Innate and plastic mechanisms in auditory cortex for maternal behavior

Jennifer K. Schiavo[1,2,3,4], **Silvana Valtcheva**[1,2,3,4], **Chloe J. Bair-Marshall**[1,2,3,4], **Soomin C. Song**[1,2,3,4], **Kathleen A. Martin**[1,2,3,4,5], **Robert C. Froemke**[1,2,3,4,5,6,*]

[1]Skirball Institute for Biomolecular Medicine, New York University School of Medicine, New York, NY, USA

[2]Neuroscience Institute, New York University School of Medicine, New York, NY, USA

[3]Department of Otolaryngology, New York University School of Medicine, New York, NY, USA

[4]Department of Neuroscience and Physiology, New York University School of Medicine, New York, NY, USA

[5]Center for Neural Science, New York University, New York, NY, USA

[6]Howard Hughes Medical Institute Faculty Scholar

Abstract

Infant cries evoke powerful responses in parents[1–4]. To what extent are parental animals intrinsically sensitive to neonatal vocalizations, or might instead learn about vocal cues for parenting responses? In mice, pup-naive virgins do not recognize the meaning of pup distress calls, but retrieve isolated pups to the nest following cohousing with a mother and litter[5–9]. Distress calls are variable, requiring co-caring virgins to generalize across calls for reliable retrieval[10,11]. Here we show that the onset of maternal behavior in mice results from interactions between intrinsic mechanisms and experience-dependent plasticity in auditory cortex. In maternal females, calls with inter-syllable intervals (ISIs) from 75:375 ms elicited pup retrieval, and cortical responses generalized across these ISIs. In contrast, naive virgins were behaviorally sensitive only to the most common ('prototypical') ISIs. Inhibitory and excitatory neural responses were initially mismatched in naive cortex, with untuned inhibition and overly-narrow excitation. During cohousing, excitatory responses broadened to represent a wider range of ISIs, while inhibitory tuning sharpened to form a perceptual boundary. We presented synthetic calls during cohousing and observed that neurobehavioral responses adjusted to match these statistics, a process requiring cortical activity and the hypothalamic oxytocin system. Neuroplastic mechanisms therefore build

on an intrinsic sensitivity in mouse auditory cortex, enabling rapid plasticity for reliable parenting behavior.

Parents must quickly respond to neonatal vocalizations signaling physiological needs[1–4]. However, as in all vocal communication, appropriate responses are particularly difficult when cries signaling the same need are variable within and across individuals[2,4,12–14]. This requires caretakers to generalize across acoustic features for reliable responses to cries signaling hunger or discomfort. While vocal perception is typically learned[15], aspects of parental care might also be hard-wired or biased from pre-parental experience given biological imperatives to ensure offspring survival. Here, we take advantage of an experience-dependent maternal behavior in mice to examine the extent to which auditory cortex is intrinsically tuned to vocal features prior to parental experience, and what neuroplastic mechanisms underlie auditory learning for maternal behavior. Mouse mothers ('dams') retrieve isolated pups back to the nest based on distress calls emitted from lost pups[5–10]. These calls contain spectro-temporal qualities distinct from adult and other pup vocalizations (e.g., wriggling calls), and are organized into temporally-modulated bouts around 3–8 Hz with syllables in the ultrasonic range (~50–80 kHz)[10,11,16,17]. Distress calls vary within and across pups, requiring dams to generalize across feature variability for reliable retrieval independent of vocal or environmental distortions. While pup-naive virgins ('NV') do not behaviorally respond to distress calls, virgins begin to retrieve pups following cohousing with a dam and litter[5–9]. The emergence of alloparenting therefore allows us to dissect the innate (intrinsic) vs. learned components of pup call recognition for the onset of maternal behavior.

The temporal structure of pup vocalizations, such as syllable repetition rate, is category-informative and vital in eliciting maternal care[10,16,18,19]. We measured the probability of observing various inter-syllable intervals (ISIs) from distress calls, and selected calls containing ISIs around the median to serve as 'prototypes' (175±25 ms, Fig. 1a). To generate a library of calls, prototypes were morphed by adding/subtracting time between syllables, ensuring prototypes and their morphs were matched in all other features (Fig. 1a, Extended Data Fig. 1). We then devised a behavioral assay in which we could dub over cold, anesthetized pups with prototypes or spectrally-matched, temporal morphs (Fig. 1b). Experienced females retrieved cold, silent pups on fewer trials than warm, vocalizing pups (retrieval$_{(cold)}$: 32.0% of trials, retrieval$_{(warm)}$: 80.0%, Fig. 1c; dams & experienced virgins (EV) pooled, Extended Data Fig. 2a). Cold pups dubbed over with prototypes (76.7%), as well as morphs with ISIs between 75:375 ms (73.7–81.0%), were retrieved at similar rates to warm pups. In contrast, ISIs slower than 575 ms (27.3%) and single syllables (15.0%) did not effectively elicit retrieval (Fig. 1c, Extended Data Fig. 2b). Similarly, in a y-maze in which prototypes and morphs were played from competing speakers, females approached prototypes more than competing ISIs slowed beyond 350 ms (Extended Data Fig. 2c,d). Notably, ISIs that fell within the adult range (~75 ms) elicited retrieval, indicating a combination of spectro-temporal features is required for categorization. This suggests females generalize across any ISI that naturally occurs in pup calls, since ISIs 575 ms, which essentially never occurred in pup call bouts (Fig. 1a), were ineffective in eliciting retrieval.

To determine how experience shapes pup call encoding, we performed two-photon Ca^{2+} imaging of layer 2/3 excitatory neurons in left auditory cortex of experienced and naive virgins[20] (Fig. 1d, Extended Data Fig. 3a–d). Experience-dependent temporal processing has been previously reported in mouse auditory cortex[21], and inactivating left, but not right, auditory cortex disrupts pup retrieval[8]. We observed robust responses to prototypes in ~16% of excitatory neurons regardless of maternal experience (Extended Data Fig. 3e). While prototype-responsive neurons in naive virgins exhibited reduced responses to the same call morphed in the temporal domain, prototype-responsive neurons in experienced virgins responded robustly across morphs with ISIs between 75:375 ms (example cells, Fig. 1e; example populations and virgins, Extended Data Fig. 3f–h). To compare across cells, we normalized within each neuron by comparing morph-evoked to prototype-evoked responses; higher normalized values indicated more similar responses between prototypes and morphs ('normalized ΔF/F'). Compared to naive virgins, temporal tuning was broader in experienced virgins across ISIs that reliably elicited pup retrieval (Fig. 1c,f). Interestingly, the narrow tuning in naive cortex did not extend to temporally-modulated pure tones (Extended Data Fig. 3i–m), and broad temporal tuning was left-lateralized in experienced females (Extended Data Fig. 4a–c). Temporal tuning in auditory cortex may therefore reflect the behavioral salience of pup call ISIs. Call-evoked ΔF/Fs in left auditory cortex were correlated with ISI probability (Extended Data Fig. 4d), and, qualitatively, we observed similar tuning in a lactating dam (Extended Data Fig. 4e). Broad temporal tuning in experienced cortex could enable generalization across calls for reliable retrieval, as tuning in virgins retrieving on 10–30% of trials was sharper before the onset of robust retrieval on 100% of trials (Extended Data Fig. 4f,g).

In contrast, we wondered whether narrow tuning to pup calls might reflect an intrinsic bias for prototypical calls in naive virgins. Cortical inhibition rapidly depresses with repeated stimulation[22], providing a potential mechanism for intrinsic tuning to prototypes in naive cortex. In line with this, whole-cell recordings in pup-naive auditory cortical slices revealed that IPSCs adapted out faster than EPSCs at prototypical repetition rates, while PSCs were similarly depressed at faster or slower stimulation rates (Fig. 1g). To test the hypothesis that prototypes are especially behaviorally-salient to naive virgins, we developed an operant paradigm in which virgins listened to continuous playback of prototypes, fast morphs (ISI:75 ms), or slow morphs (ISI:575 ms), and could turn off playback for the duration of a lever press (Fig. 1h). Only virgins listening to prototypes exhibited a significant increase in press duration by session 8 (p=0.03) (Fig. 1h,i, Extended Data Fig. 5). We hypothesize this could be a form of active avoidance resulting from heightened attentional/motivational processes engaged by prototypes[23]. This intrinsic neurobehavioral bias might accelerate maternal behavior onset, as pups frequently emit prototypical calls during virgin-pup interactions.

Next, we examined whether interneurons in left auditory cortex were also differentially tuned to pup call ISIs. While interneurons were broadly tuned regardless of experience (examples, Fig. 2a,b; summary, Fig. 2c, Extended Data Fig. 6a,b), tuning curve slopes at the transition between ISIs eliciting retrieval (75:375 ms, Fig. 1c) and ignored ISIs (575:975 ms) were significantly sharper in experienced virgins (p=0.03, Fig. 2d). Excitatory and inhibitory neuronal tuning were also matched in experienced virgins, responding similarly across

behaviorally-salient ISIs and with negative slopes between 375:975 ms (Fig. 2e, Extended Data Fig. 6c,d). We hypothesize that a sharper slope in tuning could enhance discrimination between salient and non-salient ISIs[24]. In contrast, while interneurons in naive cortex were more broadly tuned than excitatory neurons, the lack of a sharp boundary suggests interneurons were still relatively untuned to relevant statistics (Fig. 2f, Extended Data Fig. 6c,d).

These data demonstrate that experience with pups adjusts the relative balance of excitatory and inhibitory tuning at the output level. To examine the synaptic basis of temporal tuning, we performed in vivo whole-cell voltage-clamp recordings from layer 2/3 (left) auditory cortical neurons in isoflurane-anesthetized females[8] (Fig. 2g). In experienced females, prototype- and morph-evoked EPSCs were similar in magnitude and negatively correlated with ISI duration (Extended Data Fig. 6e–h; dams and EVs pooled). In naive cortex, morph-evoked EPSCs were weaker than prototype-evoked EPSCs, and EPSCs were uncorrelated with ISI duration (Extended Data Fig. 6e,f). To test the hypothesis that maternal experience enhances excitatory drive, we compared EPSC magnitudes between groups. While prototype-evoked EPSCs were similar, responses to non-prototypical ISIs were enhanced in experienced females (Fig. 2h), suggesting narrowly-tuned excitatory neurons in naive cortex inherit their tuning at least partially from narrowly-tuned excitatory inputs. In contrast, IPSCs were broadly tuned in both groups (Extended Data Figure 6i–k), albeit weaker in naive virgins (Fig. 2i).

We hypothesized that inhibition in experienced cortex could define the pup call boundary, and calculated the difference in excitatory-inhibitory magnitudes for each cell ('IPSC-EPSC ratio': $\frac{IPSC-EPSC}{IPSC+EPSC}$). In experienced females, this ratio was positively correlated with ISI duration (Pearson's r=0.47, p=0.02, Fig. 2j) and 575-evoked IPSCs were significantly stronger than EPSCs (p=0.03, Fig. 2k; Extended Data Fig. 6l), indicating that E:I balance shifted towards inhibition for slower ISIs. Thus, while interneuron tuning sharpens after maternal experience (Fig. 2b–d), net postsynaptic inhibitory impact on excitatory neurons is enhanced. This allows inhibitory presynaptic cell firing to track pup call statistics, while postsynaptic neurons adjust synaptic strengths depending on their specific computations for pup call recognition.

To observe how these response profiles emerge during cohousing, we expressed GCaMP6f in either excitatory or inhibitory neurons in pup-naive left auditory cortex, and monitored neuronal populations throughout cohousing. Virgins were tested for retrieval every 12 hours and imaged at least every 24 hours (Fig. 3a). Using spatial cross-correlations to identify the same neurons over days[25] (Fig. 3b), we observed that prototype-responsive interneurons were relatively stable, but there was substantial variability in which excitatory neurons were prototype-responsive before and after retrieval onset (Extended Data Fig. 7a–c). Despite variability in single-cell dynamics (Extended Data Fig. 7d,e), we identified clear population-level changes in temporal tuning (examples, Fig. 3c,d; summary, Fig. 3e, Extended Data Fig. 8a,b). Excitatory population tuning broadened between 75:375 ms (Fig. 3e,f), elaborating on the initial preference for prototypes. Interneurons did not broaden their tuning (Fig. 3e,g), but both excitatory and inhibitory tuning selectively sharpened between 375:975 ms

(individual virgins, Fig 3h; single-cell tuning curves, Fig. 3i). These dynamics resulted in matched excitatory-inhibitory neuronal tuning to represent calls over the range of behaviorally-relevant ISIs (+24 hours, Fig. 3e). Importantly, non-cohoused controls only exposed to calls during consecutive imaging sessions showed no systematic tuning changes (Extended Data Fig. 8c,d).

