# Stimulus-specific prediction error neurons in mouse auditory cortex

Nicholas J. Audette[1], David M. Schneider[1]

[1.] Center for Neural Science, New York University, 4 Washington Place, New York, NY 10003, USA

## Abstract

The detection and signaling of prediction errors is central to the theory of predictive processing. Experiments that alter the sensory outcome of an animal's behavior reveal enhanced neural responses to unexpected self-generated stimuli, including many neurons that do not respond to the same stimulus heard passively. These neurons may reflect the violation of a learned sensory-motor prediction, but could also emerge due to a combination of sound and movement in a way that is independent from expectation. Here, we train mice to expect the outcome of a simple sound-generating lever behavior and record neural responses to the expected sound and sounds that deviate from expectation in multiple distinct dimensions. Our data reveal suppression of expected sound responses that is specific across multiple stimulus dimensions simultaneously and stimulus-specific prediction error neurons that depend on sensory-motor expectations.

## Introduction

Sensory responses in the cerebral cortex are influenced by an animal's behavior[1–10] and can reflect an expectation for the sensory consequences of movement[11–22]. This dynamism is consistent with the theory of predictive processing, which posits that cortical activity prioritizes representing deviations from expectation over directly representing features of the external world[23,24]. Indeed, experiments that alter the sensory outcome of an animal's behavior have consistently found that neural responses to unexpected outcomes are enhanced relative to expected outcomes[11,16,17,20,22]. This finding can be attributed in part to the recruitment of neurons that are active only when an animal experiences unexpected self-generated outcomes (e.g. 'prediction error' neurons)[11,20,22]. Prediction error neurons are believed to represent the violation of a learned sensory-motor expectation, but could alternatively arise from the mixing of movement and sensation signals in a way that is unrelated to expectation[25]. If prediction error neurons are driven by expectation violation, it remains unknown whether they encode a generic error signal or whether they reflect the identity of the unexpected stimulus. Most previous experiments have intentionally violated expectations using a small number of stimuli, and typically only in animals that experienced sensorimotor coupling, making it difficult to distinguish between these possibilities. Here, we used a simple sound-generating behavior in mice to establish the presence of stimulus-specific prediction error neurons with features consistent with a cortical origin for predictive computations.

## Results

To assess how movement-based predictions impact neural responses to expected and unexpected sounds, we trained head-fixed mice to produce a simple sound-generating behavior[11]. Mice pushed a lever past a fixed threshold to trigger a water reward (on 25% of trials) when the lever was returned to the home position (Fig 1A). During training, a pure tone (8 kHz) was presented at a consistent position early in each movement, and mice were free to initiate trials ad libitum. Mice rapidly learned to perform the task and averaged more than 2000 sound-generating trials per session.

Following 10-12 days of training with the lever producing a predictable self-generated sound, we made large channel-count electrophysiological recordings from the auditory cortex while mice executed the learned lever behavior and heard either the expected sound (93% of trials) or one that varied in one of several different acoustic dimensions (1% of trials each) (Fig. 1B). On unexpected probe trials we did one of the following: substituted a sound shifted 1.1 octaves from the expected sound (Frequency), played an unexpected frequency simultaneously with the expected sound (Composite), changed the intensity of the expected sound by +/-15 dB (Quiet or Loud), changed the spatial origin of the sound (Origin), played the expected sound at the wrong lever position (Position), or omitted the sound altogether. Each of these sounds was also played in a passive listening context during which the lever was removed from the animal's reach.

We first asked how auditory cortical neurons responded to self-generated sounds that violated a mouse's expectations along different acoustic dimensions by comparing neural responses to each sound heard in the passive versus the self-generated conditions. For each sound, we identified the subset of neurons from the entire recorded population (n = 1016, N = 5) that responded when that sound was heard in either of the passive or self-generated conditions (n = 285, N = 5). In the passive listening condition, we observed strong neural responses to each sound, including the expected sound (Fig. 1B). In the self-generated condition, neural responses to the expected sound were strongly suppressed compared to the same sound heard passively. In contrast, responses to the unexpected probe sounds were not suppressed at all (Quiet, Loud, Origin), were mildly suppressed (Position), or were enhanced relative to the passive listening condition (Frequency). We next quantified the number of recorded neurons that were responsive to each sound in the passive and self-generated conditions. A relatively consistent number of neurons were responsive to each sound in the passive condition (Fig. 1C). When mice heard the expected self-generated sound, only a small subset of passive-responsive neurons responded ('Shared'), as did a small number of new neurons that were not responsive in the passive condition ('Active'). In contrast, when mice heard any sound that violated expectation, a substantially larger number of neurons responded including many neurons that were unresponsive to these same sounds heard passively ('Active').

To assess neural responses to expected and unexpected self-generated sounds at the individual neuron level, we computed a modulation index that compared a neuron's response to a sound heard in the active and passive conditions[11]. The vast majority of neurons had weaker responses to the expected sound when it was self-generated compared to when it was heard passively (negative modulation values) (Fig. 1D). In contrast, neurons responded equally strongly on average to probe sounds when they were self-generated and heard passively, with some neurons enhanced, some suppressed, and many cells responding equally across the two conditions. The notable exception was a substantial number of neurons that had enhanced responses to the self-generated frequency probe, consistent with large population-level neural responses (see Fig. 1B).

A simple circuit model in which somatic inhibition decreases the spiking of neurons tuned to the expected stimulus would lead to comparable inhibition in these same neurons following probe sounds that varied from expectation[6,13,26–28]. Our data are inconsistent with such a simple model. Despite substantial overlap in the population of neurons responsive to the expected sound and unexpected sounds when heard passively (60% ± 15%), we observed little or no suppression to unexpected sounds as measured by aggregate responses, responsive population size, or individual neuron modulation. Instead, we observed many individual neurons that experienced strong suppression of responses to the expected sound during movement while experiencing weaker suppression or even enhancement in response to other self-generated sounds (Fig 1E). These data suggest a more subtle and targeted form of inhibition that can filter neural responses to an expected sensory outcome in a way that is specific across multiple feature dimensions simultaneously.

