Systems/Circuits

Task Engagement Improves Neural Discriminability in the Auditory Midbrain of the Marmoset Monkey

Luke A. Shaheen,[1] Sean J. Slee,[2] and Stephen V. David[1]

[1]Oregon Hearing Research Center, Oregon Health and Science University, Portland, Oregon 97239-3098, and [2]Micro Systems Engineering Inc, Lake Oswego, Oregon 97035

While task-dependent changes have been demonstrated in auditory cortex for a number of behavioral paradigms and mammalian species, less is known about how behavioral state can influence neural coding in the midbrain areas that provide auditory information to cortex. We measured single-unit activity in the inferior colliculus (IC) of common marmosets of both sexes while they performed a tone-in-noise detection task and during passive presentation of identical task stimuli. In contrast to our previous study in the ferret IC, task engagement had little effect on sound-evoked activity in central (lemniscal) IC of the marmoset. However, activity was significantly modulated in noncentral fields, where responses were selectively enhanced for the target tone relative to the distractor noise. This led to an increase in neural discriminability between target and distractors. The results confirm that task engagement can modulate sound coding in the auditory midbrain, and support a hypothesis that subcortical pathways can mediate highly trained auditory behaviors.

Key words: Inferior Colliculus; auditory discrimination; subcortical plasticity

---

Significance Statement

While the cerebral cortex is widely viewed as playing an essential role in the learning and performance of complex auditory behaviors, relatively little attention has been paid to the role of brainstem and midbrain areas that process sound information before it reaches cortex. This study demonstrates that the auditory midbrain is also modulated during behavior. These modulations amplify task-relevant sensory information, a process that is traditionally attributed to cortex.

Introduction

Sound encoding by the auditory system is behavior-dependent. Changes in behavioral state, such as task engagement, arousal, attention, and motor activity, can modulate auditory processing (Fritz et al., 2003; Otazu et al., 2009; Lee and Middlebrooks, 2011; Niwa et al., 2012; Schneider et al., 2014). Most previous work on task-dependent changes has focused on the cortex, and the midbrain has been viewed as relatively static. However, a small number of studies have reported that engaging in an auditory behavior can modulate sound-evoked activity in the inferior colliculus (IC) (Ryan and Miller, 1977; Metzger et al., 2006; Slee and David, 2015) and thalamus (Jaramillo et al., 2014; Williamson et al., 2015). Moreover, lesion and inactivation studies suggest that the auditory cortex (ACtx) may be unnecessary

for performing some auditory behaviors (Guo et al., 2017a). Together, this work suggests that subcortical pathways can perform the necessary computations to transform auditory inputs into behavioral decisions.

Both anatomic and physiological evidence suggests that behavioral state can influence processing in the auditory midbrain (AM). In the ascending pathway, the IC is a convergent site for several brainstem nuclei, each specialized to process different sound features (Malmierca, 2004). The IC also receives a substantial top-down projection from cortex and numerous neuromodulatory inputs (Winer, 2005; Hurley, 2019). Thus, it is anatomically well positioned to integrate information about internal state into sound processing. Previous work in the ferret showed that engaging in a tone-versus-noise discrimination task suppressed responses to distractor sounds in both central IC (ICC) and noncentral IC (NCIC) (Slee and David, 2015). The magnitude of this suppression was similar to that found in cortical neurons during a similar task, suggesting that task-dependent suppression may, in part, be inherited from the IC. However, performing a sensory discrimination behavior requires an emergent representation of task categories, distinct from peripheral spectrotemporal representations. Previously, enhanced discriminability of task categories has only been reported in cortex (Tsunada et al., 2011; David et al., 2012; Shepard et al., 2015; Christison-Lagay and Cohen, 2018; Elgueda et al., 2019; Liu et

al., 2019; Xin et al., 2019), and it is not clear whether the same changes are observed in the IC.

To study the impact of task engagement on auditory neural discriminability, we moved to the marmoset monkey auditory model. Marmosets are appealing for studies of auditory behavior because they have a rich vocal repertoire (Agamaite et al., 2015; Eliades and Tsunada, 2019), their ACtx is well studied (Lu et al., 2001; Wang, 2018), and they have a core-belt-parabelt cortical organization homologous to that of humans (Hackett, 2011). We found the marmoset to be more flexible in the timing of its behavioral responses, allowing us to separate neural responses to target and distractor stimuli from confounding motor activity associated with target responses. We recorded single-unit activity in marmoset IC while they performed a tone-in-noise detection task, and compared neural responses to task stimuli during behavior and during passive presentation. In contrast to ferret, a smaller percentage of ICC neurons were modulated by task engagement, and responses tended to be enhanced rather than suppressed. However, in a substantial percentage of midbrain neurons outside the ICC, task engagement enhanced discriminability of target from reference sounds. This change was explained largely by increased responses to target sounds.

Materials and Methods

Surgical procedure. All procedures were approved by the Oregon Health and Science University Institutional Animal Care and Use Committee and conform to the National Institutes of Health standards. Two young adult marmosets (1 female, 1 male) were obtained from an animal supplier (Wisconsin National Primate Research Center). Normal auditory thresholds were confirmed by measuring auditory brainstem responses. Marmosets were then gradually habituated to a semi-restraint device over a 4 week period as described previously (Slee and Young, 2013). After habituation, a sterile surgery was performed (isoflurane anesthesia, 0.5%-2%) to mount a post for head fixation and to expose a small portion of the skull for the neurophysiological recording. The headpost was surrounded by Charisma composite, which bonded to the skull and also to a set of stainless-steel screws embedded in the skull. During a 2 week recovery period, animals were treated daily with antibiotics (Baytril, 2.5 mg/kg), and the wound was cleaned and bandaged. Analgesics (buprenorphine, 0.005 mg/kg; Tylenol, 5 mg/kg; and lidocaine, topical 2%) were given to control pain.

Acoustics and stimuli. All behavioral and physiological experiments were conducted inside a custom double-walled sound-isolating chamber (Professional Model, Gretch-Kenn) with inside dimensions of 8' × 8' × 6' (l × w × h). A custom second wall was added to the single-walled factory chamber by building a wooden frame to which ¾ inch MDF board was attached. The air space between the outer and inner walls was 1.5 inches. The inside wall was lined with 3 inch sound-absorbing foam (Pinta Acoustics). The chamber attenuated sounds > 2 kHz by >60 dB. Sounds from 0.2 to 2 kHz were attenuated by 30-60 dB, falling off approximately linearly on a logarithmic plot of level versus frequency.

Sound presentation and behavior were controlled by custom software written in MATLAB (The MathWorks). Source code is available at https://bitbucket.org/lbhb/baphy/src/master/. Sounds were digitally generated, converted from digital to analog (100 kHz, National Instruments model PCI-6229), and presented over a Manger sound transducer (model W05) driven with a Crown amplifier (model D-75A). The speaker was placed 1 m from the animal's head 30° contralateral to the IC under study. Sounds were calibrated using a ½ inch microphone (B&K, model 4191). All stimuli were presented with 10 ms linear onset and offset ramps.

The stimuli used in this study were pure tone targets (sine waves) masked by 1/*f*-spectrum noise and random spectral shape (RSS) distractor references (see Fig. 1*A*) (Yu and Young, 2000). RSS stimuli provide an efficient means to construct linear and/or nonlinear models of a neuron's spectral encoding (see Data analysis). In prior work, we used

temporally orthogonal ripple complexes (TORCs) as reference noises (Slee and David, 2015). Both stimuli span ~5 spectral frequency octaves. While TORCs contain complex temporal dynamics that can be used to reconstruct a neuron's temporal modulation tuning (Klein et al., 2000), RSS noise spectra do not change over time (Young and Calhoun, 2005). Since spectral and temporal receptive fields are separable in most ICC neurons (Qiu et al., 2003), and since task engagement generally affects spectral but not temporal receptive fields (Fritz et al., 2003), we used RSS noise for this study.

The RSS stimuli were similar to those used in previous studies (Young and Calhoun, 2005). Each stimulus consisted of a sum of tones spaced logarithmically at 1/64th octave with randomized phase. The stimuli were arranged in 1/8th octave bins (8 tones of the same level) that spanned 1.25-33.125 kHz (see Fig. 1*B*). The level of the tones in the bin centered at frequency f was fixed at $S(f)$ dB, drawn from a Gaussian distribution with zero mean and SD of 12 dB relative to a reference sound level. Targets were masked by broadband, 1/*f*-spectrum noise that was constructed using the same procedure as for the RSS noises, but with $S(f) = 0$ dB at all frequencies (i.e., the average reference sound level). Because frequency bins are logarithmic, the noise has a 1/*f* spectrum. In order to discourage marmosets from using timbre cues to detect targets, the 1/*f*-spectrum noise used to mask targets was also presented randomly interleaved with RSS stimuli (1/10 distractors), serving as a catch stimulus. RSS and catch stimuli were scaled by a single factor such that the average level across the set was 50 dB SPL.

Behavioral training. After each marmoset recovered from surgery, its access to water was limited 24 h before training began. Each marmoset normally received 20 g of Mazuri Callitrichid High Fiber Diet (5M16) with Harlan Vitamin D Premix, mixed to 50% water content, twice per day. During training times, the food was partially dehydrated (~35% moisture content), and water bottles were removed from the home cage. Juice rewards (Strawberry Nesquik, 27.5 g dissolved in 250 ml distilled water) were delivered through a spout ~5 mm away from the marmoset's mouth. Water delivery was controlled electronically with a solenoid valve. Licking was monitored by breaking a beam formed by an infrared LED and photodiode placed across the spout.

At the beginning of training for Marmoset C, a 1 s pure tone (50 dB SPL) was paired with a small juice reward. The frequency of this target tone was held constant within a session but varied from session to session. If the animal successfully licked during the presentation of the tone, additional juice was delivered. The marmoset quickly learned to associate the tone with a reward. At this point, the juice reward was delivered only if the marmoset licked within the target response window: 0.5-1.5 s following tone onset. Next, a random number (2-5) of 1 s RSS stimuli were presented before the tone with a 1 s interstimulus interval. A false alarm was recorded if the marmoset licked the spout before the target response window and was punished with a 4 s timeout during which the chamber lights were extinguished. A miss was recorded if the animal did not respond before the end of the target response window. Initially, the RSS stimuli were presented at 20 dB SPL (a 30 dB signal-to-noise ratio). The level was gradually increased to 50 dB SPL (matching the level of the target tone). Finally, the masking noise was added to the target tone (0 dB signal-to-noise ratio). Training was complete once the marmoset learned to complete this task with an average false-alarm rate of <25%. The entire training procedure took 3-4 weeks. For Monkey F (64 of 113 neurons), the same shaping procedure was used. However, to increase the diversity of stimuli presented during behavior, the duration of all stimuli was reduced from 1 to 0.3 s, the interstimulus interval was reduced from 1 to 0.7 s, and the target response window was shifted from 0.5-1.5 s to 0.3-1 s following target onset. No significant differences were observed in performance between animals (see Fig. 1). To control for possible differences in neural response dynamics, only the first 0.3 s of sound-evoked activity was analyzed for data collected from both animals (see below).

