nature communications

Article

https://doi.org/10.1038/s41467-025-58034-8

Location-specific neural facilitation in marmoset auditory cortex

Received: 21 January 2024

Accepted: 6 March 2025

Published online: 20 March 2025

Check for updates

Chenggang Chen[1,2], Sheng Xu[1,2], Yunyan Wang[1] & Xiaoqin Wang[1]

A large body of literature has shown that sensory neurons typically exhibit adaptation to repetitive stimulation. However, adaptation alone does not account for the ability of sensory systems to remain vigilant to the environment in spite of repetitive sensory stimulation. Here, we investigated single neuron responses to sequences of sounds repeatedly delivered from a particular spatial location. Instead of inducing adaptation, repetitive stimulation evoked long-lasting and location-specific facilitation (LSF) in firing rate of nearly 90% of recorded neurons. The LSF decreased with decreasing presentation probability and diminished when sounds were randomly delivered from multiple spatial locations. Intracellular recordings showed that repetitive sound stimulation evoked sustained membrane potential depolarization. Computational modeling showed that increased arousal, not decreased inhibition, underlies the LSF. Our findings reveal a novel form of contextual modulation in the marmoset auditory cortex that may play a role in tasks such as auditory streaming and the cocktail party effect.

It has been well established that responses of sensory neurons are influenced by stimulus context. Contextual effects are typically suppressive when the same stimulus is repeatedly presented. For instance, the presence of a preceding sound usually suppressed a neuron's responses to a succeeding sound of the same feature in auditory cortex[1–3]. Similarly, visual cortical neurons exhibited repetition suppression to the image that is presented twice[4,5]. Repetitive whisker stimulation also reduced the responses of somatosensory neurons[6,7]. In addition, neural activities in the olfactory pathway habituated to the repeated odor presentation[8,9]. An important property of repetition suppression in sensory neurons is stimulus-specific adaptation (SSA) to standard but not deviant stimuli. SSA has been studied extensively in various modalities, including auditory[10–12], visual[13,14], somatosensory[15,16] and olfactory[8,17] systems. SSA has attracted much interest due to its potential correlation with mismatch negativity (MMN)[10,18], as well as its role in deviance detection[3,19], predictive coding[20], and efficient coding[21,22].

However, facilitation instead of SSA could be evoked by repetitive auditory stimuli under certain circumstances[23,24]. Facilitation of repeated stimuli has also been reported in the visual system[25,26] and the somatosensory system[27,28]. In the auditory system, studies have identified different stimulus paradigms that could elicit facilitation to repetitive or high-probability stimuli (Supplementary Fig. 1). For example, a stronger response to sequential rather than random sinusoidal amplitude-modulated stimuli was observed[29,30]; a stimulus that normally evoked below-average response can evoke near-average responses if this stimulus was presented with a higher probability than other stimuli[31,32]; facilitation to particular combinations of masker-probe stimuli were observed[1,33,34]; responses to repeated tones were higher than the responses to random tones[23,24]. Furthermore, a small number of inhibitory neurons were found to exhibit facilitation to repetitive tones[2,35].

Despite this evidence, repetition facilitation in sensory neurons appears to be a rare phenomenon, as the vast majority (over 90%) of sensory neurons exhibit repetition suppression. It is also unclear whether repetition facilitation is neuron-specific or stimulus-specific and, if it is the latter, which types of repetitive stimuli could evoke facilitation. In this study, we investigated spatial contextual modulations of auditory cortex neurons in awake marmoset monkeys and discovered a unique type of repetition facilitation, referred to as location-specific facilitation (LSF). By stimulating neurons with a sequence of repetitive sounds delivered from a single spatial location,

[1]Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [2]These authors contributed equally: Chenggang Chen, Sheng Xu. e-mail: xiaoqin.wang@jhu.edu

we found, to our surprise, that repetitive stimulation from spatial locations away from the center of a neuron's spatial receptive field evoked long-lasting facilitation instead of adaptation as would be expected from the well-documented SSA literature. Our results revealed that nearly 90% of recorded neurons in the marmoset auditory cortex exhibited LSF. In general, repetitive stimulation delivered from nonpreferred spatial locations increased firing rates, whereas stimulation from preferred spatial locations decreased firing rates. Intracellular recordings further showed that repetitive stimulation from nonpreferred spatial locations evoked sustained membrane potential depolarization that gave rise to the observed firing rate facilitation. Computational modeling captured the findings from intra- and extracellular recordings based on the hypothesis of arousal level changes, rather than short-term synaptic plasticity. Two behavioral experiments using the same stimulus paradigms confirmed the models' hypothesis. Overall, our findings shed further light on contextual modulation in the auditory cortex and suggest that this form of spatial contextual modulation may play a role in functions such as regularity detection and sound stream segregation.

Results

Repetitive sound stimulation evoked neural firing facilitation

We tested a total of 123 single neurons in the auditory cortex on four hemispheres of three awake marmosets using wideband stimuli, including wideband noises, amplitude-modulated wideband noises, and marmoset vocalizations. Each neuron was extensively studied using the techniques previously described[36]. The stimulus duration was fixed in each session to a value between 200–2200 ms (200 ms in > 80% of the sessions). We placed fifteen equally spaced speakers on a semi-spherical surface centered around an animal's head and above the horizontal plane (Fig. 1a, b).

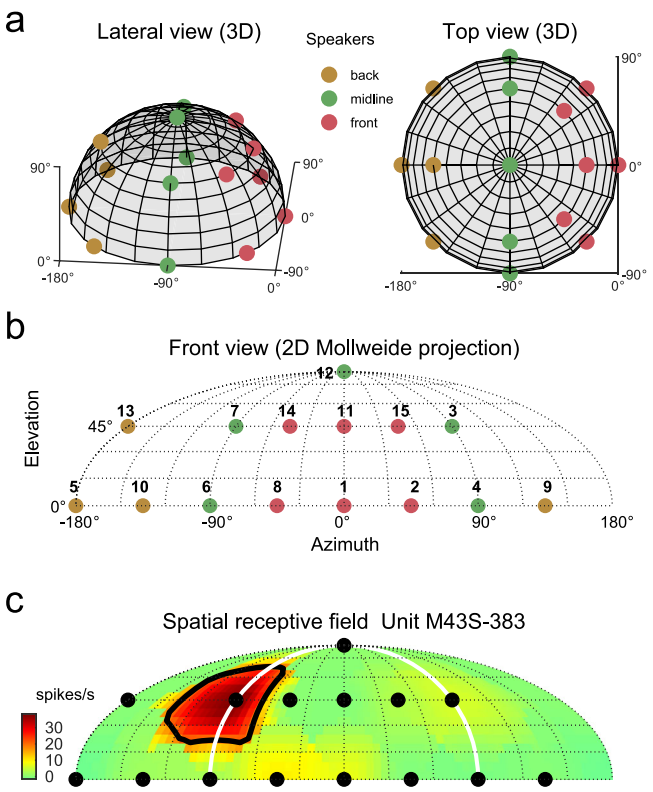

Fig. 1 | Speaker layout and equal-probability sound stimulation. a Fifteen equally spaced speakers were placed on a semi-spherical surface centered around the animal's head and above the horizontal plane. View from the front, lateral 35° and elevated 30° (left), and view from directly on top (right). Red dots indicate six front speakers, green dots indicate five midline speakers and brown dots indicate four back speakers. **b** Three-dimensional front-back space was projected to a two-dimensional plane around the midline for displaying purposes. **c** Spatial receptive field of example Unit M43S-383. The white semicircle is the boundary of the front-back space. The black line is the threshold which is defined as the half-maximum firing rate. Black dots indicate fifteen speaker locations.

In each test session, we first delivered a wideband noise stimulus from each speaker location in a randomly shuffled order. We refer to this stimulation paradigm as the "equal-probability presentation" mode for which all locations have the same occurrence probability of 1/15. Stimuli were repeated at least 5 times at each location (median: 25 repetitions). A spatial receptive field (SRF) was constructed for each neuron using the averaged firing rate of the responses to wideband noises in the equal probability presentation mode. Figs. 1c and 2a show the responses of an example neuron obtained in the equal-probability presentation mode. This neuron had an SRF centered around speaker #7 (Fig. 1c) and responded to this location with sustained firing throughout the stimulus duration, and to other locations with onset or transient firing (Fig. 2a). We ranked 15 speakers based on their firing rates obtained in the equal-probability presentation mode. The speaker with the highest (lowest) firing rate was ranked 1st (15th). Fig. 2b shows the firing rate versus speaker number and speaker rank number for this neuron obtained in the equal-probability presentation mode. Speaker #7 evoked the highest firing rate (38.5 spikes/sec), followed by speaker #8 (10.5 spikes/sec). Speakers #14, #13, #9, and #11 evoked a near-zero firing rate, and a negative firing rate was observed for speakers #2, #4, and #5. Because the firing rate was calculated by subtracting the spontaneous firing rate from the total firing rate, a negative firing rate indicated inhibition.

After the characterization of a neuron's SRF, we further tested each neuron with the "100%-probability presentation" mode in which a stimulus was repeatedly delivered from a single target speaker location (at least 50 times, median: 200 repetitions), with each repetition separated by an inter-stimulus interval of fixed or variable length (range: 500–5200 ms, see below). Fig. 2c, d show the responses of the same neuron depicted in Fig. 2a, b to 300 repetitions of a 200 ms wideband noise stimulus delivered from speaker #14 in the 100%-probability presentation mode. Speaker #14 evoked 0.8 spikes/sec firing rate in the equal-probability presentation mode (Fig. 2b, blue dot) which was considered as the baseline firing rate of this speaker location. In contrast to the expectation from previous auditory cortex literature on adaptation, the response of this neuron to the repeated presentations of a wideband noise stimulus from the same speaker #14 showed epochs consisting of consecutive trials with substantially higher firing rates than the baseline firing rate (median: 16 spikes/sec, Fig. 2d). Elevated firing rates above the baseline could be observed in trials long after the first trial (e.g., between 150th and 250th trials). Also note that during the trials with elevated firing rates, the firing patterns were sustained throughout the stimulus duration. Fig. 2e–h shows responses of another example neuron which displayed four epochs of elevated firing rate in the 100%-probability presentation mode.

Quantification of facilitation trials evoked by repetitive stimulation

To quantify the facilitated response in the 100%-probability presentation mode, trials with a firing rate exceeding a facilitation threshold were considered facilitation trials. We defined the facilitation threshold as one standard deviation (σ) above the baseline mean firing rate (μ). In Fig. 3a, we marked the facilitation threshold and facilitation trials in the responses of the two example neurons shown in Fig. 2d and h. Note that both neurons exhibited consecutive long-lasting facilitation trials (Fig. 3a). We computed the proportion of facilitation trials in each of the 725 sessions (Fig. 3b and Supplementary Fig. 2).

We also identified facilitation trials using the thresholds of 0.5, 1.5, or 2 standard deviations (σ) above the mean (μ) firing rate of the same

https://doi.org/10.1038/s41467-025-58034-8

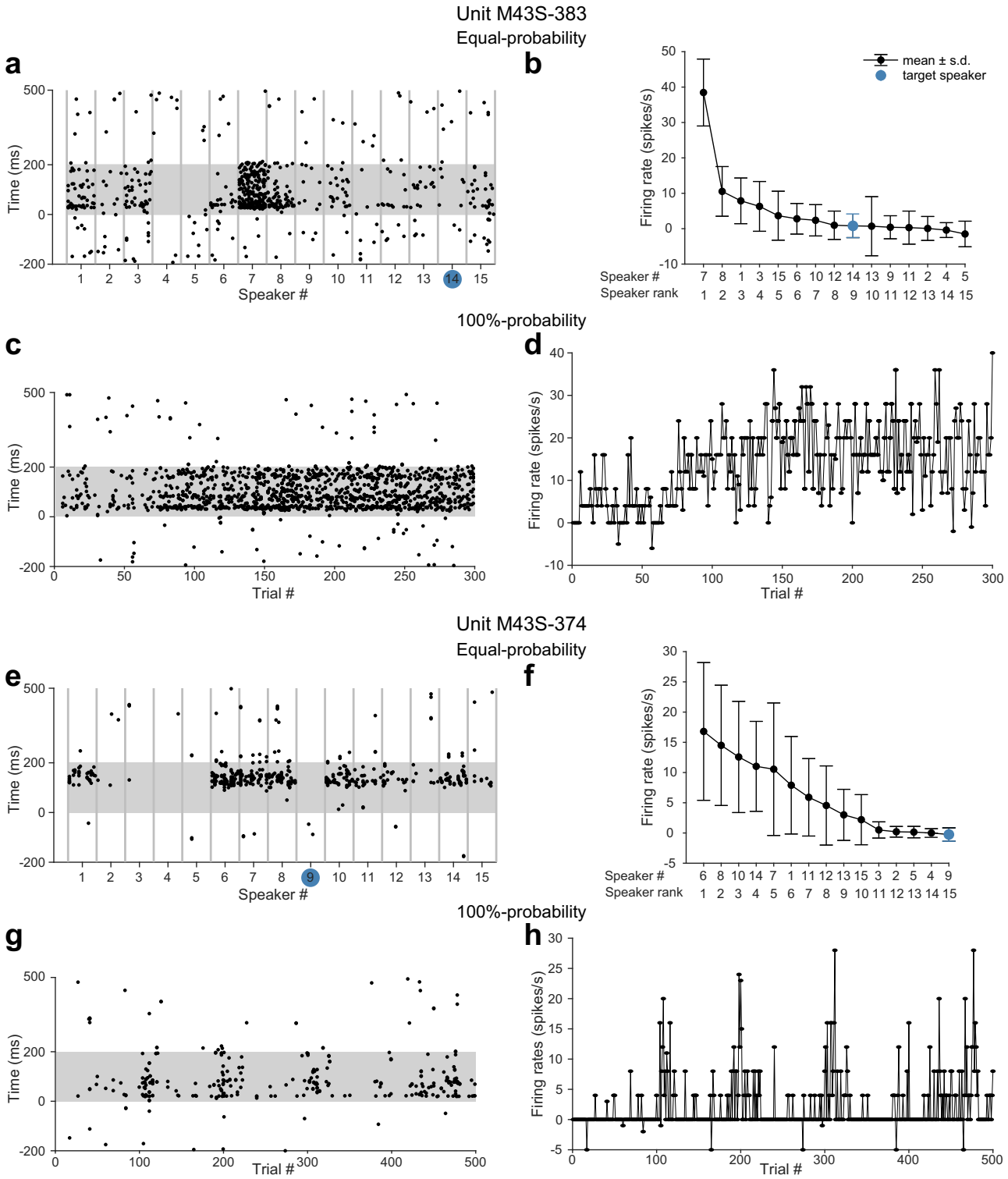

Fig. 2 | Repetitive sound stimulation evoked neural facilitation. a Spike raster plot of the previous example unit M43S-383 at fifteen speaker locations under equal-probability presentation mode. Stimuli at each speaker location (i.e., dots in different X-axis positions but within two vertical gray bars) were randomly presented ten times. **b** Firing rate versus speaker number (top row) and speaker rank number (bottom row). The blue dot indicates the target speaker (i.e., the same stimuli were delivered from the same speaker location under two presentation modes) (speaker #14, ranked 9th). Stimuli at each location were presented ten times. **c** Spike raster tested at speaker #14 ranked 9th under the 100%-probability presentation mode. **d** Trial-by-trial firing rate. **e**–**h** Similar to (**a**–**d**), but for a different unit M43S-374. All stimuli are wideband noise.