Our data indicate that strong neural responses to unexpected, self-generated sounds are driven in part by the recruitment of new neurons that do not respond to the same sounds heard passively (See Fig 1C)[11,21,22]. We investigated the abundance of these neurons for different self-generated stimuli to determine whether and how they signal the violation of a learned expectation. First, we identified a subset of neurons (termed 'active only') that responded to a particular stimulus in the active condition but not during quiet movements or to the same stimulus heard passively (Fig 2A, n = 100, 9.8% of sound-responsive population). Next, we used UMAP dimensionality reduction to assist in the visualization of these 'active only' neurons to different probe stimuli across the population (Fig 2A)[29]. Two features of these 'active only' neurons stood out: (1) different numbers of neurons were recruited by different stimuli and (2) 'active-only' neurons were generally non-overlapping across stimuli. Consistent with prior data (Fig1C), we observed a near absence of 'active only' neurons that responded to the expected, self-generated sound (N, Fig 2A). All other stimuli had far more 'active only' neurons, with the most neurons observed following the frequency probe, composite probe, and loud probe. Quantification of neural specificity across stimuli confirmed that most 'active only' neurons (67%) only signaled one type of outcome, with 90% of 'active only' neurons signaling two or fewer different outcomes (Fig 2B). Pairs of stimuli typically only shared a few 'active

only' neurons, with the largest overlap occurring between the frequency and composite probe, which only differ in the presence or absence of the expected sound (Fig 2C).

These abundant and specific 'active only' neurons appear to encode a deviation from expectation, but also could result from the combination of sound and movement in a way that was independent from a learned prediction[25,30]. To test this, we also measured the number of 'active only' neurons in a separate cohort of mice trained to make identical lever movements, but in silence. We recorded neural responses to a variety of sound frequencies in silent-trained animals in both the active and passive condition and found that a very small fraction of the sound-responsive population in each animal fit our 'active only' criteria (Fig 2D). Given the enhancement of 'active only' neurons in sound-trained animals and their specificity for unexpected, but not expected, outcomes, we establish the presence of neurons that reflect the violation of a learned sensory-motor expectation, which we call prediction error neurons moving forward.

We next asked whether the number of prediction error neurons recruited by a stimulus was related to how different the stimulus was from "expected." We measured stimulus similarity using a population-level neurometric approach, computing the absolute difference between each neuron's response to the expected sound and a probe sound across all cells in each animal. Because we aimed to understand how neural response similarity related to the number of prediction error neurons recruited by a particular probe stimulus, we removed all prediction error neurons from this analysis. We found that the difference in population-level neural responses from the expected sound in the passive condition was correlated with the number of prediction error neurons responding to that probe sound (Fig. 2E). Probe sounds that drove population responses similar to the expected sound had few prediction error neurons, whereas probe sounds with population activity patterns that differed from the expected sound had more prediction error neurons. We performed a similar analysis using the magnitude of a sound's response in the passive condition and found no significant correlation with the number of prediction error neurons (Fig 2F), indicating that the engagement of stimulus-specific prediction error neurons reflects the difference between the representation of the experienced sound and the expected sound.

Our definition of a prediction error neuron requires that it is responsive to a self-generated probe sound and unresponsive to the same sound heard passively, but it does not exclude neurons that might respond to other sounds in either of the two behavioral contexts. We therefore examined whether prediction error neurons only responded to motor-sensory discrepancies, or whether they also responded to other sounds. The majority of prediction error neurons were highly specific for just one or two sounds in the active condition and were unresponsive to any of the passive sounds (Fig 3A, left panel; B), leading us to conclude that prediction error neurons are largely distinct from neurons that encode sounds during passive listening. We also observed a small number of neurons that that were unresponsive to all passive sounds but were responsive to many of the self-generated sounds (Fig. 3A, middle panel). Notably, these were the only neurons that responded to the expected sound in the active context but not passive context. Finally, we observed a smaller fraction of neurons that were identified as prediction error neurons – because at least one sound evoked responses in the active but not passive condition – but were also responsive to a subset of other passively heard and self-generated sounds (Fig 3A, right panel). Even when visualizing results from the entire sound-responsive population, the most common neural phenotype was neurons that responded to a single active tone but no passive tones, highlighting the prevalence of specific prediction error neurons throughout the auditory cortex (Fig 3B). The presence of feature-specific prediction error neurons is reminiscent of the receptive field tuning of mismatch neurons observed in the visual cortex[20]. While it is possible that some of the prediction error neurons we identified could respond to passively heard sounds that were not used in this study, our findings are consistent with these neurons playing a role in encoding expectation violations, rather than encoding passive sound properties.

Neurons that signal a discrepancy between expectation and experience could do so via local integration of sound and expectation signals in the auditory cortex or through the computation of prediction errors at a higher cortical level that are transmitted back to the auditory cortex[23]. We reasoned that these two mechanisms would result in error signals with different latencies; top-down errors would arise later than would local computations. We therefore quantified the onset

latency of prediction error neurons measured as the time to first spike following stimulus onset (Fig 3C). Error responses to the frequency probe were as rapid as neural responses to passively heard sounds and responses to self-generated sounds by non-prediction error neurons (Fig. 3D). Onset latencies were similar for prediction error neurons that signaled expectation violations across a broad number of stimuli (3 or more) and those that were specific for a small number of stimuli (1 or 2) (Fig 3E). Given the specificity and early onset of prediction error signals following unexpected sounds, it is unlikely that these neurons are driven by feedback of a general error signal calculated elsewhere in the brain.