Electrophysiology. At the beginning of the neurophysiology experiments, a small (~1-mm-diameter) craniotomy was made in the skull, approximately dorsal to IC. The location of the hole was based on stereotaxic coordinates as well as superficial landmarks on the skull (e.g., bregma) marked during surgery. The exposed recording chamber
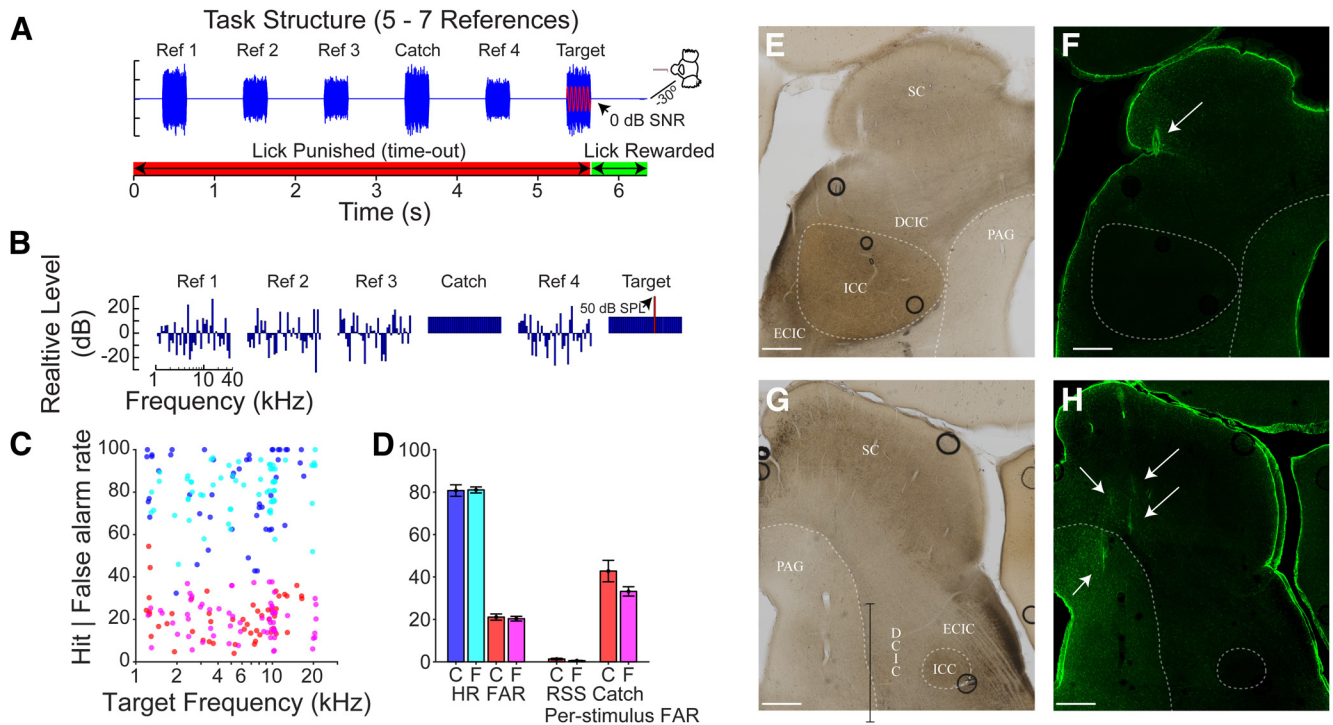
Figure 1. Tone detection task and behavioral performance. **A**, Schematic of example tone detection task trial. **B**, Example power spectra of RSS references, 1/f spectrum (catch) references, and targets (tone in 1/f-spectrum noise). **C**, Per-trial hit rate (blue represents Animal C; cyan represents Animal F) and false-alarm rate (red represents Animal C; magenta represents Animal F) as a function of target frequency. Only sessions during which physiology data were recorded are included. x axis positions were jittered slightly to improve visualization. **D**, Left, Mean ($\pm$SEM) hit rate and false-alarm rate, computed per trial. Right, Mean ($\pm$SEM) false-alarm rate for RSS (left) and catch (right) references, computed per stimulus and distinguishing between RSS and catch references. **E–H**, Coronal slices from Animal F stained with cytochrome oxidase (**E,G**) and GFAP (**F,H**). Scale bars, 500 μm. **E**, **F**, Slice is ~1.4 mm rostral to the caudal pole of the IC. **G**, **H**, Slice is 0.7 mm further rostral. **F**, GFAP marks location of single tetrode guide tube in the left superior colliculus, just dorsal to the ICC (white arrow). Reactivity to the electrode was visible in more posterior slices (data not shown). **H**, GFAP reveals four track marks dorsomedial to the right IC created by a 4-shank chronic tetrode implant (white arrows). The two GFAP-positive puncta at the top of the image curved in more rostral slices, and therefore do not mark electrode locations. **G**, Thin black line indicates the approximate depth range of sound-responsive neurons encountered with this implant (11-12.5 mm).

surrounding the craniotomy was covered with polysiloxane impression material (GC America) between recording sessions; and after many penetrations (usually > 30), the hole was filled with a layer of bone wax and dental acrylic before another craniotomy was made to provide access to other regions of the IC on the same hemisphere. Multiple craniotomies were performed on both hemispheres to target different subfields of the IC. After experiments were completed, animals were killed and perfused for histologic evaluation.

On each recording day, one tungsten microelectrode (FHC or A-M Systems, impedance 1-5 MΩ) or one tetrode (Thomas Recording; 1-2 MΩ) was slowly advanced through the craniotomy with a motorized microdrive (Alpha-Omega). The electrode was positioned (Kopf Instruments) approximately in the frontal plane at angle ±10° ML and ±10° AP from vertical. Depending on the angle of the dorsal approach to the IC, the electrode traversed 9-11 mm of brain tissue before reaching the IC.

In Marmoset C, all data were collected using acutely inserted electrodes (N = 49). The first 21 of these neurons were collected using a stimulus paradigm that did not include catch references. In Marmoset F, data from 16 neurons were collected with acutely inserted electrodes. Data from 32 neurons were collected from the right IC using a chronically implanted tetrode array (Neuralynx 5-drive). The drive contained three tetrodes and one tungsten electrode, each sheathed in 540 μm OD stainless-steel guide tubes, which were arranged in a square pattern with 700 μm center-to-center spacing. The array was implanted under ketamine/xylazine anesthesia. Electrodes were advanced slowly over the course of a month, generally 50 μm per day. Auditory responses were encountered on all four electrodes at depths 11-12.5 mm ventral to the cortical surface, but clear tonotopy was not. Given the margin of error on depth estimation because of scar tissue overlying the cortex, these depths are consistent with neuron locations at or near the IC (Hardman

and Ashwell, 2012). In one electrode, onset responses to room lights were found 10 mm ventral to the cortical surface, which is consistent with the superior colliculus being located at this depth. This provides further evidence that the auditory-responsive neurons 1 mm ventral were at a depth consistent with the IC. Histology revealed that these electrodes passed rostromedial to the IC, likely sampling from neurons in the dorsal cortex of the IC (DCIC) and the lateral periaqueductal gray (PAG; see Fig. 1G,H; see Histology and track labeling). Subsequently, data from 12 neurons were collected from the left IC using a single chronically implanted tetrode (TSD-2, Thomas Recording). To improve targeting, this electrode was implanted while the animal was awake. GFAP immunoreactivity at the ventral extent of the guide tube, just dorsal to the ICC, and a tonotopic progression of best frequencies confirmed that this electrode sampled ICC neurons (see Fig. 1E,F).

Stimulus presentation, animal monitoring via video camera, and electrode advancement were controlled from outside the sound booth. Only well-isolated single neurons were studied. Raw neural signals were bandpass filtered (0.3-10 kHz), amplified (10k, A-M Systems, 1800 or 3600 AC amplifier), digitized (20.83 kHz, National Instruments, PCI-6052E), and stored on a computer for offline analysis (details below). Recording sessions were terminated after 2-4 h, or earlier if the animal showed signs of discomfort.

Neurons were isolated using pure tones and/or wideband noise bursts (50 ms duration, 4 Hz) of variable level. Upon isolation, a target frequency was chosen to match the neuron's best frequency (BF). In cases where neurons did not respond to tones (some AM neurons), target frequency was set to 1.25, 2.5, 5, or 10 kHz. For tetrode recordings, if isolated neurons had different best frequencies, the target was chosen so it matched one neuron, and additional blocks with a target matching the other neuron were played as time permitted. To prevent long-term learning effects, target frequencies were varied widely from day to day

(often within the same day). The mean of a 5 d moving range of target frequencies the animal experienced was three octaves. This is comparable to the mean ferrets experienced in Slee and David (2015) (four octaves). Marmosets then listened passively while the exact stimuli used during behavior were played (passive block). The lickspout was present, but no reward or punishment was given if the marmosets licked. Following this passive recording, a short juice reward was paired with the target tone to cue the marmoset, and neural responses were collected while the marmosets behaved as described above (active block). Because in some cases new neurons became isolated during the active block, in 6 of 113 of the neurons presented here, the passive block was recorded following the active block. During the active block, false alarm trials were repeated on the next trial, and miss trials were repeated later in the block (inserted into a randomized location in the list of trials to be played). These trial repeats ensured that matched sets of responses during active hit trials and passive trials were collected. Marmosets rarely attempted to engage in behavior on passive blocks. If they did engage, they usually stopped licking after a few trials, presumably because they observed it had no effect. On average, they licked 0.3 times per trial during passive blocks, but 5.5 times per trial on active blocks. The average ratio of paired passive block-over-active block licks was 0.06.

Recordings were made in both the central nucleus (ICC) and regions surrounding the ICC (external and/or DCIC). Neurons likely recorded from the ICC were classified using the following criteria: (1) were recorded within the tonotopic map, (2) had strong responses to pure tones, and (3) had short latencies consistent with previous studies ($\sim$5-20 ms). Physiologic identification of neurons belonging to noncentral divisions of the IC is more challenging. Therefore, all neurons that did not meet these criteria were grouped into a single class labeled AM. For some analyses, AM neurons were separated based on latency to the maximum firing rate following sound onset. For a subset of penetrations, we were able to relate recording sites to the targeted region of IC (see Histology and track labeling).

Behavioral analysis. Behavioral performance during each trial of the detection task (2-5 references, 1 target) was scored as a hit (target response), false alarm (response preceding the target response window), or a miss (no response). The per-trial hit rate was calculated as the number of hits divided by the sum of hits and misses. Similarly, the per-trial false-alarm rate was the number of false alarms divided by the total number of trials. Per-stimulus false-alarm rates were also calculated, separately for RSS and catch references, as the number of false alarms to that stimulus, divided by total number of presentations of that stimulus. The first lick time was used to identify the stimulus that caused a false alarm using a time window *re* that stimulus' onset that matched the target window (i.e., 0.5-1.5s for Animal C and 0.3-1s *re* target onset for Animal F).

Neurophysiological spike extraction. Putative spikes were extracted from the continuous signal by collecting all events $\geq$4 SDs from zero. Spikes were detected from the events using principal component analysis and *k*-means clustering (David et al., 2009). Stability of single-unit isolation was verified by examining waveforms and interval histograms. If isolation was lost during a behavioral block, only activity during the stable period was analyzed. For tetrode recordings, the *Catamaran* clustering program (kindly provided by D. Schwarz and L. Carney) was used to separate single units from the electrode signal (Schwarz et al., 2012). In both cases, single units were defined based on visual inspection of traces and by having <1% of interspike intervals <0.75 ms.

Effects of task engagement on discrimination and mean responses. For each presentation of a given stimulus (target or reference), average responses were calculated as the driven rate (absolute rate – spontaneous rate) over 90-300 ms after stimulus onset (regardless of stimulus duration, which was 1 s for Animal C, 0.3 s for Animal F). A single spontaneous rate was calculated for each active and passive block. For the active block, only responses during hit trials were included, which were then compared against responses to the same set of stimuli collected during the passive block. To measure behavior-induced changes in neural discriminability, separation between the distributions of target and reference responses was quantified using a standard neural discrimination index, d' (Green and Swets, 1966) as follows:

$$d' = \frac{\mu_t - \mu_r}{\sqrt{\frac{1}{2}\left(\sigma_t^2 + \sigma_r^2\right)}},$$

where μ and σ are the mean and SD, respectively, of responses to targets (t) or references (r). Differences between active and passive responses for either target or references were quantified similarly by computing the z score, as $\dfrac{\mu_a - \mu_p}{\sqrt{\frac{1}{2}\left(\sigma_a^2 + \sigma_p^2\right)}}$, where subscripts a and p indicate responses during the active and passive state, respectively.

