Primary Auditory Cortex is Required for Anticipatory Motor Response

Jingcheng Li1,2, Xiang Liao1, Jianxiong Zhang1, Meng Wang1,3, Nian Yang2, Jun Zhang2, Guanghui Lv3, Haohong Li3, Jian Lu1, Ran Ding1, Xingyi Li1, Yu Guang1, Zhiqi Yang1, Han Qin1, Wenjun Jin1, Kuan Zhang1, Chao He2, Hongbo Jia4, Shaoqun Zeng3, Zhian Hu2, Israel Nelken5 and Xiaowei Chen1,6

1Brain Research Center and State Key Laboratory of Trauma, Burns, and Combined Injury, Third Military Medical University, Chongqing, China, 2Department of Physiology, Third Military Medical University, Chongqing, China, 3Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei, China, 4Brain Research Instrument Innovation Center, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, Jiangsu, China, 5Department of Neurobiology, Silberman Institute of Life Sciences and the Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem, Israel and 6CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Address correspondence to Xiaowei Chen and Xiang Liao, Brain Research Center, Third Military Medical University, Chongqing 400038, China. Email: xiaowei.chen@tmmu.edu.cn (X.C.); xiang.liao@aliyun.com (X.L.)

Abstract

The ability of the brain to predict future events based on the pattern of recent sensory experience is critical for guiding animal’s behavior. Neocortical circuits for ongoing processing of sensory stimuli are extensively studied, but their contributions to the anticipation of upcoming sensory stimuli remain less understood. We, therefore, used in vivo cellular imaging and fiber photometry to record mouse primary auditory cortex to elucidate its role in processing anticipated stimulation. We found neuronal ensembles in layers 2/3, 4, and 5 which were activated in relationship to anticipated sound events following rhythmic stimulation. These neuronal activities correlated with the occurrence of anticipatory motor responses in an auditory learning task. Optogenetic manipulation experiments revealed an essential role of such neuronal activities in producing the anticipatory behavior. These results strongly suggest that the neural circuits of primary sensory cortex are critical for coding predictive information and transforming it into anticipatory motor behavior.

Key words: anticipatory motor response, predictive coding, primary auditory cortex, rhythmic sound stimulation, 2-photon imaging
Introduction

The brain is capable of predicting expected events based on the recent history of sensory stimuli, a process critical for cognition, decision-making, and eventually for guiding ongoing and imminent actions (Sumbre et al. 2008; Nelken 2012; Summerfield and de Lange 2014). For example, following rhythmic sensory stimulation that normally triggers a movement, animals often “over-react” after the cessation of simulation by moving once or more at the expected time of the next stimulus (Rao et al. 1997; Sumbre et al. 2008). Impaired ability to anticipate stimuli in the immediate future, as indicated by both brain activity measurements and neuropsychological performance, is often found in psychiatric disorders, including, for example, schizophrenia and anxiety (Baldeweg et al. 2004; Gao et al. 2013). However, almost nothing is known about the underlying neural circuitry and the brain regions that are required for both coding expectation and generating anticipatory motor behavior. Such information may be relevant for understanding circuit malfunction in these disorders (Buhusi and Meck 2005).

Neural responses to expected events following repetitive sensory stimuli have been previously observed in different sensory systems of many species, from fish to rodents and to humans (Bullock et al. 1994; Schwartz et al. 2007; Sumbre et al. 2008; Gao et al. 2009). For example, in the visual system, these expected responses, also termed omitted stimulus potentials, can be detected as early as in retina and also occur in the downstream brain regions following rhythmic sensory stimulation (Bullock et al. 1994; Schwartz et al. 2007; Sumbre et al. 2008). In the auditory system, omitted stimulus-evoked responses have been detected in guinea pig thalamus (Gao et al. 2009) and in human and monkey cortex (Busse and Woldorff 2003; Jaaskelainen et al. 2004; Halassa et al. 2013; Andreou et al. 2015). Although these neural correlates of expectation have been found broadly in multiple sensory areas in mammals, it is unclear whether any of these regions are key elements in the brain circuits relevant for transforming the expected neuronal responses into anticipatory behavioral consequences; alternatively, the activities in these regions could be epiphenomenal, without direct behavioral consequences.