Does cortical re-tuning occur over a pre-determined range of intervals, or can relevant statistics be altered by manipulating the set of calls heard during cohousing? To determine if co-carers are sensitive to exemplar statistics, we cohoused pup-naive virgins with a dam and litter with a speaker over the nest. Cohoused virgins heard vocalizations from live pups as well as slow morphs (ISI:575 ms) presented from a speaker every three hours in alternating 12-hour blocks ('CH+575', Fig. 4a). Neural and behavioral responses were assessed 24 hours after first retrieval. Compared to virgins with standard cohousing experience ('CH'), prototype-responsive neurons in virgins that heard slow exemplars were more broadly tuned to 575 ms ISIs (Fig. 4b, Extended Data Fig. 8e). Similarly, virgins in the CH+575 group retrieved cold pups dubbed over with slow morphs (ISI:575 ms; 75% of trials) more reliably than controls (14.3%, Fig. 4c; latencies, Extended Data Fig. 8f). Photoinhibiting left auditory cortex (ACtx) exclusively during slow morph playback selectively blocked the learning of slow ISIs compared to shams (CH+575$_{sham}$: 66.7%; CH+575$_{opto-ACtx}$: 18.8%) without affecting responses to prototypes (Fig. 4d) or retrieval latencies (Extended Data Fig. 8g). Co-caring virgins can therefore learn to associate slower intervals with pup distress if heard frequently enough during cohousing given the proper engagement of auditory cortical circuits. However, not all sounds were readily learned as distress calls; virgins hearing single syllables (ISI:30s) during cohousing were unable to associate these calls with isolated pups, indicative of an intrinsic bias for temporally-modulated stimuli (Extended Data Fig. 8h,i).

Finally, we wondered whether oxytocin, a neuropeptide implicated in maternal behavior and cortical plasticity[8,26,27,28], was required for cortical re-tuning during exemplar learning. Consistent with previous work[27], oxytocin enhanced spiking in response to temporally-modulated pulses in auditory cortical slices (Extended Data Fig. 9a,b). We cohoused pup-naive virgins expressing halorhodopsin (eNpHR3.0) in oxytocin neurons (Extended Data Fig. 9c) with a dam and litter, and optically suppressed oxytocin neurons during slow morph playback as in Fig. 4d. Although photoinhibition only occurred during playback, we observed deficits in retrieval rates for warm pups (CH+575$_{sham}$: 92.3% vs. CH+575$_{opto-OT}$: 43.8% of trials), as well as cold pups dubbed over with prototypes (83.3% vs. 36.8%) and slow morphs (83.3% vs. 33.3%, Fig. 4e; latencies, Extended Data Fig. 9d). These behavioral deficits were associated with uncharacteristically narrow temporal tuning in auditory cortex 24 hours after first retrieval (Fig. 4f,g, Extended Data Fig. 9e). Single-cell tuning curves in CH+575$_{opto-OT}$ virgins were significantly narrower compared to reliably retrieving virgins (CH$_{(100\%)}$: p<0.0001), yet similar to tuning curves in virgins retrieving unreliably on 10–30% of trials (CH$_{(10–30\%)}$: p>0.99, Fig. 4h), providing further evidence that broad cortical tuning may enable reliable retrieval. Taken together, these data suggest that the central oxytocin system is chronically activated throughout cohousing (possibly during virgin-pup interactions), such that even brief perturbations of oxytocinergic regulation disrupts cortical re-tuning and retrieval learning. While oxytocin disinhibits cortical networks[8], proper spike

timing in relation to pup vocalizations is also required for the induction of long-term plasticity, such as for the pairing of reliable pre- and post-synaptic activity[27] (Fig. 4d).

In summary, the onset of pup retrieval in virgin females results from interactions between an intrinsically-primed cortex and experience-dependent learning and neuromodulatory processes (Extended Data Fig. 10). It is unclear how this bias is initialized in virgin auditory cortex, but it is unlikely to result from experience with adult vocalizations which do not occur at prototypical repetition rates. We speculate that several potential priming mechanisms could contribute to this bias, such as hardwired circuits and/or early sensory experience. As recently suggested[29], synergy between innate and learned processes may be critical for fast, efficient, and flexible learning in complex environments. Given the variability of vocalizations, a mixed strategy in which the brain is sensitized to the most common, but not all possible, stimulus statistics may direct learning while leaving flexibility for circuits to wire around relevant statistics depending on individual variability in offspring and the child-rearing environment.

Methods

Ethics.

All procedures were approved under NYU Langone Institutional Animal Care and Use committee protocols.

Stimulus library and vocalization analysis.

Pup vocalizations were recorded from isolated pups at postnatal day (PND) 1–8 using an ultrasonic microphone (Avisoft Bioacoustics CM16/CMPA, sampling rate: 200 kHz). Analysis of ISIs was performed in Adobe Audition. 5,355 inter-syllable intervals were measured from 37 audio recordings (~2:50–3:30 minutes each, total time: ~111 minutes) and a lognormal frequency distribution was generated (Fig. 1a). Pup calls used in imaging and behavioral experiments were de-noised and matched in peak amplitude (Adobe Audition). Prototypical calls contained 4–5 syllables and had an average ISI of 150–200 ms (bin:175±25 ms). Prototypes were morphed in the temporal domain by either adding (+50, +100, +200, +400, +800 ms) or subtracting (−50, −100 ms) time between syllables to generate a morph set for each prototypical call. This resulted in seven morphs with the following bin centers: 75, 125, 225, 275, 375, 575, and 975 ms (bin size ±25 ms; Extended Data Fig. 1).

Behavior and cohousing.

6–48-week-old C57BL/6 virgin females (Taconic Biosciences; Jackson Laboratory) were used in all experiments, housed post-weaning in a separate acoustically-isolated room that did not contain dams or pups. Animals were raised under standard conditions (temperature: 70–74°F, humidity: 30–70%, 12-hour light/dark cycle). Lactating dams (~3–5 months old) with pups ranging from postnatal days 1–8 were used for cohousing and behavioral testing. Prior to experimentation, all virgins were tested for retrieval to establish a baseline as described previously[8]. Briefly, females were given 15–30 minutes to acclimate to a novel behavioral arena (38×30×15 cm). At least three pups were placed in the corner surrounded

by nesting material. A trial was initiated by placing one pup in the far corner. Females were given two minutes to retrieve the pup back to the nest. If virgins failed to retrieve, the pup was placed back in the nest and another trial was initiated for a total of 5–10 trials. Virgins that did not retrieve on any trials were housed with age-matched cagemates until further testing ('naive virgins'). Experienced virgins were housed with a dam and litter for at least 72 hours before further experimentation.

To test retrieval of anesthetized pups dubbed over with auditory stimuli (Fig. 1b), experienced females were housed with pups in a behavioral arena (38×30×15 cm), modified with an electrostatic speaker in the far corner and an adjacent door. Animals were given 12–24 hours to acclimate to the arena. When testing a cohoused virgin, the dam was removed prior to testing and vice versa. All testing was done under red-light conditions. To begin, 3–5 pups were removed from the nest and anesthetized on ice for 10–15 minutes to prevent vocalizations. A trial was initiated when the female remained in the nest for at least two minutes. The corner door was then opened and a warm or cold/anesthetized pup was placed in front of the adjacent speaker. On warm trials, pups vocalized as expected (verified by ultrasonic microphone (Avisoft Bioacoustics)). On cold pup trials, no stimuli were played and an ultrasonic microphone was used to ensure no vocalizations were emitted. On trials in which cold pups were dubbed over with stimuli, a series of 5–8 prototypical calls or morphs were emitted from the corner speaker (inter-bout interval: 2–3s). For single syllable trials, syllables taken from a prototypical call were played every 30 seconds. Mice were given four minutes to retrieve the warm or cold pup back to the nest. If the pup was retrieved, the latency to retrieve was recorded. If the pup was not retrieved, the trial was marked a failure and another trial was initiated upon the mouse remaining in the nest for at least two minutes. Testing continued until each animal performed 8–16 trials, or testing could not continue as the result of nursing, cannibalization, or failure to enter the nest after three hours. Trials in which cold pups began vocalizing or warm pups failed to vocalize were excluded from analysis. Data from lactating dams and experienced virgins were pooled as there was no difference between groups (Extended Data Fig. 2a)

A y-maze was used to assess approach towards pup calls in the absence of live pups (Extended Data Fig. 2c). Dams and their litters, along with cohoused virgins, were placed in a modified y-maze box (Plexiglas; 47×36×28 cm) 24 hours prior to testing. When testing a cohoused virgin, the dam was removed prior to testing and vice versa. At the start of each trial, three pups were placed in the left, right, and center chambers. After retrieval of all three pups, the speakers in the left and right chambers were simultaneously switched on to play competing pup call bouts (inter-bout interval: 1s). A prototypical call was played from one speaker, while the same call morphed in the temporal domain was played from the other. The calls played continuously until a chamber was chosen or the timeout period (two minutes) elapsed. Only the first room entrance was counted. Prototypical and morphed calls alternated left and right speakers across trials in a pseudorandom order. Room entry (% trials) was calculated by dividing the number of trials a female approached a stimulus by the total number of trials that given stimulus was played.

Operant conditioning was used to assess behavioral responses to pup calls in naive virgins (Fig. 1h,i). The operant chamber (25×20×15 cm) consisted of an electrostatic speaker and

metal lever connected to a capacitive touch sensor (SparkFun). Pup call presentation was controlled by custom-written programs in MATLAB interfacing with an RZ6 Multi-I/O processor (Tucker-Davis Technologies). Naive virgins were in one of three groups, listening to either prototypical pup calls (ISI:175±25 ms), fast morphs (ISI:75±25 ms), or slow morphs (ISI:575±25 ms). At the start of a session, virgins were placed in the arena and immediately calls or morphs began continuously playing. Stimulus playback was only paused via lever press. Brief touches resulted in a pause in stimulus presentation for a set duration, which scaled with the ISI duration to control for discrimination between the inter-syllable interval and pause across groups (pause: 15–20x ISI duration; prototypes: ~3s, fast: ~1.5s, slow: ~9s). For example, in the prototype group a brief lever press for less than 3 seconds resulted in a pause of prototype playback for 3 seconds. As press duration increased beyond this threshold, playback was paused for the entire duration of the press.

Cranial window implantation & head-posting.

For two-photon calcium imaging, cranial window implantation over left or right auditory cortex was performed as previously described[20]. Females were anesthetized with isoflurane (1.0–2.5%) and a 3 mm craniotomy was centered 1.75 mm anterior to the lambda suture. Dexamethasone (0.01–0.025 mL) was injected subcutaneously to reduce intra-cranial swelling. Adeno-associated viruses encoding GCaMP6f (0.75–1.0 μL) were injected in the center of the craniotomy at a depth of ~1000 μm using a 5 μL syringe (33-gauge needle, Hamilton). To restrict expression of GCaMP6f to excitatory neurons, we used AAV9.CamKII.GCaMP6f (UPenn Vector Core or Addgene). For expression of GCaMP6f in inhibitory interneurons, we injected rAAV.mDLX.GCaMP6f[30] (gift of Jordane Dimidschstein and Gordon Fishell) into wild type mice or injected AAV1.Syn.Flex.GCaMP6f (UPenn Vector Core or Addgene) into Gad2-IRES-Cre C57BL/6J mice (Jackson Laboratory). A 3 mm glass coverslip was secured over the craniotomy using a mixture of Krazy Glue and Acrylic resin powder (Lang Dental Manufacturing). For head fixation during in vivo calcium imaging and in vivo whole-cell recordings, custom-made headposts (Ponoko) were secured to the skull with C&B Metabond dental cement (Parkell). Animals were given 2–4 weeks for viral expression and recovery.

In vivo two-photon calcium imaging.

Two photon-calcium imaging was performed in awake, head-fixed mice as previously described[20]. All imaging experiments, except for those in Extended Data Fig. 4a–c, were performed exclusively in left auditory cortex. Animals were habituated to head-fixation 2–4 days prior to data acquisition. We used a multiphoton imaging system (Sutter Instruments) and a 900 nm Ti:Sapphire laser (MaiTai, Spectra-Physics) to obtain GCaMP6f signals in auditory cortex. Images from layer 2/3 (~300 μm^2, 256×256 pixels) were collected using ScanImage (HHMI, Vidrio Technologies) at a rate of 4 Hz (0.26 s/frames). This scanning rate did not affect our ability to detect responses to stimuli faster than 4 Hz (Extended Data Fig. 3d). Auditory stimuli were played through an electrostatic speaker (10–12 cm from the contralateral ear) connected to an RZ6 Multi-I/O processor and controlled by RPvdsEx (Tucker-Davis Technologies) interacting with ScanImage (HHMI). Half octave pure tones ranging 4–64 kHz (500 ms, 10 ms cosine ramp) at 70 dB sound pressure level (SPL) were played in a pseudorandom order (repetition rate: 0.2 Hz). For temporally-modulated pure

tones, a tone was selected based on the best frequency for the region and a series of tone pips at this frequency (70 dB SPL, 50–80 ms, 2 ms cosine ramp) were then played with the following ISIs: 75, 175, 375, and 575 ms (±25 ms).