Multiple linear regression was used to determine how strongly each component of the d' metric contributed to discriminability changes. Regressors were the active-passive difference in target mean, reference mean, target SD, and reference SD. All inputs were divided by a common normalizing factor, the SD of responses pooled across targets and references and both active and passive conditions: $\sigma_p = \sqrt{\dfrac{\sum_{i=1}^{N}(n_i - 1)\sigma_i^2}{\sum_{i=1}^{N}(n_i - 1)}}$, where $N = 4$ and n_i and σ_i are the number of elements and SD of each distribution. Leave-one-out cross-validation was used to estimate coefficients using subsets of the neural population and evaluate errors on the held-out neurons.

Effects of task engagement on global gain and offset. Behavior-dependent changes in overall excitability were determined by comparing driven firing rates between behavior conditions (Slee and David, 2015). To obtain unbiased measures of global gain change, the average firing rate to each reference sound was calculated in 30 ms bins for both passive and active blocks over a time window 90-300 ms relative to onset (7 bins). This resulted in a set of 189-2611 (median 1029) paired passive versus active rates per neuron. For unbiased estimates of changes between passive and active conditions, we rotated passive versus active responses, so that offset and gain were fit for the difference between conditions relative to the mean across conditions. Responses were sorted by mean rate and binned in ascending order, with 50 samples in each bin. We used linear regression to find the minimum mean-squared error fit for a line to the difference as a function of mean. The y intercept of this line indicated change in offset firing rate, a constant change for all stimuli, and the slope of the line indicated change in gain (i.e., a change that scaled with the strength of the response to each stimulus). For ease of visualization, data are shown without rotation; values are computed with rotation as described above. Rotation steps were included because simulations revealed that without them, minimization of squared error resulted in a bias toward a slope of 0.

Linear spectral weighting model. Spectral tuning was measured from responses to the RSS stimuli that served as distractors in each trial (typically 100-200 per neuron). The response to an RSS stimulus was fit using a linear spectral weighting model, as follows:

$$r_j = R_0 + \sum_{i=1}^{n} w_i S_j(f_i)$$

Where r_j is the mean rate over a time window 20 ms after stimulus onset to 20 ms after stimulus offset in response to stimulus j, $S_j(f_i)$ is the stimulus level in a bin centered on frequency f, w_i is the linear weight for each frequency bin [in spikes/(sec $\times$ dB)], and R_0 is the average rate computed across the RSS set. The weights were estimated by minimizing the mean square error between the rates predicted by Equation 1 and the empirical rates r_j (Young and Calhoun, 2005). Because the model depends on the parameters linearly, this is a well-understood optimization problem, which is solved using the method of normal equations. The weights characterize linear spectral tuning and are often similar to a pure tone tuning curve.

Spectral tuning changes at BF during behavior. Behavior-dependent spectral tuning changes were determined from the difference between the spectral weighting functions measured during passive listening and behavior measured at the peak of the RSS tuning curve (peak weight; see Fig. 7A). The frequency of the peak weight corresponds to the neuron's BF. The weight difference (active-passive) was normalized by the peak
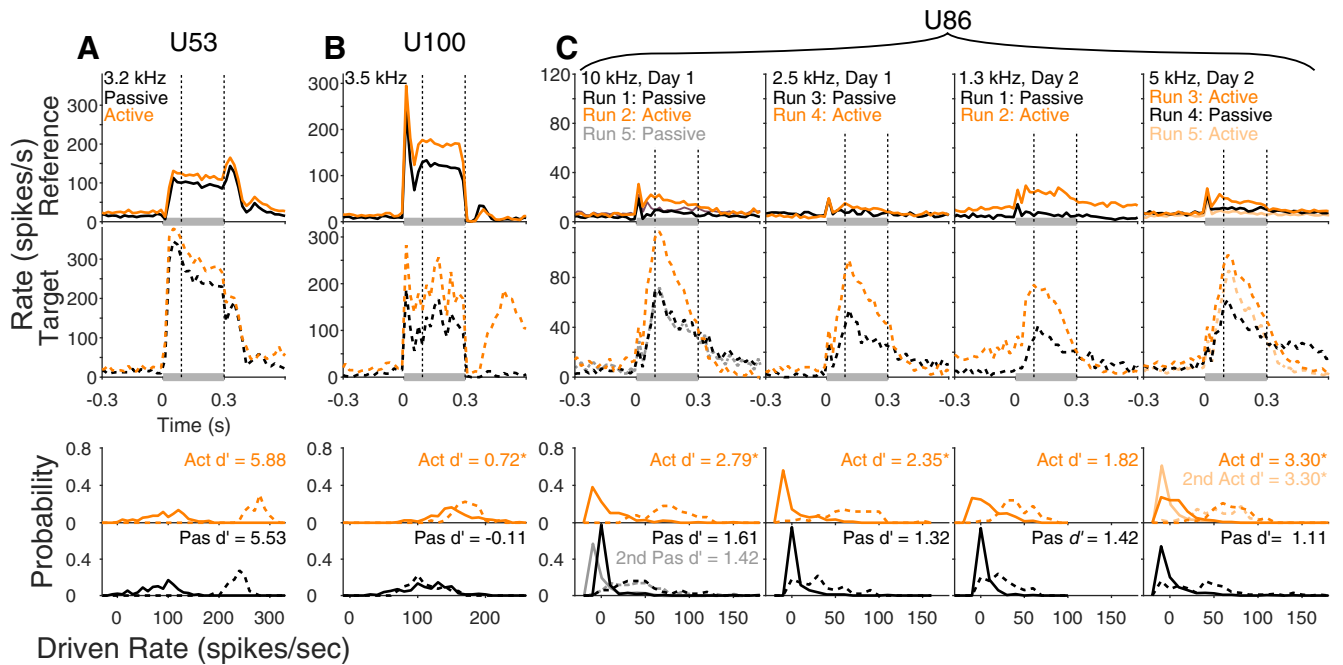
Figure 2. Example task engagement effects on reference and target responses. Top, Peristimulus time histograms (PSTHs) of rate in response to references in passive (black) and active (orange) blocks. Gray bars represent stimulus duration. Dashed lines indicate analysis window. Titles indicate the unit number assigned to our database ("Unn"). Middle, PSTHs in response to targets. Bottom, Histogram of mean driven response rates to references (solid) and targets (dashed) for passive and active conditions. Numbers indicate neural d' between reference and target. *Significant difference ($p < 0.05$, jackknife t test). **A**, Data for ICC neuron exhibiting no significant task engagement effects (BF 3.2 kHz). **B**, Data for ICC neuron exhibiting mild task engagement effects (BF 3.5 kHz). **C**, Data for AM neuron exhibiting strong task engagement effects (BF not defined, not tuned). Each column plots responses for different target frequencies and days, as noted. In row 1, a second passive was recorded. In row 4, an active recording was made first, followed by a passive and another active.

weight in the passive condition. This produced a measure of the local change at BF as a fraction of the passive weight.

Quantification of response latency. Response latencies were calculated based on response to the target during the active block. First-spike latency was quantified by the following: (1) binning spikes at 200 Hz, (2) averaging across trials, (3) subtracting the prestimulus spontaneous rate, and (4) linearly interpolating to find the time at which the response crossed 3 SDs of the spontaneous rate above the spontaneous rate. Latency to 50% of the maximum driven rate was quantified by the following: (1) binning spikes at 200 Hz, (2) averaging across trials, (3) smoothing 3× with a 3-point sliding window, (4) subtracting the prestimulus spontaneous rate, and (5) linearly interpolating to find the time at which the response crossed 50% of its range. For units that were suppressed in response to targets, the same procedures were used, but interpolation was used to find the point at which the response fell the same threshold amount below the spontaneous rate.

Poststimulus activity regression model. Multiple regression was used to investigate the source of poststimulus spiking, specifically to dissociate effects of auditory inputs from effects of motor activity related to licking and from effects related to reward delivery. The model predicted the difference between active and passive time-varying spike rate on single trials (40 Hz sampling). The regressors were reference offset times (RSS and catch references treated identically), target offset times, and lick times. The regression was fit over only poststimulus time periods (for targets: 100 ms to 3 s *re* offset; for references: 100 ms *re* offset to the onset of the next reference). To determine how much of the active-passive difference was uniquely attributable to licks, we measured how much better a full model (using both licks and stimulus times) performed at predicting single-trial activity than a model using only the stimulus times. Conversely, to determine how much of the active-passive difference was uniquely attributable to the stimulus, we measured how much better the full model performed than a model using only the licks. For each neuron, a model was fit 20 times using 20-fold cross-validation, and significance was assessed at $p < 0.05$ by paired t tests on the distributions of mean-squared errors of these model fits.

Experimental design and statistical analysis. Two animals, one of each sex, were used in this experiment. The number of neurons included in each analysis is reported in the text. For each neuron, significant changes in RSS weights, global gain, offset, average target rate, average reference rate, and d' during behavior for each neuron were assessed with 20-fold jackknifed t tests (Efron and Tibshirani, 1998). Significant average effects across the subset of behavior-modulated neurons were computed with a Wilcoxon signed-rank test (sign test). Significant differences between subsets (e.g., buildup vs nonbuildup neurons) were evaluated with a Wilcoxon rank-sum test. Significant differences in the number of cells with enhanced versus suppressed d' were evaluated with binomial tests (Slee and Young, 2013). All statistics were computed using MATLAB.

Histology and track labeling. After recordings were complete, the animals were killed with an overdose of barbiturate (Euthasol 0.5 ml/kg) and transcardially perfused (0.5% PFA). In Animal F, the 4-electrode implant had been explanted 11 months before death, the single-tetrode implant was still implanted at death. The brain was sectioned (100 µm) in the coronal plane, and stained with cytochrome oxidase. Select slices were subsequently immunostained for GFAP to identify electrode tracks. Sections were blocked and permeabilized with 10% normal goat serum and 0.4% Triton X in PBS at 23°C for 2 h, then incubated with 1:3000 rabbit anti-GFAP (#Z-0334, Agilent Technologies) primary antibody in PBS with 1.5% normal goat serum at 23°C overnight. Sections were rinsed with PBS, then incubated with 1:500 Alexa-488 (#A-11008, Invitrogen) and DAPI in PBS, rinsed in PBS, and mounted. Slices were imaged 5× objective on an Axio Imager 2 upright microscope (Carl Zeiss).