While the auditory cortex responds to both self-generated sounds and passively heard sounds, it does so with different overall firing rates (e.g. for the expected and frequency probe sounds) and with different ensembles of auditory cortical cells. We conclude our investigation by asking whether these differential patterns of neural responses are sufficient to decode the identity of self-generated sounds on individual trials. For each animal, we measured neural activity following expected, unexpected, and passive sounds, as well as on omission trials and at randomly selected timepoints during behavior ('Null'). We trained a multiclass model using support vector machine (SVM) binary learners to classify sound identity for each trial based on sound responses (0-60ms) of all neurons on an individual trial (Fig 4A)[31,32]. For each test trial, the model was trained on all other trials prior to classification, and we visualized the resulting classifications as a confusion matrix averaged across animals. We find that relatively small groups of sound-responsive neurons (60 ± 13) are sufficient to classify each unique sound, and to distinguish whether each sound was heard in the self-generated or in the passive condition. (Fig 4B) This includes the reliable classification of expected sounds, which have a strongly suppressed neural response, from all other time points, including those that corresponded to silent movements (omission trials) or random timepoints during behavior.

Overall, our data reveal that movement-based predictions result in sound-suppression and violation-signaling that are specific across multiple feature dimensions. First, we observed prediction-based suppression that was specific for the frequency, intensity, timing, and spatial origin of an expected sound at the population level and within individual neurons. Second, we observe abundant prediction error neurons in the auditory cortex only when a movement has an unexpected acoustic outcome, consistent with an experience-dependent mechanism for filtering expected sounds and detecting deviants. Finally, individual prediction error neurons typically respond with short latency and to just one probe stimulus, indicating that these neurons do not reflect the feedback of a generic error signal calculated elsewhere in the brain.

## Figures

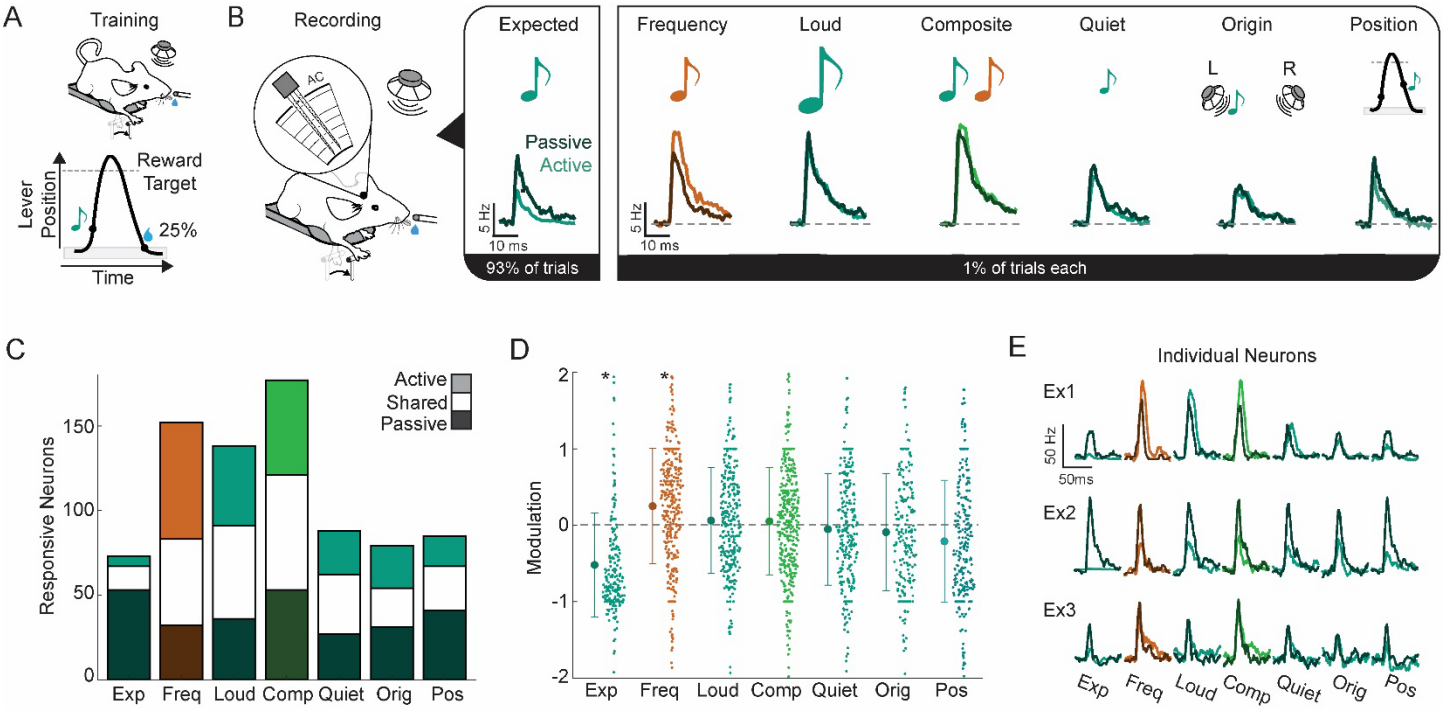


**Figure 1: Specific suppression of expected sounds across multiple acoustic dimensions.** (A) Schematic of head-fixed lever press training paradigm (Top) and stimulus and reward timing for lever movements (Bottom). Grey area indicates home position. (B) Schematic of multi-array recording sessions in trained mice (Left) and aggregate neural responses to expected and multiple unexpected sounds in the passive (Darker) and movement-evoked (Lighter) context. Neurons are included if they respond to a given sound type in either context ($p < 0.01$, 0 - 60ms post sound onset). Color differences represent sound frequency, and the fraction of each sound type is displayed in black bar. (C) Number of neurons responsive ($p < 0.01$) to a given sound in the active context (light), passive context (dark), or both (white). Total bar height represents the n value for each stimulus type for (B) and (D). (D) Modulation (See Methods) of individual neurons comparing responses in the active and passive context to each tone type. Negative values indicate weaker responses in the active condition, i.e. suppression. A one-way ANOVA detected differences amongst the groups (F-statistic $p = 2*10^{-32}$), with Exp and Freq being significantly different from all other groups (Exp, $p < 1*10^{-5}$; Freq, $p < 0.01$). (E) Average responses of three individual neurons to each tone type, showing suppression that is specific for the expected sound at the individual neuron level.