For pup call presentation, a library of 6–9 prototypical calls (2 ms cosine ramp, ISI:175±25 ms) were played in a pseudorandom order (repetition rate: 0.1 Hz). Acquired signals were screened offline to determine the presence of call-responsive neurons, and which calls evoked responses in those neurons. Temporal morphs were then played exclusively for those prototypical calls. For chronic two-photon imaging experiments (Fig. 3), cohoused virgins were tested in the standard pup retrieval assay every 12 hours. Every other time point, retrieval testing was followed immediately by calcium imaging. Neuronal responses were always assessed at first retrieval and 24 hours after first retrieval. Retrieval onset ranged from 12–109 hours of cohousing in N=13 virgins.

Image processing and analysis.

Images were aligned frame-by-frame to a stable set of frames using TurboReg for x-y movement correction (Fiji/Image J, NIH). Regions of interest (ROIs) were manually drawn on an average image for extraction of raw calcium signals (Fiji/Image J, NIH)[20]. Cells were deemed as responsive to any of the 6–9 prototypes if the integral of the fluorescence signal during the stimulus epoch (250 ms post-call onset, 1 s in duration) significantly increased from baseline (F_0=1 s prior to stimulus onset; $p<0.05$, Student's paired one-tailed t-test) and was +1.5 standard deviations from the baseline. Neurons responsive to prototypes were then further examined for their responses to temporally-morphed calls. F/F (%) was calculated as the average change in fluorescence during the stimulus epoch:

$\Delta F/F \quad (\%) = \left(\dfrac{F_t - F_0}{F_0}\right) * 100$. For temporal morphs, frames were added or subtracted to the

stimulus epoch for each 500 ms added or subtracted to the total call duration, respectively. This window was shifted according to the peak of the signal for all stimuli on a cell-by-cell basis.

For each neuron, the similarity between a prototype- and morph-evoked response was calculated by normalizing to the prototype. Normalized F/F was calculated as:

$1 - \left| \dfrac{\Delta F/F_{prototype} - \Delta F/F_{morph}}{\Delta F/F_{prototype} + \Delta F/F_{morph}} \right|$. The slope of the tuning curve at the transition between

behaviorally-salient (75:375 ms) and non-salient ISIs (575:975 ms) was calculated by taking the slope of the line between 375 to 975 ms. Tuning width was calculated by for single-cells and populations by averaging across the normalized F/F values for behaviorally-salient ISIs (75:375 ms). In chronic two-photon imaging experiments, spatial cross-correlations were used to assess whether or not a prototype-responsive cell was present over several days, as previously described[25]. Briefly, following manual inspection, a 100×100-pixel image surrounding a given ROI was obtained for each imaging session. Spatial correlations were calculated for a given ROI across days ('within ROIs') and with other ROIs in the population ('across ROIs'). Cells with spatial correlations below the 95th percentile for the null distribution ('across ROIs') were eliminated from the correlation analysis in Extended Data Fig. 7b,c.

To assess whether neuropil contamination had any significant effect on call-evoked ∆F/F, we performed corrections on two data sets as previously described[31]. Briefly, the true cell body fluorescence was calculated as: $F_{true}(t) = F_{raw}(t) - r * F_{neuropil}(t)$, in which (t) is time, F_{raw} is the raw fluorescence averaged within an ROI, $F_{neuropil}$ is the neuropil fluorescence surrounding a given ROI, and r is a ratio of contamination $\left(\dfrac{F_{blood\_vessel}}{F_{neuropil}}\right)$. Neuropil correction had no significant effect on call-evoked ∆F/Fs in excitatory or inhibitory neurons (Extended Data Fig. 3a–c).

∆F/F for pure tone-responsive neurons was calculated similarly to pup calls. For temporally-modulated pure tones, only cells that had statistically significant responses to the best frequency and the 5 Hz tone sequence (the prototypical repetition rate) were included in further analysis (p<0.05, Student's paired one-tailed t-test). Normalized ∆F/F was calculated similarly to pup calls, such that the response evoked by each temporally-modulated sequence was normalized to the response evoked by the 5 Hz sequence (ISI:175 ms).

In vitro whole-cell recordings.

In vitro recordings were performed in acute slices of auditory cortex prepared from pup-naive C57Bl/6 wild-type mice. Animals were deeply anesthetized with 5% isoflurane and decapitated. The brain was rapidly placed in ice-cold dissection buffer containing (in mM): 87 NaCl, 75 sucrose, 2.5 KCl, 1.25 NaH_2PO_4, 0.5 $CaCl_2$, 7 $MgCl_2$, 25 $NaHCO_3$, 1.3 ascorbic acid, and 10 D-Glucose, bubbled with 95%/5% O_2/CO_2 (pH 7.4). Slices (250–300 μm thick) were prepared with a vibratome (Leica P-1000), placed in warm artificial cerebrospinal fluid (ACSF, in mM: 124 NaCl, 2.5 KCl, 1.5 $MgSO_4$, 1.25 NaH_2PO_4, 2.5 $CaCl_2$, and 26 $NaHCO_3$,) (33–35°C) for <30 min, then cooled to room temperature (22–24°C) for at least 30 minutes before use. Slices were transferred to the recording chamber and superfused (2.5–3 ml/min) with oxygenated ACSF at 33°C. Somatic whole-cell voltage-clamp or current-clamp recordings were made from layer 2/3 pyramidal cells with an Multiclamp 200B or 700B amplifier (Molecular Devices) using IR video microscopy (Olympus). Data were filtered at 2 kHz, digitized at 10 kHz, and acquired with Clampex 10.7 (Molecular Devices). Data were analyzed using custom-written Matlab code (MathWorks) and Clampfit 10.7 (Molecular Devices).

In order to assess short-term plasticity of EPSCs and IPSCs, voltage-clamp recordings were acquired while focal extracellular stimulation (0.5 ms, 3–700 μA) was applied with a bipolar glass electrode at a rate of 1 stimulation every 75, 175, or 575 ms (±25 ms). Patch pipettes (3–8 MΩ) were filled with the following intracellular solution (in mM): 130 Cs-methanesulfonate, 1 QX-314, 4 TEA-Cl, 0.5 BAPTA, 4 MgATP, 0.3 Na-GTP, 10 phosphocreatine, 10 HEPES, pH 7.2. EPSCs were acquired at −70 mv, while IPSCs were acquired from −40 to 0 mV. The peak amplitude of evoked EPSCs and IPSCs were measured and normalized to the event evoked by the first stimulation on each trial (S_n / S_1).

To assess the modulatory effects of oxytocin, whole-cell recordings were acquired in current-clamp configuration with patch pipettes (4–8 MΩ) containing the following intracellular solution (in mM): 130 Cs-methanesulfonate, 4 TEA-Cl, 10 Phosphocreatine, 0.5 EGTA, 10 HEPES, 1 QX-314, 4 MgATP, 0.3 NaGTP, pH 7.2. Cells were injected with

current to raise the membrane potential near spiking threshold while focal extracellular stimulation was applied with a bipolar glass electrode at a rate of 1 stimulation every 550–575 ms. Threshold was empirically determined for each cell as the membrane potential that resulted in evoked action potentials on 40% of trials. Once a stable baseline was established for 5–10 minutes, 1 μM oxytocin (Tocris) in ACSF was washed on for 10–15 minutes followed by a washout period. To control for continuous extracellular stimulation, the same protocol was run in the absence of oxytocin. Spike probability was calculated as the probability of evoking a spike in response to a given stimulation 15–30 minutes following wash onset (or following baseline acquisition for controls).

In vivo whole-cell recordings.

Animals were anesthetized with 1.5–2.0% isoflurane and head-fixed using custom-made stainless steel headbars (Ponoko). A small craniotomy was performed over left auditory cortex and whole-cell voltage-clamp recordings were obtained from layer 2/3 (200–400 μm from pial surface) with a Multiclamp 700B amplifier (Molecular Devices). Borosilicate glass pipettes (Sutter) with resistance 5–7 MΩ contained (in mM): 130 Cs-methanesulfonate, 1 QX-314, 4 TEA-Cl, 0.5 EGTA, 4 MgATP, 0.3 NaGTP, 10 phosphocreatine, 10 HEPES, pH 7.2. Ri: 206.20±57.09 MΩ (s.d.). Once whole-cell configuration was obtained, five prototypical calls were played via electrostatic speaker (10–12 cm from the contralateral ear) to determine the best call for each cell (i.e., the call that evoked the maximal EPSC). Prototypes were then played in conjunction with a set of temporal morphs (ISIs: 125, 225, 575 ms). Cells were held at −70 mV for excitatory currents, and between 0 to +40 mV for inhibition. Recordings were analyzed using Clampfit. Only neurons that exhibited at least five significant prototype-evoked E/IPSCs were used for further analysis. Synaptic responses were calculated as the PSC area (pA*ms) divided by the duration of the analysis window (ms); the baseline window was 1 second, and prototypical and morphed calls were analyzed from the start of the call to 250 ms following call offset. Absolute magnitudes (pA) from the largest 5–10 evoked trials were used for analysis. Evoked synaptic responses were calculated by subtracting the baseline response from the stimulus response ($PSC_{stimulus} - PSC_{baseline}$). The relative difference in the evoked IPSC and EPSC ('I-E ratio') within a neuron was calculated for each stimulus as $\left(\frac{IPSC-EPSC}{IPSC+EPSC} \right)$. Data from dams and experienced virgins were pooled as these groups had similar response profiles (Extended Data Fig. 6g,h).

Pup call exposure.

Pup-naive virgins were cohoused with a dam and litter. An electrostatic speaker was placed ~15 cm above the nest. 30–60 minutes after beginning cohousing, auditory stimuli were played every ~3–4 hours for 12-hours, interleaved with 12-hour blocks without stimuli. This was meant to mimic long blocks of silence followed by brief periods of vocalizing observed during long-term audio recordings. Virgins were tested in the standard pup retrieval assay every 12 hours. For playback of slow morphs, a set of 5–8 temporal morphs (ISI:575±25 ms) were repeated 10–20 times each in a pseudorandom order (one morph presented every 8–12 seconds). For single syllable exposure, a series of single syllables was presented with an ISI of 30 seconds. All virgins underwent two-photon calcium imaging and/or behavioral testing in the cold pup retrieval assay starting 24 hours after first retrieval.

Inhibitory optogenetics.

Surgical preparation was performed as described above (Methods: Cranial window implantation). For optogenetic inhibition of left auditory cortex, opsins (0.75–1.0 μL) were injected in the center of a small craniotomy (1.75 mm anterior to the lambda suture) at a depth of ~1000 μm using a 5 μL syringe (33-gauge needle, Hamilton). The auditory cortex was inhibited in two ways: 1) the hyperpolarizing opsin halorhodopsin (eNpHR3.0) was expressed in excitatory neurons of wild-type C57BL/6J virgins (AAV1.CaMKIIa.eNpHR.EYFP; Addgene), or 2) the depolarizing opsin channelrhodopsin-2 (AAV1.EF1a.DIO.hChR2.EYF; Addgene) was expressed in a Cre-dependent manner in interneurons using Gad2-IRES-Cre C57BL/6J mice (Jackson Laboratory). A 2 mm long optic fiber (400 μm Core, 1.25 × 6.4 mm Ceramic Ferrule, 0.39 NA; ThorLabs) was then implanted at a depth of 200–400 μM and secured to the skull with C&B Metabond dental cement (Parkell). Animals were given 2–4 weeks for viral expression and recovery, followed by habituation to the patch cable (400 μm Core, 0.39 NA, FC/PC; ThorLabs). Virgins were cohoused and exposed to slow morphs on the same schedule as described in Fig. 4a; the opsins were stimulated during morph playback to inhibit auditory cortex (halorhodopsin: 532 nm wavelength, 1–3 mW/mm^2; channelrhodopsin-2: 473 nm wavelength, 1–3 mW/mm^2). There were two sham conditions: 1) halorhodopsin-expressing excitatory neurons with no optical stimulation (AAV1.CaMKIIa.eNpHR.EYFP; Addgene; N=1) and 2) YFP-expressing excitatory neurons with optical stimulation (AAV.CamKII(1.3).eYFP; Addgene; 532 nm wavelength, 1–3 mW/mm^2; N=2).

Optogenetic inhibition of oxytocin neurons was performed similarly to cortical inhibition. Briefly, 1.0–1.5 μL of a Cre-dependent halorhodopsin (AAV5.Ef1a.DIO.eNpHR.EYFP; Addgene) was injected via Hamilton syringe in to the left paraventricular nucleus of hypothalamus (AP: −720μM, ML: +120 μM; DV: 4,500–4,750 μM) of Oxt-IRES-Cre C57BL/6J mice (Jackson Laboratory). Transgenic mice expressing halorhodopsin in oxytocin$^+$ neurons were also used (Oxt-IRES-Cre x Ai39 mice; Jackson Laboratory). In a subset of mice, GCaMP6f (AAV1.Syn.Flex.GCaMP6f; Addgene) was also injected in auditory cortex as described above (Methods: Cranial window implantation). Sham animals were injected with a Cre-dependent YFP as a control (AAV5.Ef1a.DIO.EYFP; Addgene). A 5 mm optic fiber (200 μm Core, 1.25 × 10.5 mm Ceramic Ferrule, 0.39 NA; ThorLabs) was then implanted at a depth of ~4,250–4,500 μM and secured to the skull with C&B Metabond (Parkell). During cohousing, oxytocin neurons were inhibited during call playback on the same schedule as in the cortical inhibition experiments (532 nm wavelength, 1–5 mW/mm^2).