Results

We studied the effects of task engagement on the sound encoding properties of single neurons in the IC. Two marmosets (Animals C and F) were trained on a go/no-go task to report tone targets (masked by 0 dB SNR broadband noise) and ignore broadband noise distractor references (RSSs) (Yu and Young,
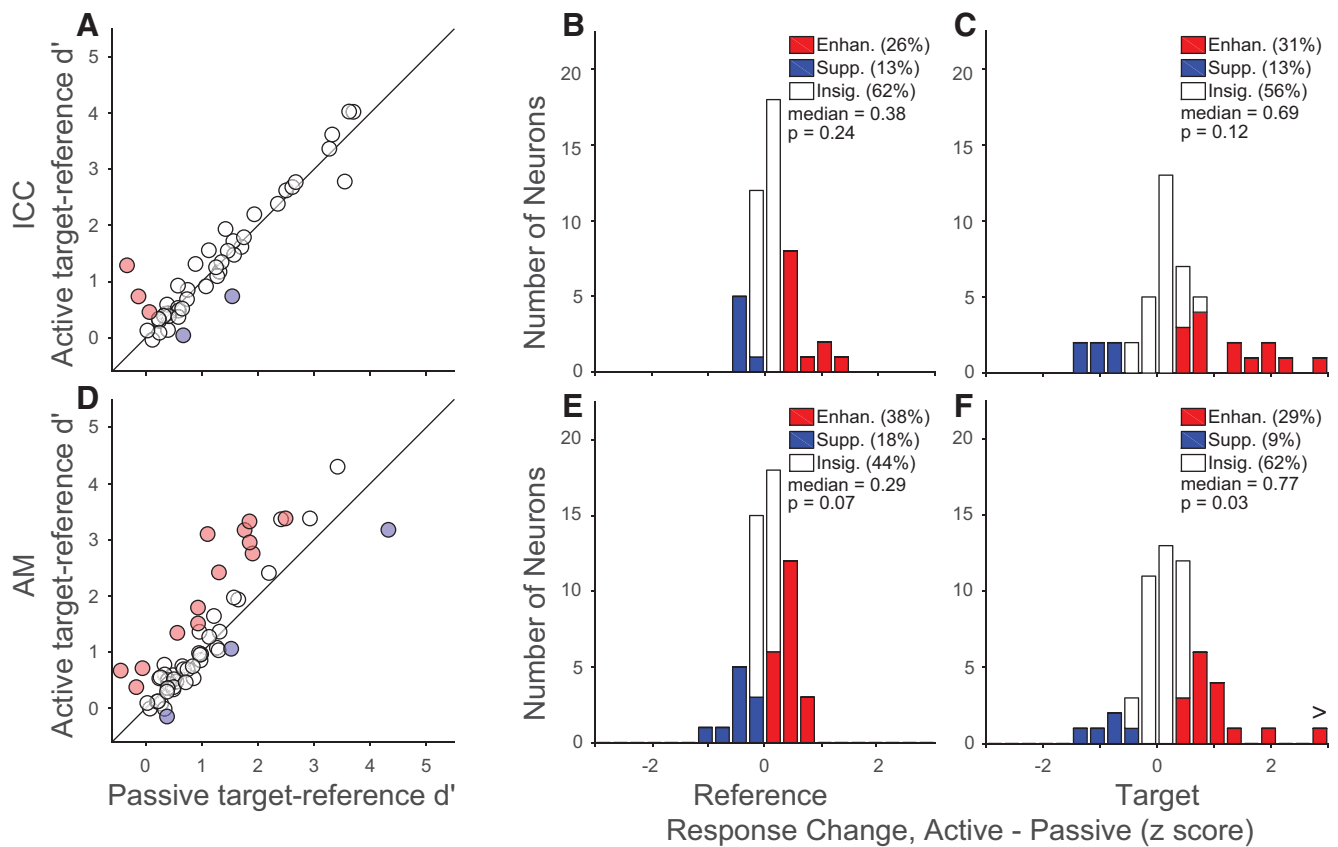
Figure 3. Task engagement increases neural discriminability of target from reference in NCIC neurons. **A**, Scatter plot compares neural discriminability (d') between target and reference sounds for passive versus active blocks. Each point represents a single ICC neuron. Red and blue represent significant increases and decreases ($p < 0.05$, jackknife t test), respectively. **B**, Histogram of difference in mean reference response between active and passive conditions for ICC neurons. Changes are normalized by the pooled SD of the response to target and reference sounds across behavior conditions. Red and blue represent significant increases and decreases ($p < 0.05$, jackknife t test), respectively. Text indicates the median change and significance (sign test) across these modulated neurons. **C**, Histogram of difference in mean target responses by ICC neurons, plotted as in **B**. **D–F**, Discriminability and mean response changes for AM neurons, plotted as in **A–C**.

2000) (Fig. 1A,B). The animals were required to withhold licking from a water-spout during presentation of the references and were given a juice reward for licking during the response window after target onset. Both animals learned this task after 3-4 weeks of training. Once training was complete, they performed reliably, with hit rates well separated from false-alarm rates (Fig. 1C,D). To prevent animals from using spectral differences between the tone-masking noise and RSS noise to detect targets, tone-masking noise samples were pseudorandomly interleaved between the references. False-alarm rates to these catch references were higher than those to RSS references, but they remained substantially lower than hit rates to targets, indicating that animals were selectively responding to the tone targets ($p < 1e-5$ for both animals, paired t test; Fig. 1D). On average, once trained, Marmoset C completed an average of 100 ± 58 (mean $\pm$ SD) correct trials per day, and Marmoset F completed 183 ± 70. By comparison, two ferrets performing a similar task completed a similar number of trials, averaging 100 ± 42 and 172 ± 62, respectively.

We made recordings in several subregions of the IC while the animals performed the task. The IC was targeted using standard physiological criteria and confirmed postmortem by histologic evaluation (see Materials and Methods). Most recording sites were located in the central nucleus (ICC; Fig. 1E,F) or noncentral, shell regions around the ICC. For the tetrode array implanted in Animal F, histology revealed that the four probes in this array passed by on the medial edge of the IC, likely sampling both from neurons in the DCIC and in the lateral PAG (Fig. 1G,H).

Neurons recorded from all the probes had robust auditory responses, many of which were highly dependent on behavioral state. Because we were unable to unambiguously assign neurons to NCIC or PAG, we have labeled all non-ICC neurons as being in the AM. Possible tuning differences between AM neurons are considered below. In sum, these data contain 49 ICC and 60 AM neurons.

Task engagement increased neural discriminability of target from reference in NCIC

For each neuron, we compared spiking activity when animals were passively listening to activity when they were engaged in the task. Identical stimuli were presented in passive and behaving conditions. Stimuli were presented in nearly the same order, with the exception that trials in which the animal missed were repeated at the end of the block. In most neurons, responses to both references and targets were enhanced during task engagement (e.g., Fig. 2). The relative change in these responses varied widely. To assess the extent to which these changes reflect the marmoset's ability to perform the tone detection task, we measured neural discriminability between target and reference sound categories for each neuron using d' (Green and Swets, 1966), and compared d' between passive and active states. For many neurons, especially in ICC, task engagement enhanced both target and reference responses equally, resulting in no change to target-reference discrimination (e.g., Fig. 2A). However, in some
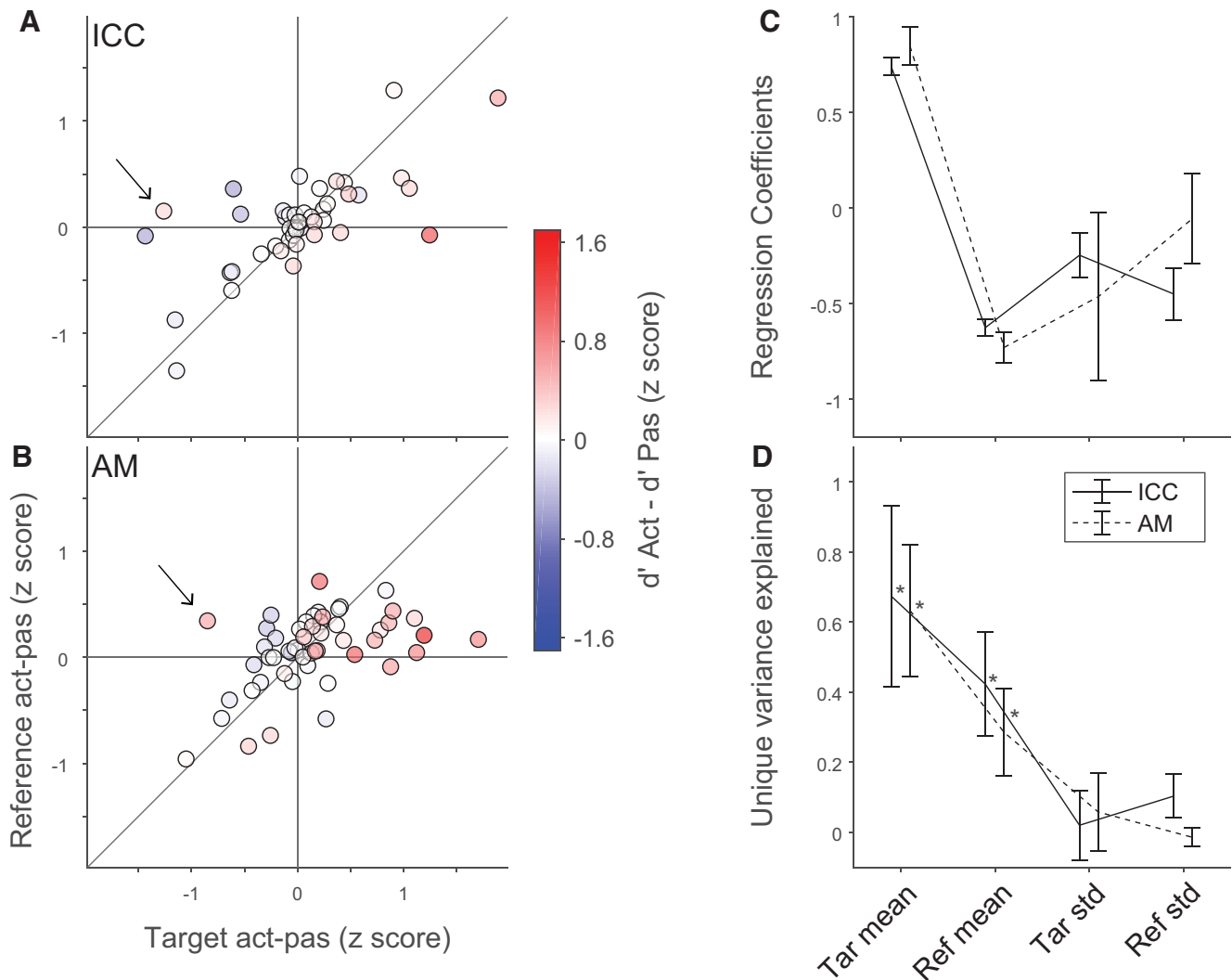
Figure 4. Changes in neural discriminability are predominantly driven by increased target responses. **A**, **B**, Difference in mean active versus passive evoked response to references (vertical axis) and targets (horizontal axis) for ICC (**A**) and AM (**B**). Differences are normalized by SD of each neuron's response across all four stimuli and condition combinations. Color scale represents difference between active and passive neural d'. Arrows indicate neurons for which d' changes are not explained by target or reference mean changes; they instead are explained by changes to variance. **C**, Regression coefficients for each factor contributing to d' difference across the population of neurons, computed separately for ICC and AM. Error bars indicate jackknife SE. **D**, Unique variance explained by each factor, quantified as the fraction increase in mean-squared error resulting from removing that coefficient from the regression. Error bars indicate SEM. *Significant difference from zero ($p < 0.05$, rank-sum test).

neurons, task engagement enhanced target responses more than reference responses, resulting in improved discrimination, which could sometimes be substantial (Fig. 2B,C).

We compared changes in d' for neurons in ICC and AM. Across the ICC population, d' was significantly enhanced by task engagement in 6% of neurons, and decreased in 4% ($p < 0.05$, jackknifed t test; Fig. 3A). In contrast, d' was enhanced in 22% of AM neurons, and suppressed in 5%. Among neurons showing changes in discriminability, the number in which d' was enhanced was significantly greater than the number in which it was suppressed in AM ($p = 0.011$) but not ICC ($p > 0.1$, binomial test; Fig. 3D). Thus, engaging in the tone detection task enhanced neural discriminability between target and reference categories in the AM.