Previous studies have extensively investigated how sensory cortical circuits code ongoing sensory inputs in mammals. These studies revealed a general scheme by which sensory information is transmitted by the main recipients of thalamic inputs, layer 4 (L4) neurons, into the neurons in the upper cortical layers, L2/3. These, in turn, project to the cortical outputs, L5 neurons, as well as to other cortical areas (Hubel and Wiesel 1979; Gilbert 1983; Callaway 1998; Douglas and Martin 2004). The coding of expected sensory events in sensory cortex at the level of cellular resolution remains unknown.

To address these questions, we studied the coding of expected auditory events in L2/3, L4, and L5 of mouse primary auditory cortex (Au1). We applied a variety of approaches including in vivo 2-photon cellular Ca2+ imaging, fiber photometry, and optogenetic gain- and loss-of-function manipulations in combination with a simple auditory associative learning task. Our results provide the first demonstration of the necessity of mammalian Au1 in the brain networks for generating anticipatory behavioral response following rhythmic auditory stimulation.

Materials and Methods

Animals

C57/BL6j male mice (2–3 months old) were provided by the Laboratory Animal Center at the Third Military Medical University. Thy1-ChR2-YFP and Pvalb-cre mice were from the Jackson Laboratory, and GAD67-GFP mice (No. RBRC03674) were from RIKEN BioResource Center. All experimental procedures were performed in accordance with institutional animal welfare guidelines and were approved by the Third Military Medical University Animal Care and Use Committee.
average power delivered to the brain was in the range of 30–160 mW, depending on the depth of imaging. Images of 600 × 600 pixels were acquired at 40 Hz frame rate in all Ca²⁺ imaging recordings. The typical size of field-of-view was ~200 μm x 200 μm. More than one planes of imaging were often acquired from each mouse.

**Auditory Stimulation**

We used an Ed1 electrostatic speaker driver and a free-field ES1 speaker (both from Tucker Davis Technologies) to present sounds, as described in detail previously (Chen et al. 2011). The speaker was placed ~6 cm away from the contralateral ear of mouse. Background noise was dominated by low-frequency components, comparable to the setup used previously (Chen et al. 2011, 2012, 2013). Broadband noise was generated by custom-written software based on LabVIEW 2012 (National Instruments), and transduced to analog voltage via a PCI 6731 card (National Instruments). The duration of each sound stimulus throughout this work was 100 ms.

We measured the sound levels by placing the microphone ~6 cm away from the loudspeaker. All sound levels were calibrated with a 1/4 inch pressure prepopulated condenser microphone system (377A01 microphone, 426B03 preamplifier, 480E09 signal conditioner, PCB Piezotronics Inc.), as previously described (Lou et al. 2014). The data were sampled at 1 MHz by a high-speed data acquisition board USB-6361 from National Instruments and analyzed by customized Labview program. Broadband noise (bandwidth 0–50 kHz) bursts were applied at about 76 dB SPL (sound pressure level), which was substantially louder than the background noise. For pure tone stimulation, the nominal 0 dB attenuation level corresponded to a sound level of 79 dB SPL for frequencies between 1 and 10 kHz, and to a sound level of 81 dB SPL for frequencies up to 40 kHz.

