Article

Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex

Highlights

- Enhanced categorical representations in higher auditory areas during behavior

- Distinct sensory and categorical information flows during categorization tasks

- Separate neuronal groups with intrinsic versus task-induced categorical responses

- Similar dynamics in different categorization tasks with distinct acoustic features

Authors

Pingbo Yin, Dana L. Strait,
Susanne Radtke-Schuller,
Jonathan B. Fritz, Shihab A. Shamma

Correspondence

pyin@umd.edu

In Brief

Yin et al. show neuronal responses become progressively more categorical in higher auditory and frontal cortical areas during passive listening in trained ferrets, and more so during engagement in categorization tasks. The dynamics of categorical responses exhibit a cascading top-down modulation pattern opposite to the rapid bottom-up sensory flow.

CellPress

Article

CellPress

Dynamics and Hierarchical Encoding of Non-compact Acoustic Categories in Auditory and Frontal Cortex

Pingbo Yin,[1,4,*] Dana L. Strait,[1] Susanne Radtke-Schuller,[1] Jonathan B. Fritz,[1,2] and Shihab A. Shamma[1,3]

[1]Institute for Systems Research, Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA
[2]Center for Neural Science, New York University, New York, NY 10003, USA
[3]Department of Cognitive Studies, Laboratoire des Systèmes Perceptifs, École Normale Supérieure, 75005 PSL Paris, France
[4]Lead Contact
*Correspondence: pyin@umd.edu
https://doi.org/10.1016/j.cub.2020.02.047

SUMMARY

Categorical perception is a fundamental cognitive function enabling animals to flexibly assign sounds into behaviorally relevant categories. This study investigates the nature of acoustic category representations, their emergence in an ascending series of ferret auditory and frontal cortical fields, and the dynamics of this representation during passive listening to task-relevant stimuli and during active retrieval from memory while engaging in learned categorization tasks. Ferrets were trained on two auditory Go-NoGo categorization tasks to discriminate two non-compact sound categories (composed of tones or amplitude-modulated noise). Neuronal responses became progressively more categorical in higher cortical fields, especially during task performance. The dynamics of the categorical responses exhibited a cascading top-down modulation pattern that began earliest in the frontal cortex and subsequently flowed downstream to the secondary auditory cortex, followed by the primary auditory cortex. In a subpopulation of neurons, categorical responses persisted even during the passive listening condition, demonstrating memory for task categories and their enhanced categorical boundaries.

INTRODUCTION

A fundamental aspect of listening is the cognitive ability to flexibly assign sensory stimuli into discrete, distinct, and behaviorally relevant categories that are task or context dependent, allowing the selection of an appropriate behavioral response to novel sensory stimuli within a known category. For example, spoken utterances can be recognized for their meaning irrespective of the speaker's voice (semantic or verbal categories), or alternatively, can be associated with the voice of a specific speaker regardless of their meaning (individual voice categories). Utterances can also be categorized by language, accent, emotional tone, speed of enunciation, and along many other

dimensions, allowing us to select an appropriate behavioral response to novel utterances from diverse speakers. This study investigates the nature of acoustic category representations for non-compact stimulus sets [1], in an ascending series of ferret auditory cortical fields up to frontal cortex, and specifically the dynamics of this representation during passive listening and active retrieval from auditory memory during task performance.

There are only a few previous neurophysiological studies on categorical representations in the auditory system of behaving animals [2–8] or anesthetized preparations following training [9, 10]. In one pioneering study, gerbils were trained to categorize frequency-modulated tones as "upward" or "downward" [2], while cortical potentials were recorded during task performance. As gerbils acquired the categorization rule, their neural activity patterns changed from initially reflecting stimulus acoustic properties to categorical membership. Similar categorical responses were observed during phoneme categorization in the primate ventral prefrontal cortex (vPFC) [3] and in the anterolateral belt area of the superior temporal gyrus [4–6]. Monkeys were trained to make a "same or different" judgment based on sequential presentation of two speech sounds ("bad" versus "dad"). At a behavioral level, monkeys perceived the range of morphed stimuli in a categorical fashion [3, 4], consistent with many earlier animal and human studies of such perceptual boundaries. Neurons in the auditory belt cortex exhibited categorical sensitivity depending on their morphology [4–6]. Responses in the primary auditory cortex (A1) also showed modulation during the performance of another auditory categorization task [7]. A recent study, utilizing 2-photon imaging in layers II and III of mouse auditory cortex [8], demonstrated dynamic task-driven modulation of single neurons and population response profiles that enhanced responses at the boundary between two trained tones during the performance of a tone discrimination task.

In comparison with the auditory system, there have been many behavioral neurophysiological studies of sensory categories in the visual system [11–19]. In a recent study [18], monkeys were trained to categorize the same set of moving, colored stimuli either into "up" versus "down" categories (when cued to attend to direction of motion) or into red versus green categories (when cued to attend to color of the moving dots). One of the main findings of this study was the discovery of a cross-current dynamic flow of categorical information in opposite directions, comprising a sequence of a transient, rapid, and bottom-up

sweep of sensory activity, followed by a reverse current of top-down categorical information flow.

The current study of categorization in the ferret auditory cortex confirms several aspects of these previous experimental results in the monkey visual cortex as outlined in the Results and Discussion, although it differs in some fundamental ways in our experimental approach. For instance, while monkeys were trained on a visual 2AFC task, the ferrets were trained on two different Go-NoGo negative reinforcement auditory categorization tasks [20]. We explored responses in the primary and secondary auditory cortex and frontal areas that parallel aspects of the ascending visual pathways in monkeys. The stimuli comprised relatively simple tones and amplitude-modulated (AM) noise that allowed us to dissect the sensory-to-category transformation and change in stimulus representations from their acoustic features (veridical representation) to task categories (categorical representation). A key feature of our experiments was the comparison of responses during passive listening with responses during task performance, which revealed task-dependent modulation.

RESULTS

Two ferrets were each trained over a period of 6 months on two different auditory categorization tasks. In both tasks, a trial consisted of the presentation of a single stimulus. The animals learned to lick the waterspout for water reward if the stimulus was a member of the Go Category, or to refrain from licking to a stimulus in the NoGo category (Figures 1A and 1B). Both animals successfully learned to categorize tone frequencies in one task (tone task) and learned to categorize amplitude modulation (AM) rates of a noise stimulus in the other task (AM task) (Figure S1A). Stimulus parameters (frequencies and rates) were organized such that there were three acoustic zones for the stimuli in each task [20, 21]. Since the two zones of Go stimuli bracketed the NoGo zone (Figure 1B), this allowed us to compare the responses to a disjunctive set of stimuli that belonged to the same Go behavioral category yet differed substantially along their feature dimension. Animals learned the tasks with a set of 6 stimuli and were able to generalize their performance to novel stimuli [20] (Figure S1B). Physiological recordings commenced after the animals attained criterion performance levels as detailed in STAR Methods. Animals displayed stable task performance during the physiological recording sessions with an average discrimination rate (DR) of 55.2 for the tone task and 50.8 for the AM task (left panels in Figure 1C) and demonstrated clear behavioral boundaries along the trained feature dimensions (middle and right panels in Figure 1C).
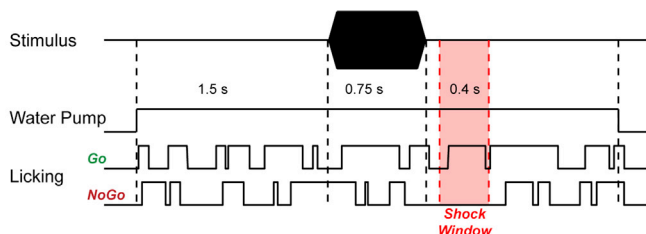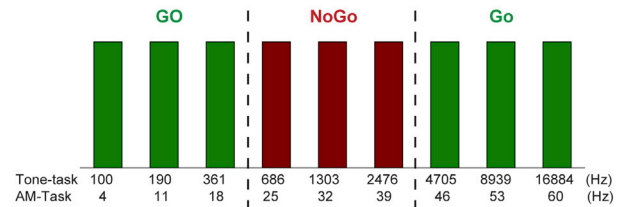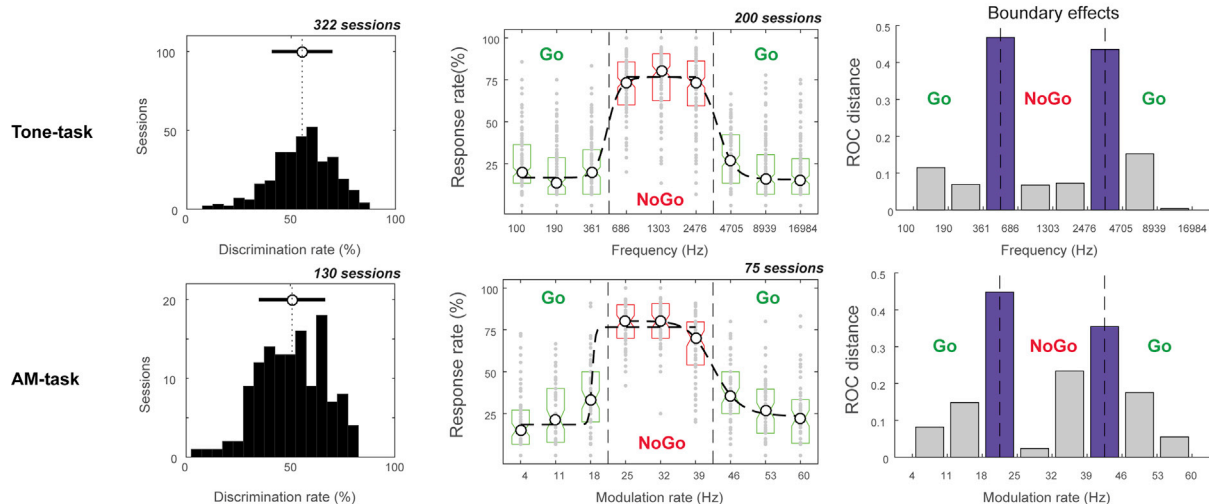
Multiple single units were isolated with *four* simultaneously inserted independently moveable electrodes in each session. Recordings were made over a period of 12–18 months from A1 and dPEG, uniformly covering all the tonotopically identified areas of these fields (Figure S2B). In each recording, responses were characterized in terms of their tuning curves (characteristic frequency [CF] and bandwidth) and spectrotemporal receptive fields (STRFs). Based on the tonotopic maps derived from our recordings, it is very likely that the great majority of our recording sites in dPEG were in the anterior area PPF, with only a few recording sites in the more posterior PSF [22, 23]. However,

because no functional differences have yet been observed between the secondary areas PPF and PSF [24], we broadly refer to the location of our pooled recordings in PPF and PSF as dPEG. We also recorded from the dorsolateral frontal cortex (dlFC) during the performance of the same tasks. At the conclusion of all recording sessions (128 sessions in ferret Guava and 95 sessions in ferret Gong), we placed HRP markers or made iron deposits (four sites shown in Figure S2A) for subsequent histological examination of the recorded areas to confirm their location based on neuroanatomical features (see STAR Methods). Post-mortem examination of brain sections revealed the exact locations of the markers, and these were used to confirm the relative position of all recording penetrations. The neuroanatomical locations of the two HRP deposits in auditory cortex were confirmed in A1 and dPEG, based on the ferret brain atlas [25] (sites 3 and 4 in Figure S2C) and the associated neurophysiologically defined tonotopic map (Figure S2B). The two recording sites with iron deposits in dlFC were identified in post-mortem histology to be in the rostral part of the anterior sigma gyrus (ASG) (sites 1 and 2 in Figure S2C). Based on the ferret atlas [25], both sites were located in the premotor cortex (PMC). We are using the dlFC nomenclature here to be consistent with the terminology used in earlier studies [26–28].

Single-Unit and Neuron Population Responses Show Enhanced Categorical Contrast during Behavior

In this report, we analyzed the responses of 1,269 isolated single units with auditory responses (recorded from two ferrets), all of which showed activation to some of the acoustic task stimuli in a passive or active experimental epoch. Of this total, 1,140 neurons were tested in the tone task (346 in A1 [181,165 cells from each animal], 430 in dPEG [233,197], and 364 in dlFC [228,136]), and 571 neurons were tested in the AM task (149 in A1 [91,58], 192 in dPEG [95,97], and 230 in dlFC [191,39]). In general, response contrast between Go and NoGo stimuli was less modulated by task performance in A1, compared to response contrast between Go and NoGo in dPEG and dlFC (Figure 3A). This is illustrated by two examples of single-unit responses from each of these three areas, shown for the tone task (Figures 2A–2F) and AM task (Figures 2G–2L) stimuli. In all panels of this figure, two key stimulus properties are highlighted: (1) category—i.e., Go versus NoGo (by color: green versus red); and (2) behavioral state—passive listening (dashed curves in left and right plots) versus task-engaged active listening (solid curves in middle plots).