speaker location in the equal-probability mode (Supplementary Fig. 2). We computed the proportion of facilitation trials of each session for each firing rate threshold. There were more sessions in the 100% than in the equal-probability mode when the proportion of facilitation trials was larger than 25% (Fig. 3b, threshold of $\mu + 1\sigma$), and 40%, 15% and 10% (Supplementary Fig. 2, thresholds of $\mu + 0.5\sigma$, $\mu + 1.5\sigma$, and $\mu + 2\sigma$). Furthermore, the summed proportion of sessions in the 100%-probability mode when the proportion of facilitation trials were larger than

https://doi.org/10.1038/s41467-025-58034-8

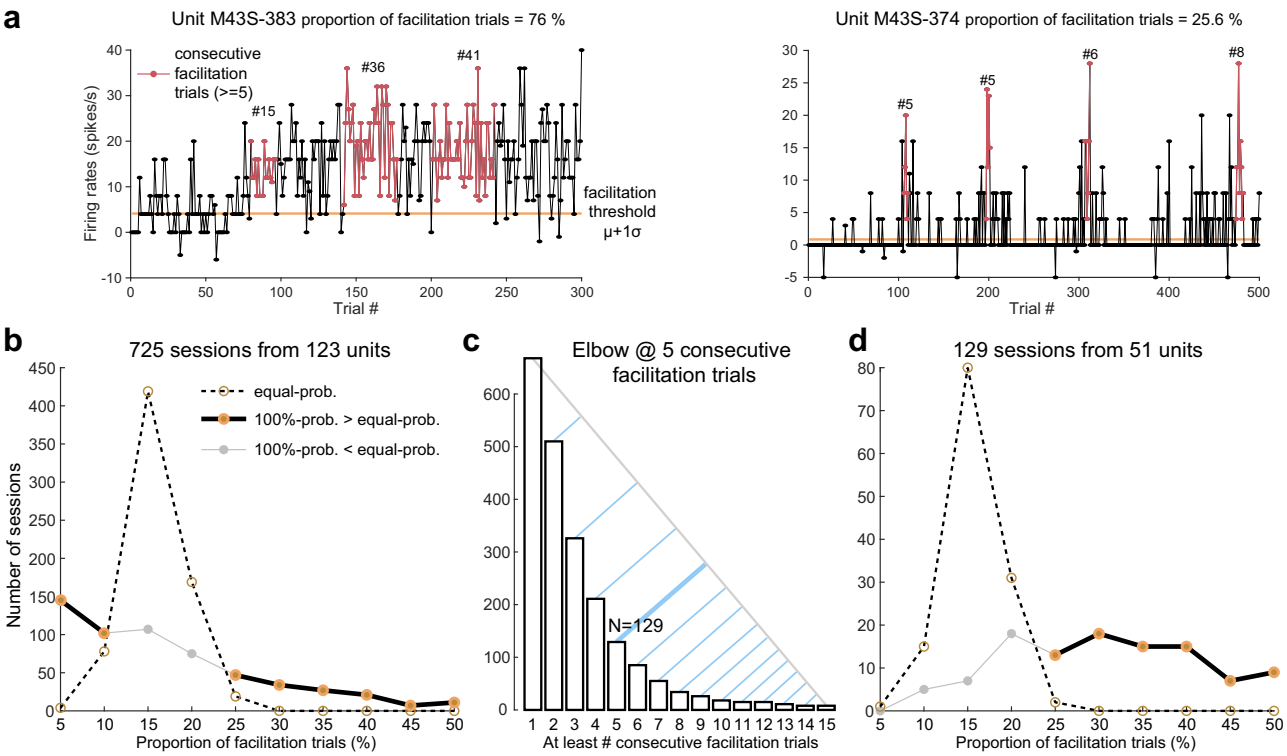

Fig. 3 | Quantification of facilitation trials evoked by repetitive stimulation. **a** Trial-by-trial firing rate of two example units in 100%-probability presentation mode (same as Figs. 2d and h). The orange bar indicates the facilitation threshold (mean + one standard deviation). Red dots and lines indicate the long-lasting facilitation (i.e., at least five consecutive trials with firing rates exceeding the facilitation threshold). **b** Number of sessions that have the proportion of facilitation trials (round to nearest 5%) at the threshold of one standard deviation for the 100%- (solid lines) and equal- (dashed lines) probability presentation modes. Stimuli either come from a single target speaker (100%-probability) or all 15 speakers (equal-probability) in one session. For measuring the proportion of facilitation trials in the equal-probability mode, we generated firing rate distributions that have the same mean and standard deviations of the firing rate of the target speaker in equal-probability mode and have the same number of trials as the corresponding 100%-probability mode. Note that in the Gaussian distribution, there are 16% of points are above one standard deviation. Consistent with this, over 400 sessions in the equal-probability mode have 15% of facilitation trials. Values at 0% and above 50% in the *X*-axis are truncated for display purposes. Supplementary Fig. 2 displays similar plots but have four thresholds with full-range *X*-axis. **c** Histogram of the number of sessions (*Y*-axis, linear) versus minimum length of consecutive facilitation trials (*X*-axis) at the threshold of one standard deviation. The gray line connects two bars at #1 and #15 trials. Blue lines measure the perpendicular distance from the peak of the thirteen bars (#2 to #14) to the gray line. The thick blue line at the #5 trial has the longest distance. **d** Similar to (**b**) but only includes 129 sessions that have at least five consecutive facilitation trials.

40, 25, 15, and 10% was similar among four thresholds (24, 27, 25, and 23%). Thus, the threshold affects the proportion of facilitation trials in each session but not the proportion of sessions that display more facilitation in the 100%-than-equal-probability mode. We focused on the threshold of μ + 1σ in the further analyses.

Consecutive facilitation trials were observed in a number of sessions tested by the 100%-probability mode. The number of sessions (*y*-axis) at a given number of minimum consecutive facilitation trials (*x*-axis) is shown at the thresholds of μ + 1σ (Fig. 3c) and μ + 0.5σ, μ + 1.5σ and μ + 2σ (Supplementary Fig. 3a), respectively. To examine whether the consecutive facilitation trials evoked by the repetitive stimuli resulted from random fluctuations in a neuron's firing, we shuffled the trial sequence and counted the number of consecutive trials where the firing rates surpassed the thresholds. We found that shuffling the firing rates markedly reduced the proportion of sessions that have a minimum number of consecutive facilitation trials (Supplementary Fig. 4). For further analyses (Supplementary Figs. 5, 6), we focused on the sessions that exhibited at least five consecutive facilitation trials (129 sessions from 51 neurons). We chose five based on the "elbow" in Fig. 3c, where the number of sessions (*y*-axis) shows a smaller decrease after #5 when we further increased the number of consecutive facilitation trials (*x*-axis). Population statistics of the length of the consecutive facilitation trials are shown in Supplementary Fig. 3b which shows that consecutive facilitation trials could persist as long as 45 trials. In the majority of sessions, 200 trials were tested (Supplementary Fig. 3c). In most sessions, it took fewer than 100 trials to achieve the first neural facilitation that lasting at least five consecutive trials (median: 44 trials) (Supplementary Fig. 3d).

There were more sessions in the 100%- than in the equal-probability mode when the proportion of facilitation trials was 10, 5, and 0% (Fig. 3b, thick line on the left and Supplementary Fig. 2). This was consistent with previous studies which showed that auditory neurons mainly exhibit suppression (adaptation) to repetitive sound stimulation[2,35]. However, for the 129 sessions with at least five consecutive facilitation trials (Fig. 3c), there were more sessions in the 100%-probability mode than in the equal-probability mode only when the proportion of facilitation trials was larger than 25% (Fig. 3d, 99 out of 129 sessions, thick line). Notice that our threshold, a single occurrence of five consecutive facilitation trials (65 out of 129 sessions), did not guarantee a higher proportion of facilitation trials because it only contributed to 2.5% (5 out of 200 median trials) of facilitation trials in one session. Therefore, long-lasting neural facilitation defined here generally excluded sessions with repetition suppression.

We also tested the effects of different inter-stimulus intervals (ISI). Neural facilitation was observed at both short ISI (e.g., Fig. 2, 700 ms) and long ISI (e.g., Supplementary Fig. 6g, 1500 ms). In addition to ISI with a fixed length, we also tested random ISIs in a subset of sessions. An example is shown in Supplementary Fig. 5a. This neuron not only showed facilitation at three constant ISIs but also at random ISIs. Across all 129 test sessions, three groups of ISIs were tested: short

(500 ms and 700 ms, 78 sessions), long (>1000 ms, 36 sessions), and random (700 ms to 2200 ms, 15 sessions). The proportion of facilitation trials was similar among the three ISI groups (Supplementary Fig. 5b; 35%, 41%, and 34%, $p = 0.1244$, one-way ANOVA).

We conducted control tests to see if the observed facilitation depended on visual inputs. We found that the facilitation was not limited to speakers located within a marmoset's visual field (< 90 degrees)[37] and could be induced from speaker locations both in front and behind an animal (Supplementary Fig. 6a–c). Across all 129 test sessions, the proportions of facilitation trials were similar between the front (64 sessions) and back (65 sessions) speaker locations (Supplementary Fig. 6d; 38% vs. 35%). The facilitation could still be observed when an animal was tested in the darkness (7 sessions) (Supplementary Fig. 6e–g). These results suggest that visual inputs are not required to induce the facilitation. Interestingly, compared to "simple" stimuli (wideband noises), a slightly larger proportion of facilitation trials was observed with "complex" stimuli (amplitude-modulated wideband noises or marmoset vocalizations) (Supplementary Fig. 6h; 30% vs. 41%, $p = 0.0054$, rank-sum test).

To investigate whether neural facilitation altered a neuron's SRF, we compared the SRF measured before and after repetitive stimuli. An example neuron is shown in Supplementary Fig. 7a (same neuron as in Fig. 1c). After presenting 300 trials of wideband noise at the same location repetitively, the SRF (bottom) appeared similar to the SRF measured before (top). More examples are shown in Supplementary Fig. 7b. To quantitatively characterize SRF changes, we calculated tuning and direction selectivity, which were similar after repetitive sound stimuli (Supplementary Fig. 7c). Furthermore, we computed the correlation between each pair of responses at fifteen locations (Supplementary Fig. 7d). Seventy-five percent (197/262) of the paired sessions had a correlation coefficient greater than 0.7. These analyses show that repetitive stimuli from the same location does not alter the SRF of the neuron being tested.

Neural facilitation depends on sound locations

We observed both neural facilitation and suppression (Supplementary Fig. 8a) at different speaker locations in the 100%-probability mode. Suppression threshold was defined as one standard deviation below the baseline firing rate (Fig. 4a, blue bar). To investigate the contextual effect caused by speaker locations, we measured the facilitation and suppression trials in all 725 sessions in the 100%-probability mode (Supplementary Table 1). As mentioned earlier, we assigned each speaker a rank number based on its baseline firing rate obtained under the equal-probability mode. Speaker ranked 1st exhibited the highest firing rate among all tested speakers and was located at or near the center of a neuron's SRF. The lowest-ranked speaker usually fell far outside of the SRF and evoked a response not significantly different from the spontaneous activity. The speaker rank of example neurons were shown in Figs. 2b, f, and 4a. We found that speakers with lower ranks usually elicited more facilitation trials than speakers with higher ranks when tested under the 100%-probability presentation mode, as shown by the example neuron in Fig. 4b, orange dots. Note that in this neuron, no facilitation was induced by the frontal speaker (speaker #1, rank 5th), and only one facilitation trial (out of 200 trials) was induced by the speaker at the contralateral 90° location (speaker #6, rank 4th). These two locations were commonly used in previous SSA studies (see Discussion). When the test speaker was ranked 15th, the average proportion of facilitation trials was 21.6%. In comparison, when the test speaker was ranked 1st, this statistic dropped to 7.3% (Fig. 4c, orange dots). At higher-ranked speakers, more suppression trials were observed than at lower-ranked speakers as shown by an example neuron in Fig. 4b, blue dots. Note that 70% of trials (140/200) exhibited suppression at the frontal speaker (speaker #1, rank 5th), and 52.5% of trials (105/200) showed suppression at the contralateral 90° location (speaker #6, rank 4th). The average proportion of suppression trials across all tested sessions was 26.1% when the test speaker was ranked 1st, whereas that number dropped to 8.9% when the test speaker was ranked 15th (Fig. 4c, blue dots). This trend was the opposite of facilitation (Fig. 4c, orange dots).

We also analyzed the proportion of facilitation and suppression trials using thresholds of 0.5, 1.5, and 2 standard deviations (σ) above and below the mean (μ) firing rate of the same speaker under equal-probability mode (Supplementary Fig. 8a), respectively. In this example session, no facilitation trials passed the $\mu + 1.5\sigma$ or $\mu + 2\sigma$ facilitation thresholds. However, 23.5% and 8.5% of trials passed the μ-1.5σ and μ-2σ suppression thresholds, respectively. Across the 15 speaker ranks and three facilitation or suppression thresholds, we observed a gradual increase in the proportion of facilitation trials from the 1st to the 15th ranked speaker (Supplementary Fig. 8b, orange dots). In contrast, suppression trials showed an opposite trend (blue dots). This analysis suggests that location preference systematically affects the sign and amplitude of contextual effects across thresholds.

Together, we found that the modulation effect of repetitive stimuli from a speaker location was symmetrical in amplitude around the 8th speaker rank and changed sign from negative (suppression) for the first 7 ranks to positive (facilitation) for the remaining 7 ranks.

Location preferences modulated the overall firing rate to repetitive stimuli from a specific location

Our previous analysis focused on the trial-based firing rate changes. Next, we would like to know whether the overall firing rate among all trials within a single session is modulated by location preferences. We did not consider the facilitation thresholds and consecutive facilitation trials here since the trial-based responses were averaged. We compared the total firing rate (i.e., without subtracting the spontaneous firing rate from the total firing rate) and firing rate under two presentation modes for the preferred and nonpreferred locations. The three highest-ranked speaker locations (1st, 2nd, and 3rd) were considered as preferred locations, and the three lowest-ranked speaker locations (13th, 14th, and 15th) were considered as nonpreferred locations.

Supplementary Fig. 8c shows the total firing rate distribution of one example neuron under equal- (left) and 100%- (right) probability modes, respectively. When the probability switched from 1/15 to 100%, we observed that the nonpreferred locations (rank 15th, orange bars) evoked a higher total firing rate (rightwards arrow), whereas the preferred locations (rank 3rd, blue bars) evoked a lower total firing rate (leftwards arrow). Similar contextual effects were observed when measured using the firing rate in Supplementary Fig. 8d, with the firing rate distribution having more overlap under the 100%-probability mode. Supplementary Fig. 4d shows that it took a median of 44 trials to achieve the first neural facilitation lasting at least five consecutive trials. Here, we would like to know how long it took to achieve the minimum and maximum neural responses when tested with preferred and nonpreferred locations, respectively. Unexpectedly, we found that neural suppression and facilitation have almost the same time scales (Fig. 4d).