Several factors can contribute to changes in d'. It could be increased by an enhancement of mean target response, a suppression of mean reference response, or a decrease in trial-to-trial variability of either target or reference responses. We investigated the origin of d' changes by measuring task

engagement effects on target and reference responses separately. To compare across neurons, the difference in evoked response (active − passive) was converted to a z score. We compared median response changes among neurons that showed a significant difference for the respective stimulus category between passive and active conditions (jackknifed t test; Figs. 3B,C,E,F, shaded bars). In ICC, both reference and target responses showed a trend toward enhancement, but the median change was not significantly different from zero (reference: $p = 0.24$; target: $p = 0.12$, signed-rank test, Fig. 3B,C). In contrast, responses were enhanced in a larger proportion of AM neurons for both sound categories, and the median change (among significantly changing neurons) was significantly greater than zero for target responses (reference: $p = 0.07$; target: $p = 0.03$, Fig. 3E,F).

The data qualitatively suggested that changes in discriminability were dominated by increases in the mean target rate. We tested this quantitatively by using multiple linear regression to assess the relative contribution of each factor across the population (see Materials and Methods). In most neurons, d' changes

were explained by a combination of changes in the mean target and reference response (Fig. 4A), but in a few neurons we observed increased d' because of a reduction in reference and target response variability (e.g., Fig. 4A,B, arrows). In both ICC and AM populations, d' changes were dominated by target enhancement, with a weaker contribution of reference suppression and no contribution from changes to trial-to-trial variability (Fig. 4D).

Neural discriminability between targets and RSS references was better than discriminability between targets and catch references

Animals were more likely to false alarm to catch stimuli than to the more common RSS reference stimuli (Fig. 1D). This behavioral bias for catch stimuli suggests that neural discriminability might also be weaker between targets and catch stimuli than for other references. To test for a differential effect on reference responses, we split nontarget data into RSS and catch stimuli and recomputed response changes separately for each category. Only neurons with significant behavior-dependent changes in reference or target response were included in this analysis (ICC: $n = 20$; AM: $n = 29$). Responses to catch references showed a trend toward a greater average increase during behavior than those to RSS references in both areas, but the difference was not significant in either (ICC: $p = 0.06$; AM: $p = 0.26$, signed-rank test on paired differences; Fig. 5A).

We also considered whether changes in d' were different for catch versus RSS references. In this case, we focused only on neurons with significant task engagement changes in target-reference d'. The number of ICC neurons with significant effects and datasets that included catch references was too small to make a meaningful comparison ($n = 2$). Among AM neurons, target-RSS d' was greater than target-catch d' in both passive ($p = 0.005$) and active states ($p = 0.001$, $n = 29$, signed-rank test; Fig. 5B). Average target-RSS d' was significantly increased during task engagement ($p = 0.022$), but target-catch d' was not ($p = 0.13$). Moreover, within neurons, the average d' change for RSS references was significantly larger than for catch references (median paired difference 0.39 SDs, $p = 0.027$, signed-rank test).

A similar, though weaker, pattern was observed when we considered the entire set of neurons, regardless of behavior-dependent changes. For this larger set, target-RSS d' was significantly higher than target-catch d' in both passive and active states for AM neurons (median paired differences 0.58 and 0.93 SDs, both $p < 1e-5$), but not for ICC neurons (median paired differences 0.01 and 0.31 SDs, $p = 0.85$ and $p = 0.21$, respectively). However, d' was not significantly different between active and passive state for either RSS or catch references in either complete sample of neurons (ICC: median paired differences of 0.04 and -0.05, $p = 0.27$ and $p = 0.35$; AM: median paired differences of -0.01 and -0.03, $p = 0.16$ and $p = 0.87$). Active-passive d' changes for RSS references were significantly larger than those for catch references in AM (median paired difference 0.23 SDs, $p = 0.018$, signed-rank test) but not ICC (0.09 SDs, $p = 0.16$). Thus, in AM, changes in neural discriminability for the catch references are weaker than for RSS stimuli, reciprocating the behavioral effects.

Largest discriminability increases occurred in AM neurons with slow, buildup responses

Some AM neurons gave distinctive slow, buildup type responses to targets (e.g., Fig. 6B, inset). Neurons were classified as buildup if their latency to 50% maximum response to the target was >30 ms greater than their first-spike latency (Fig. 6A,B). Task
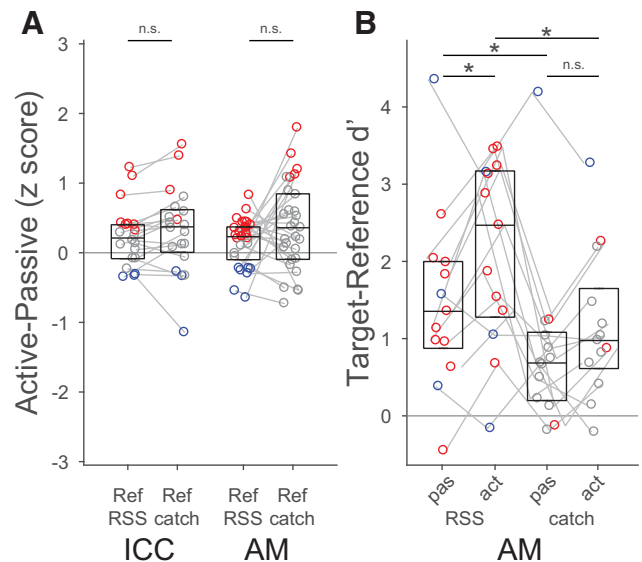

Figure 5. Neural discriminability between targets and catch references was lower than between targets and noncatch references. **A**, Change in z-scored response to RSS and 1/f spectrum catch references between active and passive conditions ($n = 20$ ICC, $n = 29$ AM neurons with significant engagement effects on reference or target responses and presented at least 2 catch stimuli per behavior condition). **B**, Target-versus-RSS reference d', and target versus catch d' in passive and active states ($n = 29$ AM neurons with significant engagement effects on target-vs-reference d' and with at least 2 catch stimuli). Exclusion criteria eliminated all but 2 ICC neurons, so only AM data are shown. **A, B**, Red represents a significant rate increase during task engagement. Blue represents a significant decrease. Boxes represent median and interquartile range. *Pairs of conditions with a significant median difference ($p < 0.05$, signed-rank test). n.s. Not significant.

engagement changes were significantly greater in buildup AM neurons ($n = 10$) than nonbuildup AM neurons ($n = 50$) for both target rate ($p = 0.0001$) and target-reference d' ($p = 0.0002$; Fig. 6C). These changes were also greater in buildup AM than ICC neurons ($p = 0.004$ and $p < 0.0001$, respectively, $n = 49$). Task engagement changes in reference responses were not significantly different between any groups. The majority of buildup neurons (9 of 10) were recorded using a chronically implanted tetrode array in Animal F. In a few neurons, the stability of the array allowed investigation of the stability of task engagement changes across days. We found strong target rate and target-reference d' increases during behavior over multiple days with different target frequencies (e.g., Fig. 2C). Buildup neurons were found at depths ranging from 11 to 12 mm re cortical surface, and they were interspersed with nonbuild neurons (9 of 32 neurons recorded with this array were buildup). This suggests that a subpopulation of neurons in the medial IC may be especially prone to behavior-dependent changes in sound encoding.

Task engagement did not change frequency tuning in ICC

Previous recordings in IC of ferrets engaged in a similar go/no-go tone detection task found a selective suppression of reference responses at the target frequency (Slee and David, 2015). The RSS stimuli used here can be used to measure frequency tuning and test whether this same effect held for marmosets performing tone-in-noise detection (Young and Calhoun, 2005). For each neuron where the target was within 1/8th of the peak RSS weight, spectral tuning to the RSS stimuli was compared between passive and active states (Fig. 7A). A few neurons showed significant changes in RSS weight at the target frequency, but most neurons exhibited no change. Across all ICC neurons, the median
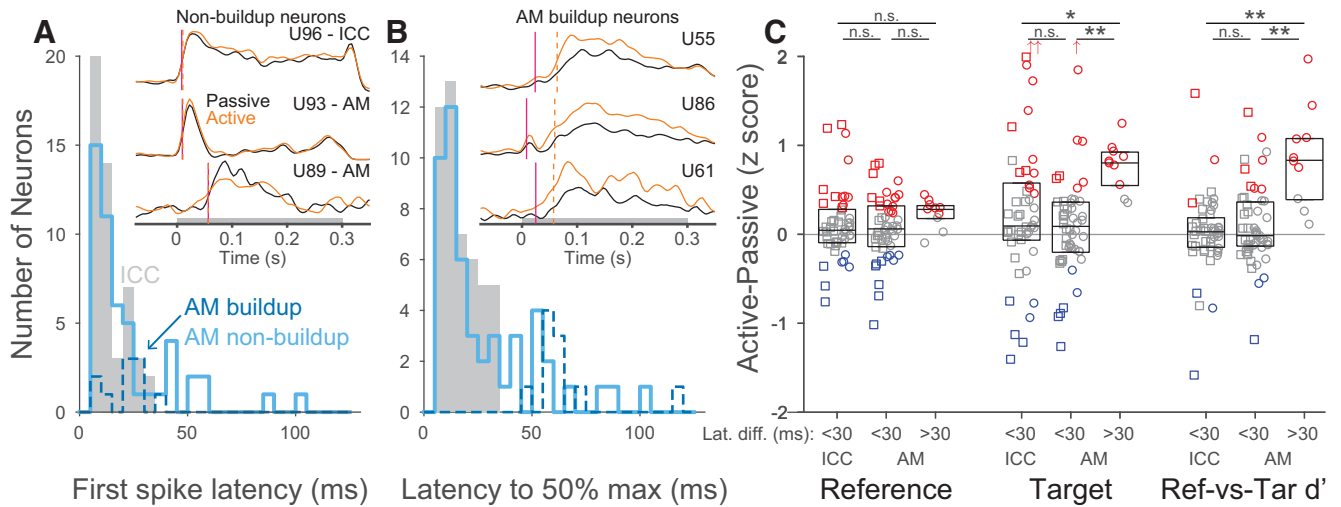
Figure 6. AM neurons with slow, buildup responses show strong enhancement of target responses during behavior. *A*, Histogram of first-spike response latency and (*B*) latency to 50% of maximum response to targets during the active block for ICC (gray), nonbuildup AM (light blue), and buildup AM (dark blue, dashed) neurons. Buildup neurons were defined as having a latency to 50% maximum driven rate > 30 ms longer than their first-spike latency. These buildup responses only occurred outside of ICC. Insets, PSTH responses to targets for 6 example neurons, non-buildup neurons in *A* and buildup neurons in *B*. Magenta lines indicate first-spike latency. Dashed orange lines indicate latency to 50% maximum response. Gray bar represents target duration. *C*, Task-related changes in reference, target, and target-reference *d′* in neurons grouped by buildup category (nonbuildup: <30; buildup: >30). Boxes represent median and interquartile range. Significant difference in median: ∗*p* < 0.05; ∗∗*p* < 0.001; rank-sum test. Squares represent neurons recorded in Marmoset C. Circles represent neurons recorded in Marmoset F. Off axis points are all from Marmoset F and have values: Target ICC: 2.4, 2.9; Target AM <30 ms: 3.9. n.s. Not significant.
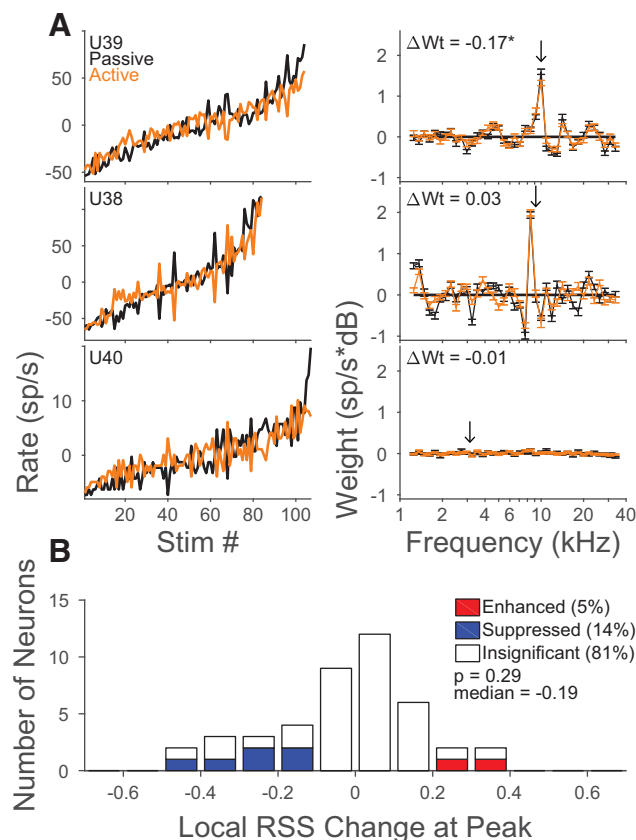