For tonal stimuli (Figures 2A–2F), the passive responses in A1 and dPEG (dashed curves) reflected the unit's frequency tuning curve (Figures 2M and 2N). Neurons from auditory cortex (Figures 2A–2D) showed clear responses to tonal stimuli and in these examples were tuned within the NoGo range (Figure 2M). In contrast, there were often *no* responses to tonal stimuli in dlFC in the passive condition (Figures 2E and 2F), and, when present, responses were not tuned. However, during task performance, responses (solid curves) increased for NoGo stimuli (red) relative to the Go responses (green), thus enhancing the contrast between responses to the two categories. This contrast enhancement in favor of the NoGo responses became progressively more pronounced in dPEG and dlFC, as was reported earlier [26, 27]. The same pattern of contrast enhancement was found

Figure 1. Task Design and Categorical Performance during Physiological Recording Sessions

(A) Go/NoGo conditioned avoidance behavioral paradigm. A trial was started by initiating water flow (~1.2 mL/min) to a water spout. The ferret could freely lick water from the waterspout. An auditory stimulus was presented 1.5 s after the onset of water flow. Animals learned to continue licking (for a Go trial stimulus) or to briefly refrain (for 400 ms) from licking (for a NoGo trial stimulus). A NoGo stimulus (middle-range tones or AM noise; see Figure 1B) was followed by a shock window that began 0.1 s after stimulus offset and lasted for 0.4 s. Animals learned to withhold licking of the waterspout during the shock window in order to avoid a mild electric shock on the tongue (when in the free-run training box) or on the tail (when head-restrained in the head-fixed holder) after a NoGo stimulus. False alarms (cessation of licking of the waterspout following a Go stimulus) resulted in a variable 5- to 10-s timeout penalty, applied at the end of the trial. Trials ended 2.0 s after sound offset by turning off water flow. Following the completion of one trial, the ferret had to cease licking the waterspout for at least 1.0 s in order to initiate the next trial.

(B) The 9-stimulus set for tone task and AM task during neurophysiological recordings. The acoustic stimuli were partitioned into three ranges along the frequency or AM rate axis.

(C) Task performance during neurophysiological recording (sessions were combined across animals). Left: behavioral performance is quantified by discrimination rate (DR) measure. The distributions of discrimination rate (DR) during performance of the tone task or AM task in all neurophysiological sessions are shown with the mean of DR indicated by a vertical dashed line and standard deviation of the DR indicated by a horizontal bar. Middle: the task performance is described by the behavioral response probability across stimulus parameters (tone frequency or AM rate). The performance data were pooled from all behavioral sessions from the two animals during neurophysiological recordings in auditory cortical areas (A1 and dPEG). The psychometric functions (dashed) obtained by sigmoid fitting of the behavioral response versus stimulus parameters around the boundaries between low and middle and the middle and high ranges. Right: the discrimination between adjacent stimuli along the feature dimensions obtained based on the distributions of behavioral responses to each stimulus. The resulting discrimination function exhibited peaks between category boundaries (the purple bar) along the training feature dimensions.

See also Figure S1.

for the AM task, especially in dPEG (Figures 2I and 2J) and dlFC (Figures 2K and 2L).

We next examined how the entire population of cells in each of the 3 cortical regions (A1, dPEG, dlFC) represented the contrast between NoGo and Go responses, and how this contrast evolved over time during passive listening and task performance. Figure 3A shows that the population contrast function (defined as NoGo-Go population responses) for both tasks broadly resembled the response patterns already described for the single-unit examples (Figure 2). Specifically, the population contrast rapidly and significantly increased during task performance compared to

the population contrast in the passive listening context (solid versus dashed curves), especially in the higher cortical regions (dPEG and dlFC). The contrast enhancement during behavior in favor of the NoGo stimuli (Figure 3A) can be traced to changes in the overall tuning function of the population responses as illustrated in Figure 3B. Initially the pre-passive responses are barely "tuned" to the NoGo stimuli (dashed lines are all fairly flat). However, during task performance (solid curves), the population response to the two categories *diverge*, becoming relatively enhanced for the NoGo tones and AM rates but in strikingly different ways in the three cortical areas (see Figures S3A and
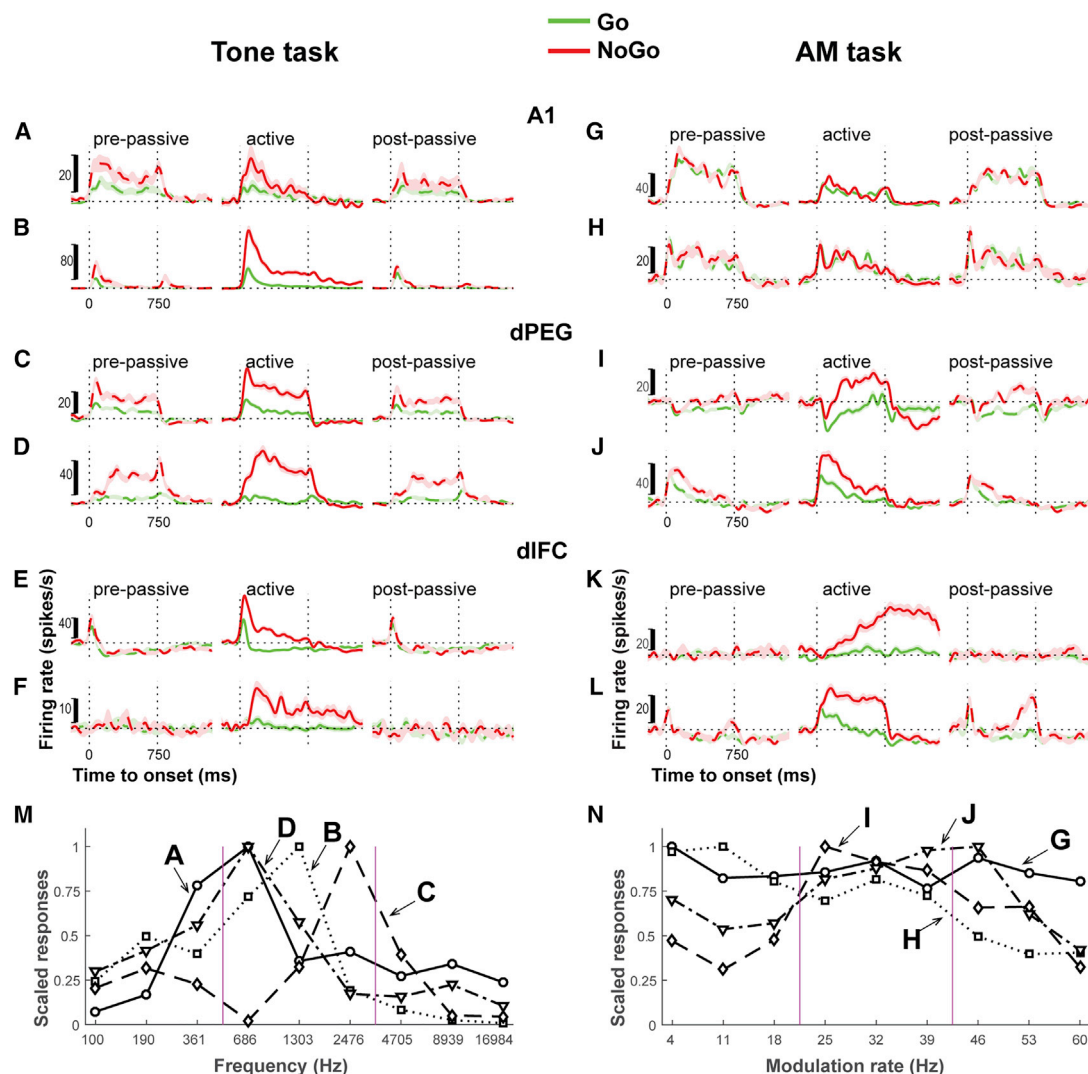
Figure 2. Examples of Single-Unit Responses during Passive Listening and Task Performance

(A–N) The examples illustrate single-neuron PSTH responses to task stimuli during passive listening and behavior in the task-performance context, when animals were engaged in either the tone task (A–F) or AM task (G–L). The green and red lines are population averages of spiking responses for trials grouped by stimulus behavioral meaning (either Go or NoGo). The shaded areas around the lines indicate the standard error of the responses. The responses to NoGo stimuli changed during behavior in all example neurons. The responses were enhanced at all levels in the tone task (A–F) and in most of the neurons in the AM task but were slightly suppressed in A1 cells during behavior (G and H). The bottom panels are the frequency response curves (M) and the modulation transfer functions (N) of the example neurons from auditory field (A1 and dPEG), which were computed from onset response to tones or the averaged sustained response to AM noises.
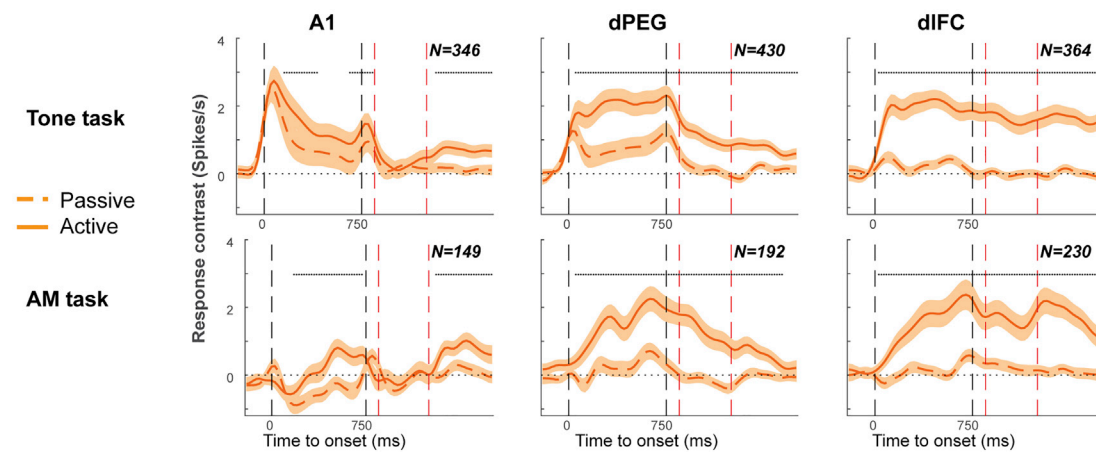
S3B for details). Go responses in A1 are *suppressed* during the task (blue downward arrows), whereas in dPEG and dlFC the NoGo responses *increase* significantly compared to the Go responses (red upward arrows). These changes during task performance enhance the difference between the responses to the two categories of stimuli and hence lead to the formation of prominent and distinct categorical responses and hence provide a neural substrate for categorical decision-making and behavioral choice.