Figure 4e (top) shows that the total firing rates of preferred and nonpreferred locations under the 100%-probability mode (y-axis) were lower (below the diagonal line) and higher (above the diagonal line), respectively, than the total firing rate under equal-probability mode (*x*-axis). Figure 4f (top) shows the number of sessions and neurons with higher total firing rates under 100%-(orange) or equal- (blue) probability modes. The preferred locations significantly suppressed the total firing rate of 90% of sessions and 86% of neurons (black bars). Conversely, the nonpreferred locations significantly facilitated the total firing rate of 89% of sessions and 88% of neurons. To investigate whether the changes in total firing rate are neuron-specific or location-specific, we analyzed 34 neurons that were tested with both preferred and nonpreferred locations (two bottom bars in each panel). Fourteen

https://doi.org/10.1038/s41467-025-58034-8

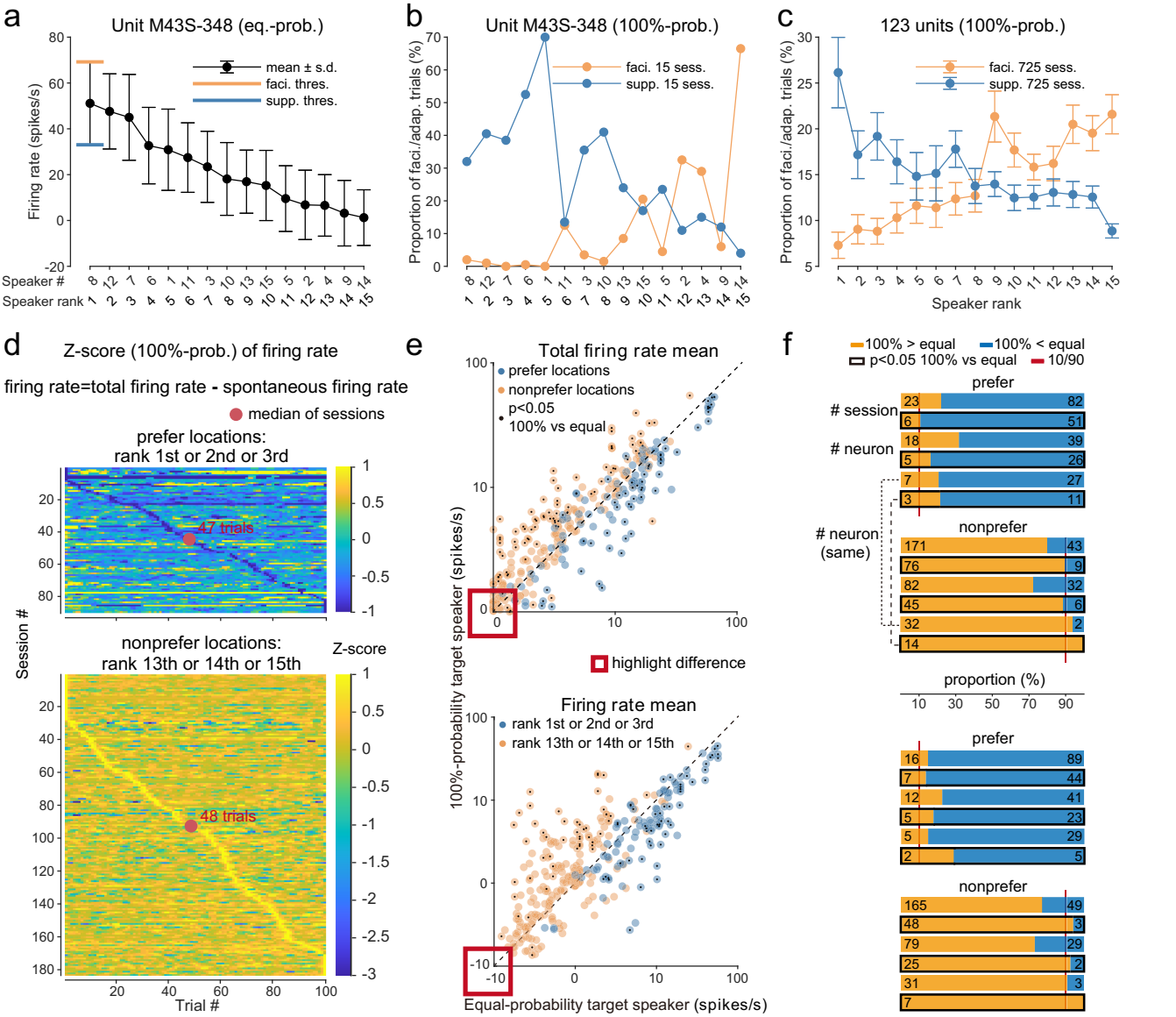

Fig. 4 | Neural facilitation depends on sound locations. a Each speaker was assigned a rank number based on its baseline firing rate obtained under the equal-probability presentation mode. Speaker ranked 1st had the highest firing rate. Orange and blue bars indicate facilitation and suppression thresholds, respectively. Dots and error bars indicate mean ± standard deviation. Stimuli at each location were presented ten times. **b** Proportion of facilitation (orange dots and lines) and suppression (blue dots and lines) trials under 100%-probability presentation mode at different speaker ranks for the same example unit. Wideband noise stimuli at each speaker location were tested 200 times. **c** Proportion of facilitation (orange dots and lines) and suppression (blue dots and lines) trials of the population data. All 725 sessions from 123 neurons were shown, regardless of the length of consecutive facilitation trials. Error bars indicate the standard error of the mean. **d** Z-score of the firing rate is sorted based on the normalized minimum (top) and maximum (bottom) value of each session. **e** Scatter plot of the total firing rates mean (top) and firing rates mean (down) obtained from 100%- (*y*-axis) vs from equal- (*x*-axis) probability presentation mode at the same target speaker. Red squares highlight the difference of the two neural responses. Blue and orange dots indicate sessions belong to high (1st, 2nd, and 3rd) and low (13th, 14th, and 15th) rank speakers, respectively. Black dots indicate sessions with significant (uncorrected) differences. Supplementary Fig. 8e shows the results after the Bonferroni correction. Two-sided *t* test. **f** Number of sessions and neurons that have higher total firing rate (top) and firing rate (bottom) under the 100%- (orange) or equal- (blue) probability modes. The significantly different (uncorrected, Supplementary Fig. 8f shows the results after Bonferroni correction) session or neuron is encircled by a black square. Prefer: responses to preferred sound stimuli at 1st, 2nd, or 3rd ranked speakers. Nonprefer: responses to nonpreferred sound stimuli at 13th, 14th, or 15th ranked speakers. 34 neurons are tested with both preferred and nonpreferred stimuli, and 14 neurons are tuned to both preferred and nonpreferred stimuli. Two-sided *t* test.

out of 34 were significantly modulated by both locations, and all of them showed increased total firing rates in response to nonpreferred locations. Among these 14 neurons, 11 (79%) also showed decreased total firing rates to preferred locations, indicating that the modulation of total firing rates was location-specific.

We observed a similar contextual effect when measured using firing rate (Fig. 4e, bottom). One advantage of using the firing rate is that we could observe the inhibitory effect, i.e., the total firing rate below the spontaneous firing rate. The preferred locations significantly suppressed the firing rate of 86% of sessions and 82% of neurons (Fig. 4f, bottom), while the nonpreferred locations significantly facilitated the firing rate of 94% of sessions and 93% of neurons. Seven out of 34 were significantly modulated by both locations, and all of them showed increased firing rates in response to nonpreferred locations. Among these 7 neurons, 5 (71%) also showed decreased firing rates to preferred locations.

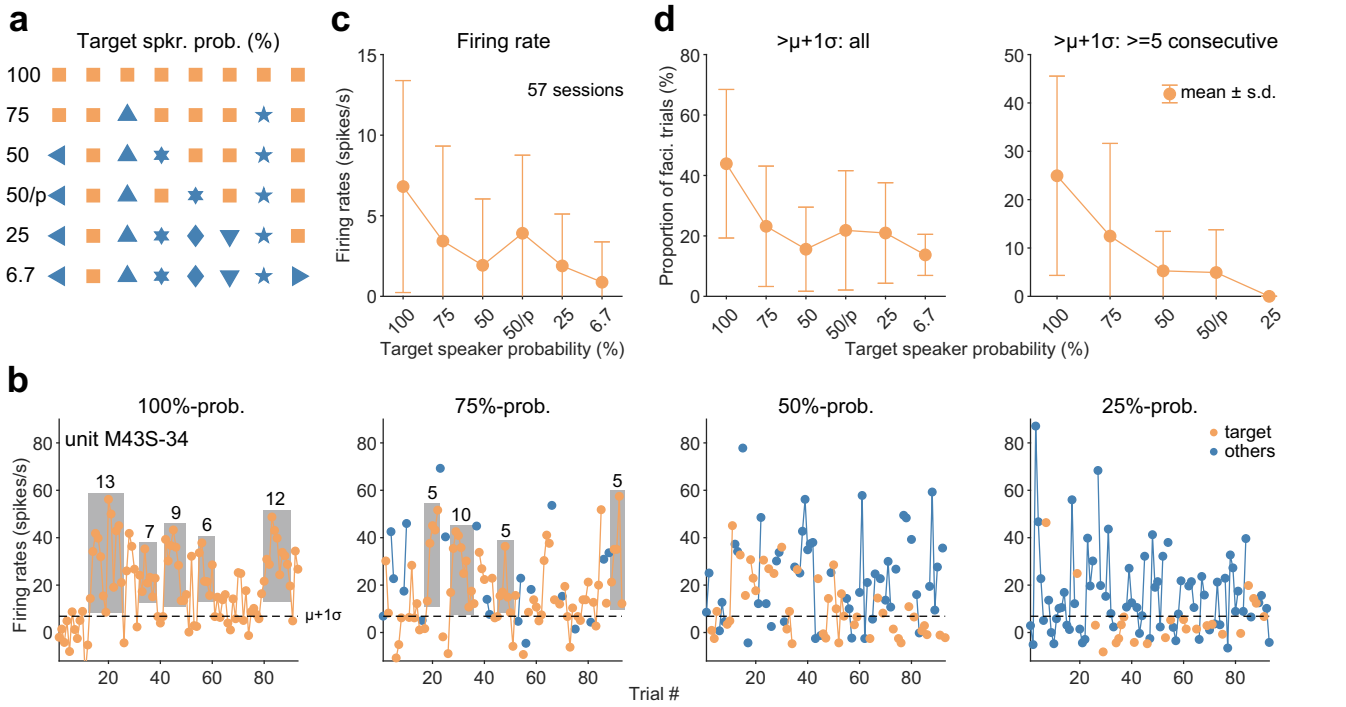

Fig. 5 | Dependence of neural facilitation on location continuity. a Six stimulus presentation modes were used in our studies. 75, 50, and 25% probabilistic presentation modes play sounds from all fifteen speakers in a randomly shuffled order while giving the target speaker (orange square) a presentation probability higher than other speakers (blue left-pointing triangle, upward-pointing triangle, hexagram, diamond, downward-pointing triangle, pentagram, right-pointing triangle). For the 50%-probability periodic presentation mode (50/p%), the target speaker was interleaved with other speakers. Therefore, the sequence of the target speaker was periodic instead of random. Equal-probability presentation mode equals to 6.7%-probability presentation mode. **b** Firing rate of target speaker (orange dots) and other speakers (blue dots) of example unit under 100% (left), 75% (middle-left), 50% (middle-right), and 25% (right) probability presentation modes. The Black dashed line indicates the μ + 1σ facilitation threshold. Gray bars indicate the long-lasting (>=5 consecutive) facilitation trials. **c** Firing rate of target speaker for six presentation modes. Dots and error bars indicate mean ± standard deviation. **d** Proportion of all facilitation trials (left) and trials that lasting at least five consecutive times (right) above the facilitation threshold at the target speaker. There are 57 sessions in total.

In summary, we found that preferred locations decreased neural activities in 84% (total firing rate: 86%, firing rate: 82%) of neurons, while nonpreferred locations increased neural activities in 91% (88%, 93%) of neurons. Furthermore, the modulation effect of repetitive stimuli was locations-specific, with 75% (79%, 71%) of neurons exhibiting increased activity to nonpreferred locations and also decreased activity to preferred locations.

Dependence of neural facilitation on location continuity

As we have shown above, neural facilitation can be induced in the 100%-probability presentation mode in which the probability of stimuli delivered from a target speaker is 100% (none from other speakers). In the equal-probability presentation mode, the presentation probability for each speaker is equal to 1/ 15 (number of speakers). Therefore, the 100%-probability presentation mode provides the location continuity for sound delivery, whereas the equal-probability presentation mode does not. We further investigated in a subset of units whether the sound location continuity was necessary to induce neural facilitation by changing the presentation probability of the target speaker from 100% to 75%, 50%, 25%, and 6.7% (Fig. 5a, target speaker: orange square; other speakers: blue color shapes). An example neuron is shown in Fig. 5b. The decrease of the target speaker's presentation probability to 75%, 50%, and 25% resulted in increasingly weaker responses to the target speaker (100%: 20, 75%: 20, 50%: 10, and 25%: 6 spikes/sec). Across all 57 sessions in 12 neurons, the decrease of the presentation probability of the target speaker from 100% to 75%, 50%, 25% and 6.7% resulted in weaker response of the target speaker (6.81, 3.43, 1.93, 1.89 and 0.88 spikes/sec, Fig. 5c) and the reduction of the proportion of facilitation trials (Fig. 5d, left, 44%, 23%, 16%, 21% and 14%). The proportion of facilitation trials lasting for at least 5 consecutive times showed a similar trend (Fig. 5d, right, 25%, 12%, 5%, and 0%).

Previous stimulus-specific adaptation (SSA) related studies found that neurons in the auditory cortex of anesthetized rats were sensitive to statistical regularities: standard and deviant tones in random sequences both evoked larger responses than the same tones in periodic sequences[23,38]. To investigate whether neural facilitation is also sensitive to the statical regularity of the target speaker, we played two different types of sequences with 50%-probability of the target speaker. In one sequence, stimuli from the target speaker and other speakers were randomly arranged, similar to the 75, 25, and 6.7%-probability modes (Fig. 5a, 50). In the other sequence, stimuli from the target speaker were interleaved with stimuli from other speakers, so that the stimuli from the target speaker were periodical (Fig. 5a, 50/p). In contrast to SSA, we found that the periodic target speaker sequence evoked relatively stronger, but not significant (Supplementary Table 2), than the random target speaker sequence (Fig. 5c, 3.91 vs. 1.93 spikes/sec). The proportions of facilitation trials were similar between random and periodic target speaker sequences with 50%-probability (Fig. 5d).

Together, our findings indicate that target speaker continuity was necessary to induce spiking facilitation, and the magnitude of the facilitation was roughly proportional to the probability of the target speaker.