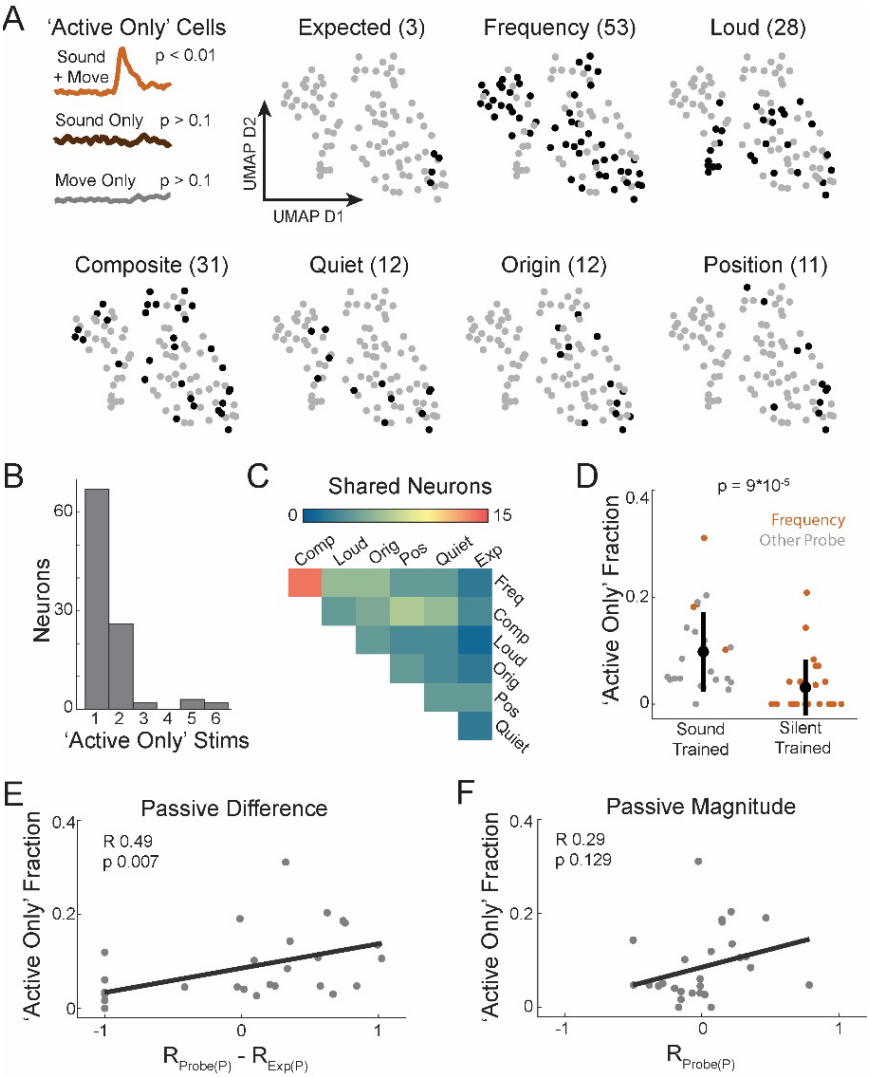


**Figure 2: Non-overlapping prediction error neurons reflect a learned expectation.** (A) Schematic depicting the identification of 'active only' neurons which respond to a given stimulus type in the active context, but not in the passive context or during movement (Left). Example neuron displays qualifying responses to the frequency probe. Visualization of 'active only' neuron quantity (dark dots) for each stimulus type plotted on a consistent UMAP distribution generated using the neural response properties of all 'active only' neurons (n = 100). (B) Quantification of the number of neurons that have a qualifying 'active only' response to a given number of stimuli. (C) Color-coded matrix showing the number of 'active only' neurons that are shared across pairs of stimuli. (D) Quantification of the number of 'active only' neurons to a given stimulus in trained animals, and animals trained on an identical but silent lever task). Each dot represents the fraction of total number of neurons in a recording session that met the criteria for 'active only' neurons for a particular stimulus. During physiology, silent trained animals hear multiple sound frequencies at an identical intensity and duration to the expected sound in sound-trained animals. Orange dots for sound trained animals denote the frequency probe for comparison, error bars represent standard deviation. (E) Linear regression of the number of 'active only' neurons (as in D) compared to the difference between the response to a given sound heard passively and the expected sound heard passively (See Methods). Each dot represents a stimulus for an animal (N = 4), and difference values were mean-normalized within animal for comparison. (F) Identical comparison as (E) but using the absolute magnitude of a neuron's response to a passive sound.

