshold, the fit was accepted and the neuron was classified as putative category-selective (Figure S3A) or frequency-tuned neurons (Figure S3C). For a ‘frequency-tuned’ neuron, if the BF during task was less than 0.2 octave from behavior boundary, it was considered as selective to near-boundary stimuli.

### Factor analysis

Factor analysis was performed using MATLAB Statistics and Machine Learning Toolbox. 15 common factors were used in our analysis and the first three factors were plotted for visualization (Figure 7A). For each imaging session, the  $\Delta F/F$  trace for individual trial was first smoothed using a moving average method (30 frames window). Then single session data was reshaped into a T (Number of frames acquired in a session) by N (number of neurons) matrix. After projecting the raw data into state space (each dimension is one common factor), spatial-temporal patterns of two categories in the state space was defined as the averaged trajectories of all correct left choice or right choice trials. Then for each trial, we evaluated its similarity with the low- and high-frequency category trajectories by calculating a selectivity index at each time points. The selectivity index was calculated based on the Euclidean distances ( $D$ ) between single trial trajectory and category trajectories at each time points.  $SI = (D_{\text{to low-freq category traj}} - D_{\text{to high-freq category traj}}) / (D_{\text{to low-freq category traj}} + D_{\text{to high-freq category traj}})$ .  $SI = 1$  indicates that current trajectory was overlapped with the trajectory indicates high-frequency category, while  $SI = -1$  indicates that current trajectory was overlapped with the trajectory of low-frequency category. Generally speaking, a negative selectivity index value means that the distance to low-frequency category trajectory is less than the distance to high-frequency category trajectory, indicating current population activity is more likely to represent a low-frequency perception, and vice versa.

We further generated a neurometric function based on the selectivity index. For each frequency, we first averaged the SI traces across all trial repeats. Then for each averaged trace, the peak or vale value within 1 s time window after stimulus onset was used as the selectivity index for corresponded stimuli. The neurometric function was then constructed from the normalized SI (scaled by the behavioral rightward choice probability) as a function of frequency, by fitting with the same sigmoidal function used for psychometric function.

To compare categorization characteristics independent of curve fitting, we calculated a categorization index (CI, [Figure 7F](#)). The CI was defined as the difference between averaged rightward probability for stimuli in the high-frequency category and that for stimuli in the low-frequency category, such that a zeros CI value indicates a flat selectivity independent of stimulus; while a CI value close to 1 indicates a near perfect categorization. The neurometric CI was calculated using the normalized SI values.

### **Population decoding**

We also used support vector machine (SVM) based algorithm for population decoding analysis, linear kernel and  $C = 1$  was used for all SVM classifiers. The minimum population size of all 21 sessions have 69 neurons. We first evaluated population discrimination of paired stimulus. For two stimulus  $S_1$  and  $S_2$ , single neuron response from individual trials were arranged as two matrix  $M_1 \times N$  and  $M_2 \times N$ , whereas  $M_1$  and  $M_2$  indicates the repeated number for  $S_1$  and  $S_2$  respectively,  $N$  indicates the neuron number. Elements within matrix were neuron response in individual trials, which was the averaged response within 1 s time window after stimuli onset. A binary classifier was then trained to classify the two dataset. Classification accuracy was calculated using 10-folds cross-validation method. When comparing the decoding accuracy for between-category stimulus pairs versus within-category stimulus pairs, to control for the factor that greater frequency difference in certain between-category stimulus pairs may contribute to higher decoding accuracies, we only compared the decoding accuracy for pairs with the frequency differences (octave steps) co-exists for between-category or within-category pairs. In our case, only tone pairs that were 0.4 or 0.8 octave different were included for comparison. For example, the decoding accuracy for a between-category pair, 13.9 kHz and 24.3 kHz (distance, 0.8 octave), is included since there are within-category pairs with the same octave distance (such as 8 kHz and 13.9 kHz, 0.8 octave distance). In contrary, the decoding accuracy for a between-category pair, 8 kHz and 18.4 kHz (distance, 1.2 octave), was not included since no such frequency difference was existed in within-category pair.

To predict single trial choice using population neuron activity, session data was arranged into a  $M \times N$  matrix, whereas  $M$  indicates the total trial number within a session. Population prediction of single trial choice was calculated using leave-one-out method, so that each trial could be predicted once. Since we have a complete prediction of all individual trials, we could also generate a neurometric function using predicted choices. In order to calculate the trial-by-trial correlation of animal performed choice and population predicted choice, we considered only the worst performed stimuli within each session. The performance of this stimuli was around chance level ( $0.53 \pm 0.16$ , one-sample t test with the mean value of 0.5,  $p = 0.44$ ), so we could have largest choice variability for the same stimuli. And by using single stimuli for analysis, we could minimum the choice bias caused by frequency difference.