Repetitive stimuli from nonpreferred sound locations induced sustained and reliable membrane potential depolarization

We next asked what are the cellular mechanisms underlying repetition facilitation and suppression, and why facilitation is long-lasting while

https://doi.org/10.1038/s41467-025-58034-8

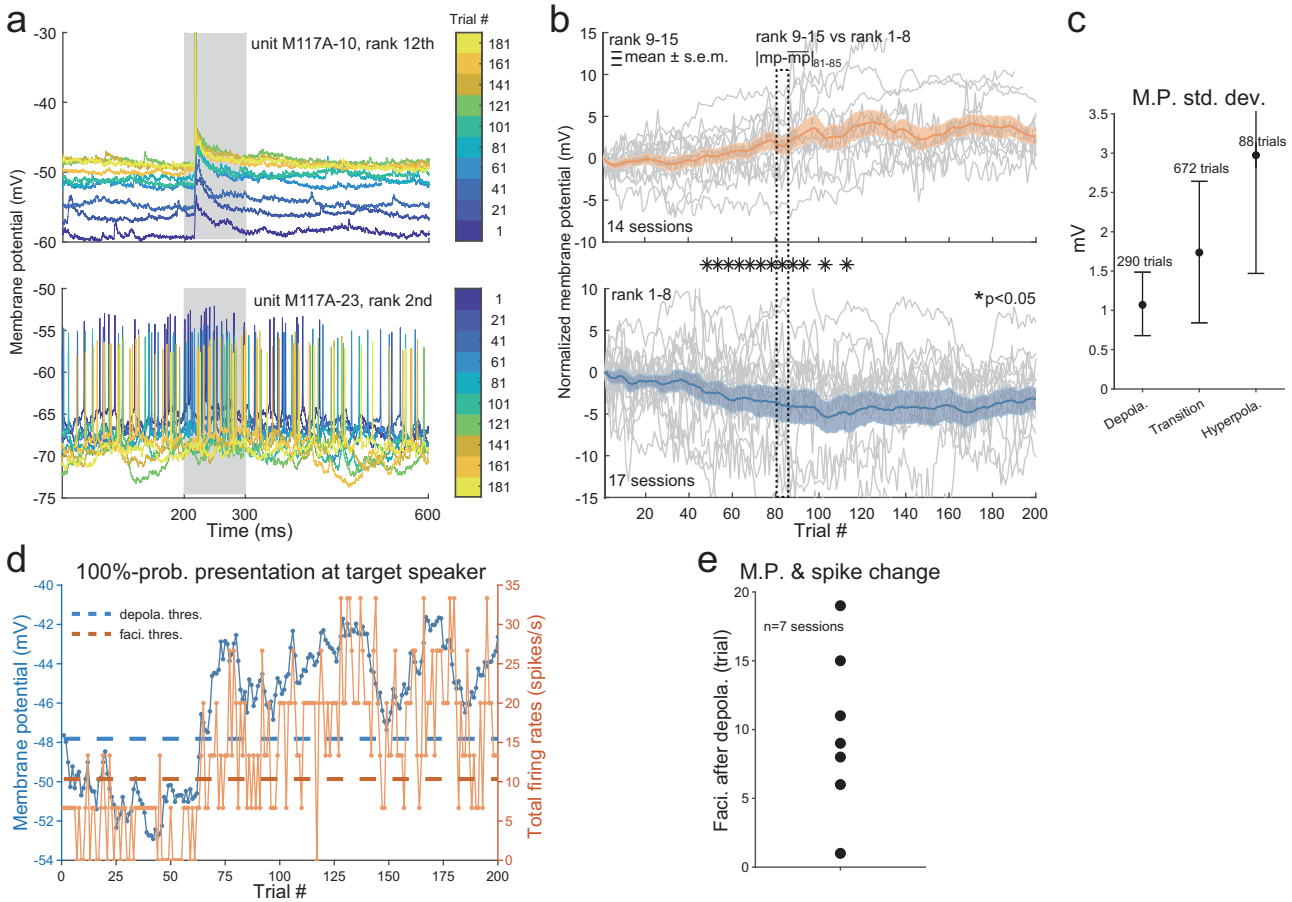

Fig. 6 | Repetitive sound stimulation induced sustained membrane potential depolarization and was less variable than hyperpolarization. a Top, membrane potential traces of unit M117A-10 recorded under the 100%-probability presentation mode for the 12th ranked session. Each color line indicates one trial from the start (blue) to the end (yellow). The vertical gray bar indicates sound stimulus from 200 ms to 300 ms. Bottom, membrane potential traces of unit M117A−23 recorded under the 100%-probability presentation mode for the 2nd ranked session. Notes the color bar direction is reversed. All stimuli are wideband noise. **b** Top, raw (thin gray), and averaged (thick orange) membrane potential across 200 trials for 14 lower-ranked sessions (9th to 15th). Membrane potential is normalized by subtracting its value at the first trial. Membrane potential in each trial is the spike-removed, averaged membrane potential before sound stimulus (1 ms to 200 ms). The error bar indicates the standard error of the mean. Note that some sessions have less than 200 trials. Middle, the absolute values of mean subtracted membrane potential across all sessions within 5 trials were compared, and the significant (Bonferroni correction, multiply the *p*-values by 40) difference between high and low-ranked sessions is indicated with stars. Bottom, raw (thin gray), and averaged (thick blue) membrane potential traces for 17 higher-ranked sessions (1st to 7th). **c** The standard deviation of membrane potential for depolarization, transition, and hyperpolarization trials across all sessions. The measure at the center of the error bars represents the mean. **d** Membrane potential (blue line and dots) and total firing rate (i.e., not subtracting the spontaneous firing rate, orange line, and dots) changes during the 100%-probability mode. Threshold of facilitation (dashed orange line) and depolarization (dashed blue line) were calculated in equal-probability mode. **e** Number of trials required to achieve the first membrane potential depolarization and spiking facilitation, with both phases lasting for at least five consecutive trials.

suppression is not (Supplementary Fig. 3). To address these questions, we performed intracellular recordings in awake marmoset[39,40] in a subset of experiments, measuring both membrane potential and spiking activity during both equal- and 100%-probability modes. In Fig. 6a, the top panel shows the membrane potential traces of an example neuron at the 12th-ranked speaker during the 100%-probability mode. The membrane potential showed a steady increase from − 60 mV and stabilized around − 50 mV. The increase in total firing rate followed a similar trajectory (Supplementary Fig. 9). The bottom panel shows the membrane potential traces at 2nd ranked speaker, which decreased from − 65 mV and stabilized around − 7 0 mV. Furthermore, membrane potentials at the high-ranked speaker were more variable than the low-ranked speaker. Notice that sustained depolarization of membrane potential is not reflective of changes in neuronal health during these long recording times. Supplementary Fig. 10 shows that neurons remained healthy in the next session after we presented repetitive nonpreferred stimuli. Furthermore, most neurons even show a hyperpolarized membrane potential after a break period of seconds to minutes.

We recorded intracellularly from 31 sessions in 14 neurons under the 100%-probability mode. We divided the target speakers into a low-ranked group (9th to 15th, Fig. 6b, top) and a high-ranked group (1st to 8th, Fig. 6b, bottom) based on the firing rate obtained under the equal-probability mode. The low-ranked group exhibited a gradual depolarization and subsequent stabilization of the resting membrane potential under the 100%-probability mode, while the high-ranked group showed the opposite trend. Moreover, in addition to the sign of membrane potentials, the variance of membrane potentials also differed between the two groups. The hyperpolarized membrane potentials in the high-rank group exhibited larger variability than the depolarized membrane potentials in the low-rank group, as indicated by the width of the error bars in Fig. 6b. We further calculated the membrane potential standard deviation of each trial. Trials were classified into depolarization, hyperpolarization, or transition groups

based on whether they passed the depolarization threshold (one standard deviation above the baseline membrane potential), hyperpolarization threshold (one standard deviation below), or neither. Depolarization trials had the lowest standard deviation, and hyperpolarization trials had the highest standard deviation (Fig. 6c, 1.07 vs. 2.97, $p < 0.0001$, one-way ANOVA).

Our previous analyses of the spiking activity measured the facilitation trial to characterize trials with firing rates exceeding the facilitation threshold (one standard deviation above the baseline firing rate). Here, we measured the depolarization trial to characterize trials with membrane potential exceeding the depolarization threshold. The membrane potential steadily increased after 40 trials during the 100%-probability mode (Fig. 6d, blue line), which was accompanied by an increase in spiking activity (Fig. 6d, orange line). For each session, we compared the difference between the number of trials needed to achieve the first long-lasting (≥ 5) membrane potential depolarization and spiking facilitation. We found that the depolarization of membrane potential always preceded the facilitation of spiking activity (Fig. 6e, median: 9 trials).

Our findings suggest that sustained membrane potential depolarization and hyperpolarization may underlie the cellular mechanisms of repetition facilitation and suppression, respectively.

A model neuron with increased arousal, but not decreased inhibition, explained intra- and extracellular recording data

We sought to understand the neural mechanisms underlying sustained membrane potential depolarization. To address this question, we added a new Hodgkin-Huxley (HH) model to explain the sustained depolarized membrane potential under 100%-probability presentation mode (Fig. 7a). Recent studies have found that higher arousal levels increase astrocyte activity[41,42] and elevate the concentration of intracellular chloride (Cl^-) in astrocytes at the onset of heightened arousal. Subsequently, this reduces the concentration of Cl^- in the extracellular space[43]. Our hypothesis is that this reduced extracellular Cl^- facilitates the efflux of Cl^- from inside the neuron. With this hypothesis ("increase arousal level"), our HH model successfully reproduced the sustained membrane potential depolarization observed in our intracellular recordings.

To explain the neural facilitation observed in our extracellular recordings, we consider two hypotheses (Fig. 7b). One hypothesis, "increase arousal level," has already explained sustained depolarization at the low-rank speaker. The other hypothesis is "release from inhibition". In this hypothesis, neurons that prefer the location (1) show strong facilitation, (2) reduce the surround suppression onto other neurons, and (3) unmask the weak responses in neurons that do not prefer the location. Fig. 7c (left) reproduced neural facilitation under 100%-probability mode based on the second hypothesis. However, after exploring the hyperparameters of the model (Fig. 7c, right), we found that the "release from inhibition" hypothesis only works within a limited range of parameter space. Importantly, this hypothesis requires that excitatory synapses have smaller adaptations than inhibitory synapses when stimulated with repetitive stimuli. However, previous whole-cell recordings found the opposite trend[44–46]. Therefore, a biologically realistic model would generate suppression instead of facilitation in response to 100%-probability stimuli.

In order to reproduce the speaker probability-dependent neural facilitation, we obtained key parameters for the model from our intracellular recordings and hypothesized that the amplitude of sustained membrane potential depolarization was proportional to the probability of the target speaker (Fig. 7d). Neural facilitation was simulated under the 100%-probability mode (Fig. 7e), where elevated firing rates were observed in tens of trials following the onset of repetitive stimuli and during epochs consisting of consecutive trials with firing rates above the facilitation threshold (Fig. 7e, gray bars). We discovered that reducing the probability of the target speaker led to a decrease in the proportion of facilitation trials in the model (Fig. 7f, gold dots). The performance of the model was in close agreement with the experimental results depicted in Fig. 5d (now displayed as Fig. 7f, green dots). We further compared the proportion of facilitation trials in our models with 129 experimentally tested sessions from 51 neurons that displayed at least five consecutive facilitation trials. The results were also similar at 100%- (Fig. 7f, gold vs red, leftmost, 41% vs 36%) and equal- (Fig. 7f, gold vs red, rightmost, 18% vs 14%) probability modes.

There are two mechanistic insights from our models. For intracellular recording data, our model suggests that changes in arousal level modulate the membrane potential. For extracellular recording data, our model excludes the "release from inhibition" hypothesis and supports the "state modulation" hypothesis. These two insights naturally link our extracellular and intracellular recording data to behavioral experiments.

Repetitive stimuli increase arousal level and shift attention

We conducted behavioral experiments in head-fixed and free-moving monkeys to confirm our model's "state modulation" hypothesis about repetitive stimuli. Numerous studies have shown that deviant or task-related sensory stimuli evoke much larger pupil size than standard or task-unrelated stimuli[47–49]. In addition, the difficulty and performance of sensory detection tasks correlate with pupil size[50,51]. Therefore, if this phenomenon is important for sensory detection, we should expect to observe larger pupil size under 100%-probability mode than under equal-probability mode.

In two head-fixed monkeys passively listening to two modes of sound stimuli, we monitored the eye and pupil using infrared or room light (Supplementary Fig. 11a–c and Supplementary Movie 1, 2). It is noteworthy that, unlike mice and other species, marmosets frequently blink and sometimes close their eyes in the head-fixed condition. To accurately track these movements, we used DeepLabCut[52], a marker-free, and deep learning-based motion tracking software. This tool allowed us to track the eyelids, pupil boundary, and corneal reflection (Fig. 8a and Supplementary Fig. 11). These three variables were then used to calculate the eye-opening area, pupil area, and saccade coordinates on the X-Y plane (Fig. 8b). Consistent with our expectations, larger eye-opening and pupil areas are observed under 100%- than equal-probability presentation mode (top and middle rows, Fig. 8c). We also observed fewer saccades under 100%-probability mode (bottom row, Fig. 8c).

Is neural facilitation important for behavior? One way to answer this question is to examine whether animals exhibit any behavioral changes to repetitive stimuli. Location-specific facilitation requires repetitive sound stimuli from one fixed location. This location could be considered to be fixed either relative to the head (egocentric) or to the world (allocentric) reference frame. Although one study[53] found some allocentric neurons in the ferret auditory cortex, which are tuned to the locations in the world instead of the head reference frame, the proportion is very low (6/92, 6.5%). Therefore, we could consider that the majority of the location continuity we referred to is egocentric. If the animals move their heads, then the SRF changes, and the neural facilitation will become weaker or even disappear (Fig. 5). If location-specific facilitation is important for behavior, we should expect to observe a relatively stable head orientation angle under 100%- than equal-probability presentation mode. Notice that it is not necessary to orient the head towards the repetitive sound stimuli.