Viral expression was confirmed using immunohistochemistry. Briefly, animals were perfused with 4% paraformaldehyde following experiments. Brains were removed and post-fixed in 4% PFA for 24 hours at 4°C, followed by immersion in 30% sucrose for 48 hours at 4°C. Brains were embedded in Optimal Cutting Temperature compound and stored at −80°C prior to sectioning. 50 μm thick slices were cut using a cryostat and stained using standard immunohistochemistry histological methods. Primary antibody (1:500): Rabbit anti-GFP (AB290, Abcam). Secondary antibody (1:1000): Goat anti-rabbit, Alexa Fluor® 488 (AB150077, Abcam).

Statistics & reproducibility.

Sample sizes for all experiments were based on sample sizes in related publications[8,20,32]. Blinding and randomization was not performed as an animal's group was determined based on their baseline behavior, and experiments required researchers to track animals over several days of behavioral experimentation. Power analysis was performed for cold pup behavioral experiments to determine sample size for statistical significance with a power of 0.8; this assay required at least n=7 trials (Fig. 1c). For whole-cell recordings *a priori* power analysis was not performed, however post-hoc analysis was used to ensure we obtained a power 0.8. For sample size reporting, the total number of single-cell tuning curves is reported since a neuron may have responded to multiple prototypes. The following comparisons are repeated as follows: 1) Behavioral data from experienced virgins in Fig. 1c also serves as the control for Fig. 4c and Extended Data Fig. 8f,h,i; 2) neuronal tuning from Fig. 3e (+24 hours) serves as the control for Fig. 4b (Note: Fig. 3e=single-cell tuning, Fig. 4b=population tuning); 3) All retrieving virgins imaged in this manuscript are used for the comparisons in Fig. 4g; 4) Data from Extended Data Fig. 4g are present in Fig. 4h. One- or two-tailed Student's t-tests were performed appropriately based on hypothesis testing. The Bonferroni method was used to correct for multiple comparisons when appropriate for all ANOVA testing. False discovery rate (FDR) was corrected for in Fisher's exact tests with multiple comparisons using the Benjamini-Hochberg method. Error bars and shading on line plots denote ±s.e.m unless otherwise stated. * represents $p<0.05$ and ** represents $p<0.01$ throughout the manuscript.

Data availability statement.

All source data are present in the accompanying excel sheet. The data that support the findings of this study are further available at figshare (https://www.nih.figshare.com) using https://doi.org/10.35092/yhjc.c.5037737 and from the corresponding author upon reasonable request.

Code availability statement.

Source code is available at figshare (https://www.nih.figshare.com) using https://doi.org/10.35092/yhjc.c.5037737.
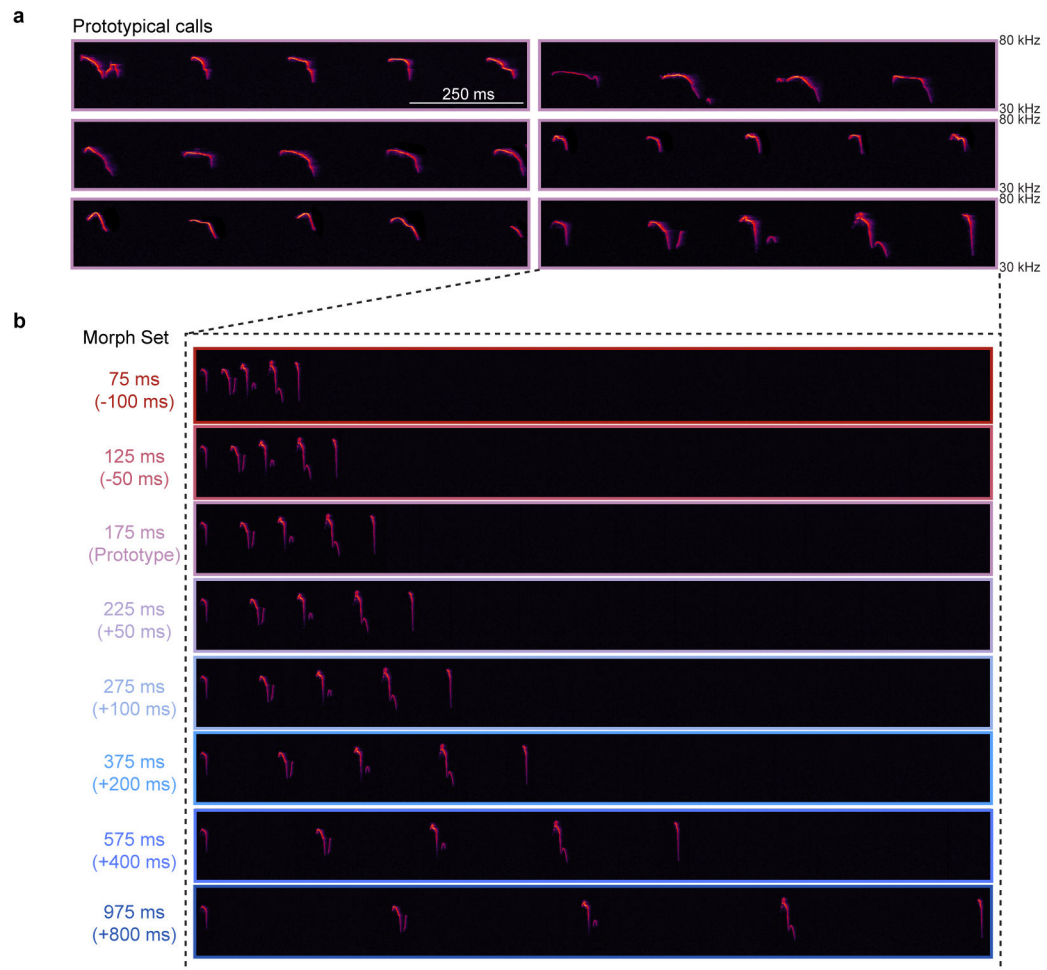
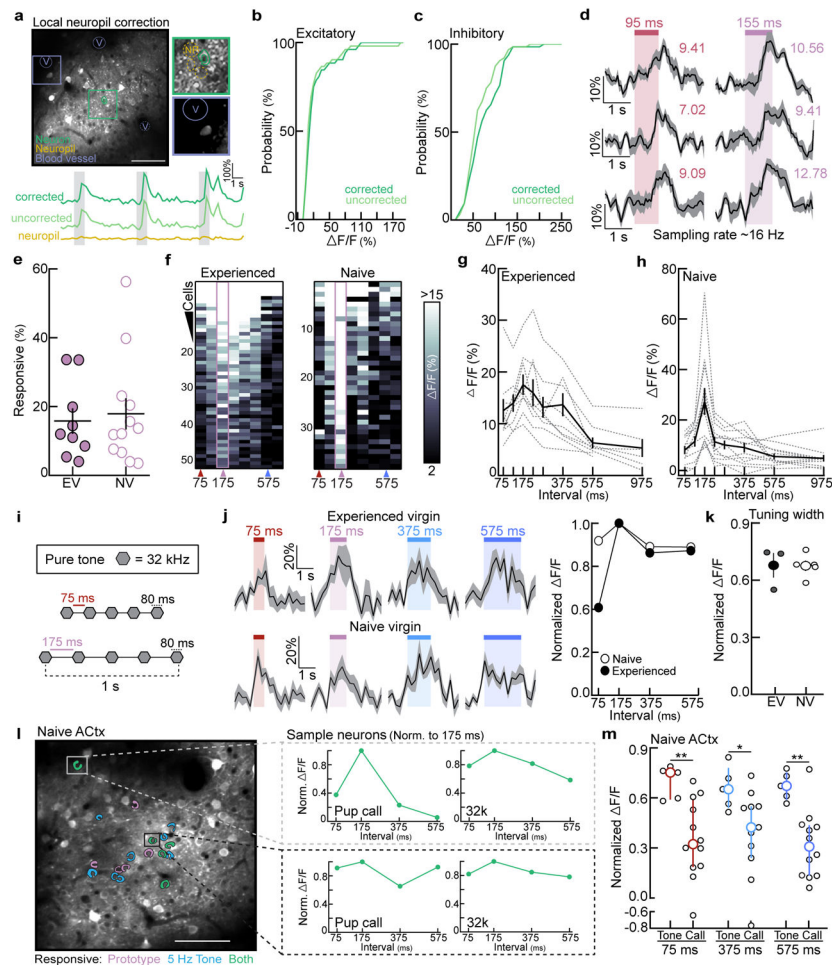## Extended Data

Extended Data Figure 1. Stimulus library of prototypical and morphed pup calls.
a, A set of six prototypical pup calls were selected from a library of pre-recorded calls. Prototypical calls had 4–5 syllables and an average ISI of 150–200 ms (bin:175±25 ms). Total duration of each prototype was ~1 s. **b,** Example of one prototypical call morphed in the temporal domain. Time was added or subtracted from the ISI to slow down or speed up the calls, respectively. Other features (e.g., frequency content) remained the same across all temporal morphs. A set of seven morphs was generated for each prototypical call (bin center: 75, 125, 225, 275, 375, 575, 975 ms; bin size: ±25 ms), resulting in a library of 42 pup call sounds. Color indicates overall speed/duration of ISIs, with red representing faster calls with shorter ISIs and blue representing slower calls with longer ISIs.
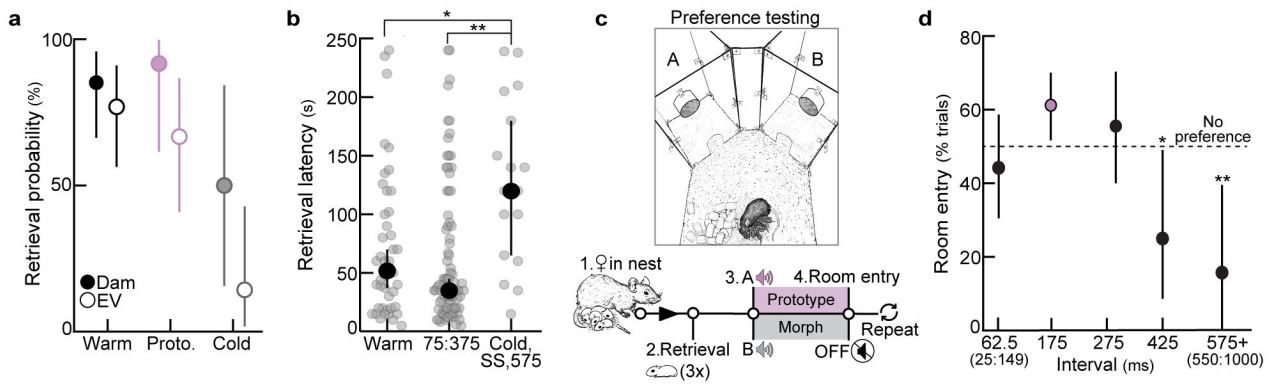
Extended Data Figure 2. Pup call ISIs drive retrieval and approach behavior in experienced females.

a, Retrieval in cold pup assay from Fig. 1c across dams and experienced virgins ('EVs'). Dams and EVs did not differ in their retrieval behavior. Warm pups: dams (85.2%, n=27 trials) vs. EVs (77.0%, n=26), p=0.50; prototypes ('proto'): dams (91.7%, n=12) vs. EVs (66.7%, n=18), p=0.19; cold pups: dams (25.0%, n=12) vs. EVs (38.5%, n=13), p=0.67 (two-tailed Fisher's test). Retrieval rate±95% binomial CIs. **b,** Latency to retrieve on successful retrieval trials from Fig. 1c. Latencies were binned based on whether ISIs elicited retrieval at similar rates to warm pups. Latencies to retrieve cold pups and cold pups dubbed over with slow morphs (single syllables ('SS') and 575 ms ISIs) were longer than for warm pups (p=0.02) or cold pups dubbed over with morphs containing ISIs between 75:375 ms (p=0.001) (Kruskal–Wallis H-test with Dunn's correction). Median±95% CIs. **c,** A y-maze was used to assess approach towards speakers playing pup calls versus temporal morphs in the absence of a live pup. A speaker in one room played a prototypical pup call, while a competing speaker in the other room played its spectrally-matched, temporal morph. Mice were given two minutes to enter a room. Data from dams and experienced virgins were pooled. **d,** When competing morphs contained ISIs between 25:149 ms (62.5 ms: 44.2%, n=52 total trials) or 201:350 ms (275 ms: 55.6%, n=45), experienced females showed no significant preference between the prototype and morph (compared to chance (0.50) using two-tailed binomial test; 62.5 ms, p=0.49; 275 ms, p=0.46). Mice showed a significant preference to approach prototypes when ISIs were slower than 350 ms. 425 ms (25.0%, n=20), p=0.04; 575+ ms (15.8%, n=19), p=0.005. Data binned ±75 ms except 575+. Retrieval rate±95% binomial CIs. Stats: *p<0.05, **p<0.01