Figure 7. Task engagement did not change frequency tuning in ICC. *A*, Example RSS tuning in three neurons: ICC neuron showing suppression at the RSS peak (top), ICC neuron showing no change (middle), and AM neuron showing no tuning to RSS stimuli (bottom). Left, Mean firing rate in response to the same set of RSS stimuli for passive (black) and active (orange) blocks. Right, Spectral weights, computed by minimizing mean-squared error of rate predictions. Arrow indicates target frequency. Text indicates fraction change in peak weight (ΔWt). *B*, Histogram of fraction change in peak weight across all ICC neurons. Only neurons where the target frequency was within 1/8 octave of the RSS tuning peak are included (*n* = 43).

fraction change in RSS peak during behavior was not significant (0.0032, *p* = 0.6, rank-sum test). Among the 19% of neurons exhibiting significant changes at the target frequency, suppression was more common (*n* = 6 vs 2 neurons of 43 ICC neurons). The median change among these neurons was −0.19 and was not significantly different from zero (*p* = 0.29, rank-sum). Most AM neurons were not tuned to the RSS stimuli according to the linear model (e.g., Fig. 7*A*, bottom) so an analysis of tuning changes was not applicable in this population.

Changes in response at the target frequency were much less common in marmoset than previously observed in ferrets, where they occurred in 62% of ICC neurons, with a median fraction change of −0.32 (Slee and David, 2015). We considered the possibility that the marmoset dataset might have lower signal to noise in tuning measurements, thereby masking a similar suppression. In the present data, the minimum detectable fractional change in peak RSS weight was estimated for each neuron as 2 jackknifed SEs on the spectral tuning curve weight at target frequency. Across 43 ICC neurons, the median was 0.2. In 84% of neurons, the detectable change was smaller than the median change of −0.32 for ferret neurons. Therefore, if selective target suppression was as strong in marmoset IC as in ferret, it would have been detected in the current dataset.

Comparison with task-dependent changes in ferret IC
The absence of task-dependent tuning changes in marmoset ICC suggests a different pattern of plasticity from that observed in ferrets (David et al., 2012; Slee and David, 2015). For a more detailed comparison of task-dependent changes between studies, we reanalyzed data from the previous study using the same approach as the current study. The ferret data were collected from ICC and NCIC during a similar tone detection task. These data differed from the current study in that reference sounds were broadband rippled noise and the target was a pure tone, not masked by noise. Despite these differences, the mean reference response analysis could be applied identically. In ferret, we found that mean response rates were more often suppressed during
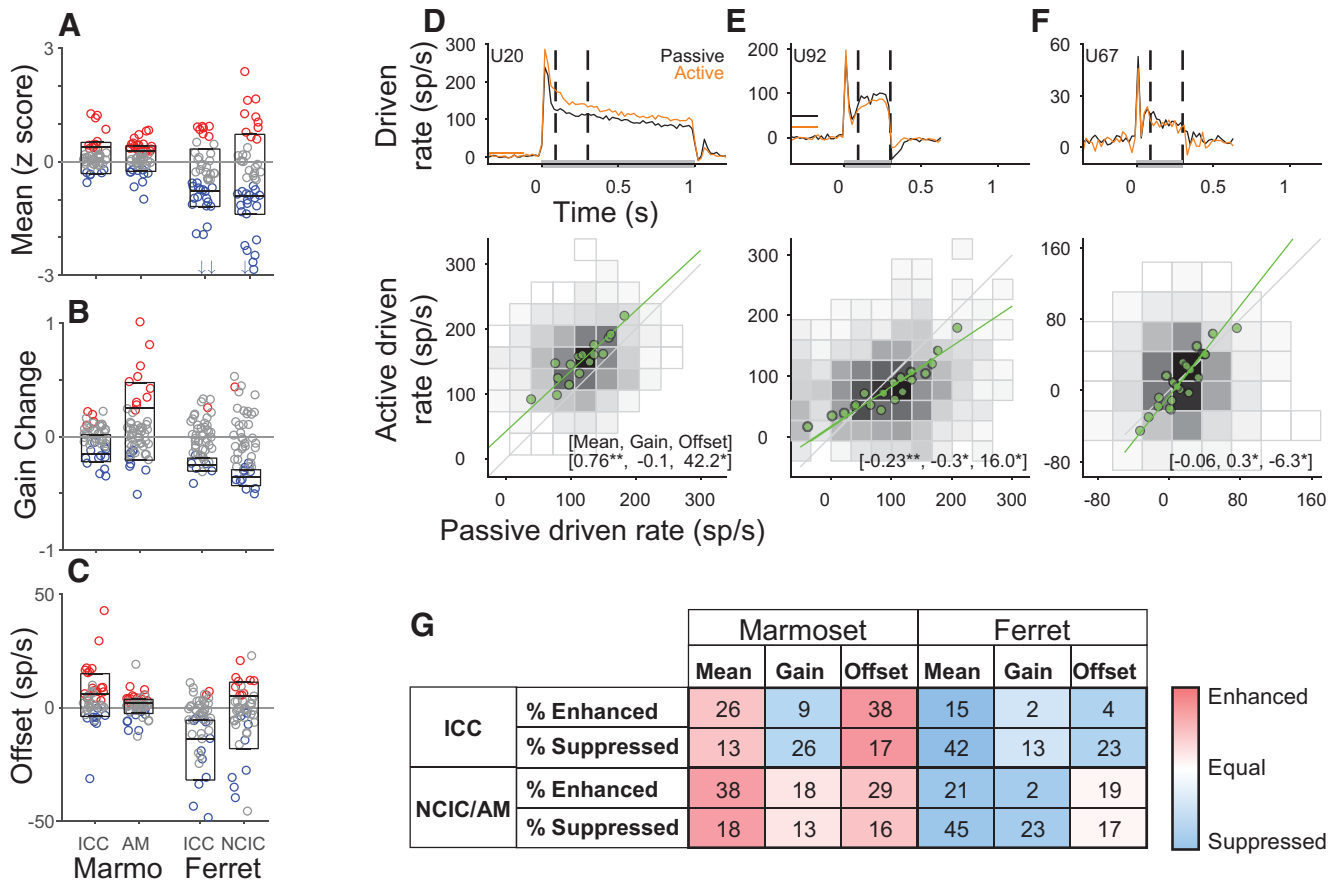
Figure 8. Comparison of task-related changes in reference noise responses between marmoset and ferret IC. Swarm plots of rate difference (**A**), gain (**B**), and offset (**C**) for each species, subdivided by brain area as indicated. All neurons were analyzed with a time window 90-300 ms following reference onset. Red and blue represent neurons with significant task-related increases or decreases ($p < 0.05$, jackknife t test), respectively. Boxes represent median and interquartile range calculated among these modulated neurons. **D–F**, Example offset-gain computations in 3 neurons. Top, PSTHs of evoked response to references in passive (black) and active (orange) blocks. Horizontal lines indicate spontaneous rate, U67 (**F**) had rates of 80 and 72 for passive and active, respectively. Gray bars represent stimulus duration. Dashed lines indicate analysis window. Bottom, Passive versus active driven rates (plotted as 2 d histograms, gray scale normalized to maximum) were binned (green points) and fit with a line by linear regression to compute gain and offset. Text indicates mean rate difference (z score), gain change, and offset for each neuron. $*p < 0.05$; $**p < 0.01$; jackknife t test. **G**, Percentage of neurons showing significant task engagement effects on reference responses in marmoset and ferret IC. Color scale represents the difference between the percentage of significantly enhanced neurons and the percentage of significantly suppressed neurons.

task engagement in both ICC and NCIC (Fig. 8A). In marmosets, slight enhancement was more common. Setting aside differences in task stimuli (see Discussion), these results indicate that reference suppression dominates in ferret IC while enhancement is more common in marmoset IC.

The previous work in ferret performed a more detailed characterization of changes in sensory activity by separately measuring changes in response gain and offset, the scaling and additive factors required to best match responses to each stimulus between active and passive conditions (David et al., 2012; Slee and David, 2015). When gain and offset changes both have the same sign, their effects are consistent with overall suppression or enhancement measured in mean response rate. However, if gain and offset have opposite signs, their relationship to changes in mean response is less predictable. In the marmoset data, behavior-dependent changes in reference response gain and offset sometimes had opposite signs. Most commonly, gain decreased and offset increased during behavior. Because these changes shifted spike rates in opposite directions, they did not consistently predict changes in mean response rate. Some neurons with gain suppression and offset enhancement exhibited a positive rate difference (overall enhanced responses; Fig. 8D), but others exhibited a negative rate difference (Fig. 8E). In a few neurons, gain was enhanced and offset was suppressed (Fig. 8F).

Consistent with the examples, gain was predominantly suppressed across marmoset ICC, while offset was enhanced (Fig. 8B,C,G). In contrast, both gain and offset were predominantly suppressed in ferret ICC. In Marmoset AM, both gain and offset were predominantly enhanced; in ferret NCIC, gain was mostly suppressed and the number of neurons with significantly enhanced/suppressed offset was nearly equal. Comparing two example populations illustrates why analysis of gain alone does not reveal a complete picture of task-related effects. Instead, it motivates the analysis of overall mean rate change used in the results reported above: in ferret NCIC, gain suppression dominated over nearly equal offset changes to produce an overall mean rate suppression; in marmoset ICC, gain suppression was eclipsed by offset enhancement to produce an overall mean rate enhancement. In sum, the current results indicate differences between population patterns of gain, offset, and mean rate change. Thus, for these stimuli, neither gain nor offset is a reliable predictor of mean rate changes.