In two free-moving monkeys that passively listened to two modes of sound stimuli, we monitored the head orientation angle using room light (Fig. 8d, e). We labeled five markers to track head orientation, and three of them have a distinct white fur (Fig. 8d and Supplementary Movie 3). The longest head-fixation times exceeding five seconds under the 100%-probability presentation mode are 26, 20, 25, and 35 s for the four example sessions. In contrast, under the equal-probability presentation mode, only the second and fourth sessions have stable

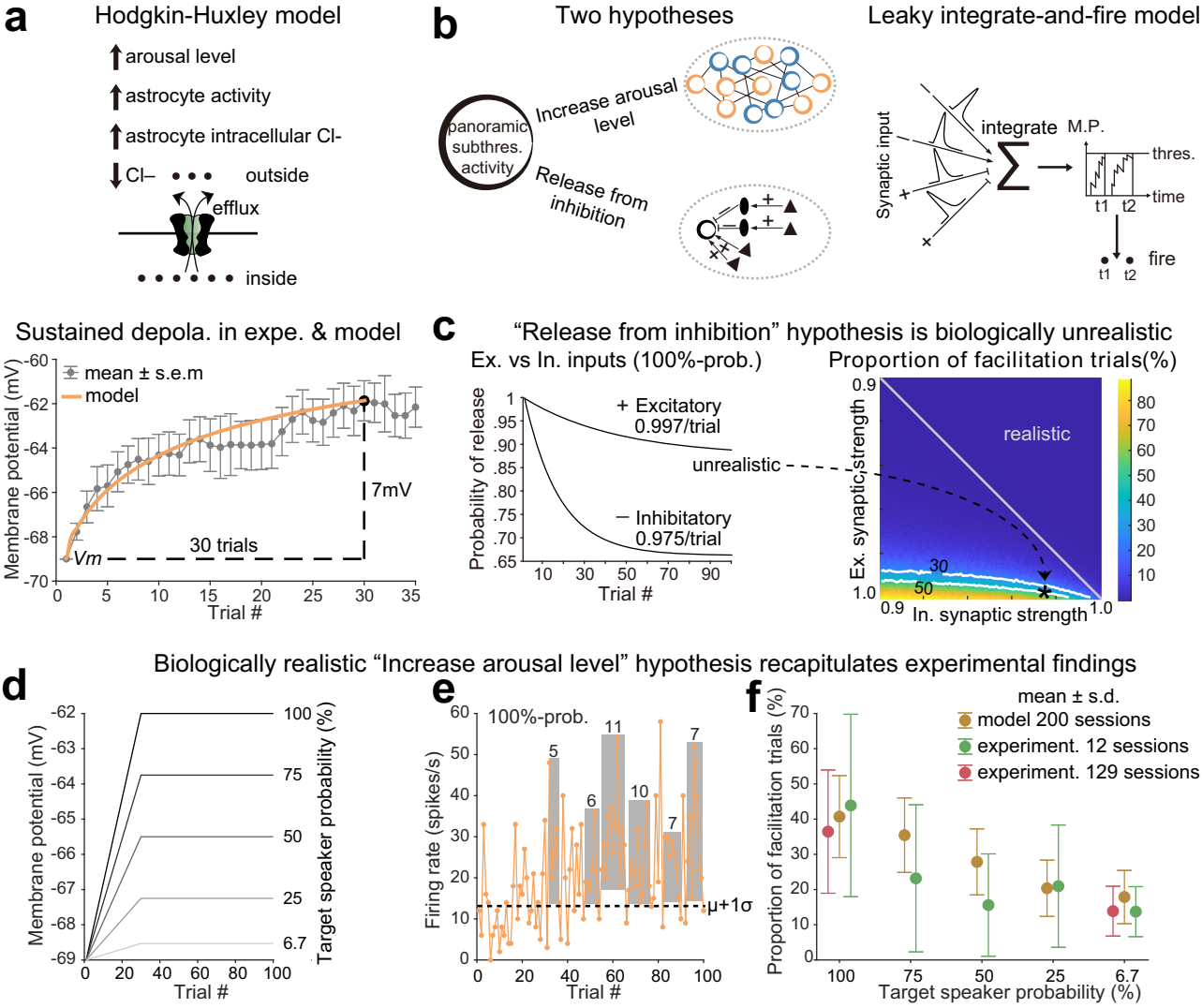

Fig. 7 | Increased arousal, but not decreased inhibition, in model neurons explained intra- and extracellular recording data. a Top: Mechanism where a higher arousal level leads to increased astrocyte activity, resulting in a lower concentration of Cl- outside the neuron. Consequently, the higher intracellular concentration of Cl- flows out through the leak channels. Bottom: Results from the Hodgkin-Huxley (HH) model, where increasing the leak conductance of Cl- leads to sustained depolarization of the membrane potential (orange curve). **b** Left: Under equal probability presentation mode, the model neuron exhibits panoramic subthreshold activity across all spatial locations. The width of the circles is proportional to the amplitude of subthreshold activity. Middle-Top: The neural activity of the network is homeostatically regulated. Middle -Bottom: Neural activity at the non-preferred location is inhibited. Right, the LIF neuron integrates multiple excitatory and inhibitory synaptic inputs and fires a spike (t1, t2) whenever the membrane potential passes the threshold. **c** Left: Difference in adaptation between excitatory and inhibitory synapses during each trial or stimulus. The excitatory synapse has a smaller amplitude of adaptation (0.003) compared to the inhibitory synapse (0.025). With 100%-probability stimuli, the probability of release for the two types of synapses will gradually diverge and stabilize over time. Right: Proportion of facilitation trials when varying the strengths of excitatory and inhibitory synapses. The diagonal line indicates equal adaptation strength for excitatory and inhibitory synapses. The upper triangular matrix represents scenarios where the excitatory synapse has a larger adaptation than the inhibitory synapse, which is biologically realistic. Two white contour lines indicate the 30% to 50% proportion of facilitation trials. **d** Membrane potential depolarization amplitude at five probability presentation modes is proportional to the target speaker probability. **e** Trial-by-trial firing rate of a LIF neuron under 100%-probability presentation mode. The thick black line indicates the facilitation threshold. Vertical gray bars indicate trials that belong to the long-lasting facilitation trials, and values atop the bars indicate the trial number. **f** Decreasing the presentation probability of the target speaker results in a smaller proportion of facilitation trials for the LIF neuron (brown dots). Green dots show the data from five probability presentation modes (same as Fig. 5d). Red dots show the data from 100%- and equal-probability presentation modes.

head-fixation longer than five seconds, and these durations (6 s and 10 s) are much shorter than those under the 100%-probability presentation mode (Fig. 8e and Supplementary Fig. 13a, b).

The head fixation angles do not align or correlate with the sound locations. Consistent with our expectations, animals displayed less head movement under 100%- than equal-probability presentation mode (Supplementary Fig. 13c). Furthermore, considering eye position also modulates SRF[54,55], the fewer saccades we observed (bottom row, Fig. 8c) further strengthens the view that SRF should be consistent, and location-specific facilitation should be present when animals pay attention to the repetitive stimuli.

Together, our modeling-inspired behavioral experiments demonstrate that repetitive stimuli are meaningful for the passive-listening monkeys.

Discussion

In this study, we investigated extra-, intracellular, and behavioral responses to repetitive sound stimulation in the auditory cortex of

https://doi.org/10.1038/s41467-025-58034-8

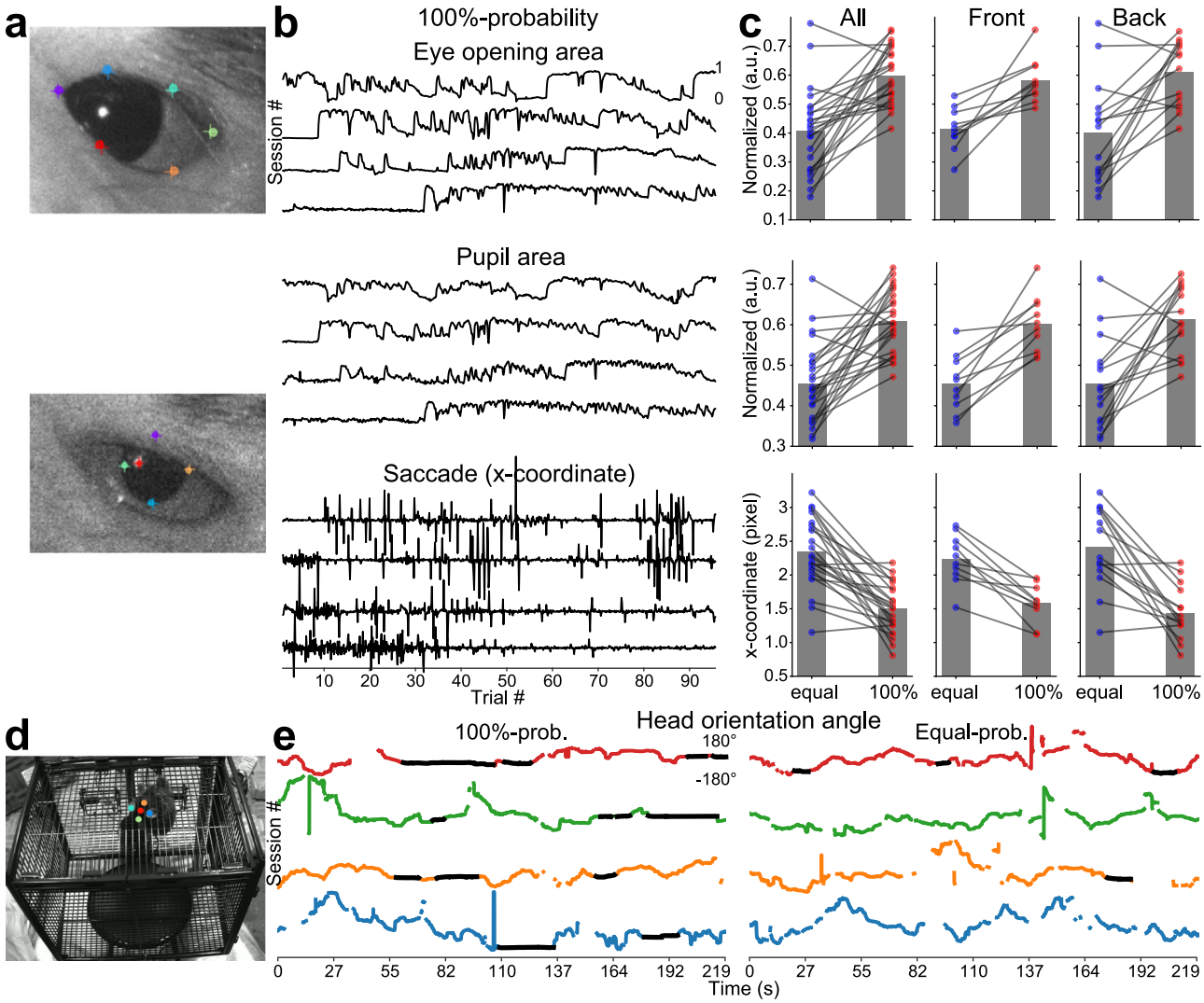

Fig. 8 | Repetitive stimuli increase arousal level and shift attention. a top, six labels are applied to track the eyelids: two on the upper eyelids, two on the lower eyelids, and two on the nasal and temporal edges. Bottom, four labels are applied to track the boundary of the pupil (up, down, nasal, and temporal), and one label (corneal reflection) is used to track saccades. Dots indicate manually labeled markers, and plus signs indicate model-predicted markers. Videos are provided in Supplementary Movie 1, 2. **b** normalized eye-opening area and pupil area (top and middle), and saccades on the *X*-axis (bottom) from four example sessions (#14, #15, #16, and #19). Supplementary Fig. 12 shows the results from all 25 sessions. **c** comparisons between the two presentation modes for all 25 sessions (left), 10 sessions with front target speakers (middle), and fifteen sessions with back target speakers (right). **d** Five labels are applied to track the head orientation. Two labels are white color ear puffs, one label is white color fur on the forehead, and two labels are black-colored fur on the middle and back of the head. Video is provided in Supplementary Movie 3. **e** Head orientation angle under the 100%-probability (left) and equal-probability (right) presentation modes. Four different colors indicate four different sessions. Thick black lines indicate any continuous head fixation of at least five seconds with less than 1% fluctuation in angle. Notice that the angle has an artificial change at −180 or 180 degrees.

awake marmoset monkeys. The major finding of this study is the observation of a novel location-specific facilitation (LSF) which is dependent on sound location and stimulus presentation mode. LSF is a different phenomenon than the well-studied stimulus-specific adaptation (SSA) and other stimulus paradigms. Computational models based on the increased arousal reproduce membrane potential depolarization and LSF. To the best of our knowledge, this is the first study to observe repetition facilitation in a location-specific manner in the majority of neurons, and behavioral changes due to repetitive stimuli. In addition, this is the first study that comprehensively explains repetition facilitation using intracellular recordings, computational models, and arousal tracking.

Comparison with previous studies on repetition facilitation

LSF is a phenomenon that has implications for both auditory spatial processing in specific and repetition facilitation in general. Here, we first discuss the four observations of LSF (excluding neural mechanisms) that contribute to repetition facilitation.

First, repetition facilitation is not just an outlier but a widely existing contextual effect. Previous studies using masker-probe and standard-oddball paradigms have reported that less than 10% of auditory cortical neurons exhibit repetition facilitation to the sound that is presented twice or have a higher firing rate to the repetitive tone than the same tone in the equal-presentation mode, respectively[1,23,24,34]. In contrast, our study demonstrated that 91% of cortical neurons exhibit repetition facilitation, which is much higher than the < 10% reported in previous studies using the masker-probe and standard-oddball paradigms. If there are only 10% of repetition facilitation neurons, it could not explain the repetition enhancement observed in functional magnetic resonance imaging (fMRI) studies because enhanced responses from 10% of neurons would be masked by the suppressed responses from the remaining 90% of neurons[26,56]. Our

data explain the repetition enhancement in fMRI as the relative strength of repetition facilitation being comparable or even stronger than the repetition suppression.

Second, repetition facilitation is not neuron-specific, as 91% of neurons exhibited facilitation and 83% exhibited suppression, with 75% of neurons showing both effects. Instead, the modulation effect is stimulus-specific, as the most nonpreferred and preferred stimuli (13th to 15th and 1st to 3rd, respectively) modulate neural activity through a push-pull mechanism, and the sign of contextual modulation reverses around the midpoint of the seven preferred and nonpreferred stimuli. This discrepancy with previous studies may be due to the fact that previous studies focused on the preferred stimuli. In the masker-probe paradigm, the probe stimulus is typically fixed at the frequency or location that maximally activates the neuron under study[1,34,57,58]. In the standard-oddball paradigm, both close-field and free-field sound stimulation were used. In close field stimulation, the sound was delivered through a sealed speaker into the contralateral ear[38,59–61] or preferred ear[62]. In free-field stimulation, the sound was played from the location contralateral to refs. 2,3,35,63, in front of ref. 64 or above[65] an animal. To the best of our knowledge, no previous studies have systematically tested standard oddball stimulation across spatial locations or at the ipsilateral location or ear.

Third, repetition facilitation is mainly evoked by nonpreferred stimuli. Preferred stimuli do evoke facilitation in 17% of neurons. This explains why previous masker-probe and standard-oddball paradigms, using preferred stimuli, either did not observe or found less than 10% of repetition-facilitated neurons. Facilitation to nonpreferred stimuli is consistent with previous studies that showed that degraded stimuli, such as low-amplitude modulated sound or natural images, evoked facilitation but not suppression[30,66]. It's important to note that while some studies using the high-probability paradigm have observed facilitation to nonpreferred stimuli[31,32], this paradigm is fundamentally different from our study, as well as the standard-oddball and masker-probe paradigms. In the high-probability paradigm, all stimuli, both preferred and nonpreferred, are presented shortly in a single long-duration trial without an inter-stimulus interval (stimulus and trial duration: 1.25 ms and 5s[67]; 5 ms and 5s[32]; 50 ms and 420s[68]; 100 ms and 120 s[31]). As a result, neurons are facilitated by these long-duration uninterrupted groups of dynamic stimuli that center on nonpreferred stimuli rather than the single repeated and static nonpreferred stimuli themselves.