Extended Data Figure 3. Two-photon calcium imaging of auditory cortical responses to pup calls and pure tones.

a-c, Neuropil correction on example data sets (N=1 region containing excitatory neurons; N=1 region containing inhibitory neurons). **a,** Top, correction was performed by measuring background fluorescence in the neuropil ('NP') surrounding each ROI (green) and local vasculature ('V'). Bottom, example ΔF/F traces from an excitatory neuron before and after correction (see methods). **b,c,** Neuropil correction had no significant effect on prototype-evoked ΔF/F (%) in excitatory neurons (**b,** n=52 neurons, p=0.57) or inhibitory neurons (**c,** n=64 neurons, p=0.14; two-tailed unpaired Kolmogorov-Smirnov test). **d,** Example ΔF/F traces from three neurons acquired at 16 Hz; colored number= ΔF/F (%). Responses from this experienced virgin were consistent with neuronal responses acquired at 4 Hz (Fig. 1e). **e,** Percentage of prototype-responsive excitatory neurons in experienced (N=9 mice) and naive virgins (N=12; p=0.74; two-tailed unpaired t-test). **f,** Example heatmaps of prototype-responsive neurons from an experienced (left) and naive virgin (right). **g,h,** Raw neuronal tuning from experienced (N=9) and naive virgins (N=12) summarized in Fig. 1f. **i-m,** Tuning to temporally-modulated tone sequences. **i,** Example stimulus set: five sequential tone pips (e.g., 32 kHz, 80 ms) with the following ISIs: 75, 175, 375, or 575 ms (ISI bin ±25 ms). **j,** Left, example ΔF/F traces evoked by temporally-modulated sequences of 32 kHz tones. Right, sample cell quantification. ΔF/F (%) normalized to the prototypical stimulus. **k,**

Tuning width (normalized F/F averaged across all stimuli). Experienced: N=3 mice, n=94 neurons; naive: N=4, n=45; p=0.97 (unpaired two-tailed t-test). **l,** Sample imaging region from a naive virgin. Prototypical calls and temporally-modulated 32 kHz tones with ISIs ~175 ms (5 Hz) activated a subset of the same cells (green). These neurons could have distinct temporal tuning to ISIs (inset). **m,** We observed higher normalized F/Fs to temporally-modulated tones (N=5 mice) than to pup call morphs (n=11–12) in naive auditory cortex (75 ms, p=0.009; 375 ms, p=0.04; 575 ms: p=0.004; unpaired two-tailed Mann-Whitney test). Median±interquartile. All data shown are mean±s.e.m. except **(m)**; Stats: *p<0.05, **p<0.01
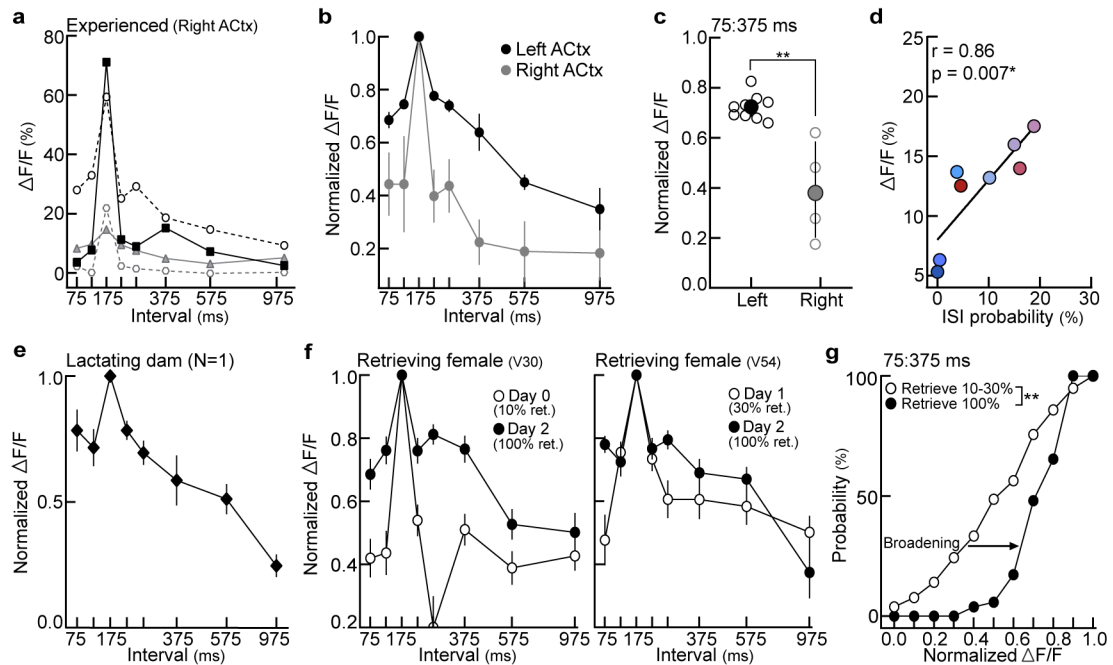
Extended Data Figure 4. Temporal tuning to pup calls in left auditory cortex reflects behavioral-salience of ISIs and retrieval probability.

a-c, Excitatory neuronal tuning in left vs. right auditory cortex of experienced virgins ('EVs'). **a,** Individual animal tuning (right auditory cortex, N=4 mice). **b,** Tuning normalized to prototypes in the left (N=9 mice from Fig. 1f) and right (N=4) auditory cortex of EVs. Mean±s.e.m. **c,** Tuning width (75:375 ms) in left vs. right auditory cortex from animals in **(b)** (p=0.003; unpaired two-tailed Mann-Whitney test). Median±interquartile. **d,** In EVs (N=9 mice), evoked F/Fs were correlated with ISI probability from the distribution in Fig. 1a (Pearson's r=0.86, p=0.007; two-tailed). Colors reflect ISI bins reported in Fig. 1a. **e,** Qualitatively, we observed broad temporal tuning in the left auditory cortex of a lactating dam (N=1 dam, n=10 single-cell tuning curves). Mean±s.e.m. **f,** Two example EVs that exhibited unreliable retrieval behavior on 10% (left) and 30% (right) of trials in a standard pup retrieval test. Temporal tuning at baseline (open circles) broadened following the onset of reliable retrieval behavior (100% of trials, closed circles). 'Days' denote days of cohousing with a dam and litter. Left, day 0: n=21 neurons, day 2: n=26. Right, day 1: n=57, day 2: n=26. Mean±s.e.m. **g,** Cumulative distribution of temporal tuning widths before (N=2 mice, n=78 neurons) and after the onset of reliable retrieval (n=52). A larger proportion of

Author ManuscriptAuthor ManuscriptAuthor Manuscript

Author Manuscript

neurons were more broadly tuned (higher normalized F/Fs) when EVs retrieved on 100% of trials (p<0.0001; two-tailed unpaired Kolmogorov-Smirnov test). Stats: *p<0.05, **p<0.01.
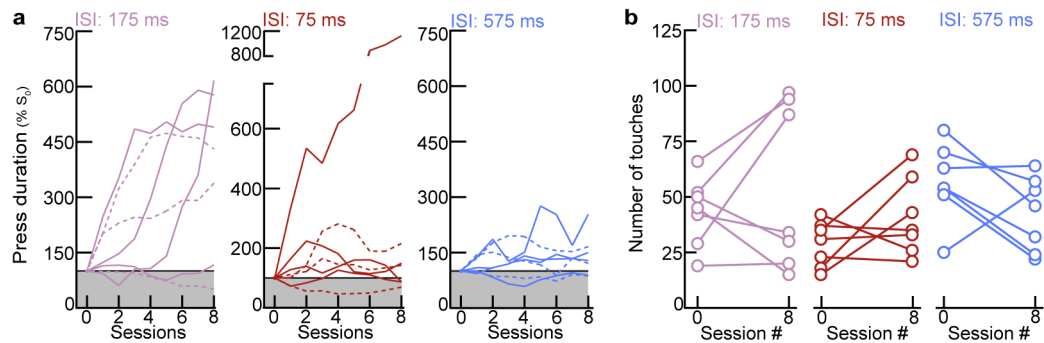
Extended Data Figure 5. Pup-naive virgins did not increase the number of times they pressed a lever to turn off prototypes or morphs.
**a,** Normalized learning trajectories in the operant test from Fig. 1h,i (N=7 mice per group). Each line denotes an individual virgin's learning curve. **b,** The total number of levers presses in a session (which turns off continuously playing prototypes or morphs) did not increase by session 8, regardless of the stimulus group (N=7 mice per group; 175 ms, p=0.44; 75 ms, p=0.19; 575 ms, p=0.14; paired two-tailed t-test).

Nature. Author manuscript; available in PMC 2021 April 07.
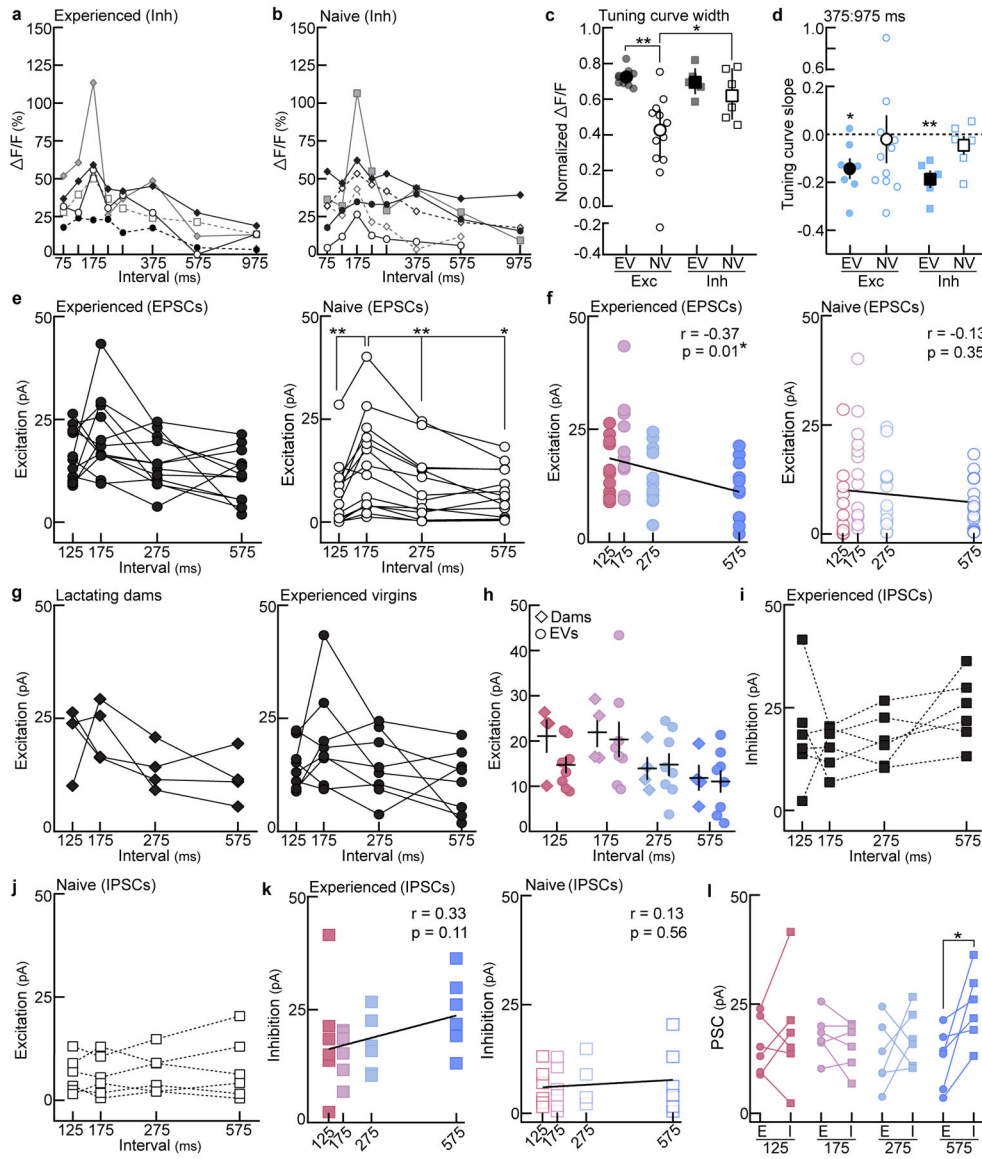
Extended Data Figure 6. Experience-dependent neuronal and synaptic temporal tuning in auditory cortex.
a, Interneuron temporal tuning curves from experienced virgins ('EVs'; N=5 mice). Each line=individual animal. **b,** Same as (a) in naive virgins ('NVs'; N=6). **c,** Tuning width across behaviorally-salient ISIs 75:375 ms (average normalized ΔF/F). Excitatory neuronal tuning curves were significantly broader in EVs (N=9 mice) than in NVs (N=12; p=0.001), and interneuron tuning was broader than excitatory tuning in naive cortex (Inh: N=6; p=0.04). Interneuron tuning width did not differ between experienced (N=5) and naive virgins (p>0.99; one-way ANOVA, Bonferroni correction). Median±interquartile. **d,** Slope of population tuning curves at the behavioral transition (375:975 ms). Slopes in EVs were significantly negative. EV$_{exc}$: N=7 mice, p=0.02; EV$_{inh}$: N=5, p=0.007; NV$_{exc}$: N=11, p=0.84; NV$_{inh}$: N=6, p=0.29 (one-sample t-test to 0.0). Mean±s.e.m. **e,** Excitatory synaptic tuning in experienced (left, n=12 cells) and naive cortex (right, n=14 cells). EV: all comparisons, p>0.05. NV: 175 vs. 125, p=0.006; 175 vs. 275, p=0.003; 175 vs. 575, p=0.02