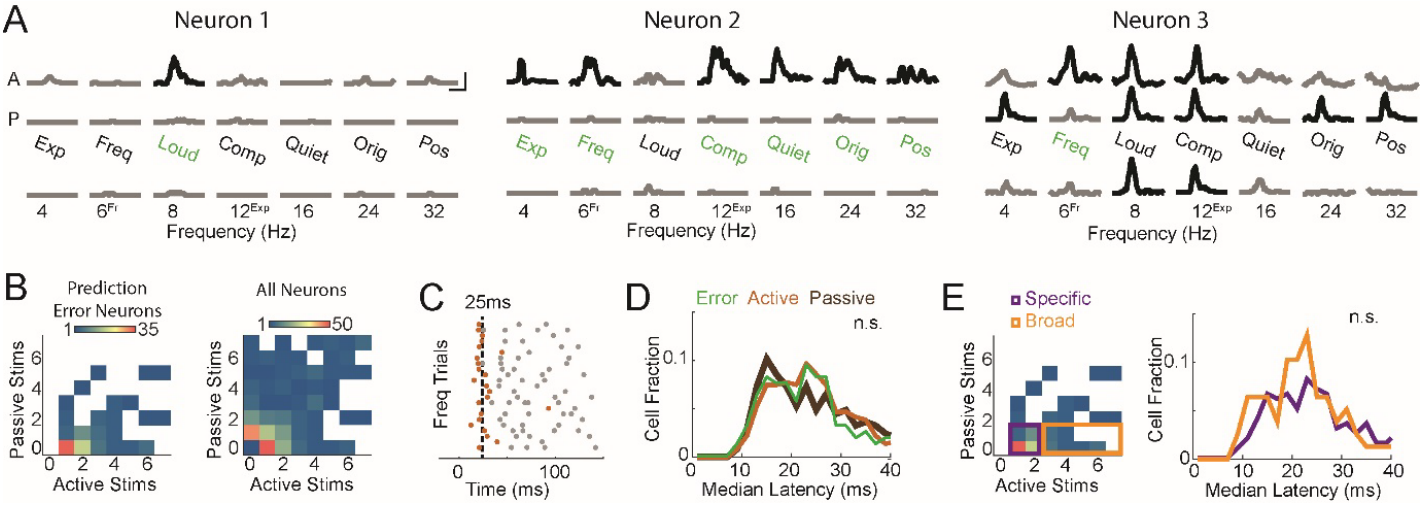


**Figure 3: Error responses are short latency and stimulus-specific.** (A) Responses of three example neurons to sounds heard actively (Top) and passively (Middle), as well as responses to a range of sound frequencies heard passively (Bottom). Black PSTHs show significant responses ($p < 0.01$), and green text indicates the stimuli for which a neuron has an error response. Superscripts denote the trained and probe frequency for each neuron and scale bar is 25ms by 20hz. (B) Map showing the number of neurons that have a significant response ($p < 0.01$) to a given number of active and passive task stimuli for prediction error neurons (Left) and the entire sound-responsive population (Right). (C) Raster of example neuron showing action potential timing following frequency probe, with the first spike on a given trial (orange) used to calculate an average onset latency. (D) Histogram of average onset latency following frequency probe trials for prediction error neurons (Green), all neurons responsive to the frequency probe (Orange), and latency of neurons responsive to the passive frequency probe following passive presentation. No difference between prediction error neuron latencies and all active response latencies ($p = 0.97$) or to passive sound responses ($p = 0.87$) was detected by KS test. (E) Histogram of onset latencies for prediction error neurons that respond to a small (Purple) or large (Orange) number of sounds with no difference ($p = 0.28$) detected by KS test.

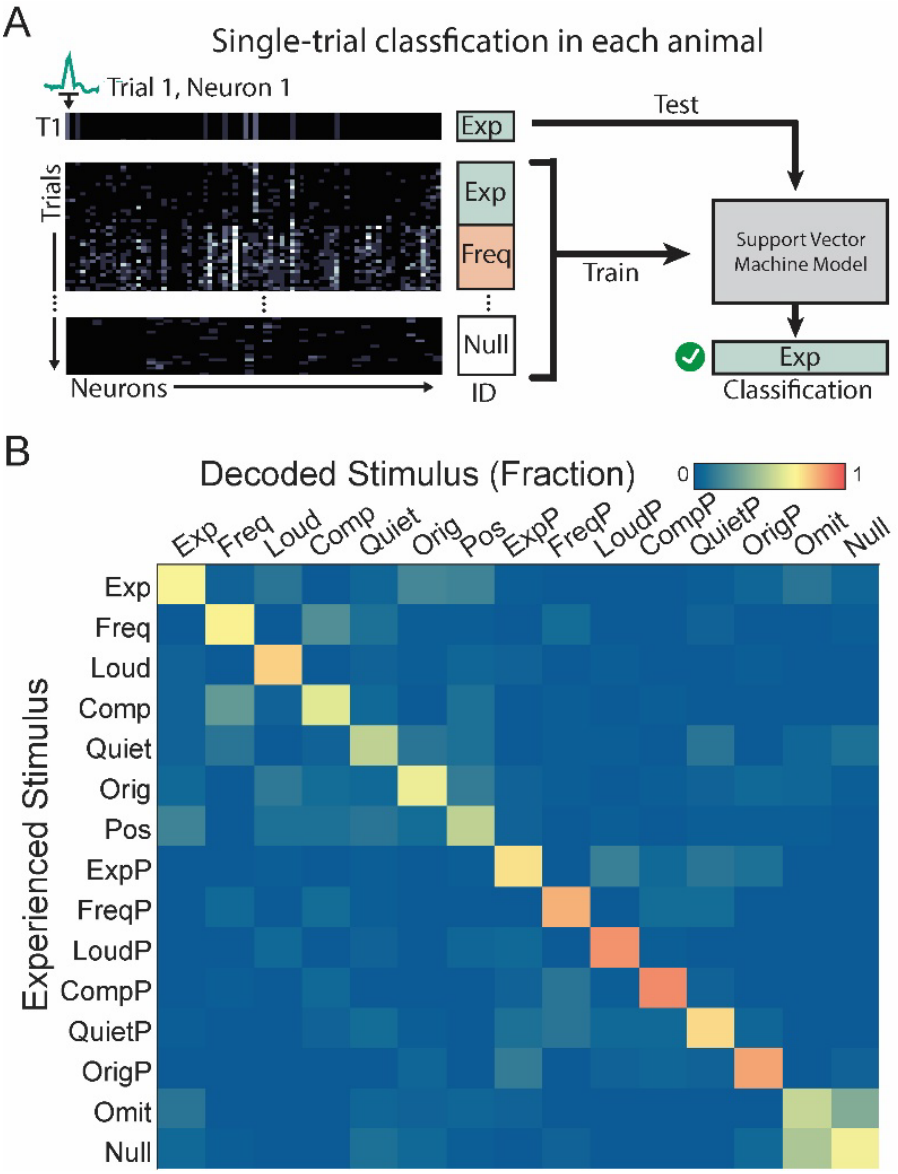


**Figure 4: Auditory cortex neurons can detect and classify self-generated sounds.** (A) Schematic of decoding data and process. For each animal (N = 4), we measured the response of each neuron to a given sound on 20 individual trials for each sound type, including active and passive experimental stimuli, omission trials, and a set of randomly selected time points during behavior that were < 0.3s away from other sounds ('Null'). Response values were organized in a trials x neuron matrix with a ground truth identifier. For each trial, we trained a support vector machine model on all other identified trials, and then used the model to classify the withheld trial. Classification yielded a confusion matrix for each animal displaying the fraction of trials that were assigned to a given class compared for each stimulus type. (B) Average confusion matrix across all animals (N = 4) showing the fraction of trials for an experienced stimulus that were identified as each stimulus type by the model.