Distinct Temporal Dynamics for Bottom-Up (Sensory) and Top-Down (Categorical) Information Flows during Behavior

The key conceptual hypothesis underlying our analysis is that the emergence of categorical representations implies that neuronal

responses in higher cortical areas become progressively more discriminative for stimuli *across* categories and less so for stimuli *within* a category, mirroring at a neuronal level the behavioral responses characteristic of category perception (Figure 1C). To test this hypothesis directly, following from Freedman and colleagues [12, 14], we defined two contrasting measures that capture these categorical versus the sensory aspects of neuronal responses. The first measure is the categorical index (CI), which is based on trial-by-trial responses and reflects the degree to which a response is driven by categorical versus sensory information (see STAR Methods for details). CI values range from 0.5 to −0.5; positive values indicate that the response is driven by categorical information, whereas negative values indicate the responses is driven by sensory information. The CI is

A Contrast (NoGo - Go)

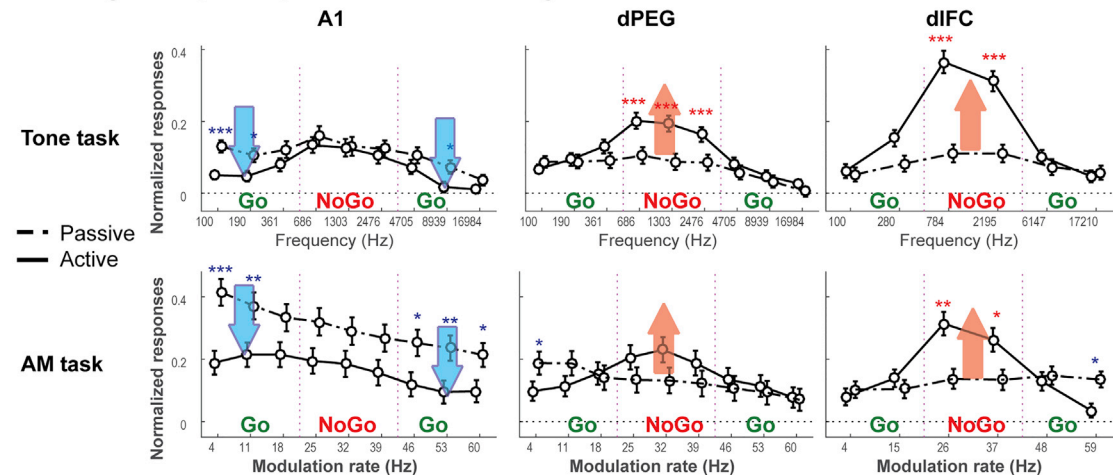**B Averaged response profiles and the change of direction**

Figure 3. Single-Unit Population PSTHs Illustrate the Enhanced Contrast between the Go and NoGo Categorical Responses during Behavior

(A) Averaged PSTHs of the response contrast between NoGo and Go-sounds in neuron populations in A1, dPEG, and dlFC during different task conditions. As already indicated in the example neurons in Figure 2, responses increased for the NoGo stimuli during active task performance (solid lines) compared to pre-task passive listening conditions (dashed lines), especially in dPEG and dlFC. Significance of the difference between the two conditions was assessed by the Wilcoxon signed-rank test between the two categories in each PSTH bin (p < 0.05 in three consecutive bins). The presence of statistically significant differences between responses in active and passive conditions is indicated by the gray dashed lines above the PSTHs. Average spontaneous activity before stimulus presentation was subtracted from all PSTHs.
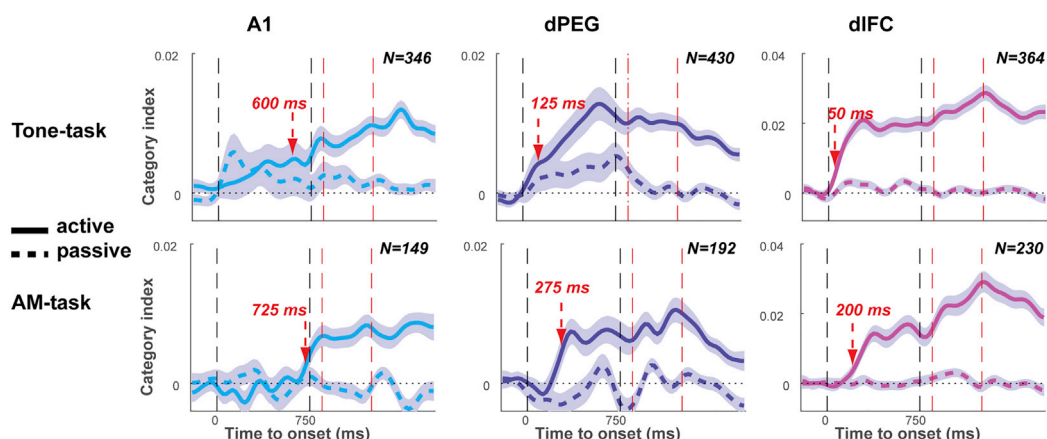
(B) Average changes between responses during the passive (dashed) and active (solid) states were evaluated from responses at each tone frequency and AM rate in A1, dPEG, and dlFC. Stars indicate significant changes assessed by the Wilcoxon signed-rank test (*p < 0.05, **p < 0.01, and ***p < 0.001, respectively). The broad red and blue arrows highlight the overall direction of the response changes in different ranges (low, middle, and high). Overall, in both tone and AM tasks, the responses enhanced the contrast between the two categories by either increasing in the Middle range (NoGo) or decreasing to both Low and High ranges (Go) (Figure S3 depicts the detailed changes of Go and NoGo response PSTHs during task performance).

computed from the overall response evoked by stimuli in passive listening or during task performance (Figures S3A and S3B). It is also computed at each time bin so as to estimate its temporal dynamics; hence, CI is a time function over the duration of a trial (Figure 4A). We averaged this index across the neuronal population in each area and during different behavioral states, which allowed us to examine both the extent and dynamics of the global categorical responses in each area as shown in Figure 4A. A second complementary measure was defined as the sensory index (SI), which captures the purely sensory aspects of the responses. It is the proportion of response variance explained by
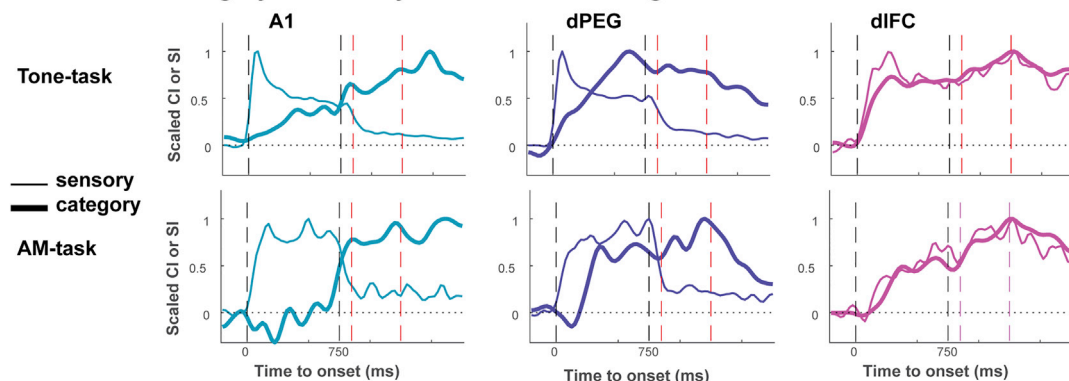
individual stimuli over the total variance. SI varies from 0 to 1.0, where a value near 1.0 reflects a highly sensory selective response (see STAR Methods for more details), e.g., the response of a cell that is finely tuned to one tone. As with CI, SI can be computed for each unit at each time bin throughout a trial in both passive and active conditions (Figure S4B).

All cortical regions exhibited weak categorical responses in the pre-passive state (dashed curves of Figure 4A with CI close to 0). Nevertheless, some pre-passive tone responses in A1 and dPEG were significantly categorical (CI > 0), indicating that some responses in the neuronal population during the passive
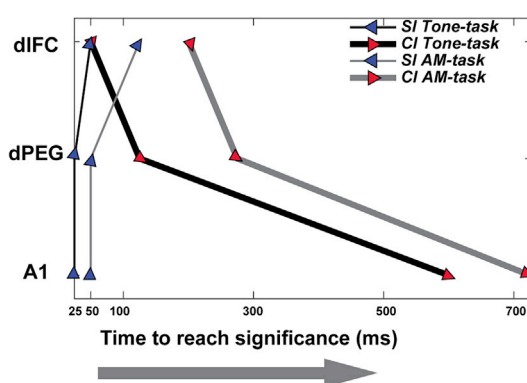
A Time course of Category index (CI)

B Normalized Category *vs* Sensory information in during task

C Category/sensory information flow during task

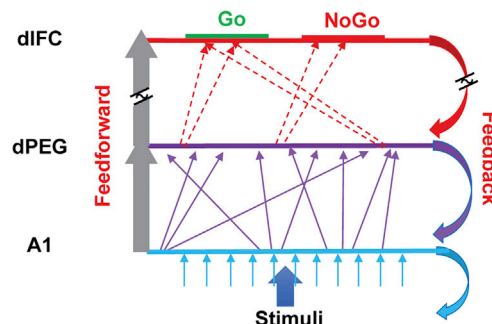**D Schematic feedforward and feedback Information flows**

Figure 4. Population Responses Illustrate the Different Temporal Dynamics of the Bottom-Up (Sensory) and Top-Down (Category) Information Flow during Behavior in the Three Cortical Areas (A1, dPEG, and dIFC)

(A) Plots illustrate the time course of the averaged categorical index (CI, see Figure S4A for the further details of CI computation) in passive (dashed) and active (solid) task conditions, with the shading indicating the standard error at each time point. Response dynamics of the tone- and AM-task responses (top and bottom panels) are shown in A1, dPEG, and dIFC (left to right panels in different colors). The arrow in each panel marks the categorical latency, the time following stimulus onset at which the CI is significantly above zero in the active state. This latency is longest in A1 and progressively *decreases* in the secondary auditory cortex and frontal cortex. The abstract nature of these dynamics is highlighted by the similarity between responses in the tone and AM tasks. The vertical dashed lines in all panels indicate either the stimulus period on/off (black lines) or the shock period on/off (red).

(B) Contrasting the categorical (CI, the data from Figure 4A) and sensory (SI, see Figure S4B) responses in the active state. Each panel contrasts the categorical versus sensory aspects of the responses during task performance (bold versus thin lines, respectively). The SI responses rise rapidly in all areas, but with a slightly increasing latency from A1 to dIFC. The CI by contrast emerges first in the dIFC, and only much later in A1. Both CI and SI data are scaled to their peaks.

(legend continued on next page)

condition exhibited categorical responses prior to task performance. However, during task performance (solid lines), the CI increased substantially relative to the passive state. In both tasks, task-related CI enhancement became larger in ascending the hierarchy from A1 toward the secondary dPEG areas and to the dlFC (left to right panels of Figure 4A).

An interesting feature of the development of these categorical responses is the dynamics of their rise during the active state. Specifically, the arrows in each panel mark the time (relative to stimulus onset) when CI became significantly >0. For both the tone task and AM task, this ''categorical latency'' was *longest* in the primary region (A1) and *shortened* gradually toward the higher areas of the dPEG and dlFC. This dynamic feature of the categorical responses is highlighted in Figure 4B where we plot in each panel the categorical versus sensory aspects of the responses (CI versus SI) curves (for comparison, both curves were scaled to the peak value). The SI (sensory) bottom-up responses exhibited the basic familiar pattern of short latency in primary sensory cortex (A1) that increased slightly in higher areas (dPEG and to dlFC) (see Figure S4B for detailed SI information). In contrast, the CI showed more gradual dynamics and the opposite trend, exhibiting shortest latencies in higher areas and longer latencies in the primary cortical sensory area. This pattern of dynamics suggests a bidirectional temporal information flow that is remarkably similar across these two very different tasks and stimuli. This point is emphasized in Figure 4C where we plot the SI and CI response latencies (x axis) during task performance, as a function of their hierarchical cortical origin (A1, dPEG, dlFC). In both tasks, the shortest latencies appear in A1, the origin of the bottom-up sensory flow of information (SI, blue arrowheads) from A1 toward the dlFC (all in a few tens of milliseconds), and then the top-down ''reverse'' flow of categorical (CI, red arrowheads) information toward the dPEG and back to A1 (occurring over a period of hundreds of milliseconds). These opposite temporal dynamics of information flow during categorization tasks are considered in more detail in the Discussion.

Categorical Response Patterns Distinguish Two Neuronal Populations with Intrinsic versus Task-Induced Categorical Responses

It is evident from Figure 4A (dashed lines) that there is a neuronal population, particularly in A1 and dPEG, that exhibited categorical responses in the *passive* state in the trained animals. To explore the characteristics of this population and its implications for the formation and encoding of categorical perception in auditory cortex, we first segregated the cells in each cortical area into two groups based on an *averaged* CI computed during the prepassive state ($CI_{passive}$) using the *onset responses* to the task stimuli (0–250 ms in the tone task and 0–375 ms in the AM task), rather than using the spiking waveform width [5] (see Figure S7). We examined the CI dynamics separately for each of these two groups of cells as shown in Figure 5A. In the first group (left panels), referred to as ''intrinsically categorical cells,'' it exhibited categorical responses in the passive condition ($CI_{passive} > 0$). The CI from this neuron group increased further during the task, and with the same rapid form and short latencies as seen in the passive state, and it maintained this high value well beyond the behavioral response window. The remaining cells (right panels) formed a ''task-induced categorical group,'' which had no categorical responses in the passive state ($CI_{passive} = < 0$) but exhibited a slow buildup of CI in the active state with the staggered latencies seen earlier in the whole population (Figure 4A). The proportions of these two populations within each of the lower cortical regions were roughly similar (Figure S5A). However, unlike A1 and dPEG, few dlFC neurons were intrinsically categorical, and most displayed weak or no passive responses with $CI_{passive}$ near 0. Hence, most dlFC neurons were characterized by ''task-induced'' categorical responses. The similar pattern was also evident in neuron populations when animals engaged in the AM task (bottom row of panels in Figure 5).