Poststimulus activity is predominantly lick-related
While the focus of this paper is on the early sensory response to sound, where animal movements do not complicate interpretation, we also noticed substantial long lasting, post-target activity in many neurons, including in the ICC (e.g., Fig. 9A). We
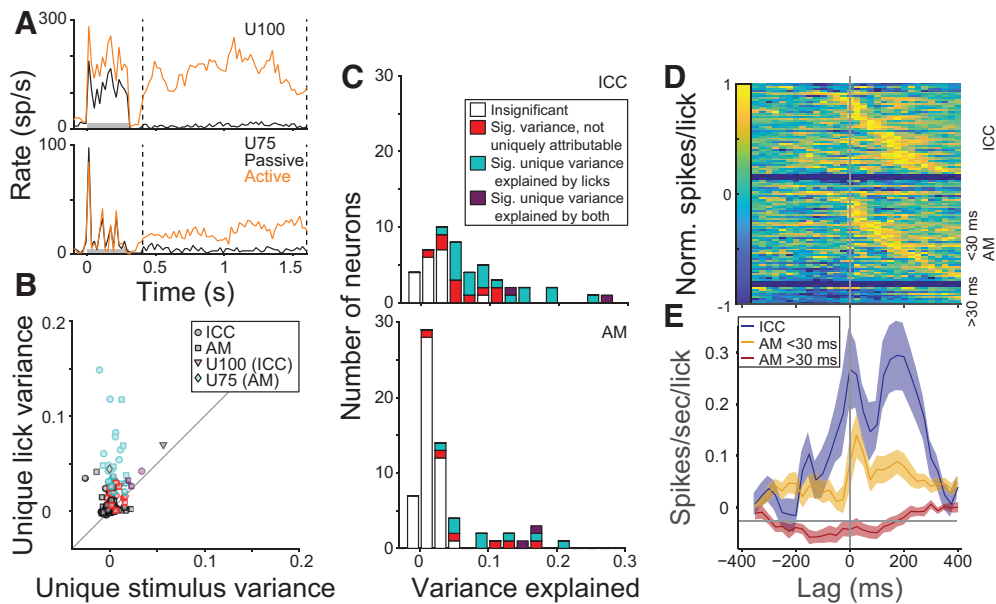
Figure 9. Post-target spiking activity is predominantly lick-related. ***A***, Mean target responses in an ICC (top) and an AM neuron (bottom), both with substantial post-target spiking. Gray bars represent stimulus duration. Dashed lines indicate analysis window. ***B***, Scatter plot compares variance of single-trial spike rate uniquely attributable to licking versus stimulus for each neuron. Area and significance indicated in legend, color scheme as in ***C***. ***C***, Histogram of variance explained in poststimulus spike rate by full multiple-regression model for ICC neurons (top) and AM neurons (bottom). ***D***, Heat map of lick response functions (coefficients of lick regressor at time lags before and after lick events). Each row represents the function for a single neuron, normalized by maximum absolute value. Rows are sorted by the time of the positive peak and grouped by areas and response latency difference (as in Fig. 6). ***E***, Mean (±SE) lick response functions (not normalized) for ICC neurons (blue), and AM neurons without (yellow) and with buildup responses (red).

speculated that this long-latency activity might not be a feedforward response to the target sound. Instead, it could reflect one of several processes: (1) premotor commands related to the decision to respond, (2) sounds generated by the animal's movement to receive the reward, or (3) signals encoding the value of the juice reward. Work in ACtx has demonstrated motor signals in behaving animals (Vaadia et al., 1982; Schneider et al., 2014; Huang et al., 2019), suggesting that motor signals may also reach the midbrain. On the other hand, late sound-evoked activity in some IC neurons in rhesus monkeys has been reported to be modulated by reward value (Metzger et al., 2006). To distinguish between activity evoked by task stimuli and these other possible sources, we took advantage of the observation that the timing of licks was not stereotyped relative to the target offset time. This variability permitted multiple linear regression to quantify the fraction of poststimulus spiking activity that could be uniquely explained by licks versus stimulus offset events. In this model, spikes evoked by premotor activity or self-generated sound (possibilities 1 and 2, above) should be predicted by the lick regressor. On the other hand, spikes driven by reward (possibility 3) should be explained by the stimulus regressor.

Across the majority of neurons, the regression analysis revealed that post-target activity could generally either be explained by licking (e.g., U75; Fig. 9B, cyan) or could not be uniquely attributed to licking or sound (red). There were significant unique contributions of stimulus in only 4 neurons (e.g., U100, purple). Therefore, poststimulus activity predominantly encodes either premotor commands or sound generated from licking or other associated movements. The proportion of neurons with significant variance explained by licks was greater among ICC than AM neurons (43 vs 16%). Furthermore, across all neurons, regardless of significance, the amount of post-target activity (quantified by the variance explained by the model) was greater among ICC than AM neurons (median 0.044 vs 0.016, $p = 0.0003$, rank-sum; Fig. 9C). Therefore, poststimulus activity is stronger in ICC neurons.

The prevalence of post-target activity in ICC neurons suggested that this activity might be related to self-generated sounds. To distinguish between effects of motor activity and responses to self-generated sound, we studied the time course of lick-related activity revealed by the regression model (Fig. 9D,E). In most ICC neurons and in AM neurons without buildup responses, lick response functions peaked after zero, indicating that spikes tended to follow licks. This temporal relationship suggests that these spikes were predominantly driven by self-generated sound. In most of the buildup AM neurons, spiking was suppressed before licking (lick response functions are negative and peak before zero; Fig. 9E, red line), which suggests that they could encode premotor activity. A few AM neurons ($n = 4/60$) exhibited prelicking spiking that was correlated with the animal's behavioral choice (choice probability; not shown). Together, these results suggest that the majority of post-target activity reflects responses to self-generated noise, although neurons outside of ICC may also encode activity associated with the licking motor response to the target sound.

Discussion

We found that engagement in a tone-versus-noise discrimination task modulated sound-evoked activity in the marmoset AM. These results confirm previous observations of task-related plasticity in ferret and macaque IC (Ryan and Miller, 1977; Slee and David, 2015). Furthermore, they support the hypothesis that the midbrain is capable of transforming auditory inputs into behavioral decisions, a role traditionally attributed to cortex. This hypothesis is consistent with lesion and inactivation studies demonstrating that ACtx may be unnecessary for performing some auditory behaviors, in particular, relatively simple, highly trained discrimination tasks (Guo et al., 2017a). While task engagement increased responsiveness to both target and distractor sounds throughout the midbrain, responses to targets were selectively enhanced outside ICC, increasing neural

discriminability between targets and distractors. These changes in selectivity could support behavioral discriminations between task categories, an emergent representation previously observed only in ACtx (Tsunada et al., 2011; David et al., 2012; Shepard et al., 2015; Christison-Lagay and Cohen, 2018; Elgueda et al., 2019; Liu et al., 2019; Xin et al., 2019).

Representation of task-related categories in midbrain

The ACtx does not appear to be necessary for some auditory behaviors (Neff et al., 1975; Heffner, 2005; Guo et al., 2017a). In species ranging from rodents to humans, ACtx lesions do not affect coarse frequency discrimination (Butler et al., 1957; Zatorre, 1988; Ono et al., 2006; Gimenez et al., 2015) but do impair more difficult tasks, including fine frequency discrimination (Harrington et al., 2001; Tramo et al., 2002), discrimination of frequency sweeps (Kelly and Whitfield, 1971; Harrington et al., 2001), and pitch perception (Whitfield, 1980; Zatorre, 1988). Permanent lesions might lead to compensatory plasticity, where new circuits form to make up for loss of function. However, this pattern holds even after accounting for long-term plasticity: in mouse, optogenetic inactivation of ACtx has little effect on pure tone discrimination but abolishes the ability to discriminate a pure tone from a frequency sweep (Ceballo et al., 2019).

To our knowledge, the impact of inactivating ACtx on detection of tones in noise has not been tested. In humans, temporal lobe lesions impair speech discrimination in background noise more than in quiet (Heilman et al., 1973; Olsen et al., 1975). Whether this deficit represents a general inability to suppress background noise or is specific to complex foregrounds, such as speech, is unclear. Invariance of neural coding to background noise is stronger in ACtx (Narayan et al., 2007; Moore et al., 2013; Mesgarani et al., 2014), but some invariance has been reported in IC (Rabinowitz et al., 2013).

The present data suggest that invariance to background noise in IC may be increased by task engagement. Responses to tones are specifically enhanced relative to background noise. Future experiments might introduce foreground sounds other than tones to test whether this enhancement is specific to tones or reflects a general increase in noise invariance.

AM neurons with slow, buildup responses were particularly affected by task engagement

Data from 1 animal included neurons on the IC-PAG border. In contrast to the strong onset responses of other AM neurons, responses of these neurons built up slowly. They were also highly task-modulated, responding strongly to sounds during task engagement, but weakly during passive listening. Their spike rates tended to suppress before licking, suggesting that they encode premotor signals. The substantial task-related plasticity suggests that this area contributes particularly to discrimination of targets from distractors.

The precise anatomic location of these task-modulated neurons remains unclear. First-spike latency cannot be used to determine their location, as latencies as short as 10 ms have been reported for both DCIC and PAG neurons (Syka et al., 2000; Marshall et al., 2008; Johansen et al., 2010). However, from the limited data available, it appears that PAG neurons have strong onset responses to broadband noise (Johansen et al., 2010, their Fig. 5d), while at least some DCIC neurons have "pure sustained" responses similar to the buildup responses we observed (Syka et al., 2000, their Fig. 5). DCIC axons constitute a dominant source of input to the dorsal medial geniculate body (Wenstrup, 2005), which in turn projects to ACtx belt areas (Mothe et al., 2006). In

marmosets, these same belt areas (rostromedial and caudomedial) project back to rostromedial DCIC (Mothe et al., 2006, their Fig. 10), likely the area where strong task modulation was observed in this study. This network may operate in parallel to the lemniscal stream of ICC, ventral medial geniculate body, and primary ACtx (Bartlett and Wang, 2011; Mellott et al., 2014). One report demonstrated that DCIC lesions impair auditory attention without impairing discrimination (Jane et al., 1965). Further study of dorsomedial IC is needed to clarify its role in the auditory attention network.

Comparison with previous studies

The task-related plasticity we observed in marmoset IC replicates changes in ferret IC during a similar behavior (Slee and David, 2015). There were, however, some differences. Unlike in ferrets, only a relatively small fraction (30%) of marmoset ICC neurons exhibited task-dependent plasticity. Responses in both species were modulated by behavior in noncentral regions. However, sound-evoked activity in marmosets tended to be enhanced during behavior, rather than suppressed, as in ferret.

There are two major methodological differences that could explain the divergent effects. In the ferret study, TORCs were used as reference sounds; the current study used RSS noises. These stimuli have similar spectral bandwidth but different temporal dynamics. Therefore, the dynamics of activity evoked by TORCs versus RSS noises may have led to complex differences in excitatory versus inhibitory network activity and subsequent differences in overall response magnitude.

Another difference between the tasks was that for marmosets targets were masked by distractor noise, while for ferrets targets occurred in isolation. A difference in task difficulty could be responsible for the discrepancy, as difficulty has been shown to affect plasticity in ACtx (Atiani et al., 2009). Moreover, in providing a mask for targets, broadband noise was associated with both negative (timeout) and positive (juice) reward values. For ferrets, noise was only associated with negative values. There is growing evidence that reward and motor associations can impact sensory coding in cortex (Vaadia et al., 1982; Brosch et al., 2011; David et al., 2012; Jaramillo et al., 2014; Guo et al., 2019; Huang et al., 2019). Perhaps the reward systems guiding learning also impact coding in midbrain.

These results could also be because of species differences. A study of IC during tone detection (without masking noise or distractors) in rhesus macaques also found enhancement of responses (Ryan and Miller, 1977). Cortical organization and cortico-collicular feedback differ between primates and other mammals (Wenstrup, 2005; Winer, 2005; Mothe et al., 2006). In primates, the core-belt-parabelt in cortex is arguably more elaborate and differentiated than in carnivores and rodents (Hackett, 2011). Greater functional specialization of the primate brain may leave ICC more specialized for veridical auditory encoding and less affected by state changes. Finally, neuromodulatory input to IC is quite diverse (Hurley, 2019). For example, exogenously applied serotonin globally increases or decreases responses to tones in some neurons, but selectively alters the tuning to tone frequency in others (Hurley and Pollak, 1999). Thus, small differences in neuromodulatory input between species could have substantial impact on task-related changes in activity.