## Methods

### Animals

All experimental protocols were approved by New York University's Animal Use and Welfare Committee. Male and female wild-type (C57BL/6) mice were purchased from Jackson Laboratories and were subsequently housed and bred in an onsite vivarium. We used 2–4 month-old mice for our experiments that were kept on a reverse day-night cycle (12h day, 12 h night).

### Surgeries

For all surgical procedures, mice were anaesthetized under isolfurane (1-2% in $O_2$) and placed in a stereotaxic holder (Kopf), skin was removed over the top of the head, and a Y-shaped titanium headpost (H.E. Parmer) was attached to the skull using a transparent adhesive (Metabond). Mice were treated with an analgesic (Meloxicam SR) and allowed to recover for 5 days prior to training. Following training and 24 to 48 hours prior to electrophysiology, a small craniotomy was made to expose the auditory cortex (~2mm diameter, -2.5mm posterior, 4.2 mm left from bregma). Another small craniotomy was made above the right sensory cortex and a silver-chloride reference electrode was positioned atop the surface of the brain for use as a ground electrode and covered (Metabond). Exposed craniotomies were covered with a silicone elastomer (Kwik-Sil) and the mouse was allowed to recover in its home cage, and an additional training session was performed prior to electrophysiology.

### Behavioral Training and Data Collection

We adapted a custom head-restrained lever-based behavioral training paradigm where mice push a lever and hear closed-loop sounds[11]. A custom-designed lever (7cm long, 3D printed using Formlabs Form2) was mounted to the post of a rotary encoder (US Digital) 5cm from the lever handle. A magnet (CMS magnetics) was mounted to the bottom of the lever, which was positioned 4cm above a larger static magnet which established the lever resting position and provided light and adjustable movement resistance. The lever handle (top) was positioned adjacent to a tube (Custom, 3D printed using Formlabs Form2) to hold mice directly below two plate clamps (Altechna) to secure the mouse headpost. Lever and mouse apparatus was constructed with Thor-labs components. A water tube, controlled by a solenoid valve (The Lee Company), was positioned in front of the mouse. Digital signals for lever movement were collected by a data acquisition card (National Instruments) connected to a computer and logged by custom Matlab software (Mathworks, PsychToolBox) and sampled at 2Khz. Digital processing of lever movements received sufficient processing in real time to track important movement thresholds, which were used to trigger sound events based on user-defined closed-loop rules. Sound output was delivered from the computer to a sound card (RME Fireface UCX), the output of which was routed to an ultrasonic speaker (Tucker Davis Technologies) located lateral to the mouse, ~10cm from the mouse's right or left ear. We recorded sounds during test experiments using an ultrasonic microphone (Avisoft, Model # CM16/CMPA-P48) positioned 5 cm from the lever to confirm that the lever produced negligible noise (<1 dB SPL) and that experimenter-controlled sounds were delivered at a consistent volume of 50, 65, and 80 db depending on stimulus type. All training was performed in a sound-attenuating booth (Gretch-Ken) to minimize background sound and monitored in real-time via IR video.

During lever training, mice were water restricted and maintained greater than 80% of pre-restriction body weight and received all of their water (1-2ml) while performing the lever behavior. In practice, body weight was often above 90% since diminished body weight was not necessary to induce lever pressing once mice learned the task. During training, mice were head-fixed to the behavioral apparatus and presented with the lever and lick-port after ~10 minutes of quiet acclimation. Mice were then allowed to make outwards lever movements at will. For a movement to be considered valid, we required the lever to remain in the home position (~+/- 3mm from rest) for >200 ms prior to initiation. Valid movements that reached a reward threshold (~15mm from home position) elicited a small water reward (5-10uL) when the lever returned to home position. Auditory feedback in the form of a pure tone (50ms duration, 65dB, 12khz) was

delivered on all trials when the lever crossed a set threshold 1/3 of the way between the home position and reward threshold for the first time in a trial. To ensure strong coupling between movement and sound, auditory feedback was provided on all trials, regardless of whether mice obeyed the home-position requirement and would subsequently receive a reward. Initially, 100% of successful trials produced a reward, but over the course of training that number was dropped to 25% to produce more lever movements per session. The reward rate was stable for at least 5 sessions prior to recording. Overall, mice received between 18 and 22 sessions of training over 10-12 days prior to electrophysiology, with either one or two sessions per day.

**Electrophysiological Recording and Aggregate Neural Responses**

Following training, surgical opening of a craniotomy, and one subsequent training session, mice were positioned in the behavioral apparatus and a 128-channel electrode (128AxN, Masmanidis Lab) was lowered into the auditory cortex orthogonal to the pial surface[33]. The electrode was connected to a digitizing head stage (Intan) and electrode signals were acquired at 30khz, monitored in real time, and stored for offline analysis (OpenEphys). The probe was allowed to settle for at least 20 minutes, at which point the lever and lick-port were introduced and mice were allowed to make lever movements at will as in any other training session. After performing at least 30 standard lever movements, we unexpectedly began a probe session in which mice heard several different sounds. 93% of sounds were as expected ('Exp', 12khz, 65db) while 1% each were a substituted frequency ('Freq', 5.6khz, 1.1 octave lower, 65db), both the unexpected and an unexpected frequency ('Comp', 5.6khz and 12khz, 65db), a higher intensity ('Loud', 12khz, 80db), a lower intensity ('Quiet', 12khz 50db), played from a different origin ('Orig', 12khz, 65db, played from a speaker on the left side of the mouse's head), played during the return phase of the lever movement ('Pos', 12khz, 65db, half way between reward threshold and the return to the home position on trials reaching reward threshold), or omitted. The requirements for reward delivery were not influenced by the identify or timing of auditory feedback. Following probe sessions, the lever was removed and tone frequencies ranging from 3 to 32kHz (0.5 octave spacing) as well as all tones presented during the active phase of the task were presented with random inter-tone intervals drawn from a flat distribution with range 1 to 2 seconds.