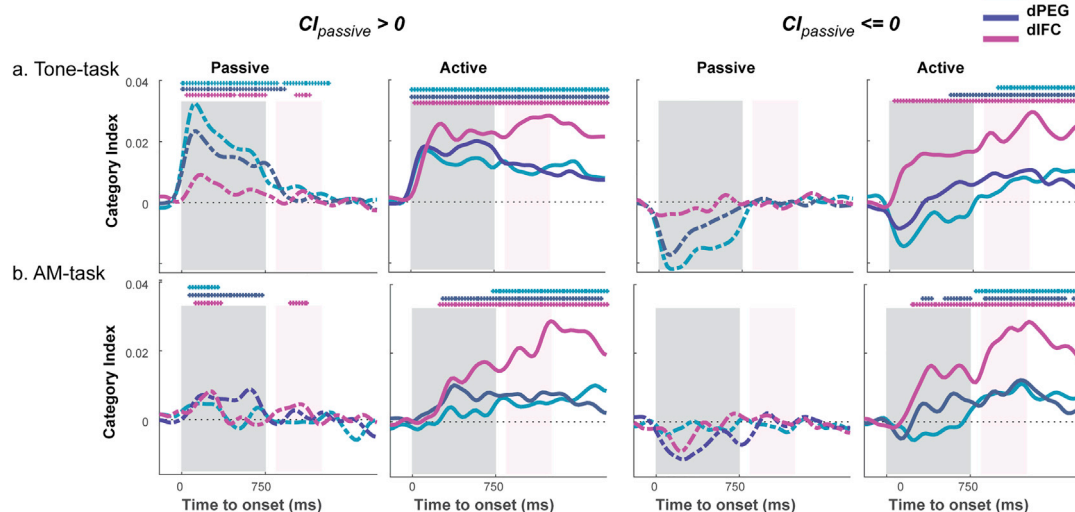

We further explored whether there are any response characteristics that might explain the origin of these two cell groups. To do so, we examined for each cell in A1 and dPEG the relationship between its CI and two basic properties (relevant mostly for the tone task): (1) the best frequency (BF) and (2) the bandwidth as quantified by the SI of the cell. To clarify this link, we simulated neuronal population responses that had a range of BFs and SIs and measured the resulting CI responses in a tone-task paradigm (simulation details in STAR Methods). The results are plotted in Figure 5B, where each dot represents a simulated cell, placed at its BF (x axis) and SI (y axis), and its CI is indicated by its color. It is evident that narrowly tuned cells near the center of the NoGo range exhibit categorical responses (red dots) because they respond only to tones within this category. Note also that the CI of very broadly tuned cells (low SIs) are green since they are broadly tuned and hence mix responses from both tone categories.

Figure 5C illustrates the resulting measured distribution of responses in A1 and dPEG, which exhibits a reasonable correspondence to the simulated results of Figure 5B. Hence, we conclude that, at least in a proportion of cells, with narrow bandwidths and appropriate BFs, their responses can appear rather ''categorical'' even *before* the task begins, because the task categories are a partition of a natural continuum of frequency. However, it is also evident in further analysis of the population responses, that *during* task performance, many cells become tuned to the categorical stimuli of the tasks (see details in Figures S5B and S5C for the changes of CI during active state) and

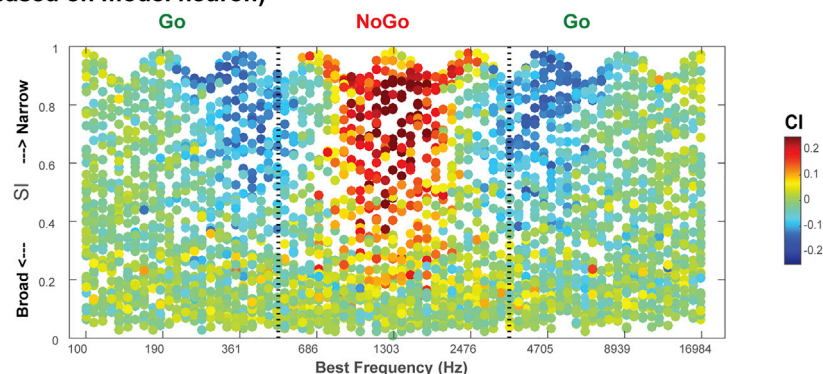

(C) Categorical (thick lines) versus sensory (thin lines) information flow is revealed by the latency of the CI versus SI in the two tasks (tone task in black and AM task in gray lines). The hypothetical flow of information is indicated by the direction of arrows along the trajectory of each task. The bottom arrow indicates the time from stimulus onset to accumulate significant category or sensory information (see text for details).

(D) A schematic of the dynamic flow of sensory and categorical information. The encoding of categorical information during their formation and retrieval is schematized by the two directional flows: a feedforward bottom-up interlayer sensory pathways (symbolized by the gray arrow on the left) from A1, to dPEG, and eventually to dlFC (through a series of intermediate fields, dashed arrows) whose activation is strongest during task performance. This pathway undergoes long-term modifications that reflect task categories. The reverse feedback top-down flow only occurs during task performance and propagates categorical information to earlier cortical areas, eventually modifying their feedforward connectivity.

A Dynamics of Category Index in different neuron populations

B Simulation (based on model neuron)

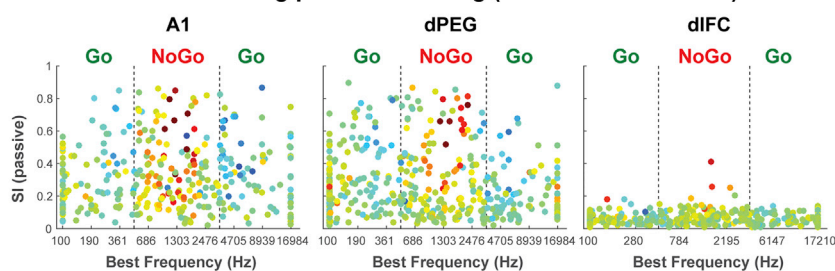## C Neruonal data measured during passive listening (0-250 ms from onset)

Figure 5. CI Dynamics in Two Distinct Neuronal Populations Grouped According to Their CIs during Passive Listening and the Simulated Categorical Responses

(A) Dynamics of the categorical responses in the two cell populations (see also Figure S5A) segregated according to their average passive categorical responses ($CI_{passive}$) to tone (0- to 250-ms onset period)/AM (0- to 375-ms onset period) task stimuli: intrinsically categorical neurons have $CI_{passive} > 0$ (left panels) and task-induced categorical neurons with $CI_{passive} \leq 0$ (right panels). During task engagement, the group of intrinsically categorical cells exhibit rapid buildup of categorical responses that resemble their passive counterparts but are stronger and last much longer over the response period. In comparison, the second group exhibits slowly rising CI in the active state, with staggered latency similar to Figure 4A (see text for more details). The similar dynamics shown in performing the tone task and AM task indicate a general processing of categorical formation/representation across involved feature dimension.

(B) Simulating the relationship between CI of a cell and its two response properties: best frequency (BF) and bandwidth parameterized by the sensory index SI of the neuron. Scatterplot depicts a large simulated neuronal population where each cell is depicted as a dot placed according to its BF and SI (x and y axes) and whose color reflects its CI computed from its simulated responses: red (blue) dots are highly (least) categorical. Cells with BFs in the middle frequency range (NoGo) and with intermediate to narrow bandwidths likely have a high CI (red). The opposite occurs for (blue) cells near the boundaries; these are broadly tuned cells (low SIs) that are also non-categorical (green).

(legend continued on next page)

exhibit highly enhanced representations of the stimuli at the categorical boundaries.

Population Analysis Reveals How Categorical Edges Are Enhanced during Category Formation

To gain a global view of the population representation of the categories and to compute distance metrics among the categorical responses, we utilized principal-component analysis (PCA) to represent the data efficiently in a reduced dimensional space (e.g., Figure S6). We first extracted randomly drawn trial responses from the total available for each stimulus in each cortical area and rearranged them to form the 3D matrix (trial × time-bin × neurons as in Figure S6A) and then applied the PCA over this matrix. The resulting PCs can then be utilized to gain a different perspective on the responses by projecting and displaying the responses of all units to the 9 tones during the pre-passive, active, and post-passive epochs (e.g., response trajectories generated from the 3 PC projections in Figure S6B for dPEG responses).

Using projections of all responses onto these PCs, we quantified in Figure 6A the pairwise distances for all tones (left) and AM noises (right) in all task epochs. For ideal categorical responses, we would expect a blocked distribution in which all responses to tones or AM noises *within* a category would be similar, while those *across* categories would be far apart. Such a blocked response is evident in the matrices of Figure 6A (in both passive and active states) but becomes especially clear during task performance (middle row of panels). These results are further summarized by the bar plots accumulating (average) distances *between* categories versus *within* categories (BC versus WC; Figure 6B). Categorical enhancement during behavior is evident by the significant increases in BC (enhanced relative height of red to blue bars); by contrast, WC remained relatively unchanged during behavior. Thus, categorical formation is mostly driven by the increased *divergence* across categorical boundaries (BC) and *not* by the *convergence* within a category (WC).

Finally, a classic signature of categorical formation is the sharpened transitions across the categorical boundaries, as revealed in Figure 6C by the pairwise distances between responses to *adjacent* stimuli during passive (dashed) and active (solid) states. Distances are relatively large near the category edges, even in the passive state (especially for the tone task) indicating the presence of permanent (category-sensitive) changes that likely reflect the extended period (months) of training. The edge effects, however, become even more pronounced during task performance (solid relative to dashed lines), as summarized by the bar plots below, which illustrate that distances *across* categories (black bars) increase substantially relative to *within* category changes (gray bars) creating a boundary that is similar to the behavioral boundary shown earlier in Figure 1C. All these and other categorical effects were more clearly seen during the tone task than during the AM task (see also Figure S6C), likely reflecting the better behavioral performance on

the tone task AM task (Figure 1C). For instance, the representation of the upper rate edge in the AM tasks is weak, mirroring the animals' poorer performance near this boundary (Figure 1C).

Categorical Information Becomes Hierarchically More Abstract across the Two Independently Trained Tasks

In many of recordings, we were able to gather responses in *both* tasks when the animal performed both tasks sequentially, thus allowing us to examine whether categorical responses in both tasks are abstracted, becoming jointly encoded in a population of cells that essentially transcended the sensory aspects of the stimuli and emphasized instead their decision-related categorical labels. Figure 7 summarizes the findings from such (jointly characterized) cells in each of the three cortical regions: A1 (125), dPEG (163), dlFC (154). First, we computed and (scatter) plotted in Figure 7A the average CI of each unit in passive and active behavioral contexts for the two tasks over the sustained portion of the response.