**Electrophysiological Recordings In Vivo**

For both cell-attached and whole-cell patch-clamp recordings in neurons of Au1 in vivo, we used shadow-patching procedure as described previously (Kitamura et al. 2008; Chen et al. 2010, 2011, 2012, 2013). Recordings were performed with an EPC10 amplifier (USB double, HEKA Elektronik). Electrophysiological data were filtered at 10 kHz and sampled at 20 kHz using Patchmaster software (HEKA). For cell-attached recordings of CaL520 AM labeled neurons, a glass electrode filled with normal ACSF containing 50 μM Alexa-594 had a tip resistance of 5–8 MΩ. For whole-cell recordings in awake mice, an electrode had a resistance of 5–8 MΩ with internal solution containing 112 mM potassium gluconate, 8 mM KCl, 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 4 mM Mg-ATP, 0.375 mM Na2GTP, 10 mM sodium phosphocreatine and 50 μM Alexa-594. The series resistance of electrode was continuously monitored and the data were rejected for analysis if the resistance was higher than 30 MΩ.

**Electrophysiological Recordings in Brain Slices In Vitro**

Brain slice preparations including primary auditory cortex or medial geniculate body (MGB) were acutely prepared from adult mice (2–3 months old) in ice-cold oxygenated dissection buffer using conventional procedures as described elsewhere (Cruikshank et al. 2002; Ji et al. 2016). Slices of 300 μm thickness were cut from the virally transduced brain hemispheres (for MGB recordings) or normal brains (for Au1 recordings) by a vibrating microtome (Leica V1000s). After being incubated in a warmed (34 °C) ACSF for at least 45 min, we transferred the slice to the recording chamber for recordings at room temperature. Whole-cell recordings were performed with an AXON amplifier ( Molecular Devices; MultiClamp 700B). Electrophysiological data were filtered at 20 kHz. For depolarizing cortical neurons, we used positive current injections with amplitude of 200–400 pA (duration: 100 ms). This stimulation pulse was applied for 20 times at an interstimulus-interval (ISI) of 2 s to test the echo responses. For optogenetic inactivation of MGB neurons, light stimulation (594 nm wavelength) was applied to the patched-MGB neurons with an intensity of ~10 mW/mm² at the tip of fiber (200 μm diameter, NA 0.53; Changchun New Industries). The fiber position was adjusted to optimize the overlap of light spot and neuron's position. To have action potential firing during light-off periods, continuous current injections with amplitude of +50 to +100 pA were often applied to MGB neurons in slice preparations.

**Virus Injection and Histology**

Mice were anesthetized by 1–2% isoflurane in oxygen, and placed in a stereotactic frame. After removing the skin, a small craniotomy (0.5 × 0.5 mm) was made at coordinates: AP: −3.10 mm; ML: 3.55 mm. For verification of the dye injection sites in L5, a micro-pipette filled with AAV2/8-syn-EGFP virus (BrainVTa) was slowly inserted into the L5 of Au1 through the craniotomy with manipulator tilted 10° from the z-axis. Injection of 60–80 nl of virus was performed at a depth of 600–700 μm from the pial surface. For injections of rAAV-CaMKIIa-ArchT-EGFP, rAAV-CaMKIIa-Chr2 (E123T/T159C)-mCherry, rAAV-EF1α-DIO-Chr2 (H134R)-mcherry or rAAV-CaMKIIa-EGFP into Au1, the micropipette was inserted into L4 (400–500 μm) and both 100–120 nl of virus was injected. After injection, the pipette was held in place for 5 min before retraction. Mice were killed for histology 4 weeks after AAV injection or immediately after optogenetic experiments. After perfusion with 4% paraformaldehyde and dehydration with 15% sucrose for one night, the brains were cut into 50 μm slices for confocal imaging (Zeiss-LSM 700). For termination of mice for post hoc histology, we first anesthetized the animals with 2% isoflurane and then injected an overdose of sodium pentobarbital (200 mg/kg) through intraperitoneal injection. For injection of rAAV-CaMKIIa-ArchT-EGFP into MGB, a small craniotomy (0.5 × 0.5 mm) was made at coordinates: AP: −3.16 mm; ML: 2.15 mm. A micropipette including the virus was down vertically 2.9 mm from the pia matter and then 60–80 nl of virus was injected.