Fourth, repetition facilitation, but not suppression, is long-lasting. Our intracellular recordings revealed that the more stable membrane potentials to the nonpreferred stimuli than to the preferred stimuli may contribute to this long-lasting effect. Although the functional implication of this long-lasting effect is unclear, previous studies have found that repetition facilitation is a top-down phenomenon that originates from the auditory cortex. Cortico-collicular inactivation led to a decrease of repetition facilitation in the inferior colliculus but have no effect over repetition suppression[24]. Cortico-thalamic inactivation blocked repetition facilitation in the medial geniculate body and switched facilitation to suppression[69]. Furthermore, repetition facilitation is also stronger in the nonlemniscal auditory areas[23]. In contrast to the short duration and irregular repetition suppression, this cortical long-lasting facilitation may play an important role in auditory streaming (see below).

Neural mechanisms underlying LSF

We investigated the neural mechanisms underlying long-lasting facilitation with six approaches. (1) Extracellular recording: We gradually decreased the presentation probability of the target speaker and observed gradually weaker repetition facilitation. (2) Intracellular recordings: We observed sustained and largely reliable membrane potential depolarization to nonpreferred stimuli. (3) HH model: We hypothesized that elevated arousal increases the leak conductance and reproduced the sustained membrane potential depolarization. (4) LIF model: We scaled the amplitude of depolarized membrane potential with stimulus probability and reproduced this long-lasting, stimulus continuity-dependent facilitation. (5) Head-fixed monkeys: We found increased eye-opening and pupil areas in response to repetitive stimuli. (6) Free-moving monkeys: We observed fewer head movements in response to repetitive stimuli. Why does the animal pay attention to the repetitive stimuli? It is very likely that switching between equal- and 100%-probability modes makes the repetitive stimuli salient. A salient auditory spectrotemporal feature could attract attention automatically[70,71]. In this study, a wideband noise was not salient when it was presented randomly from different locations to characterize SRF. However, a wideband noise became salient when it was repetitively presented from one location, while sounds at all other locations disappeared. This auditory spatial pop-out hypothesis is similar to visual saliency, where a visual item in sharp contrast with its neighboring items in a simple feature, such as color or orientation, automatically captures attention[72].

The neural circuit mechanism of repetition facilitation and LSF is still unclear and could be explored using other techniques. For example, multi-channel recordings with depth information[73] could determine whether neural facilitation and suppression occur simultaneously or in different layers. Since both effects have similar time scales and require tens of trials to build up, we might observe them in different populations of neurons simultaneously in the superficial layers. Two-photon imaging in transgenic mice[74,75] could assess the existence or strength of neural facilitation across different cell types. Somatostatin (SOM) inhibitory neurons contribute to the surround inhibition[76,77] and exhibit repetition facilitation to repetitive preferred stimuli[2,35]. It would be interesting to examine whether SOM neurons exhibit repetition suppression to nonpreferred stimuli in SOM-Cre mice.

Implications for the spatial receptive field of cortical neurons

Auditory cortex neurons in anesthetized animals exhibit predominantly broad SRFs that typically increase in size (or width) as sound level increases[78,79]. In contrast, studies in awake animals have reported restricted SRFs which do not increase or show less increase in size as sound level increases[80–83]. It has been shown that behavioral engagement could further decrease the size of SRFs and, therefore, sharpen spatial tuning of cortical neurons[84–86]. The finding of the present study further showed that the spatial tuning of auditory cortex neurons in awake marmosets is not static in that a non-preferred spatial location could become responsive under particular conditions. This suggests that cortical neurons can respond to spatial locations away from the center of SRF dynamically. When stimuli from other locations were inserted into the repetitively presented sound sequence from one location, neural facilitation was interrupted and even diminished (Fig. 5). However, neurons still preserve their original SRF after being presented with repetitive sound stimuli (Supplementary Fig. 7). In contrast, after a repetitive pure tone stimulus is presented, a neuron changes its spectral receptive field by reducing responses to the specific tone frequency[59]. A non-static SRF could play a role in spatial and binaural tuning plasticity that is observed in monaural-deprived animals[87,88].

Candidate neural substrate for auditory streaming

In a natural environment such as a cocktail party, there are often multiple sound sources generating continuous and simultaneous sounds[89]. One of the main challenges for a listener is forming auditory streams, which requires acoustic cues such as frequency, temporal regularity, and sound location[90,91]. Over the past decade, a rapidly increasing number of studies have investigated the effect of temporal regularity or repetition for streaming[92,93]. Similar to repeated stimulation, regular stimulation also induces stronger responses than random

stimulation when measured with magneto-electro encephalography (M/EEG) and fMRI in humans[94,95]. Repetition causes the target to pop out from the background and is robust to inattention[96–98]. Our findings that repetitive sound stimulation can evoke neural facilitation in passive listening animals provide a candidate single-neuron correlate of this perceptual phenomenon. Specifically, neurons that prefer the target location initially localize the target stream, but then suppress their firing rates as the location information becomes redundant after repetitive presentation. On the other hand, neurons that do not prefer the target location track the target sound stream from this unpreferred location through long-lasting neural facilitation. The highly dynamic SRF of cortical neurons enables them to localize and stream sound stimuli in complex auditory scenes. Our findings and models provide valuable new insights into the neural mechanisms of auditory streaming.

Methods

Animal preparation and experimental setup

Neurophysiological data were collected from five hemispheres of four monkeys (Monkey 1: left, Monkey 2: left and right, Monkey 3: right, Monkey 4: left). All experimental procedures were approved by the Johns Hopkins University Animal Use and Care Committee. These procedures, including head cap implantation and neurophysiological recordings, were detailed in this protocol paper[99]. A typical recording session lasted 3–4 h, during which an animal sat quietly in a specially adapted primate chair with its head immobilized. Throughout the entire recording session, the animal was closely monitored via a video camera by the researcher. The eye position was not controlled, but when the animal closed its eyes for a prolonged period, the experimenter ensured the animal opened its eyes before the next stimulus set was presented. Behavioral data were collected from four monkeys. Two monkeys used for head-fixed experiments had implanted headcaps and cranial windows for unrelated imaging studies. The two monkeys used for free-moving experiments did not undergo any invasive procedures.

Experiments were conducted in a double-walled sound-proof chamber (Industrial-Acoustics) with the internal walls and ceiling lined with three-inch acoustic absorption foam (Sonex). Fifteen free-field loudspeakers (FT28D 28 mm Dome Tweeter, Fostex) were placed on a semi-spherical surface centered around the animal's head and above the horizontal plane. The speaker setup was similar to our previous studies (Zhou and Wang, 2012), but with speakers covering the rear sphere. Eight speakers were evenly positioned at 0° elevation, five speakers were evenly spaced at + 45° elevation in the frontal hemifield, one speaker was located at + 45° elevation in the rear midline, and one speaker was located directly above the animal. The distance between the speakers and the animal's head was 1 meter.

Extracellular and intracellular recordings

Extracellular recording procedures were identical to those described in our previous publications. A sterile single tungsten microelectrode (A-M Systems) was held by a micro-manipulator (Narishige) and inserted nearly perpendicularly into the auditory cortex through a small opening on the skull (1.0–1.1 mm in diameter) and advanced by a hydraulic micro-drive (David Kopf Instruments). The tip and impedance of the electrode was examined before each recording session (2–5 MΩ impedance). Spikes were detected by a template-based spike sorter (MSD, Alpha Omega Engineering) and continuously monitored by the experimenter while data recordings progressed. The raw voltage signal was also recorded. Intracellular recording procedures were identical to those described in our previous publications[39,40]. The recordings were made in the auditory cortex through the intact dura using a concentric recording electrode and guide tube assembly. The sharp recording pipette was made of quartz glass. The guide tube was made of borosilicate glass. The sharp recording pipette was pulled by a laser puller (P-2000, Sutter Instrument), and the guide tube was pulled by a conventional puller (P-97, Sutter Instrument). The electrode assembly was advanced perpendicularly relative to the cortical surface with a motorized manipulator (DMA1510, Narishige). The electrical signals were amplified (Axoclamp 2B, Molecular Devices), digitized (RX6, Tucker-Davis Technologies), and saved using custom programs (MATLAB, Mathworks).

Behavioral tracking

We applied the same stimulus paradigms used in previous neurophysiological recordings to behavioral experiments. Videos were continuously acquired at 20 frames per second using the software provided by the camera vendors (FlyCapture2). We synchronized the video acquisition with sound stimuli by recording an LED trigger signal (MNWHL4, Thorlabs) at each sound stimulus. To reduce acoustic reflection caused by the camera, camera lens, and light source, all items were fixed on the wall to avoid blocking stimuli from any speakers.

For head-fixed recording, we primarily used infrared (IR) light (M850L3, Thorlabs) focused on the marmoset's face (SM1U25B, Thorlabs). A camera lens with a 100 mm focal length (86-410, Edmund) was used to capture clear video of the face. We also used room light in five sessions. The camera used (Chameleon3 CM3-U3-50S5C, Point Gray Research) is not an IR camera, and its sensor (IMX264, Sony) is more sensitive to visible light than IR light. Therefore, the pupil appeared unclear (though still trackable) under room light, so we turned off the room light in twenty sessions. For free-moving recordings, we used only room light and placed the monkeys inside a cage (L: 480 mm, W: 380 mm, H: 460 mm) that allowed the marmosets to jump freely. To capture the entire cage, we used a different camera (Grasshopper GS3-U3-23S6M-C, Point Gray Research) equipped with a camera lens (58-000, Edmund) with an 8.5 mm focal length.

Acoustic stimuli

Four different stimulus presentation designs were used. (1) 100%-probability: the same stimulus was repeatedly delivered from a fixed speaker location over many trials. (2) Unequal probability and random (75, 50, and 25%): stimulus was delivered from multiple speaker locations in a randomly shuffled order, but target speaker has a higher probability than others. (3) Unequal probability and periodic (50/p%): The stimulus delivered from the target speaker was interleaved with stimulus delivered from other speakers so that the stimulus delivered from the target speaker was periodical. (4) Equal probability (6.7%): stimulus was delivered from multiple speaker locations in a randomly shuffled order, and all speakers shared the same occurrence probability. For each neuron, if allowed by the experiment conditions, the equal-probability presentation mode was tested at multiple separate sessions at different time points, and the other three presentation modes were tested between the equal-probability presentation mode.

Stimuli were generated digitally in MATLAB at a sampling rate of 97.7 kHz using custom software, converted to analog signals (RX6, Tucker-Davies Technologies), attenuated (PA5, Tucker-Davies Technologies), power amplifier (Crown Audio), and played from specified loudspeaker one at a time.

The amplifiers were fine-tuned so that a 4 kHz tone emitted from the front speaker reached a sound pressure level of 95 decibels (dB SPL) at 0 dB attenuation. This measurement was taken at the anticipated position of the marmoset's head, using a half-inch free-field microphone (type 4191, Brüel & Kjær). Loudspeaker responses, assessed via Golay codes, exhibited a relatively uniform frequency response curve (varying within ± 3–7 dB) and showed minimal spectral disparity across speakers (less than 7 dB relative to the mean) within the utilized frequency range. The most significant spectral discrepancies (5–7 dB) were confined to narrow bands near the speakers' upper-frequency

limit (above 28 kHz), surpassing the initial spectral notch documented in marmoset head-related transfer functions (HRTF[100]).

The sound tokens used included unfrozen wide-band noise (1–32 kHz), frozen wide-band noise, amplitude-modulated wide-band noise, and species-specific vocalizations (mainly wide-band twitter calls). Fixed inter-stimulus intervals (ISI) were used in four presentation modes. A variety of random ISI were used only in 100%-probability presentation mode. The shortest ISI was 500 ms, and the longest ISI was 5200 ms. To determine the optimal sound level for tested neurons, a rate-level function was employed, ranging from 10 dB to 90 dB attenuation in 10 dB increments (equivalent to 5 to 85 dB SPL). The majority of neurons were assessed at their optimal sound level. For consistency, this same sound level was maintained across different presentation modes. Total firing rates were calculated over a time window beginning 15 ms after stimulus onset and 50 ms after stimulus offset. Total firing rates minus the spontaneous rate was the firing rate. In the plotted SRF, the black thick line was the half-maximum threshold of SRF, and the area encircled by the threshold was the reciprocal of spatial tuning selectivity.

Identification of cortical areas, layers and cell types

We used the best frequency (BF) of neurons to identify the subregions of the auditory cortex. For the neurons significantly responding to at least one tone stimulus played at the front speaker, we specified the frequency of the tone stimulus that evoked the maximum response rate as the neuron's BF. The marmoset auditory cortex is situated largely ventral to the lateral sulcus and exhibits a topographical frequency gradient along the rostral-caudal axis. The boundary between the primary auditory cortex (A1) and the caudal area (caudal-medial and caudal-lateral belt) can be identified by an abrupt decrease of BF at the high-frequency (caudal) border of A1. A1 was further divided into the low-frequency (≤ 8KHz) rostral A1 and high-frequency (> 8KHz) caudal A1 along the rostral-caudal axis. The first spike depth was the absolute depth where the first spike was detected from this electrode and was depended on the thickness of the dura, variations in granulation tissue, proximity to the curvature of the sulcus, and orthogonality of the electrode penetration to the cortical surface. We used the trough-to-peak duration of spike waveforms to identify putative excitatory and inhibitory neurons[101]. Before analyzing the spike duration, we calculated the signal-to-noise ratio of spike waveforms, which was defined as the action potential peak-to-peak height divided by the standard deviation of the background noise over 1 ms preceding all spikes ($20 \times \log10(AP_{peak-peak}/Noise_{SD})$).

Our analysis found that LSF can be induced among different cortical areas, in both superficial and deep layers, and in both putative excitatory and inhibitory neurons (data not shown).

Speaker ranking

In the 100%-probability presentation mode, each speaker was assigned a rank based on its firing rate obtained in the equal-probability mode. Among the 15 speakers tested, the speaker ranked 1st demonstrated the highest firing rate, while the one ranked 15th showed the lowest. Typically, for SRFs with a singular peak, the speaker ranked 1st and 15th were situated at the nearest and farthest points from the preferred direction, respectively. In cases of SRFs with multiple peaks, lower-ranked speakers were occasionally positioned close to the preferred direction. The rationale for using firing rate rank, as opposed to distance rank, was twofold: Firstly, the SRF of certain neurons was broadly spread, making it imprecise to calculate the distance between a specific speaker and the SRF center. Secondly, the SRF center was often defined by a cluster of high-ranked speakers, with minimal consideration given to lower-ranked ones. As a result, employing the distance rank would overlook the contributions of these lower-ranked speakers.

Proportion of facilitation, adaptation, depolarization and hyperpolarization trials

Facilitation trials were identified as those with firing rates at least 0.5/1/1.5/2 standard deviations above the mean rate observed under equal-probability presentation. The ratio of the count of facilitation trials to the total number of trials per session was calculated as the proportion of facilitation trials. Conversely, adaptation trials were those with firing rates at least 0.5/1/1.5/2 standard deviations below the mean rate of the equal probability mode. To assess changes in membrane potential, spikes were extracted from the voltage signal using a 3 ms window centered on the spike peak. The proportions of depolarization and hyperpolarization trials were determined in a manner analogous to facilitation and adaptation trials, by comparing the deviation from the mean membrane potential.