In the passive state (Figure 7A, top panels), there were comparable numbers of cells in each of these four groups. However, during the active state, dlFC responses changed substantially, almost all becoming categorical (142/154), with the majority being of the joint kind (82/142 red). Remarkably, the situation in A1 and dPEG was different in that the proportion of cells with shared categorical responses remained comparable to each of the task-specific CIs and stayed roughly unchanged between the passive and active states, suggesting that the basic character of the passive categorical responses in auditory cortex is consistent with the responses in the active state.
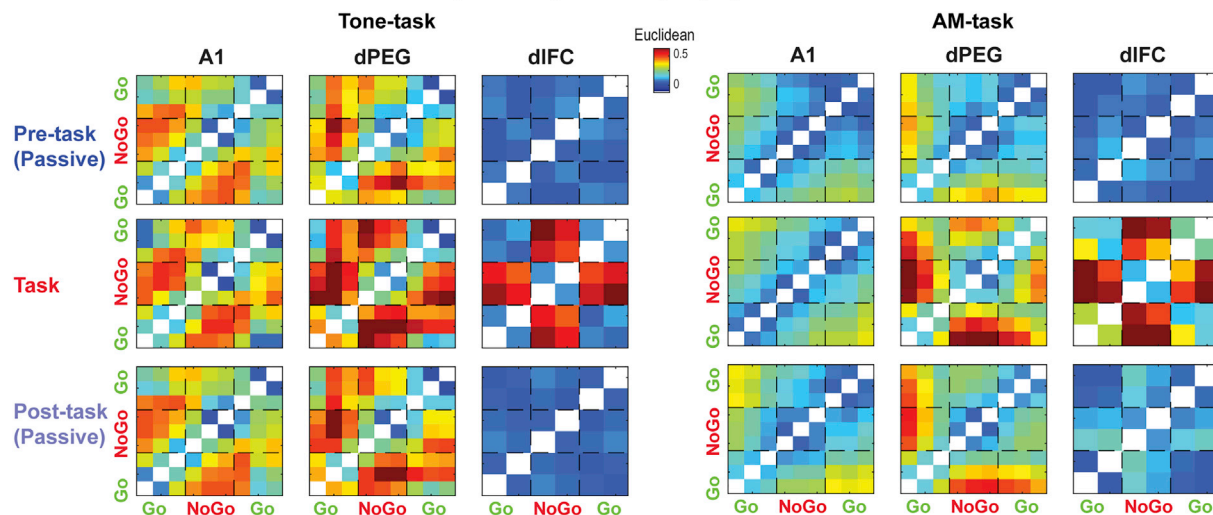
Given the large changes in the proportion of joint responses in the dlFC, and the staggered latencies of the CI responses observed earlier (Figure 4A), we wondered whether the joint CI responses A1 and dPEG might also exhibit these dynamics (which were obscured by the averaging of the CI responses in Figure 7A). This indeed was the case, as shown in Figure 7B, which displays the instantaneous *proportion* of joint cells (i.e., cells that exhibited joint categorical responses to the tone and AM tasks) near stimulus onset, during its sustained portion, and late portion. Compared to the passive state (dashed), the percentage of joint responses during the active state (solid lines) increases toward the end of the stimulus in all areas (p = 0.0194, 2.0931×10^{-5}, 7.262×10^{-13} for A1, dPEG, and dlFC, respectively) and more substantial in the dlFC (p = 2.921×10^{-8} and 1.843×10^{-14} for the onset and sustained periods) by binomial test. These results are indicative of the progressively more abstract categorical responses, i.e., joint and independent of task specific stimuli, in ascending to higher levels of the cortical hierarchy.

DISCUSSION

Responses of the auditory cortex are known to reflect many aspects of acoustic perception, ranging from the representation of

---

(C) Distribution derived from actual passive responses to tone-task stimuli in the three cortical regions. A1 and dPEG are generally responsive during passive state and have moderate to fine tuning, whereas passive dlFC responses are usually weak or not present at all, and hence its CIs are near zero with few exceptions (see text and STAR Methods for more details). The stars above the graphs indicate significant positive CI (p < 0.05 in three consecutive bins with one-sample one-tailed t test). See Figures S5B and S5C for more detailed changes of the CI during task engagement in relation to the neurons' tuning properties and also Figure S7 for populations grouped by spike waveform width.

A Euclidean distance matrix in Principal Components (PC) space

B Change of average distance: BC vs WC
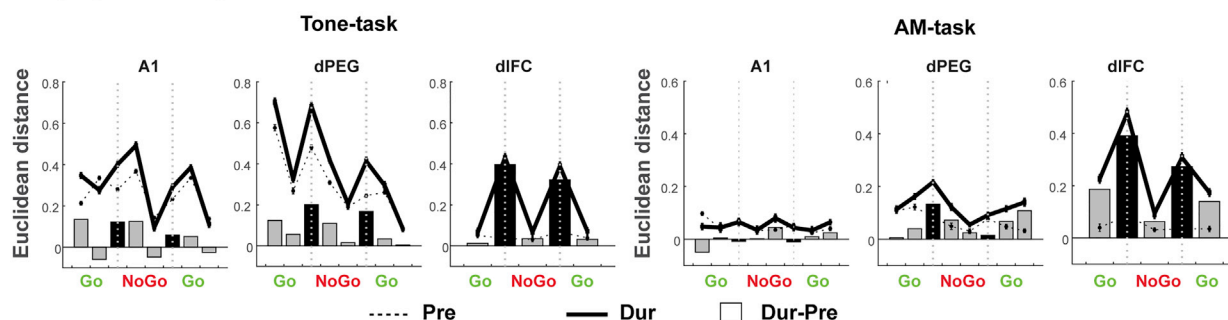

C Category boundary effects

Figure 6. Population Analysis Demonstrates the Enhancement of Categorical Boundaries during Category Formation

(A) Heatmaps display the pairwise averaged Euclidean distance matrix during pre-passive, active, and post-passive tone-task (left) and AM-task (right) conditions.

(B) Bar plots (top panels) summarize the averaged Euclidian distance between sound pairs falling within (WC) and across (BC) the behavioral categories in all task conditions (color coded). The distances in all plots are computed from the first 3 PCA components over the neuron populations. (All significance measures are based on the Wilcoxon rank-sum test: *p < 0.05, **p < 0.01, ***p < 0.001).
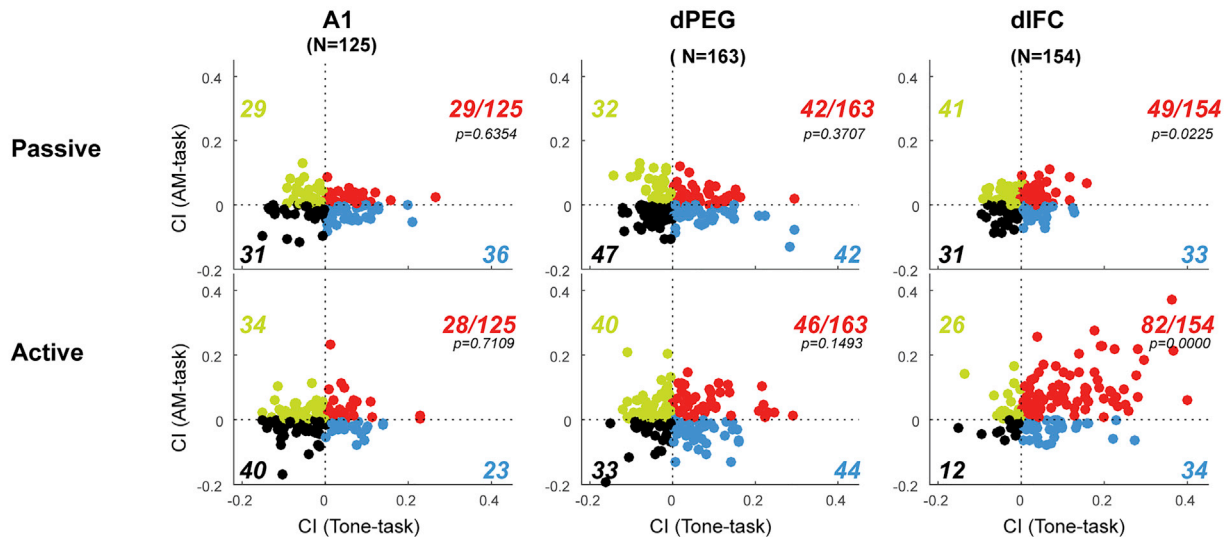
(C) Plots illustrate the categorical boundaries computed from the Euclidean distance between each neighboring stimulus pairs in the pre-passive (dashed) and active (solid) states. The bars in the plots display the averaged *difference* of the distances between the two states (Active-Passive), and the highlighted bars (black) mark the changes of the distance at the pairs across the boundaries of the behavioral Go and NoGo categories that become more sharply defined during the active state and are progressively better defined from A1 to dPEG, to the dlFC, especially during performing the tone task (left).

See also Figure S6.

the sensory attributes of a complex sound such as its pitch, timbre, and location [29], to the rapid plasticity due to engagement of cognitive functions during task performance such as memory retrieval [30–33], decision making [33–35], and

categorical perception [4, 8, 36]. The present study focused on the emergence and retrieval of categorical representations of diverse acoustic stimuli along multiple levels of the auditory cortical hierarchy during passive listening and during

A Category Index (CI) across both tasks

B Temporal progression of increasing task categoical abstraction
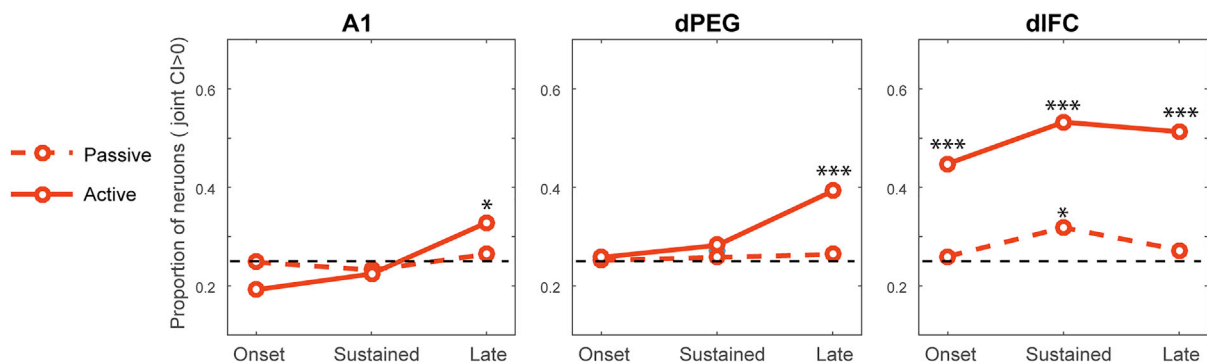

Figure 7. Category Information Becomes Progressively More General across the Two Independently Trained Tasks in Higher Cortical Areas
Neurons tested in both the tone task and AM task, in all three cortical areas.

(A) Scatterplot illustrates the correspondence between the categorical responses in each cell (represented by the dots) to both tasks: CIs from tone task (abscissa) and the CIs from the AM task (ordinate). The CI measurements were based on the sustained response (250–750 ms) evoked by stimuli in the pre-passive or active states. The p value in the panels are indicates the likelihood of observed joint CI >0 form a binomial test.

(B) Dynamics of the joint categorical responses. The proportion of cells becoming jointly categorical increases significantly during task performance, but as with the dynamics of the categorical responses in the task-induced categorical cells (Figure 5), the increase becomes evident near the onset of the stimuli in dIFC, but only later in dPEG (sustained: 250–750 ms) and A1 (late: 750–850 ms). The stars above the lines indicate the significance levels from binomial test at corresponding condition: *p < 0.05, **p < 0.01, and ***p < 0.001, respectively.

performance of two different Go-NoGo categorization tasks. Our results demonstrate that in trained ferrets: (1) single-unit and population responses become progressively more categorical as one moves up the hierarchy of auditory cortical fields from the primary to the frontal cortex. (2) The categorical response becomes significantly enhanced during task performance, especially in the higher areas. (3) However, in a sub-population of cells in the auditory cortical areas, categorical responses persist even in the (passive) non-task condition, thus exhibiting a state-independent long-term *memory* for task categories in trained ferrets. (4) Auditory categories are delineated by enhanced boundaries that match their perceptual properties, especially during task performance in higher cortical stages. (5) The observed bi-directional and temporally staggered flow of information is in

remarkable agreement with the earlier finding in the primate visual system [18]. This pattern of dynamics persists in both auditory categorization tasks, emphasizing the global and abstract nature of this cognitive function and top-down flow of category information during task performance.

While auditory categorization has previously been described in neurophysiological studies in avian, rodent, and primate cortex [2, 4, 5, 9, 10, 37–39], they differed from our approach in using either global (epidural) population cortical response measures and/or more complex stimuli (speech or FM sweeps) with acoustic features whose parametric representations are uncertain in the areas studied, which in turn complicated the task of dissecting how the categories formed and how category boundaries were enhanced during task performance. A recent 2-photon

imaging study of auditory cortex [8] demonstrated dynamic enhancement of compact category boundaries during engagement in a tone category discrimination task but was limited in temporal resolution and spatial analysis of information flow in the auditory cortical network.