We also constructed neurometric function using population decision variables ([Figure S7](#)). Population activity can be represented by high-dimensional data points, with the number of dimensions equals the number of neurons. We first trained an SVM classifier to find the optimal hyperplane to discriminate patterns corresponded to two choices. To produce a training dataset, a small subset of trials (~30) were sampled from all trials. Then trials with the same behavior choice were averaged, so we could have two high-dimensional data points corresponded to two choices. This subsampling was repeated 300 times. This dataset was then used to train a SVM classifier to find the classification hyperplane that separates two choice clusters. Then for each frequency, population response of were averaged across trials, and this averaged population activity was used as the population response pattern for that frequency (single point in the neuronal space). We then defined the population decision variable (PDV), which was the Euclidean distance of the population responses vector of each stimulus to the discriminant hyperplane ([Figure S7A](#)). In order to discriminate choice preference for each stimulus, the PDVs were given a negative sign for those points being classified as left choice category by the trained classifier. PDVs was used to represent the likelihood of each frequency to be classified as high-frequency or low-frequency category as shown in [Figure S7B](#) (Kiani et al., 2014). To generate a neurometric function, the PDVs were normalized to the minimum and maximum values of behavioral rightward probability of all stimulus ([Figures S7C](#) and [S7D](#)). The neurometric function was generated by fitting the normalized PDVs using the same sigmoidal function used to produce psychometric function.

For analysis of population decoding across time using estimated spike rate ([Figure S6](#)), classifiers were trained and tested using data in 100 ms time bins. For each time bin, we randomly selected 80% of trials to train an SVM classifier, and then use the trained classifier to predict the rest 20% of trials. This process was repeated 1000 times and the averaged prediction accuracy was used as the population decoding accuracy for each time bin.

### **Modeling selective amplification**

To understand the potential neuronal computation mechanisms underlying stimulus categorization, we incorporated a selective-amplification ([Jazayeri and Movshon, 2007](#); [Raizada and Poldrack, 2007](#)) with a modified pooling model ([Figure S8](#)). According to pooling model, we consider two neuron pools with preferred frequencies lie either in lower frequency range below the boundary or in higher frequency range above the boundary. The difference between the summation of responses of the two pools was

used as the population readout, representing the population decision (Nienborg et al., 2012; Shadlen et al., 1996). We then examined the categorization performance with or without applying selective amplification to the pooling summation difference. The tuning function of individual neuron within the pools was defined as:

$$f_{prefer}(x) = a * \exp\left(-\frac{(x - \mu)^2}{2 * \sigma^2}\right) + d$$

where the four coefficients are:  $a$ , tuning amplitude;  $d$ , baseline activity level;  $\mu$ , preferred frequency;  $\sigma$ , tuning width. For simulation, the preferred frequencies (between 8 kHz and 32kHz) were set to be evenly distributed in the neuron pools. The amplitude  $a$  was set to 1 and  $d$  was set to zeros. Neurons with tuning frequency lower than task defined boundary (16 kHz) were assigned to low category pool, and neurons with higher tuning frequency were assigned to high category pool. A neurometric decision can be defined as the difference in the pooling sum:

$$C(x) = \frac{1}{N} \sum_1^N f_R(x) - \frac{1}{N} \sum_1^N f_L(x)$$

$C(x)$  was the choice function, which was the pooling sum difference as a function of frequencies (choice function). A negative value represents a decision for low-frequency category; a positive value represents a decision for high-frequency category.

We use a Gaussian function  $G(x)$  to represent enhanced responses to near-boundary stimuli, by fitting to the data of boundary-selective neurons (Figure 3H),

$$G(x) = a_m * \exp\left(-\frac{(x - \mu_m)^2}{2 * \sigma_m^2}\right) + d_m$$

where  $a_m = 2$  was the modulation amplitude,  $\mu_m = 0$  was the category boundary,  $\sigma_m = 0.4$  was the modulation width (the octave step we have been used in current study), and  $d_m = 1$  was the modulation baseline when only considering facilitation effect. We applied the selective amplification to the choice function by:

$$P(x) = C(x) * G(x)$$

where  $P(x)$  represents the final choice function after selective amplification, which exhibit a steeper categorization than the  $C(x)$  function.

### Statistical Analysis

Statistical analysis was performed using MATLAB 2016a (MathWorks). All data are shown as mean  $\pm$  s.e.m. unless mentioned otherwise. The Comparison significance was tested using Student's t test, two-sample t test, paired-sample t test, Wilcoxon signed-rank test. N.S.,  $p > 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

### DATA AND CODE AVAILABILITY

The data generated in this study to reproduce all the results have been deposited in the Mendeley Data (<https://doi.org/10.17632/bvx47zg5mt.1>).