**Behavioral Task**

We developed a simple sound-licking task for behavioral experiments. Before training, the mouse was implanted with the same chamber used for 2-photon imaging of awake mice under isoflurane anesthesia, and then was allowed to recover for 2–5 days. After one night (20:00–08:00) of water restriction, the mouse was habituated for head fixation for 2–3 days. During habituation, the mouse received water in the behavioral setup exclusively. A drop of water (duration 20 ms) was provided from a spout 100 ms after the end of each sound stimulus (broadband noise, 76 dB SPL, duration 100 ms). The water delivery apparatus was soundproofed. The mouse was trained to associate sound with water, and thus licked the water spout in a short time window after each sound. Water reward was not
provided when no sound stimulus was present. Thus, there was no water delivery at echo time point. Licking behavior was monitored with a camera (frame rate 30 Hz) under infrared illumination. Sound stimuli were delivered without any cues at random ISIs during training sessions. Each training session included 20–40 trials, and typically 5–6 sessions were needed for reaching a stable level of performance. The mouse was used for experiments if its success rate was above 80%. The success rate was calculated by the ratio of hit trials to total number of sound stimuli. Licking responses that occurred within a time window of 500 ms following the onset of auditory stimulus were considered as hit trials.

**Simultaneous Optical Fiber and Behavior Recording**

Trained mice as described above were used in these experiments. Small craniotomies (~0.5 × 0.5 mm) were made above Au1 and S2 (the center point: Bregma +0.02 mm or +0.5 mm, 3.5 mm lateral to midline) under isoflurane anesthesia (1–2%).

The cortical neurons were bulk loaded with the calcium indicator Oregon Green BAPTA-1AM (OGB-1 AM, 500 μM) using a glass pipette. Optical fiber-based Ca2+ recording was achieved using a modular fiber optic device (model “FiberOptoMeter v1.0,” Suzhou Institute of Biomedical Engineering and Technology), modified from the previously described setup (Adelsberger et al. 2005, 2014; Grienberger et al. 2012). An optical fiber (200 μm diameter, NA 0.53, Thorlabs) was placed on the dura of each cortical site and then fixed on the skull with dental cement (Tetric EvoFlow from Ivoclar Vivadent Corporate). To excite the fluorescent dye (OGB-1 AM), continuous laser light from a 50 mW solid state laser (wavelength 488 nm, Sapphire, Coherent) was delivered with an intensity of 0.7 mW/mm² at the tip of fiber. The emitted fluorescent light was collected via the same fiber and then detected with an avalanche photodiode (APD, S5343, Hamamatsu Photonics). The licking behavior was monitored with a camera (frame rate: 30 Hz) under infrared illumination. In addition, for recording neuronal responses in MGB, we used an optical fiber with a diameter of 200 or 50 μm (for MGbv recordings: 6 mice with a diameter of 200 μm, 2 mice with 50 μm; for MGBd recordings: 4 mice with a diameter of 200 μm, 3 mice with 50 μm). For behavior tests, fiber recordings, and optogenetic experiments in awake mice, the head was kept in a horizontal position with the same custom-made chamber (head post) as used for 2-photon imaging in head-fixed mice (Supplementary Fig. 1B).

**Optogenetic Activation and Optical Fiber Recording**

In experiments with optogenetic stimulation and/or optical fiber-based activity recording, we also used the modular fiber-optic device (“FiberOptoMeter v1.0,” Suzhou Institute of Biomedical Engineering and Technology). Thy1-ChR2-YFP or virally injected or sham-control mice were trained for the sound-licking task before optogenetic experiments. On the experiment day, 2 optical fibers were placed into L5 of bilateral auditory cortices under isoflurane anesthesia. The end of fiber was stripped and then lowered into the target region slowly with manipulator (~600 μm deep from pial surface for the Thy1-ChR2-YFP mice, ~500 μm deep for the virally injected mice). The behavioral experiments started at least 2 h after the end of anesthesia.