As noted earlier, previous observations on the neural basis of categorization that are most comparable to our present study arose from pioneering experiments on visual categorization that explored responses in a hierarchy of primate visual cortical areas and higher areas in parietal and frontal cortices [11, 14, 15, 17–19, 40, 41]. However, our study of auditory categorization differs from this body of work in several ways and sheds light on new aspects of the cognitive function of categorization. For example, our behavioral tasks involve dual categorical engagements with two completely different stimulus sets (tones and AM noise). The categories are non-contiguous or non-compact in that they included stimuli that are non-neighboring along the frequency or AM-rate axis. We also describe differences in the details of the representation at different cortical levels, with categories already formed in A1, becoming more apparent in dPEG, and clearest in dlFC, especially during task engagement. We confirmed earlier reports that some of the category-selective responses in dlFC were observed only when category distinctions were task relevant but not during passive exposure to the same stimuli [17, 26]. However, by contrasting the responses during behavioral tasks versus passive listening, we demonstrate a long-term representation (memory) for the dual categories that is present even during passive listening in a sub-population of intrinsically categorical neurons and also a more malleable, task-induced representation in another sub-population that exhibits dynamic categorical responses only during task engagement, with staggered, increasing latencies toward the earlier processing levels.

Categorical Representations Are Enhanced during Behavior

In our experiments, we observed that categorical responses are enhanced during task performance, beginning in the earliest cortical regions. For both the tone and AM noise tasks, the responses of the two categories diverge, while the diverse (and non-contiguous) stimuli within a category coalesce by converging into a similar pattern that ultimately represents the category label rather than the acoustic properties of the stimuli. An extreme example is the strong dlFC response during task performance to the NoGo tones and AM noise regardless of their frequencies or rates, while all Go stimuli elicit little or no responses (Figure 3B). This pattern of NoGo enhancement and/or Go suppression emerges gradually in A1 and builds up in the dPEG (arrows in Figure 3B). This specific pattern of enhancement and/or suppression between the two categories in our tasks likely reflects the Go/NoGo structure of our task design. If the two categories were associated with similar behavioral response outcomes, as in a positive reinforcement 2AFC task, we speculate that they would have driven mutually exclusive neuronal populations with similar responses.

Curiously, while the neuronal representation of the two categories within a single task became almost completely segregated at the level of the dlFC, single-unit responses nevertheless often encoded multiple categories across both of the two

independent auditory tasks, consistent with our previous results showing multimodal joint representation across auditory and visual tasks in frontal cortex [25] and previous reports of "multi-tasking" neurons in frontal and parietal cortex responses during visual categorization tasks in primate [42, 43]. In the current study, this finding is clearly illustrated in our data in Figure 7A by the large number of dlFC neurons that jointly categorized the NoGo stimuli in the tone task and the AM task. Such joint encoding was evident in the stages of dPEG or A1 only during late period of responses (Figure 7B), which might reflect the top-down modulation from dlFC and/or other higher cortical regions.

Cortical Dynamics of Categorical Responses

Another key property associated with categorical responses concerns their dynamics. Thus, while response onsets to task stimuli were rapid (Figures 2 and 4B) and followed the expected increase in latency toward higher cortical regions (from about 20 ms latency in A1, to 50 ms in dlFC), their categorical nature as quantified by the CI exhibited far more complex dynamics and a *reverse order* staggered latency during both tasks (Figures 4A and 4B). In both tasks, the buildup was quite rapid in dlFC, reaching its peak level within 50–200 ms following stimulus onset. In dPEG, the buildup was far slower, taking over 200–500 ms, while it did not begin to increase substantially until much later in A1, increasing only after the end of the stimulus. This pattern of response dynamics that occurred in both conditioned avoidance Go-NoGo auditory tasks in the auditory system of a carnivore is remarkably similar to that described in performance of positive reinforcement 2AFC visual tasks in the primate visual system [14, 18, 19], a correspondence that highlights the fundamental nature of this phenomenon being evident in different animal models, sensory modalities, cortical regions, and experimental paradigms and demonstrates a remarkable generality of the basic principles of categorical processing.

Dual Populations of Category Neurons

Characterizing single-unit responses to task stimuli in the passively listening ferrets allowed us to discover additional properties of the complex dynamics of the categorical responses. Some of the intrinsically categorical neurons may be present before training, and we suggest that additional intrinsically categorical neurons are recruited and may reflect long-term learning due to extensive training on the two tasks. Task engagement, however, induced enhanced CIs in *both* subpopulations but with a striking difference in dynamics (Figures 5 and S5A) that was only possible to dissociate because of our reliance on single-unit recordings. Thus, in the intrinsically categorical cells, response onsets during performance were rapid throughout all cortical regions. By contrast, the additional task-induced categorical responses exhibited a staggered buildup over hundreds of milliseconds. During task engagement, a new slower response component emerged in all cells but was especially evident in the task-induced categorical subpopulation. These two components were also evident in the visual responses (Figure 2c in Siegel et al., 2015 [18]), but their dissociation into separate populations was not possible because of the multiunit nature of the recordings in their study. We conjecture that these two cell populations and their dynamics reflect, on the one hand, permanent learning that shapes the sensory selectivity of cortical

cells along the bottom-up pathway, and on the other hand, a top-down feedback that becomes effective only during task performance (Figure 4D). We note that the existence of two neuronal populations with different categorical response selectivity was previously conjectured to be associated with cells of narrow versus broad spikes, putatively corresponding to interneurons versus pyramidal cells [5, 44, 45]. However, we did not observe such a segregation of spike waveforms in our data (Figure S7).

Finally, PCA of neuronal population activity (Figure 6) provided a geometric visualization and quantification of the passive and active categorical responses in terms of a distance metric that confirmed the results garnered from the single-unit and averaged responses. The PCA, however, also revealed a key indicator of categorical perception in the neuronal responses, namely, the sharpened categorical boundaries, which were evident in the passive and active states, and show increased prominence in ascending levels of the cortical hierarchy from A1 to dlFC (Figures 6B and 6C).

Speculative Model of Category Formation

We offer a speculative view of the gradual formation of categorical perception upon repeated task engagement (initial acquisition and subsequent reinforcement training) depicted in Figure 4D that gives a schematic of the hierarchically organized cortical areas from primary to frontal cortex that resembles the typical structure of multi-layered neural network models of sensory systems [46, 47] and studies of the computational principles underlying visual categorization [47–50].

We conjecture that initially (e.g., in naive animals), all responses to the Go and NoGo stimuli recruit different populations of cells in the early cortical fields (A1) according to their passive (pre-existing) sensory selectivity. A small subset of these cells may already behave "categorically" in our Go/NoGo tone task by virtue of their BF and narrow tuning. These pre-existing (intrinsically) categorical cells therefore form the nucleus for the initial buildup of the NoGo (and suppression of the Go) categorical responses. During task performance, these cells provide the fast "feedforward" categorical input to the higher areas up to dlFC, which is responsive to NoGo stimuli *only* during engagement [26]. The dlFC activation in turn supplies a positive "feedback" signal that back-propagates to the lower areas, and we hypothesize that timing-related plasticity, arising from the two task-related activations, reinforces and gradually enhances the categorical nature of the feedforward pathway (Figure 4D). It is therefore evident why this staggered feedback is critical for the buildup of the categorical responses, especially in the broadly tuned, categorically malleable cells, which may gradually become more categorical over the time course of training as they change their input selectivity. By contrast, the intrinsically categorical neurons in the auditory cortex already exhibit large CI responses in their rapid feedforward activations and hence are not modulated further to the same degree by the feedback influences. To encode more complex categories, multiple, task dependent, selective sensory features will have to be recruited [51, 52]. In addition to timing-related plasticity, an elegant cortical circuit model suggests that reward learning mediated by plastic top-down feedback is also likely to play a role in the emergence of category tuning from initially untuned neurons [53].

To test all these speculations in the ferret model system, it would be ideal to monitor chronically the responses in all of the cortical regions along this auditory pathway to the dlFC from A1, to dPEG, and in the two other intervening cortical areas of the VPr (rostral ventroposterior field) and PSSC (pseudosylvian sulcal cortex) [23, 27] during category learning while animals are trained from the task-naive state to task proficiency. Moreover, to fully understand the development and representation auditory categories, it would also be valuable to monitor subcortical areas such as the striatum that have been shown to play an important role in auditory category formation [54]. However, with available stable, chronic recording techniques [55] it is now possible to delineate the learning-induced changes in neural representation, in the interaction of feedforward and feedback responses and their dynamics of information flow as categorical responses gradually emerge during auditory category learning.

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
  - Animals
- METHOD DETAILS
  - Experimental apparatus
  - Behavioral Tasks
  - Headpost Implant Surgeries and Neurophysiological recordings
  - Experimental Procedures and Stimuli
  - Localization of recording sites
  - Data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
- DATA AND CODE AVAILABILITY

AUTHOR CONTRIBUTIONS

P.Y., J.B.F., and S.A.S. designed the experiments. P.Y. and S.A.S. analyzed data and prepared the figures. P.Y., S.A.S., and J.B.F. wrote the manuscript. P.Y. trained the animals and performed all neurophysiological recordings. D.L.S. helped with neurophysiological recording in one animal. P.Y. and J.B.F. made the HRP injections and iron deposits. S.R.-S. processed the brains and sliced one of them (Guava). P.Y. and S.R.-S. analyzed the neuroanatomical results and prepared the neuroanatomical figures.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Heffner, C.C., Idsardi, W.J., and Newman, R.S. (2019). Constraints on learning disjunctive, unidimensional auditory and phonetic categories. Atten. Percept. Psychophys. *81*, 958–980.

2. Ohl, F.W., Scheich, H., and Freeman, W.J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. Nature *412*, 733–736.

3. Russ, B.E., Orr, L.E., and Cohen, Y.E. (2008). Prefrontal neurons predict choices during an auditory same-different task. Curr. Biol. *18*, 1483–1488.

4. Tsunada, J., Lee, J.H., and Cohen, Y.E. (2011). Representation of speech categories in the primate auditory cortex. J. Neurophysiol. *105*, 2634–2646.

5. Tsunada, J., Lee, J.H., and Cohen, Y.E. (2012). Differential representation of auditory categories between cell classes in primate auditory cortex. J. Physiol. *590*, 3129–3139.

6. Tsunada, J., and Cohen, Y.E. (2014). Neural mechanisms of auditory categorization: from across brain areas to within local microcircuits. Front. Neurosci. *8*, 161.

7. Christison-Lagay, K.L., and Cohen, Y.E. (2018). The contribution of primary auditory cortex to auditory categorization in behaving monkeys. Front. Neurosci. *12*, 601.

8. Xin, Y., Zhong, L., Zhang, Y., Zhou, T., Pan, J., and Xu, N.L. (2019). Sensory-to-category transformation via dynamic reorganization of ensemble structures in mouse auditory cortex. Neuron *103*, 909–921.e6.

9. Gentner, T.Q., and Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. Nature *424*, 669–674.

10. Jeanne, J.M., Thompson, J.V., Sharpee, T.O., and Gentner, T.Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. J. Neurosci. *31*, 2595–2606.

11. Sigala, N., and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature *415*, 318–320.

12. 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.

13. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. J. Neurophysiol. *88*, 929–941.

14. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. *23*, 5235–5246.

15. Swaminathan, S.K., and Freedman, D.J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nat. Neurosci. *15*, 315–320.

16. Nieder, A., Freedman, D.J., and Miller, E.K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. Science *297*, 1708–1711.

17. McKee, J.L., Riesenhuber, M., Miller, E.K., and Freedman, D.J. (2014). Task dependence of visual and category representations in prefrontal and inferior temporal cortices. J. Neurosci. *34*, 16065–16075.

18. Siegel, M., Buschman, T.J., and Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. Science *348*, 1352–1355.

19. Brincat, S.L., Siegel, M., von Nicolai, C., and Miller, E.K. (2018). Gradual progression from sensory to task-related processing in cerebral cortex. Proc. Natl. Acad. Sci. USA *115*, E7202–E7211.

20. Yin, P., Shamma, S.A., and Fritz, J.B. (2016). Relative salience of spectral and temporal features in auditory long-term memory. J. Acoust. Soc. Am. *140*, 4046–4060.

21. Friedrich, A., Zentall, T., and Weisman, R. (2007). Absolute pitch: frequency-range discriminations in pigeons (Columba livia): comparisons with zebra finches (Taeniopygia guttata) and humans (Homo sapiens). J. Comp. Psychol. *121*, 95–105.

22. Bizley, J.K., Nodal, F.R., Nelken, I., and King, A.J. (2005). Functional organization of ferret auditory cortex. Cereb. Cortex *15*, 1637–1653.

23. Bizley, J.K., Bajo, V.M., Nodal, F.R., and King, A.J. (2015). Cortico-cortical connectivity within ferret auditory cortex. J. Comp. Neurol. *523*, 2187–2210.