Shuffling the firing rate sequence

To assess the continuity of facilitation, we anticipated observing more consecutive trials surpassing the threshold in the experimental session than in a session with shuffled firing rates. Supplementary Fig. 4a illustrates this analysis, where repetitive sound stimuli from the 15th-ranked speaker were presented, leading to sequences of 5, 6, and 9 consecutive trials exceeding the $\mu + 2\sigma$ facilitation threshold. However, these extended sequences of facilitation trials dissipated when the firing rate sequence was shuffled. The proportion of facilitation sessions was then compared between the experimental and shuffled firing rates that had different numbers of consecutive facilitation trials (Supplementary Fig. 4b, solid vs. dashed orange lines).

Furthermore, we utilized facilitation thresholds derived not only from the equal-probability mode but also from the mean firing rate in the 100%-probability mode. This approach compensates for the general increase in firing rates seen in the 100%-probability mode, thereby mitigating potential biases due to firing rate variations between the two presentation modes. Note that this adjusted threshold was only applied when comparing the count of consecutive trials between the experimental and shuffled sessions. Even with this new set of thresholds, sequences of 3 and 4 consecutive trials exceeding the $\mu + 2\sigma$ facilitation threshold were observed exclusively in the experimental sessions, not in the shuffled ones (Supplementary Fig. 4c). Similar outcomes were noted when applying thresholds calculated under the 100%-probability presentation mode (Supplementary Fig. 4d).

Behavioral data analysis

We used MATLAB to extract the light trigger signals for synchronizing the stimuli with video frames. All the remaining analyses (behavior tracking, data analysis and visualization, and statistical tests) were conducted in Python through Jupyter Notebook. For the behavioral tracking, we tested two open-source, deep-learning-based software: SLEAP (v1.3.3, https://github.com/talmolab/sleap) and DeepLabCut (v2.3.9, https://github.com/DeepLabCut/DeepLabCut). We chose DeepLabCut because it was more accurate and robust than SLEAP for tracking pupil area using our dataset. We selected 100 video frames as our training samples and trained the frames on a single GPU (NVIDIA GTX 1080Ti, 11 GB) with either 100,000 or 500,000 iterations. Notice that the loss became stable after 100,000 iterations and extra training provided minimal benefits. Once the model was trained, it could be used to analyze all the videos quickly (less than two minutes for a 2 GB video).

For head-fixed recording, we labeled six markers for the eyelids in one training set and five markers for the pupil boundary and corneal reflection in a separate set. All markers were manually labeled and validated before being applied to further video analysis. For free-moving recording, we labeled five markers on the head: the left, right, and front markers were placed on the white fur. There are two

limitations in this dataset. First, a single camera could not capture the three-dimensional movement, causing some or all markers to become invisible at certain locations. Second, the camera was placed outside the cage, and the cage grids often blocked the markers, leading to marker loss. DeepLabCut exported an HDF5 file containing the markers' information for all frames for data analysis. In each frame, each marker included not only its X-Y coordinates but also the likelihood of this marker. The likelihood value (0 to 1) reflects the marker's tracking performance, and we chose markers with a likelihood greater than 0.95 for further analysis. We used the statistical functions from the SciPy package for statistical analysis and the Matplotlib package for data visualization.

Computational model of sustained membrane potential depolarization

In this model, we simulate the Hodgkin-Huxley neuron model with a focus on the leak conductance primarily carrying chloride ions (Cl^-). Four key parameters are: Membrane Capacitance (C_m), 1.0 μF/cm²; Sodium Conductance (g_{Na}), 120 mS/cm²; Potassium Conductance (g_k), 36.0 mS/cm²; Initial Leak Conductance ($g_{L\_initial}$), 0.3 mS/cm². Equilibrium Potentials: E_{Na}, 50.0 mV (sodium); E_K, −77.0 mV (potassium); E_L, −54.387 mV (chloride, primarily).

The time parameters are set for a total time (T) of 400 ms with a time step (dt) of 0.01 ms. The external current (I_{ext}) is set to zero throughout the simulation.

The gating variables and their rate constants are defined as follows:

$$\alpha_m(V) = \frac{0.1(V+40)}{1-\exp(\frac{-(V+40)}{10})},\ \beta_m(V) = 4.0\exp\left(\frac{-(V+65)}{18}\right)$$

$$\alpha_h(V) = 0.07\exp\left(\frac{-(V+65)}{20}\right),\ \beta_h(V) = \frac{1.0}{1+\exp\left(\frac{-(V+35)}{10}\right)}$$

$$\alpha_n(V) = \frac{0.01(V+55)}{1-\exp(\frac{-(V+55)}{10})},\ \beta_n(V) = 0.125\exp\left(\frac{-(V+65)}{80}\right)$$

The initial membrane potential (V) is set to −65.0 mV. The membrane potential (V) is updated at each time step (dt) using the following equation:

$$V(t+dt) = V(t) + \frac{dt}{C_m}(I_{ext} - I_{ion})$$

Where I_{ion} is the total ionic current:

$$I_{ion} = I_{Na} + I_K + I_L,\ I_{Na} = g_{Na}m^3h(V - E_{Na})$$

$$I_k = g_k n^4(V - E_K),\ I_L = g_L(V - E_L)$$

The gating variables (m, h, and n) are updated using the following first-order differential equations:

$$\frac{dm}{dt} = \alpha_m(1-m) - \beta_m m,\ \frac{dh}{dt} = \alpha_h(1-h) - \beta_h h,\ \frac{dn}{dt} = \alpha_n(1-n) - \beta_n n$$

The leak conductance (g_L) is incremented by a small value (0.0001mS/cm²) at each time step to achieve a gradual increase in the resting membrane potential.

Computational models of location-specific facilitation

We used a leaky integrate-and-fire (LIF) neuron to simulate the location-specific neural facilitation. One LIF neuron has different strengths of excitatory and inhibitory synaptic strength (EI-LIF), and facilitation occurred when the excitatory synaptic inputs are stronger than inhibitory ones. The other LIF neuron has sustained membrane potential depolarization (MP-LIF). Our MP-LIF model dynamically changed its resting membrane potential and spiking threshold based on the probability of the presentation speaker, e.g., 100% for 100%-probability presentation mode and 6.7% for equal-probability presentation mode. Supplementary Table 3 listed the names of parameters and corresponding values used in the two model neurons.

The membrane potential V_{t+1} of a LIF neuron at time step Δt was:

$$V_{t+1} = -\frac{\Delta t}{C}\left[g_{e_t}(V_t - E_e) + g_{i_t}(V_t - E_i) + g_{rest}(V_t - E_{rest})\right] + V_t + \sigma_s\omega_n\sqrt{\Delta t} \quad (1)$$

g_{e_t} and g_{i_t} was the excitatory and inhibitory synaptic conductance (see Eqs. 2 and 3). C, E_e, E_i and g_{rest} was the membrane capacitance, excitatory reversal potential, inhibitory reversal potential, and leak conductance. Those values were obtained from the in vivo whole-cell recording in the auditory cortex of anesthetized rats[102]. Gaussian noise $\sigma_s\omega_n$ was added to generate the spontaneous firing. Action potential was evoked when the V_{t+1} reached the spike threshold V_{spike}. V_{t+1} was reset to E_{rest} after the action potential. E_{rest} was obtained from our intracellular studies. V_{spike} is the sum of the threshold above resting potential V_{th} and E_{rest}.

Excitatory conductance g_{e_t} and inhibitory conductance g_{i_t} were:

$$g_{e_t} = P_{rel\_e}rN_e\Delta te^{\frac{\Delta t}{\tau}} + \sigma_c\omega_n \quad (2)$$

$$g_{i_t} = P_{rel\_i}N_i\Delta te^{\frac{\Delta t}{\tau}} + \sigma_c\omega_n \quad (3)$$

P_{rel_e} and P_{rel_i} were the excitatory and inhibitory synaptic release probability, respectively. P_{rel_e} and P_{rel_i} were fixed to one in the LIF model. The inhibitory and excitatory inputs have the same strength and occurred simultaneously, so the inhibitory to excitatory ratio r equal to one, and the inhibitory to excitatory delay d (not shown in the equation) equal to zero. The number of excitatory inputs N_e, inhibitory input N_i and time constant τ were fixed, and conduction noise $\sigma_c\omega_n$ was added to generate the spontaneous firing[102]. Each session is composed of randomly presented trials T_r for computing the facilitation and adaptation threshold and continuously presented trials T_c for computing the facilitation percent, adaptation percent, and firing rate.

For the EI-LIF model, in each trial T, the excitatory and inhibitory synaptic release probability $P_{rel\_e_{t+1}}$ and $P_{rel\_i_{t+1}}$ were:

$$P_{rel\_e_{t+1}} = 1 + \left(A_e P_{rel\_e_t} - 1\right)e^{\frac{-\Delta t}{P_s\tau_e}} \quad (4)$$

$$P_{rel\_i_{t+1}} = 1 + \left(A_i P_{rel\_i_t} - 1\right)e^{\frac{-\Delta t}{P_s\tau_i}} \quad (5)$$

A_e and A_i was the excitatory and inhibitory synaptic amplitude, respectively. A_e and A_i change from 0.901 to 1 with a step size of 0.001. τ_e and τ_i was the excitatory and inhibitory recovery time constant, respectively. Each session is composed of randomly presented trials T_r for computing the facilitation and adaptation threshold and continuously presented trials T_c for computing the facilitation percent, adaptation percent, and firing rate. The median firing rate in the 100%-probability mode was 12 spikes per second. Therefore, the number of stimulus counts N_{SC} was chosen so that the average firing rate in the EI-LIF model could match the firing rate in the recording data. Notice that

N_{SC} was Poisson distributed and not every stimulus input could evoke a spike output in the LIF neuron. We run the EI-LIF model for 50 sessions for every combination of A_e and A_i.

For the MP-LIF model, in each trial T, the dynamic resting membrane potential E_{rest_MP} and spike threshold V_{spike} were:

$$E_{restMP} = \begin{cases} E_{rest}, 1 \le T < T_r \\ E_{rest}: P_s M/(T_{th}-1): E_{rest} + P_s M, T_r + 1 \le T < T_r + T_{th} \\ E_{rest} + P_s M, T_r + T_{th} + 1 \le T < T_r + T_c \end{cases} \quad (6)$$

$$V_{spike} = \begin{cases} V_{th} + E_{rest}, 1 \le T < T_r \\ V_{th} + E_{rest}: P_s MS/(T_{th}-1): E_{rest} + P_s MS, T_r + 1 \le T < T_r + T_{th} \\ V_{th} + E_{rest} + P_s MS, T_r + T_{th} + 1 \le T < T_r + T_c \end{cases} \quad (7)$$

The depolarization value M and the number of trials to reach the stabilized depolarization value T_{th} were obtained from our intracellular recordings. Since lower presentation probability evoked less neural facilitation, therefore M was scaled by the presentation probability P_s. The spike threshold V_{spike} was further modulated by the spike threshold scale S. Almost no spike was evoked when S equal to one but excessive spikes were evoked when S equal to zero. Therefore, S and stimulus count N_{SC} were chosen so that the average firing rate matched the recording data (the median firing rate in the 100%-probability presentation mode was 12 spikes per second). Notice that N_{SC} was Poisson distributed and not every stimulus input could evoke a spike output in the LIF neuron. We run the MP-LIF model for two hundred sessions at every presentation probability.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw data is available at GitHub (https://github.com/ccg1988/LSF_Nature_Communications_2024) and Zenodo (https://zenodo.org/records/14737013).

Code availability

The code is available at GitHub (https://github.com/ccg1988/LSF_Nature_Communications_2024) and Zenodo (https://zenodo.org/records/14737013).

References

1. Bartlett, E. L. & Wang, X. Long-lasting modulation by stimulus context in primate auditory cortex. *J. Neurophysiol.* **94**, 83–104 (2005).
2. Natan, R. G., Rao, W. & Geffen, M. N. Cortical interneurons differentially shape frequency tuning following adaptation. *Cell Rep.* **21**, 878–890 (2017).
3. Obara, K. et al. Change detection in the primate auditory cortex through feedback of prediction error signals. *Nat. Commun.* **14**, 6981 (2023).
4. Muller, J. R., Metha, A. B., Krauskopf, J. & Lennie, P. Rapid adaptation in visual cortex to the structure of images. *Science* **285**, 1405–1408 (1999).
5. Williams, N. P. & Olson, C. R. Contribution of individual features to repetition suppression in macaque inferotemporal cortex. *J. Neurophysiol.* **128**, 378–394 (2022).
6. Ganmor, E., Katz, Y. & Lampl, I. Intensity-dependent adaptation of cortical and thalamic neurons is controlled by brainstem circuits of the sensory pathway. *Neuron* **66**, 273–286 (2010).
7. Heiss, J. E., Katz, Y., Ganmor, E. & Lampl, I. Shift in the balance between excitation and inhibition during sensory adaptation of S1 neurons. *J. Neurosci.* **28**, 13320–13330 (2008).
8. Wilson, D. A. Habituation of odor responses in the rat anterior piriform cortex. *J. Neurophysiol.* **79**, 1425–1440 (1998).
9. Kato, H. K., Chu, M. W., Isaacson, J. S. & Komiyama, T. Dynamic sensory representations in the olfactory bulb: modulation by wakefulness and experience. *Neuron* **76**, 962–975 (2012).
10. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* **6**, 391–398 (2003).
11. Anderson, L. A., Christianson, G. B. & Linden, J. F. Stimulus-specific adaptation occurs in the auditory thalamus. *J. Neurosci.* **29**, 7359–7363 (2009).
12. Malmierca, M. S., Cristaudo, S., Perez-Gonzalez, D. & Covey, E. Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J. Neurosci.* **29**, 5483–5493 (2009).
13. Hamm, J. P. & Yuste, R. Somatostatin interneurons control a key component of mismatch negativity in mouse visual cortex. *Cell Rep.* **16**, 597–604 (2016).
14. Peter, A. et al. Stimulus-specific plasticity of macaque V1 spike rates and gamma. *Cell Rep.* **37**, 110086 (2021).
15. Katz, Y., Heiss, J. E. & Lampl, I. Cross-whisker adaptation of neurons in the rat barrel cortex. *J. Neurosci.* **26**, 13363–13372 (2006).
16. Liu, C., Foffani, G., Scaglione, A., Aguilar, J. & Moxon, K. A. Adaptation of thalamic neurons provides information about the spatiotemporal context of stimulus history. *J. Neurosci.* **37**, 10012–10021 (2017).
17. Verhagen, J. V., Wesson, D. W., Netoff, T. I., White, J. A. & Wachowiak, M. Sniffing controls an adaptive filter of sensory input to the olfactory bulb. *Nat. Neurosci.* **10**, 631–639 (2007).
18. Fishman, Y. I. & Steinschneider, M. Searching for the mismatch negativity in primary auditory cortex of the awake monkey: deviance detection or stimulus specific adaptation? *J. Neurosci.* **32**, 15747–15758 (2012).
19. Polterovich, A., Jankowski, M. M. & Nelken, I. Deviance sensitivity in the auditory cortex of freely moving rats. *PLoS ONE* **13**, e0197678 (2018).
20. Carbajal, G. V. & Malmierca, M. S. The neuronal basis of predictive coding along the auditory pathway: From the subcortical roots to cortical deviance detection. *Trends Hear* **22**, 2331216518784822 (2018).
21. Wark, B., Lundstrom, B. N. & Fairhall, A. Sensory adaptation. *Curr. Opin. Neurobiol.* **17**, 423–429 (2007).
22. Kohn, A. Visual adaptation: physiology, mechanisms, and functional benefits. *J. Neurophysiol.* **97**, 3155–3164 (2007).
23. Parras, G. G. et al. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* **8**, 2148 (2017).
24. Lesicko, A. M. H., Angeloni, C. F., Blackwell, J. M., De Biasi, M. & Geffen, M. N. Corticofugal regulation of predictive coding. *Elife* **11**, https://doi.org/10.7554/elife.73289 (2022).
25. Li, L., Miller, E. K. & Desimone, R. The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* **69**, 1918–1929 (1993).
26. de Gardelle, V., Waszczuk, M., Egner, T. & Summerfield, C. Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cereb. Cortex* **23**, 2235–2244 (2013).
27. Brecht, M. & Sakmann, B. Dynamic representation of whisker deflection by synaptic potentials in spiny stellate and pyramidal cells in the barrels and septa of layer 4 rat somatosensory cortex. *J. Physiol.* **543**, 49–70 (2002).
28. Derdikman, D. et al. Layer-specific touch-dependent facilitation and depression in the somatosensory cortex during active whisking. *J. Neurosci.* **26**, 9538–9547 (2006).