The light stimulus consisted of a single pulse of 473 nm blue light, and delivered with an intensity of ~100 mW/mm² at the tip of fiber (400 μm diameter, NA 0.48; from Thorlabs or Changchun New Industries). The naive transgenic mice for this experiment did not receive training for the sound-licking task. Light stimulation was applied with a duration of 100 ms, followed 100 ms later by a drop of water (20 ms).

**Optogenetic Inhibition**

The mice with rAAV-CamKIIα-ArchT-EGFP or rAAV-CaMKIIα-EGFP expression in Au1 and the mice expressing rAAV-CamKIIα-ArchT-EGFP in MGB were trained for the sound-licking task before optogenetic experiments. On the experiment day, 2 optical fibers were slowly inserted into bilateral auditory cortices (~500 μm deep) with the tip 400 μm below pial surface or bilateral MGB regions (AP: −3.16 mm; ML: 2.15 mm; DV: 2.9 mm). The light stimulus consisted of a single pulse of 593 nm yellow light, and delivered with an intensity of ~150 mW/mm² at the tip of fiber (400 μm diameter, NA 0.48; from Thorlabs). Light stimulation was applied with a duration of 8 s immediately after the 10th sound stimulus.

**Data Analysis**

All analyses of Ca2+ imaging, electrophysiology, and video data were conducted offline by using custom-written software in LabVIEW 2012 (National Instruments), Igor Pro 5.0 (Wavemetrics), and Matlab 8 (Mathworks). To extract fluorescence signals, regions of interest (ROIs) were visually identified and drawn based on fluorescence intensity. The pixels in each specified ROI were averaged to estimate fluorescence changes (f). Relative fluorescence changes Δf/f = (f − f0)/f0 were calculated as Ca2+ signals, where the baseline fluorescence of the ROI f0 was estimated as the 25th percentile of the fluorescence within a sliding time window. Glial cells were excluded in our analysis according to their morphology and Ca2+ transients’ time course. At the end of 2-photon imaging, z-stacks of the imaged regions were acquired. The reconstructed projection images were generated from the z-stack images using ImageJ (Volume Viewer; http://rsbweb.nih.gov/ij/) (see Fig. 2A, left panel).

For both 2-photon imaging and fiber photometry data, Ca2+ transients were automatically detected using a sliding time window (1 s) as the baseline and a following 500 ms time window for detecting. Briefly, the baseline of Ca2+ signals was first detrended using a linear fit, and the noise level was set to be 3 times the standard deviation of the baseline. The mean value in a window (200 ms) around the peak value was determined as the amplitude of Ca2+ transient candidate. Moreover, the first derivative of the Ca2+ signals in the detection window was calculated to measure the rate of rise of Ca2+ transients. A signal was accepted as a transient when both the amplitude and the maximum rate of rise were greater than their corresponding noise level. After that, the trace of the detected Ca2+ transient was first extracted by exponential IIR filtering (window length: 200 ms) and then subtracted from the original signal. The residual fluorescence trace was used as the baseline for a next transient detection, which is similar to the previously described peeling approaches (Grewe et al. 2010).

Responses were defined as echo responses if they occurred within a time window of ±500 ms around the echo time point following the sound sequence. For comparison, the peak amplitude within a time window of 500 ms before the onset of echo response was defined as the baseline. We used Wilcoxon signed rank test to determine statistical significance between the spontaneous and echo response amplitudes (Δf/f). For the unpaired cases, Wilcoxon rank sum test was used to determine statistical significance. P < 0.05 was considered statistically
significant. Summary data are presented in figures as mean ± standard error of the mean (SEM).

To estimate the variability of the fraction of cells with echo response, we applied bootstrap approach, sampling with replacement from the original data set of cells 10,000 times, and calculated the bootstrap confidence interval of the fraction of cells (95% confidence interval). The significance test between the fractions of cells was performed using a permutation test.