(repeated measures one-way ANOVA, Bonferroni correction). **f,** Correlation of EPSCs with ISIs in experienced (left, n=12 cells; Pearson's r=−0.37, p=0.01) and naive cortex (right, n=14; Pearson's r=−0.13, p=0.35; two-tailed). **g,** Excitatory synaptic tuning in the auditory cortex of lactating dams (left, n=4 cells) and EVs (right, n=8) (repeated measures one-way ANOVA, Bonferroni correction). **h,** Prototype and morph-evoked EPSCs did not differ between dams (n=4 cells) and EVs (n=8) (125, p=0.11; 175, p=0.80; 275, p=0.85; 575, p=0.85; unpaired two-tailed t-test). Mean±s.e.m. **i,j,** Inhibitory synaptic tuning in experienced cortex (**i,** n=6 cells) and naive cortex (**j,** n=6 cells; repeated measures one-way ANOVA, Bonferroni correction). **k,** Correlation of IPSCs with ISIs in experienced (left, n=6 cells; Pearson's r=0.33, p=0.11) and naive cortex (right, n=6; Pearson's r=0.13, p=0.56; two-tailed). **l,** Within cell comparison of PSCs in experienced cortex (125, p=0.47; 175, p=0.26; 275, p=0.44; 575, p=0.03; two-tailed paired t-test). Stats: *p<0.05, **p<0.01.
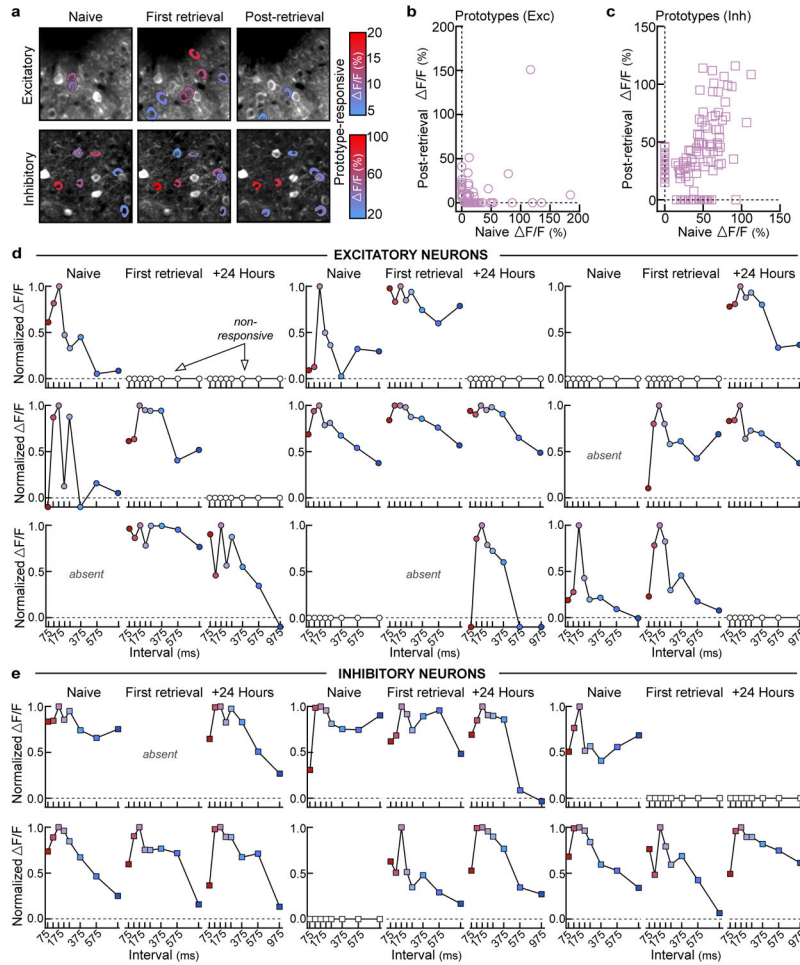
Extended Data Figure 7. Variability in prototype-responsive neurons and single-cell temporal tuning curves during cohousing.

a, Example tracking of excitatory (top) and inhibitory (bottom) neurons. Colored cell bodies denote prototype-responsive cells in that imaging session based on ΔF/F (%). **b,c,** Correlating prototype-evoked ΔF/F before (naive) and after retrieval onset (+24–96 hours). Zeros denote non-responsive cells. Of the neurons that were prototype-responsive at baseline

and successfully tracked, ~18.8% of excitatory neurons and ~72.3% of inhibitory neurons remained responsive in the final imaging session. **b,** Excitatory (n=104 neurons), Spearman's r=−0.40, p<0.0001 (two-tailed). **c,** Inhibitory (n=122 neurons), Spearman's r=0.46, p<0.0001 (two-tailed). **d,** Example excitatory neurons depicting single-cell dynamics during cohousing. Each set of three graphs is a single neuron's tuning at baseline (naive), first retrieval, and 24 hours after first retrieval (+24 hours). Open circles denote neurons that were not prototype-responsive at a given timepoint, whereas 'absent' indicates a neuron that was not present in the imaging region. Colors represent ISIs used throughout the manuscript (prototype=pink). **e,** Same as **(d)**, but for inhibitory neurons.
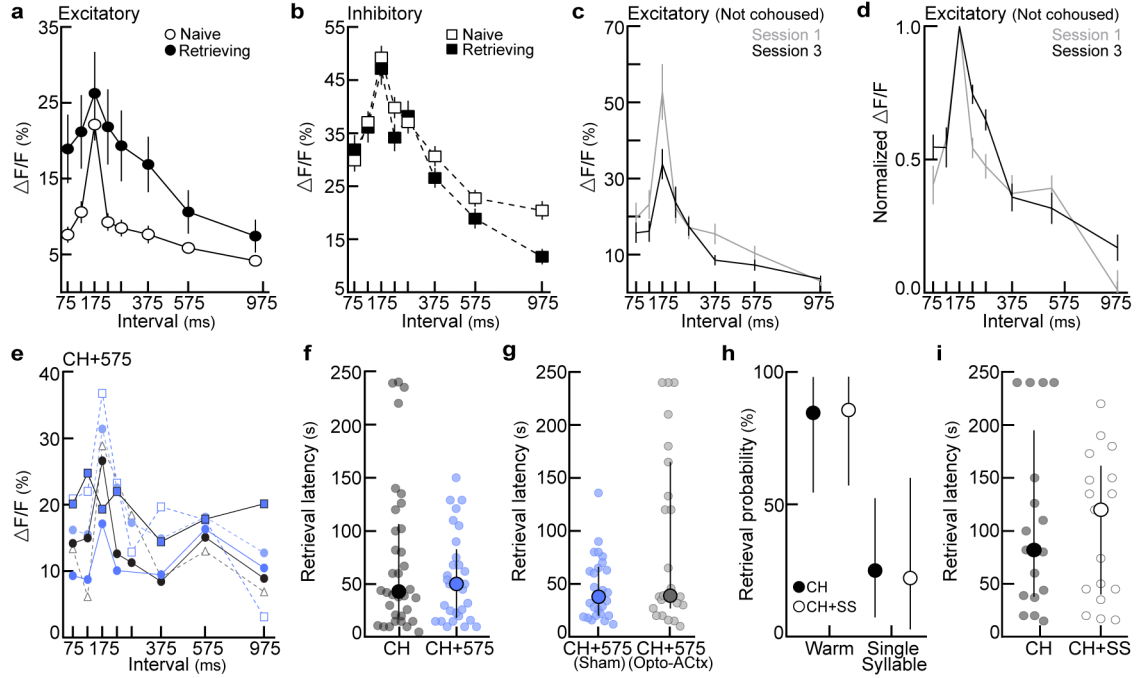
Extended Data Figure 8. Re-tuning of cortical neurons requires cohousing and reflects the statistics of pup call exemplars.
**a,** Excitatory temporal tuning before and after retrieval onset (raw data for Fig. 3e). Naive: n=165 single-cell tuning curves, retrieving: n=70. Mean±s.e.m. **b,** Same as **(a)** for interneurons. Naive: n=128, retrieving: n=86. **c,d,** To ensure that listening to pup calls while head-fixed under the two-photon microscope could not explain the broadening of excitatory tuning we observed in Fig. 3, we assessed temporal tuning to pup calls in pup-naive virgins on three consecutive imaging days without cohousing or retrieval testing. There were no systematic changes in cortical tuning in the absence of experience with pups (**c,** raw tuning; **d,** normalized tuning). Mean±s.e.m. **d,** Session 1: N=4 mice, n=53 single-cell tuning curves; session 3: N=5, n=48 (one-way ANOVA, Bonferroni correction). **e,** Raw temporal tuning from virgins in Fig. 4b. Virgins were exposed to slow (ISI:575 ms) morphs during cohousing (N=6 mice). Tuning was assessed 24 hours after retrieval onset; each colored line represents an individual animal's tuning curve. **f,** Latency to retrieve on successful trials from Fig. 4c. CH (n=34 trials) vs. CH+575 (n=29), p=0.73 (unpaired two-tailed Mann-Whitney test). Median±interquartile. **g,** Latency to retrieve on successful trials from Fig. 4d. CH+575$_{sham}$