24. Atiani, S., David, S.V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S.A., and Fritz, J.B. (2014). Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. Neuron *82*, 486–499.

25. Radtke-Schuller, S. (2018). Cyto- and Myeloarchitectural Brain Atlas of the Ferret (Springer International).

26. Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., and Shamma, S.A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. *13*, 1011–1019.

27. Elgueda, D., Duque, D., Radtke-Schuller, S., Yin, P., David, S.V., Shamma, S.A., and Fritz, J.B. (2019). State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. Nat. Neurosci. *22*, 447–459.

28. Zhou, Z.C., Yu, C., Sellers, K.K., and Fröhlich, F. (2016). Dorso-lateral frontal cortex of the ferret encodes perceptual difficulty during visual discrimination. Sci. Rep. *6*, 23568.

29. Bizley, J.K., and Walker, K.M. (2010). Sensitivity and selectivity of neurons in auditory cortex to the pitch, timbre, and location of sounds. Neuroscientist *16*, 453–469.

30. 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.

31. Fritz, J.B., Elhilali, M., and Shamma, S.A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. J. Neurosci. *25*, 7623–7635.

32. Yin, P., Fritz, J.B., and Shamma, S.A. (2014). Rapid spectrotemporal plasticity in primary auditory cortex during behavior. J. Neurosci. *34*, 4396–4408.

33. Selezneva, E., Scheich, H., and Brosch, M. (2006). Dual time scales for categorical decision making in auditory cortex. Curr. Biol. *16*, 2428–2433.

34. Niwa, M., Johnson, J.S., O'Connor, K.N., and Sutter, M.L. (2012). Activity related to perceptual judgment and action in primary auditory cortex. J. Neurosci. *32*, 3193–3210.

35. 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.

36. Cohen, Y.E., Bennur, S., Christison-Lagay, K., Gifford, A.M., and Tsunada, J. (2016). Functional organization of the ventral auditory pathway. Adv. Exp. Med. Biol. *894*, 381–388.

37. Ohl, F.W., Deliano, M., Scheich, H., and Freeman, W.J. (2003). Early and late patterns of stimulus-related activity in auditory cortex of trained animals. Biol. Cybern. *88*, 374–379.

38. Russ, B.E., Lee, Y.-S., and Cohen, Y.E. (2007). Neural and behavioral correlates of auditory categorization. Hear. Res. *229*, 204–212.

39. Engineer, C.T., Perez, C.A., Carraway, R.S., Chang, K.Q., Roland, J.L., Sloan, A.M., and Kilgard, M.P. (2013). Similarity of cortical activity patterns predicts generalization behavior. PLoS ONE *8*, e78607.

40. Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. J. Neurophysiol. *100*, 1407–1419.

41. Freedman, D.J., and Assad, J.A. (2016). Neuronal mechanisms of visual categorization: an abstract view on decision making. Annu. Rev. Neurosci. *39*, 129–147.

42. Cromer, J.A., Roy, J.E., and Miller, E.K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. Neuron *66*, 796–807.

43. Fitzgerald, J.K., Freedman, D.J., and Assad, J.A. (2011). Generalized associative representations in parietal cortex. Nat. Neurosci. *14*, 1075–1079.

44. Diester, I., and Nieder, A. (2008). Complementary contributions of prefrontal neuron classes in abstract numerical categorization. J. Neurosci. *28*, 7737–7747.

45. Viswanathan, P., and Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. *25*, 1259–1269.

46. Kell, A.J., and McDermott, J.H. (2019). Deep neural network models of sensory systems: windows onto the role of task constraints. Curr. Opin. Neurobiol. *55*, 121–132.

47. Hong, H., Yamins, D.L.K., Majaj, N.J., and DiCarlo, J.J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. Nat. Neurosci. *19*, 613–622.

48. Yamins, D.L., and DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. *19*, 356–365.

49. Chaisangmongkon, W., Swaminathan, S.K., Freedman, D.J., and Wang, X.-J. (2017). Computing by robust transience: how the fronto-parietal network performs sequential, category-based decision. Neuron *93*, 1504–1517.e4.

50. Kar, K., Kubilius, J., Schmidt, K., Issa, E.B., and DiCarlo, J.J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nat. Neurosci. *22*, 974–983.

51. Jiang, X., Chevillet, M.A., Rauschecker, J.P., and Riesenhuber, M. (2018). Training humans to categorize monkey calls: auditory feature- and category-selective tuning changes. Neuron *98*, 405–416.e4.

52. Liu, S.T., Montes-Lourido, P., Wang, X., and Sadagopan, S. (2019). Optimal features for auditory categorization. Nat. Commun. *10*, 1302.

53. Lim, S.J., Fiez, J.A., and Holt, L.L. (2019). Role of the striatum in incidental learning of sound categories. Proc. Natl. Acad. Sci. USA *116*, 4671–4680.

54. Engel, T.A., Chaisangmongkon, W., Freedman, D.J., and Wang, X.J. (2015). Choice-correlated activity fluctuations underlie learning of neuronal category representation. Nat. Commun. *6*, 6454.

55. McMahon, D.B., Bondar, I.V., Afuwape, O.A., Ide, D.C., and Leopold, D.A. (2014). One month in the life of a neuron: longitudinal single-unit electrophysiology in the monkey visual system. J. Neurophysiol. *112*, 1748–1762.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Ferret (*Mustela putorius*)	Marshall Farm	https://www.marshallferrets.com/
Software and Algorithms		
MATLAB 2010-2015	Mathworks	https://www.mathworks.com/
Adobe illustrator CS4.0 & 2019	Adobe Systems	https://www.adobe.com/

LEAD CONTACT AND MATERIALS AVAILABILITY

This study did not generate new unique reagents. Further information and requests for resources and reagents should be directed to and fulfilled by the Lead Contact, Pingbo Yin (pyin@umd.edu) or Shihab Shamma (sas@isr.umd.edu)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Two adult female ferrets (*Mustela putorius*, Marshall Farms, North Rose, NY), each weighting 600-900 g, began training at one year of age. During behavioral training and physiological recording, the animals were placed on a water-control protocol in which they had restricted access to water during the week except as reward during behavior or as liquid supplements (if the animals did not drink sufficiently during the behavioral sessions), and received *ad libitum* water freely over weekends. Animal health was monitored on a daily basis to avoid dehydration and weight loss (animals were maintained above 80% of *ad libitum* weight). Ferrets were housed in pairs or trios in facilities accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC), and were maintained on a 12-hr light-dark artificial light cycle.

All animal experimental procedures were conducted in accordance with the National Institutes of Health's Guide for the Care and Use of Laboratory Animals, and were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Maryland.

METHOD DETAILS

Experimental apparatus

Ferrets were trained in a custom-built transparent Lucite testing box (18 cm width X 34 cm depth X 20 cm height), placed within a single-walled, sound attenuated chamber (IAC). A lick-sensitive waterspout (2.5 cm X 3.7 cm) stood 12.5 cm above the floor in front of the testing box, and animals could easily lick from the waterspout to obtain water through a small opening in the front wall of the box. The waterspout was connected to a computer-controlled water pump (MasterFlex L/S, Cole-Parmer), and also to a custom-built interface box that converted the licks into a TTL digital signal feed to a computer. A loudspeaker (Manger, Germany) was positioned 40 cm in front of the testing box for sound delivery during behavioral training, and the animal's behavior was streamed with a video camera, displayed graphically trial-by-trial and continuously monitored on a computer screen.

Behavioral Tasks

Animals were trained on two different classification tasks using a conditioned avoidance paradigm [30], categorizing either tone-frequencies (Tone-task), or rates of amplitude-modulated noise (AM-task). Stimuli and tasks were previously described in detail [20] in a previous version of the task in which an appetitive paradigm was used. In both tasks, animals started with a set of 6 stimuli, which were divided into three zones along the relevant feature axis. Each zone was assigned the desired behavioral meaning (one Middle NoGo-zone and two flanking Go-zones). The frequencies of the set of 6 tones were set at **100, 280, 784, 2195, 6147, 17210** Hz with equal spacing in logarithmic scale. The 6 stimuli in the AM-task consisted of white noise modulated at six rates: **4, 15, 26, 37, 48,** and **59 Hz**. Animals were trained for one session a day (100-200 trials), after 1-2 days of habituation in the testing box in which they learned to lick on the waterspout for water.

A typical trial started by turning on the water pump (at a flow rate of ∼1.2 ml/min) upon which the animals began licking the waterspout for water. A sound (0.75 s duration) was presented 1.5 s after water flow commenced. Ferrets learned to stop licking the waterspout (Hit) when a NoGo sound was presented in order to avoid a mild tongue shock (the 0.4 s shock window began 0.1 s after offset of a NoGo sound (NoGo-trial, Figure 1A)). The animals learned to continue licking the waterspout on trials with a Go sound (Go-trial,

Figure 1A). The trial ended 2.0 s later after sound offset by turning the water pump off (i.e., each trial was 4.25 s long). Animals learned to stop licking the waterspout for 1.0 s in order to initiate the next trial. False alarms (in which animals incorrectly stopped licking the waterspout in the response window during Go-trials) resulted in a 5-10 s timeout penalty, which was applied before the next trial. The training session ended when the animal was no longer thirsty and did not lick the waterspout for 3 consecutive Go trials. Task performance was measured using the discrimination rate (DR), which was defined as:

$$\textbf{DR} \; = \; \textbf{Hit rate} \; * \; (\textbf{1 − False Alarm rate}) \; * \textbf{100}\%.$$

Animals were first trained on the Tone-task to respond according to tone frequencies (discriminating frequencies in a Go zone from frequencies in two flanking No-Go zones). The behavioral criterion was defined as the animal performing three consecutive sessions (over 100 trials) with DR > 40.0%. After reaching this criterion, animals received an additional 1-2 weeks of training to further consolidate task performance.. Animals were then trained on the AM-task and learned to respond to AM rates (discriminating AM-rates in a Go zone from AM-rates in two flanking No-Go zones). During early stages of training, an intensity cue (for NoGo sounds) was used to help shape the animals' response to the correct sound category. The Go sounds were presented at lower sound intensity levels than the NoGo sounds (up to an initial 60 dB relative attenuation as indicated by the gray scale of filled markers in Figure S1A). The DR was computed for each training session, and was used to guide the adjustment of the attenuation level applied to Go sounds. The intensity cues were gradually removed during training. The intensity cues were not necessary for training on the AM-task, and were not used since the animals had already learned the basic task paradigm following Tone-task training.

Headpost Implant Surgeries and Neurophysiological recordings

After reaching behavioral criterion on both tasks, a stainless steel headpost was surgically implanted on the ferret skull under aseptic conditions while the animals were deeply anesthetized with 1%–2% isoflurane. After recovery from surgery (about 2-3 weeks), animals were placed in a double-wall sound proof booth (IAC) to habituate to head restraint in a customized head-fixed holder. Animals were then re-trained on the head-restrained version of the task, where a mild electric shock was delivered to the tail on "miss" trials. Neurophysiological recordings were begun when animals regained a criterion level of task performance in three sequential behavioral sessions with DR greater than 0.4 in both Tone- and AM- tasks.

To gain access to the brain for neurophysiological recording, small craniotomies (1-2 mm in diameter) were made over auditory and frontal cortex. Recordings were conducted using 4 tungsten microelectrodes (2-5 MΩ, FHC) simultaneously introduced through the same craniotomy and controlled by independently moveable drives (Electrode Positioning System, Alpha-Omega). Raw neural activity traces were amplified, filtered and digitally acquired by a data acquisition system (AlphaLab, Alpha-Omega). Multiunit neuronal responses were monitored on-line (including all spikes that rose above a threshold level of 4 standard deviations of neural signals). Single-units were isolated offline by customized spike-sorting software, based on a PCA and template-matching algorithm (Meska-PCA, NSL).

During neurophysiological recordings in auditory cortex, a 9-stimulus set of frequencies or AM rates was used in each task (Figure 1B) instead of the 6-stimulus training set. However, the 6-stimulus set was used in most of the recordings from dlFC. There was no difference in behavioral performance using either the 6 or 9-stimulus set.