To determine the learning curve of the mice for the sound-licking task, the success rates of licking response were visually identified and computed for the training sessions (5–6 sessions, 20–40 stimuli per session). During the test, licking activities were semi-automatically tracked from the recorded video and quantified as a time course. The tongue movements during licking were measured at a particular location around the mouth of mouse, and a ROI around the mouth was drawn by visual inspection. The amplitudes of licking strength were calculated as ROI-based image intensity difference between the video frames.

EPSP events (or depolarization events) recorded in brain slice preparations were detected automatically by measuring the prominence of peaks: the electrophysiological data were first smoothed using exponential IIR filtering (window length: 50 ms), and then the local maxima were detected (Matlab function: findpeaks) as EPSP candidates. The differences between the peak values of the EPSP candidates and the average values of baseline data (window length: 50 ms) before the peaks were compared with a threshold (0.8 mV) to determine EPSP events. One-tailed significance was tested by comparing the event probabilities during the echo period (±0.5 s around the echo time point) against the event probabilities before that (from −1.5 to −0.5 s), and Wilcoxon signed rank test was used to evaluate the population effects.

**Results**

**Echo Responses in Au1 L2/3 Neurons in Anesthetized and Awake Mice**

We first explored the activity of L2/3 neurons of Au1 using 2-photon Ca\(^{2+}\) imaging (Stosiek et al. 2003; Bandypadhyay et al. 2010; Rothschild et al. 2010) in isoflurane-anesthetized mice, following labeling with the highly sensitive fluorescent Ca\(^{2+}\) indicator Cal-520 AM (Tada et al. 2014) (Fig. 1A; Supplementary Fig. 2A). Simultaneous electrophysiological recordings of somatic action potentials in Cal-520-labeled neurons revealed that Cal-520 signals could reliably report single action potentials with near 100% success rate (Supplementary Fig. 2B–D). We imaged activities of up to 100 neurons (average number: 51 neurons) from each imaging plane (up to 3 planes from each mouse, −100 μm apart) simultaneously while presenting isochronous broadband noise stimuli (duration 100 ms, 76 dB SPL, ISI of 2–4 s but fixed within sequence). We observed both randomly occurring spontaneous and sound-evoked Ca\(^{2+}\) transients in the imaged neurons (Supplementary Fig. 2E–G). Based on the responses to the first 10 stimuli in the sequence, we found that sound-evoked responses to broadband noise stimulation had success rates ranging from 0% to 100% across all imaged neurons (mean 38%; 486 neurons from 4 mice; Supplementary Fig. 2H): 39% of the imaged neurons showed responses with a success rate higher than 50%, 43% showed some responses but had a success rate lower than 50%, and 18% were silent. Across the population, about 85% neurons had larger success rates during sound stimulation than the spontaneous response rates (mean spontaneous rate = 16%; 486 neurons from 4 mice) calculated from time windows with the same duration.