(n=26 trials) vs. CH+575$_{opto-ACtx}$ (n=23 trials), p=0.19 (unpaired two-tailed Mann-Whitney test). Median±interquartile. **h,i,** Playback of single syllable calls during cohousing ('CH +SS') did not alter the behavioral-salience of single syllables. **h,** Retrieval rates in cold pup assay. CH (N=6 mice), CH+SS (N=4). Warm pup: CH (84.6%, n=13 trials) vs. CH+SS (85.7%, n=14), p>0.99. Single syllable: CH (22.2%, n=9) vs. CH+SS (25.0%, n=16), p>0.99 (two-tailed Fisher's exact test). Retrieval rate±95% binomial CIs. **i,** Latency to retrieve on successful trials from **(h)** (n=17 trials each; p=0.98; unpaired two-tailed Mann-Whitney test). Median±interquartile. Stats: *p<0.05, **p<0.01
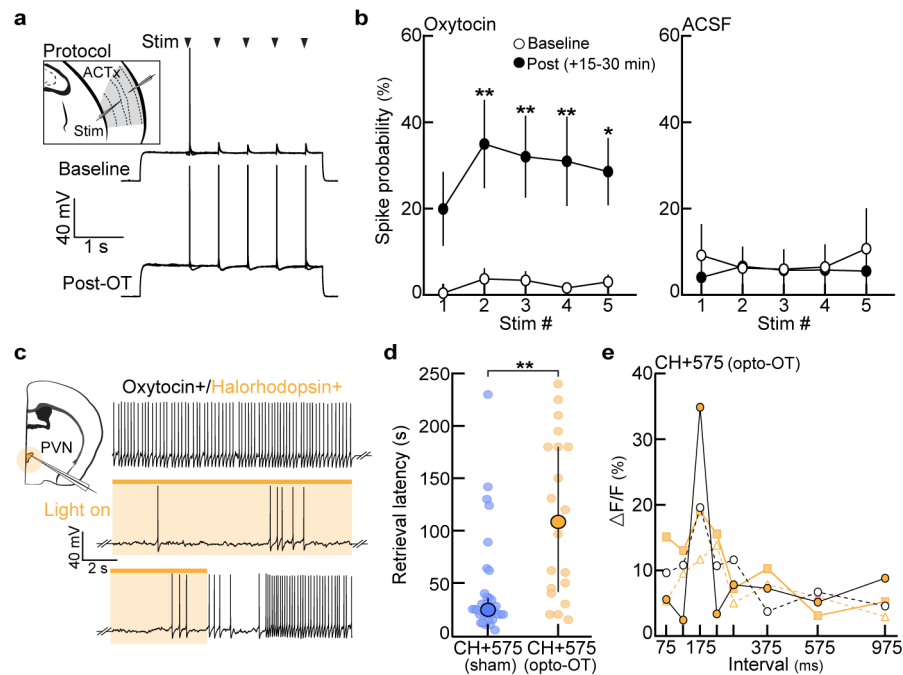
Extended Data Figure 9. Optical inhibition of OT neurons perturbs the re-tuning of auditory cortical neurons.
a, Example in vitro current-clamp recording from pup-naive auditory cortex. The probability of evoking spikes in response to 5 extracellular stimulus pulses (ISI:575 ms) was measured before and after oxytocin (1 μM) wash. **b,** Left, evoked spike probability was significantly enhanced 15–30 minutes following the onset of oxytocin wash (n=9 cells; p=0.34 (stim 1), p=0.009 (stim 2), p=0.008 (stim 3), p=0.001 (stim 4), p=0.04 (stim 5)). Right, repetitive stimulation in the absence of oxytocin (ACSF, n=7 cells) did not induce changes in spike probability (one-way Friedman test with Dunn's correction). Mean±s.e.m. **c,** Example in vitro current-clamp recording from an oxytocin neuron containing halorhodopsin (NpHR3.0). Photostimulation efficiently perturbed spiking for the duration of the light. **d,** Latency to retrieve on successful trials from Fig. 4e (CH+575$_{sham}$ (n=32 trials) vs. CH +575$_{opto-OT}$ (n=20), p=0.0003; unpaired two-tailed Mann-Whitney test). Median ±interquartile. **e,** Temporal tuning curves from CH+575$_{opto-OT}$ virgins in Fig. 4f (N=4 mice). Each colored line represents the tuning curve for one virgin.
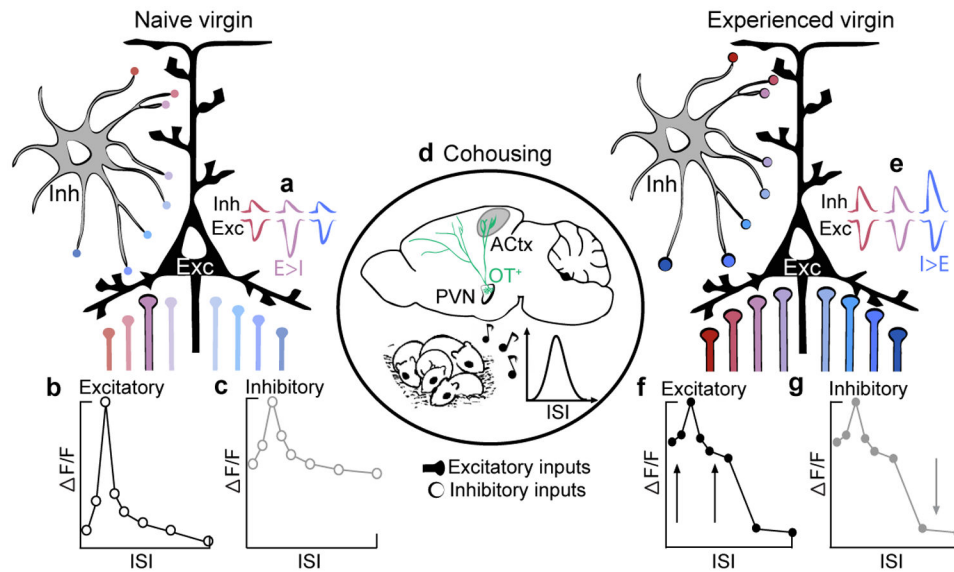
Extended Data Figure 10. Intrinsic tuning in auditory cortex acts as a scaffold for experience-dependent plasticity during cohousing.

a,e, synaptic tuning; **b,c,f,g,** neuronal tuning. Colors denote ISIs used throughout the manuscript (red=fast, pink=prototypes, blue=slow). **a,b,** In pup-naive auditory cortex, excitatory neurons respond robustly to prototypical calls **(a)** as a result of sharply tuned excitatory drive and weak, untuned synaptic inhibition **(b)**. Intrinsic tuning might result from hardwired cortical circuits or developmental experiences. **c,** Interneurons exhibit broad, unselective tuning characteristic of inhibitory populations that pool local activity. **d,** Oxytocin release during cohousing, possibly stimulated by virgin-pup interactions, may serve to transiently decrease intracortical inhibition[8]. This could enable the re-balancing of excitatory-inhibitory inputs: excitatory neurons broaden as excitatory drive increases across all ISIs. While inhibitory output tuning sharpens, net postsynaptic inhibitory drive is enhanced to (1) balance the increase in excitation across salient ISIs[8] and (2) sharpen tuning to slow ISIs (**e,** I>E for slow ISIs). As a result of oxytocin receptor lateralization[8], left auditory cortex may be particularly sensitive to exemplars frequently heard during cohousing. Whereas oxytocin may disinhibit the network, proper spike timing in relation to pup vocalizations is also required for the pairing of reliable pre- and post-synaptic activity[27] (Fig. 4d). **e-g,** In experienced virgins, excitatory and inhibitory neurons are broadly tuned to behaviorally-salient ISIs, which reflect exemplar statistics, to enable generalization and reliable pup retrieval.

Acknowledgements.

References

1. Swain JE, Kim P, Ho SS Neuroendocrinology of parental response to baby-cry. J. Neuroendocrinol 23 (11), 1036–1041 (2011). [PubMed: 21848646]

2. Lingle S, Wyman MT, Kotrba R, Teichroeb LJ, Romanow CA What makes a cry a cry? A review of infant distress vocalizations. Curr. Zool 58 (1), 698–726 (2012).

3. Dulac C, O'Connell LA, Wu Z Neural control of maternal and paternal behaviors. Science 345 (6198), 765–770 (2014). [PubMed: 25124430]

4. Zeskind PS A Developmental Perspective of Infant Crying in Lester BM, Zachariah Boukydis CF, Infant Crying: Theoretical and Research Perspectives. Springer, New York (1985).

5. Ehret G, Koch M, Haack B, Markl H Sex and parental experience determine the onset of an instinctive behavior in mice. Naturwissenschaften 74, 47 (1987). [PubMed: 3561521]

6. Koch M & Ehret G Estradiol and parental experience, but not prolactin, are necessary for ultrasound recognition and pup retrieving in the mouse. Physiol. Behav 45 (4), 771–6 (1989). [PubMed: 2780846]

7. Elyada YM & Mizrahi A Becoming a mother — circuit plasticity underlying maternal behavior. Curr. Opin. Neurobiol 35, 49–56 (2015). [PubMed: 26143475]

8. Marlin BJ, Mitre M, D'amour JA, Chao MV, Froemke RC Oxytocin enables maternal behavior by balancing cortical inhibition. Nature 520 (7548), 499–504 (2015). [PubMed: 25874674]

9. Noirot E The onset of maternal behavior in rats, hamsters, and mice: A selective review In Lehrman DS, Hinde RA, & Shaw E, Advances in the study of behavior: IV. Academic Press, New York (1972).

10. Ehret G Infant rodent ultrasounds – A gate to the understanding of sound communication. Behav. Genet 35, 19–29 (2005). [PubMed: 15674530]

11. Liu RC, Miller KD, Merzenich MM, Schreiner CE Acoustic variability and distinguishability among mouse ultrasound vocalizations. J. Acoust. Soc. Am 114, 3412–22 (2003). [PubMed: 14714820]

12. Lindová J, Špinka M, Nováková L Decoding of baby calls: Can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? PLoS One 10 (2015).

13. Weatherholtz K & Jaeger TF Speech Perception and Generalization Across Talkers and Accents. Oxford Research Encyclopedia of Linguistics (2016).

14. Holt LL & Lotto AJ Speech perception as categorization. Atten. Percept. Psycho 72 (5), 1218–1227 (2010).

15. Petkov CI & Jarvis ED Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Front. Evol. Neurosci 4 (2012).

16. Castellucci GA, Calbick D, McCormick D The temporal organization of mouse ultrasonic vocalizations. PLoS One 13 (2018).

17. Ehret G & Bernecker C Low-frequency sound communication by mouse pups (Mus musculus): wriggling calls release maternal behavior. Animal Behav. 34 (3), 821–830 (1986).

18. Uematsu et al. Maternal approaches to pup ultrasonic vocalizations produced by a nanocrystalline silicon thermo-acoustic emitter. Brain Res. 1163, 91–9 (2007). [PubMed: 17659265]

19. Gaub S & Ehret G Grouping in auditory temporal perception and vocal production is mutually adapted: the case of wriggling calls of mice. J. Comp. Physiol 191, 1131–1135 (2005). [PubMed: 16075266]

20. Kuchibhotla KV et al. Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci 20 (1), 62–71 (2017). [PubMed: 27798631]

21. Liu RC, Linden JF, Schreiner CE Improved cortical entrainment to infant communication calls in mothers compared with virgin mice. Eur. J. Neurosci 23 (11), 3087–97 (2006). [PubMed: 16819999]

22. Metherate R & Ashe JH Facilitation of an NMDA Receptor-Mediated EPSP by Paired-Pulse Stimulation in Rat Neocortex via Depression of GABAergic IPSPs. J. Physiol 481, 331–348 (1994). [PubMed: 7738829]

23. Jean-Richard-Dit-Bressel P, Killcross S, McNally GP. Behavioral and neurobiological mechanisms of punishment: implications for psychiatric disorders. Neuropsychopharmacology 43 (8), 1639–1650 (2018). [PubMed: 29703994]

24. Butts DA & Goldman MS Tuning curves, neuronal variability, and sensory coding. PLoS Biol. 4 (2006).

25. Katlowitz KA, Picardo MA, Long MA Stable Sequential Activity Underlying the Maintenance of a Precisely Executed Skilled Behavior. Neuron 98 (6), 1133–1140 (2018). [PubMed: 29861283]

26. Valtcheva S & Froemke RC Neuromodulation of maternal circuits by oxytocin. Cell Tissue Res. 375 (1), 57–68 (2019). [PubMed: 30062614]

27. Mitre M et al. A Distributed Network for Social Cognition Enriched for Oxytocin Receptors. J. Neurosci 36 (8), 2517–2535 (2016). [PubMed: 26911697]

28. Pekarek BT, Hunt PJ, Arenkiel BR Oxytocin and Sensory Network Plasticity. Front. Neurosci 14 (2020).

29. Zador AM A critique of pure learning and what artificial neural networks can learn from animal brains. Nat. Commun 10 (2019).

30. Dimidschstein J et al. A viral strategy for targeting and manipulating interneurons across vertebrate species. Nat. Neurosci 19, 1743–1749 (2016). [PubMed: 27798629]

31. Kerlin AM, Andermann ML, Berezovskii VK, Reid RC Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. Neuron 67, 858–871 (2012).

32. Tasaka G et al. The Temporal Association Cortex Plays a Key Role in Auditory-Driven Maternal Plasticity. Neuron, in press (2020).
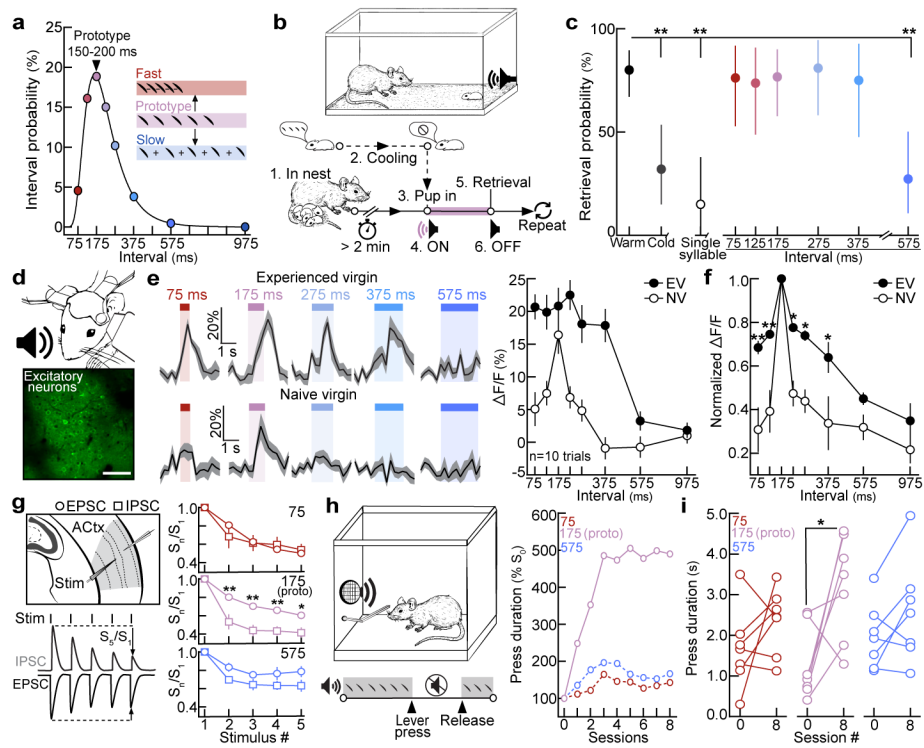
Figure 1. Temporal statistics drive behavioral and cortical responses to pup calls in naive and experienced females.