Experimental Procedures and Stimuli

At each recording site, we first used 0.5 s random tones pips of varying frequency (spanning up to 8 octaves ranges, 4 tones/octave) and intensity (20-70 dB range, 10 dB increments) to characterize neuronal properties such as characteristic frequency (CF), sharpness of tuning (by computing Q-10 dB), and minimal response latency etc. Temporally-orthogonal ripple combinations (TORCs) were also used to construct a neuron's STRF (spectrotemporal receptive field). The typical behavioral neurophysiological session at each site included at least three epochs: active task-engagement during performance of either the Tone-task or AM-task, passively listening to the identical task stimulus set before and after task performance. In each epoch, all stimuli were repeated 10-15 times. One or both Tone- and AM-task behavioral physiological sessions were carried out in each recording site depending on the animals' motivation and stability of the neuronal activity.

All acoustic stimuli were ramped with 5 ms rise-fall time and presented at 60-70 dB SPL, with the exception of the tone pips. The sounds were digitally generated using custom-made MATLAB (The MathWork, Natick, MA) functions at 40 kHz sampling rate and were converted at 16-bit resolution through a NI-DAQ card (PCI-6052E), then amplified (MA-3, Rane) and delivered through the free-field loudspeaker (Manger, German) located ∼1.2 m in front of animals' head. All behavioral performances were controlled and monitored through a custom-built MATLAB GUI. All trial events and behavioral responses were recorded and stored in the computer for further analysis and assessment of behavioral performance.

Localization of recording sites

In ferrets, auditory cortex is located ∼13-16 mm anterior to the occipital crest, the most distinctive and most easily accessible skull landmark [25], and ∼12 mm lateral to the skull midline (Figure S2A). Initial recordings began through a small craniotomy placed above auditory cortexA1 with these coordinates. The neuron's characteristic frequency (CF; defined below) was measured at each electrode penetration. Based on the gradient of the CFs, the existing craniotomy was further expanded in order to cover both primary (A1) and more lateral secondary (dPEG) auditory regions. The CFs obtained from all penetrations at different stages of the craniotomy were aligned together to form a tonotopic map for each animal (Figure S2B), referring to two landmarks, placed (in the bone cement)

on either side of the craniotomy. The locations of A1 and dPEG were confirmed, based on their tonotopic organization [22]. The CF gradient in A1 runs from high to low frequency in a dorsal-lateral direction, while showing a frequency reversal of the CF gradient at the low frequency border with dPEG. Thus, the lowest CF contour line was used as the border line to divide A1 and dPEG.

Recordings in dlFC began through a small craniotomy ∼25-30 mm anterior to the occipital crest, corresponding to a frontal cortex region including the anterior part of the ASG/PMC and the most anterior dlFC area or PRG/dPFC as shown by the green shaded regions in Figure S2A. The craniotomy was gradually expanded based on the response from the electrode penetrations. Locations of the recording electrodes relative to two reference marks placed in the bone cement surrounding the craniotomy was measured for later alignment of all electrode penetration sites in a frontal cortex craniotomy.

In order to histologically identify the recording areas, following completion of recordings, two penetration sites in each identified cortical area were selected to make either an iron electrolytic deposit (in one animal, Gong) or by placement of a small amount of HRP (in the second animal, Guava). The sites in auditory cortex were selected based on the tonotopic maps (Figure S2B) such that one HRP injection was located in A1 and another in dPEG. The two selected sites in dlFC were identified by responses evoked during task performance. The HRP deposits were made by fine glass pipette with small amounts of HRP on its tip. The iron deposits were made by passing a small constant current through a stainless-steel electrode (5∼7.5 μA for 300 s). The confirmation of injection sites in 50 μm coronal brain slices of the fixed ferret brains was done by histological localization of DBA (diaminobenzidine) reaction to visualize HRP, or by Prussian blue reaction to highlight the iron deposits. The dlFC sites were identified from iron deposits as shown in Figure S2C (sections 1 and 2). The HRP sites were successfully identified in auditory cortex, and confirmed one recording site was located in A1 (sections 4) and another in dPEG (sections 3) as shown in Figure S2C.

Data analysis
Behavioral assessment
During task performance, a behavioral response in a given trial was assessed based on the animal licking pattern in two time windows: before the start of the shock time window (Figure 1A) and during the 0.4 s shock time window. In order to score a given trial, the animals were required to lick in the time window before the shock window. Refraining from licking during the shock window was a correct response to a NoGo sound and was defined as a hit, whereas cessation of licking during the shock window was an incorrect response to a Go sound and was defined as a false alarm. The hit rate (HR) and false alarm rate (FR) were computed in each training session. The basic metric used in quantify the behavioral performance based on HR and FR was the discrimination rate (DR), defined as: DR = HR*(1-FR). The criterion for successful behavioral performance in distinguishing Go and NoGo sounds in a given training session was a DR > = 0.4. The criterion for task mastery was a DR > = 0.4 in three consecutive training sessions (> 100 trials per session) in the absence of any intensity cues for targets (i.e., all stimuli were presented at equal loudness). Animals continued training for a few weeks in order to consolidate their performance. Animals displayed a very stable behavior performance during neurophysiological recording in both Tone and AM-task (Figure 1C, left panels).

Computing characteristic frequency (CF)
In auditory cortex, neurons' CF were obtained by analyzing their response to tones pips with varying frequency (spanning up to 8 octaves ranges, 4 tone/octave) and intensity (20-70 dB range, 10 dB increment). A two-dimensional response matrix (frequency x intensity) was formed by taking the averaged evoked response to tones (for the first 100 ms after tone onset) at each frequency and intensity level. The response matrix was first normalized, by subtracting the mean and then dividing by the standard deviation of the baseline activity (averaged over a 100 ms window before tone onset). A 2-dimensional spline interpolation was applied to the normalized response matrix to obtain finer resolution of frequency (1/12 octave) and intensity (1 dB). The longest iso-response contour line in the frequency x intensity space was defined as the neuron's tuning curve (TC), and the frequency corresponded to the lowest intensity on TC was the neuron's *CF.* The site CF for each penetration site was represented by taking the median value of all isolated single units in the penetration. The site CFs were used to generate the tonotopic map, after aligning penetration site locations in the same craniotomy surface relative to the two landmarks placed around the recording region (see Figure S2B). In both animals, from dorsal to lateral, the CF gradient runs from high to low and then mirror reverses from low to high. The primary (A1) and secondary (dPEG) auditory cortical areas were identified based on the CF tonotopic maps.

ROC Analysis
The basic metrics used to determine whether a single-unit's response could discriminate between individual stimulus pairs (distance) or stimulus categories (Go versus NoGo) or behavioral choices (licking versus stop licking) were derived from signal-detection-theory-based ROC analysis. To compute the ROC area, we first computed the averaged spike rate evoked in a given observation time window in each trial. The proportion of responses to each of the paired conditions having a firing rate greater than a given criterion level was calculated, and this measurement was repeated for criterion levels along the full range of the evoked firing rate. A neural ROC was formed by plotting the paired probability functions, and the area under the ROC represented the neural discriminability between the paired conditions. ROC area has a value ranging from 0 to 1.0, and is symmetric around 0.5. A value at 0 or 1.0 indicated that the ROC-based ideal observer had a perfect prediction of the paired condition and a value at 0.5 indicated chance prediction. Several subsequent neural metrics were derived from this ROC analysis.

Categorical Index (CI)
CI is defined as: **CI = BC − WC**; where BC (between category) denotes averaged ROC areas computed from all selected stimulus pairs in which one stimulus belongs to the Go and another to the NoGo category; WC (within category) denotes the averaged ROC areas computed from all selected pairs in which the two stimuli were from the same category (either Go or NoGo). The neural

distance between any stimulus pair was represented by the absolute value of (ROC-0.5). Therefore, CI had a range from −0.5 to 0.5. Positive CI indicates a categorical response, while a negative CI indicates a non-categorical response. The permutation tests were performed to evaluate the significance of the obtained CI for each neuron in each passive and active epoch.

Sensory index (SI)

SI was used to quantify a neuron's response selectivity to stimuli, defined as the proportion of variance explained by individual stimulus over the total variance:

$$SI = SSB/SST$$

Where SSB denotes the sum of squares between individual stimuli, and SST denotes the sum of squares total. SI yields a range from 0 to 1.0, a value near 1.0 if all variance could be explained by individual stimuli (for a highly selective neuron), or a value near 0.0 if not (indicating a neuron that responds evenly to all stimuli, or has no response at all). Since a neurons' SI is positively correlated with a neuron's tuning bandwidth, it was also used as an indicator of the sharpness of its response profile to the task stimulus sets. The significance of SI value in each passive and active epoch was tested with One-way ANOVA.

Simulation of neuronal responses

Auditory responses were simulated using a bell-shape response profile across frequencies resembling a Gaussian distribution, which was parameterized for the evoked response (R) of a neuron to a given frequency (x) as:

$$R = \frac{1}{\sqrt{2\pi BW^2}} e^{-\frac{(x - BF)^2}{2BW^2}} + \varepsilon$$

Where BF is the neuron's best frequency, BW its bandwidth, and ε denotes the state-dependent noise. The CI and SI were then computed based on the simulated responses for different BF and BW ranges (e.g., between 0.25-5 octaves in 0.1 steps).

Population analysis

Population analysis included neurons with at least 10 repetitions for each stimulus to form a pseudo-population representation in each cortical area and task. For each neuron, 10 trials were randomly picked from all available trials (up to 15 trials) for each stimulus and sorted [by stimulus repetitions]. The trials from all neurons were concatenated to form a 3-dimensional population matrix ([stimulus x repetition] x time bins x neurons) as illustrated in Figure S6A. The trial data were moving-averaged with a 50 ms time window at 25 ms steps. The data from each neuron were corrected by subtracting their baseline activity, and then normalizing to the maximum of the absolute responses among all recording epochs (two passive and one active). This procedure was repeated 10 times to obtain 10 population matrices. A mean population representation matrix was obtained for each area by averaging over the 10 matrices to reduce noise. Principal component analysis (PCA) was applied over the averaged population-response matrix by treating neurons as variables with [bins x trials] as observations (rearranged as a two-dimensional matrix: [bins x trials] x neurons). PC coefficients (PC_{coeff}) were obtained by applying PCA on the data matrix from the active epoch. The PCA scores (PC_{sr}) for individual population matrix at each epoch and moment were computed through the product of the PC_{coeff} and the population data matrix at each time moment (POP_t): $PC_{sr} = POP_t * PC_{coeff}$. The dynamics of task stimulus responses were then represented in PC space by using the first three PC's in the passive and active task conditions (Figure S6B). Euclidean distance between pairs of task stimulus responses was computed based on their representation in PC space.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

The permutation test was performed to test the significance of the obtained CI for each neuron in each of passive and active epoch. We first computed the shuffled CI (CIsh) by randomly assigned stimulus labels for trials within each pair conditions, and this procedure was repeated 1000 times. The p value is the fraction of those 1000 values of CIsh that are greater than or equal to the observed CI. The neuron is assumed to have significant more categorical driven responses than purely sensory driven response if $p \leq 0.05$, or more sensory driven responses than purely sensory if $p > = 0.95$. Otherwise, the neuron was thought to be equally driven by category and sensory or lack of response at all (which is true for most of neurons from dlFC during passively listening).

One-way ANOVA was performed to test the null hypothesis that means of response from each stimulus are equal, which use for evaluating the significance of the obtained SI for each neurons in each passive and active conditions.

The significance of the population dynamics in CI or SI was assessed by One-sample t test to test the null hypothesis that the means of the CI or SI from each population are no difference with their baselines in each condition (passive and active). While the changes of the population dynamics in CI or SI or PSTH between active and passive conditions in each population were assessed by Wilcoxon signed rank test for aero median. The tests on the dynamics were performed in 100 ms time window which began from 200 ms before stimulus onset with 25 ms sliding window beyond 600 ms beyond offset of the stimulus. The significance of the bin was defined as $p < 0.05$ in three consecutive bins from the test to compensate the repeated-measure bias.

For population analysis, the differences in changes of the averaged Euclidean distance of WC and BC between passive and active conditions were assessed by Wilcoxon rank sum test to test if the changes in WC and BC come from distribution with equal medians, or the divergence across categorical boundaries (with BC increasing) and not the convergence within a category (with no or less changes in WC) during task performance.

For neuron populations with two tasks, the binomial test was employed to test if the observed joint probability of positive CI in both Tone and AM-task differs from the expected likelihood from random combination, which has chance likelihood round $0.5 \times 0.5 = 0.25$.

DATA AND CODE AVAILABILITY

The raw data related to the current study have not been deposited in a public repository because the complexity of the customized data format and the amount in size. Source data and relevant MATLAB codes for generating figures in the paper are available by reasonable request to Lead Contact, Pingbo Yin (pyin@umd.edu).