Figure 1B,C illustrate an example recording in L2/3, in which the responses of 3 representative neurons to a train of 20 broadband noise stimuli with an ISI of 2 s are shown. Following the end of the last sound stimulus, these neurons produced 1–3 extra responses at the approximate repetition rate of the sequence, although no sound was played (Fig. 1C, red). We referred to these extra responses as echo responses. Such echo responses were present in a subpopulation of neurons (Fig. 1D, top: the onsets of Ca\(^{2+}\) transients; middle: the number of Ca\(^{2+}\) transients in each 250 ms bin along the stimulation sequence; bottom: the averaged response). About 15% of the imaged neurons (14.6% ± 2.0%; Fig. 1F) in L2/3 displayed echo responses. Moreover, 3.8% of all imaged L2/3 neurons showed echo responses but were not responsive during the stimulus sequence. The majority (~72%) of echo neurons responded once after the end of the stimulation sequence (Fig. 1D). The average amplitude of echo responses for all the imaged neurons (0.051 ± 0.003 Δf/f) was significantly larger than that of spontaneous activity before the echo time point (0.019 ± 0.001 Δf/f, n = 1263 neurons from 12 mice, p < 0.001; Fig. 1G). Remarkably, rhythmic sound stimuli at another ISI, 4 s, also induced echo responses (362 imaged neurons from 3 mice; Fig. 1E, top trace, and Fig. 1G) whose latencies relative to the last stimulus matched well the ISI (Fig. 1H). In contrast, stimuli delivered at alternating ISIs of 2 and 3 s did not induce significant echo response (669 imaged neurons from 5 mice, Fig. 1E, the second trace, and Fig. 1G; see the analysis for the fraction of cells in Supplementary Fig. 3). We tested sequences of 5, 10, or 30 stimuli, and found that at least 10 stimuli were necessary for producing reliable echo responses. Interestingly, 10 stimuli also produced the largest averaged response and the maximal number of echo neurons (Supplementary Fig. 4A–C). In addition, when we compared the success rates of sound-evoked responses of echo neurons with nonecho neurons, we observed that the echo neurons had a slightly higher success rate of responses during repetitive sound stimulation (1263 neurons from 12 mice; Supplementary Fig. 4D). In addition to the ISIs of 2 and 4 s, we also tested more ISIs, including 1, 3, 6, and 10 s. We found that stimuli at an ISI of 3 s induced significant echo responses (496 neurons from 4 mice; Supplementary Fig. 5A,B). In this case, latency to last sound stimulus was significantly different from that evoked at ISIs of either 2 or 4 s (Supplementary Fig. 5C). Furthermore, the average latency of each mouse with ISI 2, 3, and 4 s was closely distributed around the corresponding ISI (Supplementary Fig. 5D). In addition, at an ISI of 3 s, 20 rhythmic stimuli produced a similar number of echo neurons to that at an ISI of 4 s (Supplementary Fig. 5E). In contrast, we did not observe significant echo response at ISIs of 1, 6, and 10 s when we tested with 20 rhythmic stimuli (Supplementary Fig. 5F). This is an interesting observation, suggesting that effective ISIs occur in the range of 2–4 s. In addition to broadband noise stimulation, we also tested the sequence of pure tones at 3 different intensities. We found that all pure tones (2, 4.9, 8.9, 12, and 40 kHz) at different intensities were able to evoke significant echo responses (Supplementary Fig. 6), indicating no dependence on stimulus type and intensity.

General anesthesia affects many aspects of cortical circuits during spontaneous activity and sensory processing (Wang et al. 2005; Alkire et al. 2008; Haider et al. 2013). We therefore repeated these experiments in awake mice. The mice were confirmed to be awake by camera monitoring under infrared illumination (Supplementary Fig. 7A). Whole-cell current-clamp recordings of
Figure 1. Au1 L2/3 neurons show echo responses following rhythmic sound stimulation in anesthetized and awake mice. (A) Schematic of experimental protocol, in which Cal-520 AM was injected into the mouse Au1. (B) Two-photon image of Cal-520-labeled L2/3 neurons. (C) Ca\(^{2+}\) responses of 3 representative neurons before, during, and after broadband noise stimuli under anesthesia, as indicated in panel B. (D) Summary of Ca\(^{2+}\) transients in all the imaged L2/3 neurons before, during, and after rhythmic sound stimulation under anesthesia. Upper, plot of onsets of Ca\(^{2+}\) transients. Middle, histograms showing the number of Ca\(^{2+}\) transients in 250 ms bins. Lower, averaged Ca\(^{2+}\) responses of all neurons. (E) Averaged Ca\(^{2+}\) responses of all the imaged neurons in different groups. The number of all imaged mice and neurons are indicated above each trace. (F) Summary of the fractions of echo neurons in anesthetized and awake mice. Permutation test, ***P < 0.01 (n = 1263 imaged neurons for anesthetized mice, and 855 neurons for awake mice). Error bars represent 95% confidence intervals. (G) Comparison of response amplitude between echo response and the corresponding baseline. Wilcoxon signed rank test, 2 s: ***P < 0.001 (df = 1262, z = −15.0141), 4 s: ***P < 0.001 (df = 361, z = −12.4011), 2 s/3 s: ns (df = 668, z = 0.9787) and 2 s awake: ***P < 0.001 (df = 854, z = −6.5261), n = 1263, 362, 669, and 855 imaged neurons, respectively. df denotes degrees of freedom, ns denotes no significance. Error bars indicate SEM unless stated otherwise throughout the study. (H) Latency of the first echo response to the last auditory stimulus (left) or to the echo time point (right). Red circles represent individual data points; black circles represent the mean values. n = 185, 2, and 184 echo neurons, respectively.
L2/3 neurons revealed that small, high-frequency membrane potential fluctuations with a unimodal distribution prevailed in the awake state (peak membrane potential: $-66.7 \pm 1.5$ mV, $n = 5$ cells) (Supplementary Fig. 7B). These differed from the large, slow fluctuations between depolarized and hyperpolarized states that were recorded in the same type of neurons under anesthesia (Chen et al. 2011, 2013), further confirming the state of mice.