After recording, electrical signals were processed and the action-potentials of individual neurons were sorted using Kilosort2.5[34], and manually reviewed in Phy2 based on reported contamination, waveform principal component analysis, and inter-spike interval histograms. Because the identification of prediction error neurons could be dramatically skewed by the loss of neural signals over the course of an experiment, we excluded any neuron that had a statistically significant difference ($p<0.05$) in baseline firing rate or the response rate to passively heard tones from the pre- and post-behavioral passive tone sessions. We analyzed neurons with non-fast-spiking waveforms, separated by plotting peak to valley ratio against action potential width. Tone-evoked average firing rate PSTHs were measured in 2 ms bins and aligned to sound onset for each neuron for each tone type (Fig 1B). PSTHs and individual neuron modulation for a given tone type include all neurons that were responsive ($p<0.01$) to a given tone in either the active or passive condition measured as an increase in firing rate from baseline (60ms prior to stimulus onset) during the sound response window (0-60ms post stimulus onset) across trials. To measure the movement-based modulation of each neuron's responses to the lever-associated or probe tones, we compared the neural sound response in our analysis window to the same sound in the active and passive condition using a radial modulation index. Radial modulation was calculated as the theta value resulting from a cartesian to polar transformation of the response strength in the active condition compared to the response strength in the passive condition. Theta values were converted to a scale of +/- 2 and rotated such that a value of 0 corresponded to equal responses across the two conditions. The fraction of neuron overlap reported in the text measures the fraction of neurons responsive to the passively heard expected sound that also respond to each probe sound.

In a subset of animals, we performed electrophysiological recording of mice that had been trained on an identical version of the lever task but without sound feedback. On experiment day mice first performed silent lever pushes for 20-

50 trials, then we delivered a range of sound frequencies (4-24khz, half octave intervals, 50ms duration, 65dB,) at the sound threshold during lever pushes, followed by presentation of the same sounds passively with the lever removed, as above.

**Prediction error Neuron Analysis**

'Active Only' neurons were defined as having a significant response in the sound response window ($p<0.01$, 0-60ms post stimulus onset compared to baseline) for a given stimulus type, but not to the same stimulus heard passively ($p>0.1$) or at the same position during movement on omission trials ($p>0.1$). 'Active Only' neurons were identified independently for each stimulus type. Visualization of 'active only' neurons for each stimulus type used a 2-dimensional representation of neural space generated by UMAP[29]. This representation was generated based on a neuron-by-stimulus type matrix where each cell contained the average response of a neuron to each self-generated tone type. 'Active only' neurons were identified in silent trained animals using the same functional definition comparing activity in the movement condition, passive sound condition, and active condition.

The fraction of 'active only' neurons (Fig 2D-F) was defined as the number of 'active only' neurons for a stimulus type divided by the total number of neurons responsive to that sound in any context (active or passive) and are presented with data points representing a single stimulus in one animal. For analyses involving individual animals (Fig2D-F, Fig 4), data was analyzed only for animals that had more than 40 sound responsive neurons in the population ($N = 4$). For regression comparisons, the neurometric difference between each probe stimulus was calculated by comparing the average response of each neuron to an unexpected sound to its response to the expected sound, both in the passive condition. The absolute value of each difference was then summated across all neurons in an animal and used to represent the average population response difference between two stimuli. These values were mean normalized within each animal to allow for comparison across animals. A similar process was used for passive response magnitude, but with average firing rates summated across all neurons in an animal instead of making a comparison to the expected sound. Onset latencies were defined for each neuron as the average of first post-stimulus spike times on each trial. Trials that did not produce an action potential in the sound response window were removed from the average. Histograms of onset latencies were created using 2ms bins.

**Decoding Analysis**

Decoding data were organized in a trials-by-neuron matrix within each animal, with each cell representing the response of an individual neuron on an individual trial. A consistent number of trials (20, randomly selected) was used for each stimulus type. Each trial, in sequence, was removed from the data set, and the remaining trials along with the ground truth identity of the experienced stimulus was used to train a multiclass error-correcting output codes model using support vector machine binary learners[31,32]. The trained model was then used to classify the withheld trial, which was then compared to the ground truth identify of the stimulus. This process was repeated for all trials in an animal, with the results visualized as a confusion matrix comparing the classification result to the ground truth identity of each trial. Each cell represents the number of trials classified as a given stimulus type divided by the number of ground truth trials for a given stimulus type. The resultant confusion matrices were then averaged across animals.

**Statistical Analysis**

Throughout, animal values are denoted by a capital N while cell values are denoted by a lowercase n. Unless otherwise reported, all averages and error bars denote mean ± standard deviation. P values are reported in text or on the relevant figure panels for all statistical comparisons. Statistical comparison of aggregate neural activity use a one-way ANOVA followed by two-sided, non-paired, non-parametric rank sum test and Bonferroni correction for multiple comparisons. Statistical comparison of onset latency across groups was performed using a Kolmogorov-Smirnov test. The relationship between the number of prediction error neurons and neural response properties was measured using linear regression and correlation coefficient analysis with p and R values reported.

**Acknowledgements**

We thank Alessandro La Chioma, Ralph Peterson, and Grant Zempolich for their thoughtful comments on the manuscript. We thank members of the Schneider lab for fruitful discussions. We thank Jessica A Guevara for expert animal care and technical support. This research was supported by the National Institutes of Health (1R01-DC018802 to DMS); a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (D.M.S); fellowships from the Searle Scholars Program, the Alfred P. Sloan Foundation, and the McKnight Foundation (D.M.S.); and an investigator award from the New York Stem Cell Foundation (D.M.S). D.M.S. is a New York Stem Cell Foundation - Robertson Neuroscience Investigator.

**Declaration of Interests**

The authors declare no competing interests.