Metrics of task-related changes: gain and offset versus mean rate

Measures of response gain and offset reveal multiplicative and additive factors that scale neural responses during behavior

(McAdams and Maunsell, 1999; David et al., 2012; Slee and David, 2015; Guo et al., 2017b). Because of the nonlinearity of spike generation, purely gain or offset changes in presynaptic inputs can be transformed into a mixture of effects on spiking output (Seybold et al., 2015; Phillips and Hasenstaub, 2016). Since this transformation depends on spiking threshold, a group of neurons with heterogeneous thresholds could have heterogeneous gain and offset changes in spiking output, despite similar presynaptic changes. Furthermore, changes in selectivity that alter the relative response to different stimuli (i.e., a tuning shift) (David et al., 2012; Slee and David, 2015) are not well described by a gain and offset model. In the present data, gain and offset changes sometimes conflicted in sign, suggesting that task engagement does not simply change overall excitability.

The changes in mean response in the current study provide a more interpretable measure of task-related effects when gain and offset changes do not agree. This approach also allows for straightforward measurement of target versus reference discriminability. Thus, a complete analysis of task-related effects might include changes in gain, offset, and mean rate.

A causal role of midbrain in behavior?

These results show, for the first time, that engaging in a discrimination task enhances discrimination of targets from distractor sounds in IC, and suggest that high-level representations of task categories begin to emerge in the midbrain. This hypothesis could be tested by simultaneous measurement of behavioral and neural tone detection thresholds while varying task difficulty or during optogenetic or pharmacological manipulation of nonlemniscal IC outputs. We now know that IC neurons are modulated by task engagement; future work will determine whether this plasticity plays a causal role in decision-making.

References

Agamaite JA, Chang CJ, Osmanski MS, Wang X (2015) A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). J Acoust Soc Am 138:2906–2928.

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

Bartlett EL, Wang X (2011) Correlation of neural response properties with auditory thalamus subdivisions in the awake marmoset. J Neurophysiol 105:2647–2667.

Brosch M, Selezneva E, Scheich H (2011) Representation of reward feedback in primate auditory cortex. Front Syst Neurosci 5:5.

Butler RA, Diamond IT, Neff WD (1957) Role of auditory cortex in discrimination of changes in frequency. J Neurophysiol 20:108–120.

Ceballo S, Piwkowska Z, Bourg J, Daret A, Bathellier B (2019) Targeted cortical manipulation of auditory perception. Neuron 104:1168–1179.e5.

Christison-Lagay KL, Cohen YE (2018) The contribution of primary auditory cortex to auditory categorization in behaving monkeys. Front Neurosci 12:601.

David SV, Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc Natl Acad Sci USA 109:2144–2149.

David SV, Mesgarani N, Fritz JB, Shamma SA (2009) Rapid Synaptic Depression Explains Nonlinear Modulation of Spectro-Temporal Tuning in Primary Auditory Cortex by Natural Stimuli. J Neurosci 29:3374–3386.

Efron B, Tibshirani RJ (1998) Introduction to the bootstrap. Boca Raton, FL: CRC.

Elgueda D, Duque D, Radtke-Schuller S, Yin P, David SV, Shamma SA, Fritz JB (2019) State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. Nat Neurosci 22:447–459.

Eliades SJ, Tsunada J (2019) Marmosets in auditory research. In: The common marmoset in captivity and biomedical research (Fox JG, Marini RP, Wachtman LM, Tardif SD, Mansfield K, eds), pp 451–475. San Diego: Academic.

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

Gimenez TL, Lorenc M, Jaramillo S (2015) Adaptive categorization of sound frequency does not require the auditory cortex in rats. J Neurophysiol 114:1137–1145.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Oxford: Wiley.

Guo L, Ponvert ND, Jaramillo S (2017a) The role of sensory cortex in behavioral flexibility. Neuroscience 345:3–11.

Guo W, Clause AR, Barth-Maron A, Polley DB (2017b) A corticothalamic circuit for dynamic switching between feature detection and discrimination. Neuron 95:180–194.e5.

Guo L, Weems JT, Walker WI, Levichev A, Jaramillo S (2019) Choice-selective neurons in the auditory cortex and in its striatal target encode reward expectation. J Neurosci 39:3687–3697.

Hackett TA (2011) Information flow in the auditory cortical network. Hear Res 271:133–146.

Hardman CD, Ashwell KW (2012) Stereotaxic and chemoarchitectural atlas of the brain of the common marmoset, Ed 1. Boca Raton, FL: CRC.

Harrington IA, Heffner RS, Heffner HE (2001) An investigation of sensory deficits underlying the aphasia-like behavior of macaques with auditory cortex lesions. NeuroReport 12:1217–1221.

Heffner HE (2005) The neurobehavioral study of auditory cortex. In: The auditory cortex: a synthesis of human and animal research (Heil P, Scheich H, Budinger E, Konig R, eds), pp 111–126. New York: Psychology.

Heilman KM, Hammer LC, Wilder BJ (1973) An audiometric defect in temporal lobe dysfunction. Neurology 23:384–386.

Huang Y, Heil P, Brosch M (2019) Associations between sounds and actions in early auditory cortex of nonhuman primates. eLife 8:e43281.

Hurley L (2019) Neuromodulatory feedback to the inferior colliculus. In: The Oxford handbook of the auditory brainstem (Kandler K, ed). New York: Oxford UP.

Hurley LM, Pollak GD (1999) Serotonin differentially modulates responses to tones and frequency-modulated sweeps in the inferior colliculus. J Neurosci 19:8071–8082.

Jane JA, Masterton RB, Diamond IT (1965) The function of the tectum for attention to auditory stimuli in the cat. J Comp Neurol 125:165–191.

Jaramillo S, Borges K, Zador AM (2014) Auditory thalamus and auditory cortex are equally modulated by context during flexible categorization of sounds. J Neurosci 34:5291–5301.

Johansen JP, Tarpley JW, LeDoux JE, Blair HT (2010) Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. Nat Neurosci 13:979–986.

Kelly JB, Whitfield IC (1971) Effects of auditory cortical lesions on discriminations of rising and falling frequency-modulated tones. J Neurophysiol 34:802–816.

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

Lee CC, Middlebrooks JC (2011) Auditory cortex spatial sensitivity sharpens during task performance. Nat Neurosci 14:108–114.

Liu ST, Montes-Lourido P, Wang X, Sadagopan S (2019) Optimal features for auditory categorization. Nat Commun 10:1302.

Lu T, Liang L, Wang X (2001) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. J Neurophysiol 85:2364–2380.

Malmierca MS (2004) The inferior colliculus: a center for convergence of ascending and descending auditory information. Neuroembryol Aging 3:215–229.

Marshall AF, Pearson JM, Falk SE, Skaggs JD, Crocker WD, Saldaña E, Fitzpatrick DC (2008) Auditory response properties of neurons in the tectal longitudinal column of the rat. Hear Res 244:35–44.

McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci 19:431–441.

Mellott JG, Foster NL, Ohl AP, Schofield BR (2014) Excitatory and inhibitory projections in parallel pathways from the inferior colliculus to the auditory thalamus. Front Neuroanat 8:124.

Mesgarani N, David SV, Fritz JB, Shamma SA (2014) Mechanisms of noise robust representation of speech in primary auditory cortex. Proc Natl Acad Sci USA 111:6792–6797.

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

Moore RC, Lee T, Theunissen FE (2013) Noise-invariant neurons in the avian auditory cortex: hearing the song in noise. PLoS Comput Biol 9: e1002942.

Mothe LA, Blumell S, Kajikawa Y, Hackett TA (2006) Thalamic connections of the auditory cortex in marmoset monkeys: core and medial belt regions. J Comp Neurol 496:72–96.

Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K (2007) Cortical interference effects in the cocktail party problem. Nat Neurosci 10:1601–1607.

Neff WD, Diamond IT, Casseday JH (1975) Behavioral studies of auditory discrimination: central nervous system. In: Auditory system: physiology (CNS). Behavioral studies psychoacoustics. (Abeles M, Bredberg G, Butler RA, Casseday JH, Desmedt JE, Diamond IT, Erulkar SD, Evans EF, Goldberg JM, Goldstein MH, Green DM, Hunter-Duvar IM, Jeffress LA, Neff WD, Yost WA, Zwicker E, Keidel WD, Neff WD, eds), pp 307–400. Handbook of sensory physiology. Berlin: Springer.

Niwa M, Johnson JS, O'Connor KN, Sutter ML (2012) Activity related to perceptual judgment and action in primary auditory cortex. J Neurosci 32:3193–3210.

Olsen WO, Noffsinger D, Kurdziel S (1975) Speech discrimination in quiet and in white noise by patients with peripheral and central lesions. Acta Otolaryngol 80:375–382.

Ono K, Kudoh M, Shibuki K (2006) Roles of the auditory cortex in discrimination learning by rats. Eur J Neurosci 23:1623–1632.

Otazu GH, Tai LH, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. Nat Neurosci 12:646–654.

Phillips EA, Hasenstaub AR (2016) Asymmetric effects of activating and inactivating cortical interneurons. eLife 5:e18383.

Qiu A, Schreiner CE, Escabí MA (2003) Gabor analysis of auditory midbrain receptive fields: spectrotemporal and binaural composition. J Neurophysiol 90:456–476.

Rabinowitz NC, Willmore BD, King AJ, Schnupp JW (2013) Constructing noise-invariant representations of sound in the auditory pathway. PLoS Biol 11:e1001710.

Ryan A, Miller J (1977) Effects of behavioral performance on single-unit firing patterns in inferior colliculus of the rhesus monkey. J Neurophysiol 40:943–956.

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

Schwarz DM, Zilany MSA, Skevington M, Huang NJ, Flynn BC, Carney LH (2012) Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. J Neurosci Methods 206:120–131.

Seybold BA, Phillips EA, Schreiner CE, Hasenstaub AR (2015) Inhibitory actions unified by network integration. Neuron 87:1181–1192.

Shepard KN, Lin FG, Zhao CL, Chong KK, Liu RC (2015) Behavioral relevance helps untangle natural vocal categories in a specific subset of core auditory cortical pyramidal neurons. J Neurosci 35:2636–2645.

Slee SJ, David SV (2015) Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. J Neurosci 35:13090–13102.

Slee SJ, Young ED (2013) Linear processing of interaural level difference underlies spatial tuning in the nucleus of the brachium of the inferior colliculus. J Neurosci 33:3891–3904.

Syka J, Popelár J, Kvasnák E, Astl J (2000) Response properties of neurons in the central nucleus and external and dorsal cortices of the inferior colliculus in guinea pig. Exp Brain Res 133:254–266.

Tramo MJ, Shah GD, Braida LD (2002) Functional role of auditory cortex in frequency processing and pitch perception. J Neurophysiol 87:122–139.

Tsunada J, Lee JH, Cohen YE (2011) Representation of speech categories in the primate auditory cortex. J Neurophysiol 105:2634–2646.

Vaadia E, Gottlieb Y, Abeles M (1982) Single-unit activity related to sensorimotor association in auditory cortex of a monkey. J Neurophysiol 48:1201–1213.

Wang X (2018) Cortical coding of auditory features. Annu Rev Neurosci 41:527–552.

Wenstrup JJ (2005) The tectothalamic system. In: The inferior colliculus (Winer JA, Schreiner CE, eds), pp 200–230. New York: Springer.

Whitfield IC (1980) Auditory cortex and the pitch of complex tones. J Acoust Soc Am 67:644–647.

Williamson RS, Hancock KE, Shinn-Cunningham BG, Polley DB (2015) Locomotion and task demands differentially modulate thalamic audiovisual processing during active search. Curr Biol 25:1885–1891.

Winer JA (2005) Decoding the auditory corticofugal systems. Hear Res 207:1–9.

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

Yu JJ, Young ED (2000) Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. PNAS 97:11780–11786.

Young ED, Calhoun BM (2005) Nonlinear modeling of auditory-nerve rate responses to wideband stimuli. J Neurophysiol 94:4441–4454.

Zatorre RJ (1988) Pitch perception of complex tones and human temporal-lobe function. J Acoust Soc Am 84:566–572.