a, ISI distribution (n=5,355). Circles=bin centers (±25 ms). **b,** Protocol for dubbing over cold pups. **c,** Retrieval probabilities from assay in (**b**). Warm (n=55 trials) vs. cold (n=25), single (n=20), 575 (n=22): p=0.0003 (two-tailed Fisher's test, FDR correction; dams and EVs pooled). Retrieval rate±95% binomial CIs. **d,** In vivo two-photon Ca^{2+} imaging. Scale, 75 μm. **e,** Example F/F traces and quantification. **f,** Excitatory neuronal tuning normalized to prototypes (EV: N=9 mice, NV: N=12; 375, p=0.050; one-way ANOVA, Bonferroni correction). **g,** In vitro whole-cell recordings in virgin auditory cortex. (75, n=10 cells; 175, n=14; 575, n=13; one-way ANOVA, Bonferroni correction). **h**, Left, operant paradigm: virgins turn off pup calls for duration of lever press. Right, example normalized learning trajectories (N=3 mice). **i,** Mean lever press duration (N=7 virgins per group; 75, p=0.24; 175, p=0.03; 575, p=0.13; paired two-tailed t-test). Data shown are mean±s.e.m. except (**c**); *p<0.05, **p<0.01.
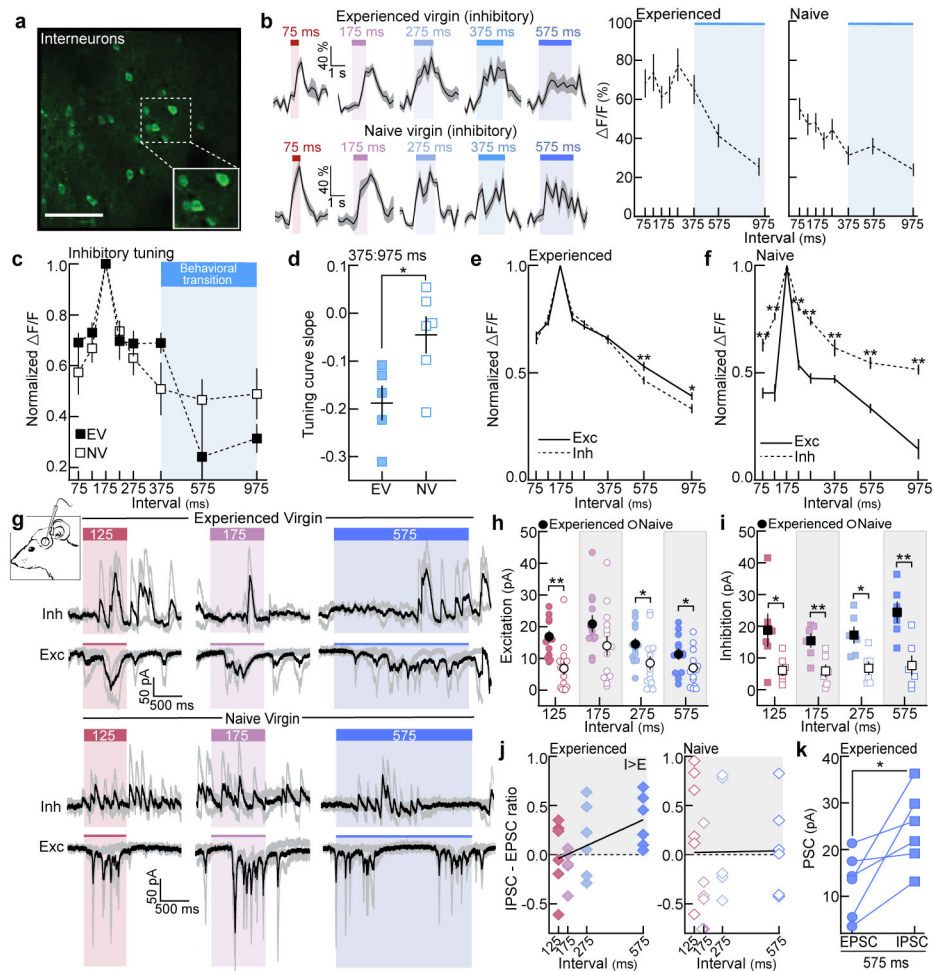
Figure 2. Excitatory and inhibitory tuning and synaptic responses are altered by maternal experience.

a, GCaMP6f in interneurons. Scale, 75 μm. **b**, Example F/F traces and quantification. Blue shading, behavioral transition from Fig. 1c. **c**, Normalized inhibitory temporal tuning (EV: N=5 mice, NV: N=6). **d,** Tuning curve slopes from **(c)** (375:975 ms, p=0.03; unpaired two-tailed t-test). **e,** Temporal tuning across neurons in EVs (exc: n=366 single-cell tuning curves, inh: n=260; 575 ms, p=0.009; 975 ms, p=0.048). **f,** Same as **(e)** for naive cortex (exc: n=268, inh: n=128). **g**, In vivo voltage-clamp recordings from auditory cortical neurons. **h,** Prototype- and morph-evoked EPSCs in experienced (n=12 cells) and naive cortex (n=14) (125, p=0.001; 275, p=0.03; 575, p=0.04; unpaired one-tailed t-test). **i,** Prototype- and morph-evoked IPSCs (n=6 cells each; 125, p=0.047; 175, p=0.009; 275, p=0.01; 575, p=0.004; unpaired two-tailed t-test). **j,** Correlation of IPSC-EPSC ratio with ISIs in experienced (left, n=6 cells; Pearson's r=0.47, p=0.02) and naive cortex (right, n=6; Pearson's r=0.01, p=0.96; two-tailed). **k,** Within-cell comparison of 575-evoked PSCs in experienced cortex (n=6 cell; p=0.03; paired two-tailed t-test). Data shown are mean±s.e.m.; Stats: one-way ANOVA, Bonferroni correction for **c,e,f**; *p<0.05, **p<0.01.
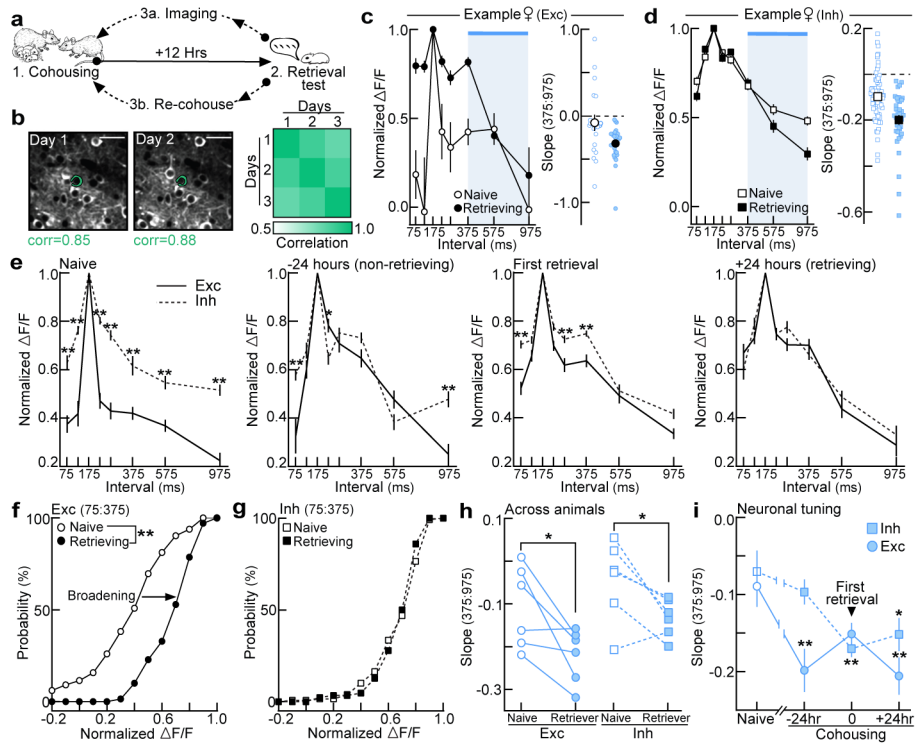
Figure 3. Cohousing with pups results in coordinated plasticity of excitatory and inhibitory neuronal tuning.

a, Two-photon imaging of left auditory cortex during cohousing. **b,** Spatial cross-correlations of pixels surrounding ROIs for single-cell identification. **c,d,** Example tuning curves in virgins before/after cohousing. Blue shading=behavioral transition (**c,** excitatory, naive: n=19 single-cell tuning curves, retrieving: n=24; **d,** interneurons, naive: n=64, retrieving: n=37). **e,** Excitatory (n=165) and inhibitory (n=128) single-cell tuning was initially mismatched, but became similar 24 hours after first retrieval. −24 hours: n=50 (exc), n=62 (inh); first retrieval: n=112, n=108; +24 hours: n=70, n=86 (one-way ANOVA, Bonferroni correction). Data aligned to first retrieval, binned ±12 hours. **f,g,** Cumulative distribution of tuning widths (normalized ΔF/F 75:375) from **(e)** before and after retrieval onset (**f**, excitatory: p<0.0001; **g,** interneurons: p=0.16; unpaired two-tailed Kolmogorov-Smirnov test). **h,** Population tuning curve slopes (375:975) before cohousing and after virgins were successfully retrieving (N=6 mice each; exc, p=0.02; inh, p=0.02; one-tailed paired t-test). **i,** Single-cell tuning curve slopes (exc: n=50–129; inh: n=49–121; compared to baseline using one-way ANOVA, Bonferroni correction). Data shown are mean±s.e.m.; *p<0.05; **p<0.01.
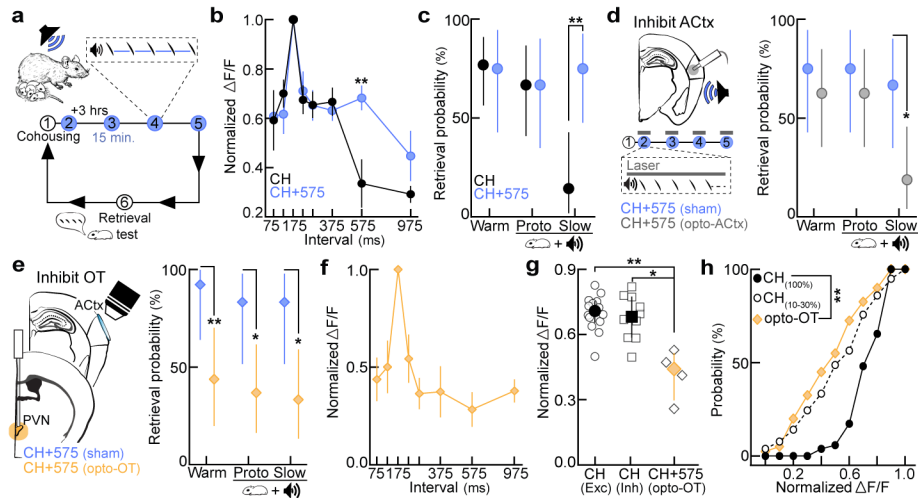
Figure 4. Auditory cortex and the oxytocinergic system are required for the re-tuning of cortical neurons during cohousing.

a, Playback of slow morphs (ISI:575 ms) during cohousing (CH+575; step 1=cohousing). **b,** Excitatory tuning in CH+575 (N=6) and CH (cohoused) virgins (N=5) (575, p=0.007; one-way ANOVA, Bonferroni correction). Mean±s.e.m. **c,** Retrieval in cold pup assay. CH (N=9–17 mice, n=14–26 trials), CH+575 (N=8–9, n=12–16); slow, p=0.001. **d,** Left, ACtx photoinhibition during morph playback on schedule in **(a)**. Right, retrieval in cold pup assay. CH+575$_{sham}$ (N=3 mice, n=12 trials each), CH+575$_{opto-ACtx}$ (N=4, n=16); slow, p=0.02. **e,** Left, photoinhibition of oxytocin neurons on schedule in **(a,d)**. Right, retrieval in cold pup assay. CH+575$_{sham}$ (N=3 mice, n=12–13 trials), CH+575$_{opto-OT}$ (N=5, n=16–19); warm, p=0.008; prototypes, p=0.02; slow, p=0.01. **f,** Temporal tuning in CH+575$_{opto-OT}$ virgins (N=4 mice; pan-neuronal GCaMP6f). Mean±s.e.m. **g,** Tuning curve widths (75:375 ms). CH +575$_{opto-OT}$ virgins from **(f)** vs. CH$_{exc}$ (N=14 mice, p=0.008) and CH$_{inh}$ (N=10, p=0.03). Median±interquartile. **h,** Cumulative distribution of tuning widths from CH+575$_{opto-OT}$ virgins (N=4 mice, n=40 tuning curves) vs. virgins retrieving on 100% (CH$_{100\%}$, N=2, n=52; p<0.0001) or 10–30% of trials (CH$_{10-30\%}$, N=2, n=78; p>0.99). Statistics: **c,d,e,** two-tailed Fisher's test, retrieval rate±95% binomial CIs; **g,h,** Kruskal-Wallis test with Dunn's correction; *p<0.05; **p<0.01.