## References


1. Clancy, K.B., Orsolic, I., and Mrsic-Flogel, T.D. (2019). Locomotion-dependent remapping of distributed cortical networks. Nat. Neurosci. *22*, 778–786. 10.1038/s41593-019-0357-8.

2. Niell, C.M., and Stryker, M.P. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. Neuron *65*, 472–479. 10.1016/j.neuron.2010.01.033.

3. Ayaz, A., Stäuble, A., Hamada, M., Wulf, M.A., Saleem, A.B., and Helmchen, F. (2019). Layer-specific integration of locomotion and sensory information in mouse barrel cortex. Nat. Commun. *10*. 10.1038/s41467-019-10564-8.

4. Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273. 10.1038/s41586-019-1787-x.

5. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science (80-. ). *364*. 10.1126/science.aav7893.

6. Zhou, M., Liang, F., Xiong, X.R., Li, L., Li, H., Xiao, Z., Tao, H.W., and Zhang, L.I. (2014). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. Nat. Neurosci. *17*, 841–850. 10.1038/nn.3701.

7. Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci. *22*, 1677–1686. 10.1038/s41593-019-0502-4.

8. McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015). Waking State: Rapid Variations Modulate Neural and Behavioral Responses. Neuron *87*, 1143–1161. 10.1016/j.neuron.2015.09.012.

9. Kuchibhotla, K. V., Gill, J. V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., and Froemke, R.C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci. *20*, 62–71. 10.1038/nn.4436.

10. Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *16*. 10.1038/nn.3464.

11. Audette, N.J., Zhou, W., Chioma, A. La, and Schneider, D.M. (2022). Precise movement-based predictions in the mouse auditory cortex. Curr. Biol., 1–16. 10.1016/j.cub.2022.09.064.

12. Schneider, D.M., Sundararajan, J., and Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. Nature *561*, 391–395. 10.1038/s41586-018-0520-5.

13. Nelson, A., Schneider, D.M., Takatoh, J., Sakurai, K., Wang, F., and Mooney, R. (2013). A circuit for motor cortical modulation of auditory cortical activity. J. Neurosci. *33*, 14342–14353. 10.1523/JNEUROSCI.2275-13.2013.

14. Leinweber, M., Ward, D.R., Sobczak, J.M., Attinger, A., and Keller, G.B. (2017). A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. Neuron *95*, 1420-1432.e5. 10.1016/j.neuron.2017.08.036.

15. Reznik, D., Guttman, N., Buaron, B., Zion-Golumbic, E., and Mukamel, R. (2021). Action-locked Neural Responses in Auditory Cortex to Self-generated Sounds. Cereb. Cortex *31*, 5560–5569. 10.1093/cercor/bhab179.

16. Rummell, B.P., Klee, J.L., and Sigurdsson, T. (2016). Attenuation of responses to self-generated sounds in auditory cortical neurons. J. Neurosci. *36*, 12010–12026. 10.1523/JNEUROSCI.1564-16.2016.

17. Eliades, S.J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature *453*, 1102–1106. 10.1038/nature06910.

18. Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E., and Knight, R.T. (2010). Single-trial speech suppression of auditory cortex activity in humans. J. Neurosci. *30*, 16643–16650. 10.1523/JNEUROSCI.1809-10.2010.

19. Knolle, F., Schwartze, M., Schröger, E., and Kotz, S.A. (2019). Auditory Predictions and Prediction Errors in

Response to Self-Initiated Vowels. Front. Neurosci. *13*, 1–11. 10.3389/fnins.2019.01146.

20. Zmarz, P., and Keller, G.B. (2016). Mismatch Receptive Fields in Mouse Visual Cortex. Neuron *92*, 766–772. 10.1016/j.neuron.2016.09.057.

21. Jordan, R., and Keller, G.B. (2020). Opposing Influence of Top-down and Bottom-up Input on Excitatory Layer 2/3 Neurons in Mouse Primary Visual Cortex. Neuron *108*, 1194-1206.e5. 10.1016/j.neuron.2020.09.024.

22. Keller, G.B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. Neuron *74*, 809–815. 10.1016/j.neuron.2012.03.040.

23. Keller, G.B., and Mrsic-Flogel, T.D. (2018). Predictive Processing: A Canonical Cortical Computation. Neuron *100*, 424–435. 10.1016/j.neuron.2018.10.003.

24. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical Microcircuits for Predictive Coding. Neuron *76*, 695–711. 10.1016/j.neuron.2012.10.038.

25. Muzzu, T., and Saleem, A.B. (2021). Feature selectivity can explain mismatch signals in mouse visual cortex Graphical abstract. CellReports *37*, 109772. 10.1016/j.celrep.2021.109772.

26. Singla, S., Dempsey, C., Warren, R., Enikolopov, A.G., and Sawtell, N.B. (2017). A cerebellum-like circuit in the auditory system cancels responses to self-generated sounds. Nat. Neurosci. *20*, 943–950. 10.1038/nn.4567.

27. Wilson, N.R., Runyan, C.A., Wang, F.L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. Nature *488*, 343–348. 10.1038/nature11347.

28. Schneider, D.M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature *513*, 189–194. 10.1038/nature13724.

29. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. bioRxiv.

30. Saleem, A.B., Ayaz, A., Jeffery, K., Harris, K.D., and Carandini, M. Integration of visual motion and locomotion in mouse visual cortex. 10.1038/nn.3567.

31. Cristianini, N., and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods (Cambridge University Press) 10.1017/CBO9780511801389.

32. Narsky, I., and Porter, F.C. (2013). Reducing Multiclass to Binary. Stat. Anal. Tech. Part. Phys. *1*, 371–379. 10.1002/9783527677320.ch16.

33. Yang, L., Lee, K., Villagracia, J., and Masmanidis, S.C. (2020). Open source silicon microprobes for high throughput neural recording. J. Neural Eng. *17*. 10.1088/1741-2552/ab581a.

34. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Kenneth D., H. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv, 061481. 10.1101/061481.