29. Cai, R., Richardson, B. D. & Caspary, D. M. Responses to predictable versus random temporally complex stimuli from single units in auditory thalamus: Impact of aging and anesthesia. *J. Neurosci.* **36**, 10696–10706 (2016).
30. Kommajosyula, S. P., Cai, R., Bartlett, E. & Caspary, D. M. Top-down or bottom up: decreased stimulus salience increases responses to predictable stimuli of auditory thalamic neurons. *J. Physiol.* **597**, 2767–2784 (2019).
31. Watkins, P. V. & Barbour, D. L. Specialized neuronal adaptation for preserving input sensitivity. *Nat. Neurosci.* **11**, 1259–1261 (2008).
32. Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L. & King, A. J. Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron* **66**, 937–948 (2010).
33. Brosch, M. & Scheich, H. Tone-sequence analysis in the auditory cortex of awake macaque monkeys. *Exp. Brain Res.* **184**, 349–361 (2008).
34. Phillips, E. A. K., Schreiner, C. E. & Hasenstaub, A. R. Diverse effects of stimulus history in waking mouse auditory cortex. *J. Neurophysiol.* **118**, 1376–1393 (2017).
35. Kato, H. K., Gillet, S. N. & Isaacson, J. S. Flexible sensory representations in auditory cortex driven by behavioral relevance. *Neuron* **88**, 1027–1039 (2015).
36. Wang, X., Lu, T., Snider, R. K. & Liang, L. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* **435**, 341–346 (2005).
37. Chaplin, T. A., Yu, H. H. & Rosa, M. G. Representation of the visual field in the primary visual area of the marmoset monkey: magnification factors, point-image size, and proportionality to retinal ganglion cell density. *J. Comp. Neurol.* **521**, 1001–1019 (2013).
38. Yaron, A., Hershenhoren, I. & Nelken, I. Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron* **76**, 603–615 (2012).
39. Gao, L., Kostlan, K., Wang, Y. & Wang, X. Distinct subthreshold mechanisms underlying rate-coding principles in primate auditory cortex. *Neuron* **91**, 905–919 (2016).
40. Gao, L. & Wang, X. Subthreshold activity underlying the diversity and selectivity of the primary auditory cortex studied by intracellular recordings in awake marmosets. *Cereb. Cortex* **29**, 994–1005 (2019).
41. Reitman, M. E. et al. Norepinephrine links astrocytic activity to regulation of cortical state. *Nat. Neurosci.* **26**, 579–593 (2023).
42. Gau, Y. A. et al. Multicore fiber optic imaging reveals that astrocyte calcium activity in the mouse cerebral cortex is modulated by internal motivational state. *Nat. Commun.* **15**, 3039 (2024).
43. Untiet, V. et al. Astrocytic chloride is brain state dependent and modulates inhibitory neurotransmission in mice. *Nat. Commun.* **14**, 1871 (2023).
44. Varela, J. A., Song, S., Turrigiano, G. G. & Nelson, S. B. Differential depression at excitatory and inhibitory synapses in visual cortex. *J. Neurosci.* **19**, 4293–4304 (1999).
45. Galarreta, M. & Hestrin, S. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nat. Neurosci.* **1**, 587–594 (1998).
46. Bridi, M. S., Shin, S., Huang, S. & Kirkwood, A. Dynamic recovery from depression enables rate encoding in inhibitory synapses. *iScience* **23**, 100940 (2020).
47. Hong, L., Walz, J. M. & Sajda, P. Your eyes give you away: prestimulus changes in pupil diameter correlate with poststimulus task-related EEG dynamics. *PLoS ONE* **9**, e91321 (2014).
48. Montes-Lourido, P., Kar, M., Kumbam, I. & Sadagopan, S. Pupillometry as a reliable metric of auditory detection and discrimination across diverse stimulus paradigms in animal models. *Sci. Rep.* **11**, 3108 (2021).
49. Saderi, D., Schwartz, Z. P., Heller, C. R., Pennington, J. R. & David, S. V. Dissociation of task engagement and arousal effects in auditory cortex and midbrain. *Elife* **10**, https://doi.org/10.7554/eLife.60153 (2021).
50. Zhao, S., Bury, G., Milne, A. & Chait, M. Pupillometry as an objective measure of sustained attention in young and older listeners. *Trends Hear* **23**, 2331216519887815 (2019).
51. McGinley, M. J., David, S. V. & McCormick, D. A. Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* **87**, 179–192 (2015).
52. Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
53. Town, S. M., Brimijoin, W. O. & Bizley, J. K. Egocentric and allocentric representations in auditory cortex. *PLoS Biol.* **15**, e2001878 (2017).
54. Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R. & Inati, S. Eye position influences auditory responses in primate inferior colliculus. *Neuron* **29**, 509–518 (2001).
55. Werner-Reiss, U., Kelly, K. A., Trause, A. S., Underhill, A. M. & Groh, J. M. Eye position affects activity in primary auditory cortex of primates. *Curr. Biol.* **13**, 554–562 (2003).
56. James, T. W. & Gauthier, I. Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Hum. Brain Mapp.* **27**, 37–46 (2006).
57. Reale, R. A. & Brugge, J. F. Directional sensitivity of neurons in the primary auditory (AI) cortex of the cat to successive sounds ordered in time and space. *J. Neurophysiol.* **84**, 435–450 (2000).
58. Zhou, Y. & Wang, X. Spatially extended forward suppression in primate auditory cortex. *Eur. J. Neurosci.* **39**, 919–933 (2014).
59. Condon, C. D. & Weinberger, N. M. Habituation produces frequency-specific plasticity of receptive fields in the auditory cortex. *Behav. Neurosci.* **105**, 416–430 (1991).
60. Hershenhoren, I., Taaseh, N., Antunes, F. M. & Nelken, I. Intracellular correlates of stimulus-specific adaptation. *J. Neurosci.* **34**, 3303–3319 (2014).
61. Nieto-Diego, J. & Malmierca, M. S. Topographic distribution of stimulus-specific adaptation across auditory cortical fields in the anesthetized rat. *PLoS Biol.* **14**, e1002397 (2016).
62. Ulanovsky, N., Las, L., Farkas, D. & Nelken, I. Multiple time scales of adaptation in auditory cortex neurons. *J. Neurosci.* **24**, 10440–10453 (2004).
63. Chen, I. W., Helmchen, F. & Lutcke, H. Specific early and late oddball-evoked responses in excitatory and inhibitory neurons of mouse auditory cortex. *J. Neurosci.* **35**, 12560–12573 (2015).
64. Natan, R. G. et al. Complementary control of sensory adaptation by two types of cortical interneurons. *Elife* **4**, https://doi.org/10.7554/eLife.09868 (2015).
65. Farley, B. J., Quirk, M. C., Doherty, J. J. & Christian, E. P. Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. *J. Neurosci.* **30**, 16475–16484 (2010).
66. Turk-Browne, N. B., Yi, D. J., Leber, A. B. & Chun, M. M. Visual quality determines the direction of neural repetition effects. *Cereb. Cortex* **17**, 425–433 (2007).
67. Kvale, M. N. & Schreiner, C. E. Short-term adaptation of auditory receptive fields to dynamic stimuli. *J. Neurophysiol.* **91**, 604–612 (2004).
68. Dean, I., Harper, N. S. & McAlpine, D. Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* **8**, 1684–1689 (2005).
69. Kommajosyula, S. P., Bartlett, E. L., Cai, R., Ling, L. & Caspary, D. M. Corticothalamic projections deliver enhanced responses to medial geniculate body as a function of the temporal reliability of the stimulus. *J. Physiol.* **599**, 5465–5484 (2021).

70. Kayser, C., Petkov, C. I., Lippert, M. & Logothetis, N. K. Mechanisms for allocating auditory attention: an auditory saliency map. *Curr. Biol.* **15**, 1943–1947 (2005).
71. Huang, N. & Elhilali, M. Push-pull competition between bottom-up and top-down auditory attention to natural soundscapes. *Elife* **9**, https://doi.org/10.7554/eLife.52984 (2020).
72. Yan, Y., Zhaoping, L. & Li, W. Bottom-up saliency and top-down learning in the primary visual cortex of monkeys. *Proc. Natl. Acad. Sci. USA* **115**, 10499–10504 (2018).
73. Song, X., Guo, Y., Chen, C., Lee, J. H. & Wang, X. Tonotopic organization of auditory cortex in awake marmosets revealed by multi-modal wide-field optical imaging. *Curr. Res. Neurobiol.* **6**, 100132 (2024).
74. Chen, C. & Song, S. Differential cell-type dependent brain state modulations of sensory representations in the non-lemniscal mouse inferior colliculus. *Commun. Biol.* **2**, 356 (2019).
75. Chen, C. & Song, S. Distinct neuron types contribute to hybrid auditory spatial coding. *J. Neurosci.* **44**, https://doi.org/10.1523/JNEUROSCI.0159-24.2024 (2024).
76. Keller, A. J. et al. A disinhibitory circuit for contextual modulation in primary visual cortex. *Neuron* **108**, 1181–1193 (2020).
77. Kato, H. K., Asinof, S. K. & Isaacson, J. S. Network-level control of frequency tuning in auditory cortex. *Neuron* **95**, 412–423 (2017).
78. Middlebrooks, J. C. & Pettigrew, J. D. Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *J. Neurosci.* **1**, 107–120 (1981).
79. Mrsic-Flogel, T. D., King, A. J. & Schnupp, J. W. Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. *J. Neurophysiol.* **93**, 3489–3503 (2005).
80. Mickey, B. J. & Middlebrooks, J. C. Representation of auditory space by cortical neurons in awake cats. *J. Neurosci.* **23**, 8649–8663 (2003).
81. Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E. & Recanzone, G. H. Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *J. Neurophysiol.* **96**, 3323–3337 (2006).
82. Zhou, Y. & Wang, X. Level dependence of spatial processing in the primate auditory cortex. *J. Neurophysiol.* **108**, 810–826 (2012).
83. Remington, E. D. & Wang, X. Neural representations of the full spatial field in auditory cortex of awake marmoset (Callithrix jacchus). *Cereb. Cortex* **29**, 1199–1216 (2019).
84. Lee, C. C. & Middlebrooks, J. C. Auditory cortex spatial sensitivity sharpens during task performance. *Nat. Neurosci.* **14**, 108–114 (2011).
85. van der Heijden, K., Rauschecker, J. P., Formisano, E., Valente, G. & de Gelder, B. Active sound localization sharpens spatial tuning in human primary auditory cortex. *J. Neurosci.* **38**, 8574–8587 (2018).
86. Amaro, D., Ferreiro, D. N., Grothe, B. & Pecka, M. Source identity shapes spatial preference in primary auditory cortex during active navigation. *Curr. Biol.* **31**, 3875–3883 (2021).
87. Popescu, M. V. & Polley, D. B. Monaural deprivation disrupts development of binaural selectivity in auditory midbrain and cortex. *Neuron* **65**, 718–731 (2010).
88. Keating, P., Dahmen, J. C. & King, A. J. Complementary adaptive processes contribute to the developmental plasticity of spatial hearing. *Nat. Neurosci.* **18**, 185–187 (2015).
89. Cherry, E. C. Some experiments on the recognition of speech with one or two ears. *J. Acoust. Soc. Am.* **25**, 975–979 (1953).
90. McDermott, J. H. The cocktail party problem. *Curr. Biol.* **19**, R1024–R1027 (2009).
91. Shamma, S. A. & Micheyl, C. Behind the scenes of auditory perception. *Curr. Opin. Neurobiol.* **20**, 361–366 (2010).
92. Bendixen, A., Denham, S. L., Gyimesi, K. & Winkler, I. Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* **128**, 3658–3666 (2010).
93. Andreou, L. V., Kashino, M. & Chait, M. The role of temporal regularity in auditory segregation. *Hear Res.* **280**, 228–235 (2011).
94. Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J. & Chait, M. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc. Natl. Acad. Sci. USA* **113**, E616–E625 (2016).
95. Southwell, R. *et al*. Is predictability salient? A study of attentional capture by auditory patterns. *Philos. Trans. R Soc. Lond B Biol. Sci.* **372**, https://doi.org/10.1098/rstb.2016.0105 (2017).
96. McDermott, J. H., Wrobleski, D. & Oxenham, A. J. Recovering sound sources from embedded repetition. *Proc. Natl. Acad. Sci. USA* **108**, 1188–1193 (2011).
97. Masutomi, K., Barascud, N., Kashino, M., McDermott, J. H. & Chait, M. Sound segregation via embedded repetition is robust to inattention. *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 386–400 (2016).
98. Mehta, A. H., Feng, L. & Oxenham, A. J. Neural auditory contrast enhancement in humans. *Proc. Natl. Acad. Sci. USA* **118**, https://doi.org/10.1073/pnas.2024794118 (2021).
99. Gao, L. & Wang, X. Intracellular neuronal recording in awake nonhuman primates. *Nat. Protoc.* **15**, 3615–3631 (2020).
100. Slee, S. J. & Young, E. D. Sound localization cues in the marmoset monkey. *Hear. Res.* **260**, 96–108 (2010).
101. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55**, 131–141 (2007).
102. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442–446 (2003).

Acknowledgements

This research was supported by the National Institute of Health grant DC003180 to X.W. We thank A. Pistorio, and J. Estes for assistance with animal care, X. Song, X. Liu, H. Mandal and M. Osmanski for their comments on the earlier versions of the manuscript.

Author contributions

C.C., S.X., and Y.W. performed experiments, C.C. analyzed data and carried out computer modeling, and C.C. and X.W. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58034-8.

Correspondence and requests for materials should be addressed to Xiaoqin Wang.

Peer review information *Nature Communications* thanks Michael Pecka and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.