We observed echo responses following the same rhythmic stimuli in awake mice as well (855 imaged neurons from 8 awake mice, Fig. 1E-H; Supplementary Fig. 7C,D). In awake mice, as in anesthetized mice, echo responses occurred mostly (~88%) once (Supplementary Fig. 7C) and their latencies were approximately equal to the ISI (2 s) (Fig. 1H). Echo responses in awake mice (21.5% ± 2.7%, $n = 855$ imaged neurons from 8 mice; Fig. 1F) were slightly more prevalent than in anesthetized mice (permutation test, $P = 0.0015$). Thus, neurons in L2/3 of Au1 exhibited significant echo responses corresponding to the ISI of the preceding sound sequence in both anesthetized and awake mice.

To test whether the neurons showing echo responses were excitatory or inhibitory, we performed the same experiments in GFP knock-in mice, in which almost all cortical GABAergic neurons were GFP positive (Tamamaki et al. 2003). We observed that a subpopulation (12%) of GFP$^+$ neurons showed echo responses, comparable to the fraction (19%) of GFP$^-$ neurons with echo response (430 neurons from 4 mice; Supplementary Fig. 8A-C). Therefore, excitatory and inhibitory neurons are roughly proportionally represented among echo neurons in L2/3.

**Echo Responses in L5 and L4**

One key target of L2/3 in Au1 is L5, which provides the cortical outputs to many downstream targets, including several motor-related regions (Perales et al. 2006; Znamenskiy and Zador 2013; Hackett 2015). We imaged responses of L5 neurons to the same stimuli. To achieve optimal imaging quality for deep layer neurons, we minimized fluorescence from superficial layers by injecting Cal-520 AM at a depth of ~600 µm (Supplementary Fig. 9A). The injection location in L5 was confirmed by injecting AAV-synapsin-EGFP (enhanced green fluorescence protein) at the same depth and then by post hoc histological staining of both EGFP and DAPI (4',6-diamidino-2-phenylindole) (Supplementary Fig. 9B). This procedure allowed for labeling cells within a spherical volume with a diameter of ~300 µm yet achieving no or only weak staining in superficial layers (see many unlabeled cell bodies appearing as dark holes at the depth of 169 µm in Fig. 2A).

Ca$^{2+}$ imaging was restricted to the neurons located between 550 and 750 µm, corresponding to the cells within L5 (Anderson et al. 2009). L5 neuronal somata were visualized with sufficient contrast and dynamic signal during both spontaneous activity and sound stimulation (Supplementary Fig. 9C-E), comparable to the previous work using regenerative amplification 2-photon microscopy (Mittmann et al. 2011) or using a red-shifted fluorometric Ca$^{2+}$ dye (Tischbirek et al. 2015). As illustrated in an example (Fig. 2B,C) as well as in the summary of all imaged neurons (Fig. 2D), we found echo responses in a fraction (33.5% ± 2.4%; 503/1500 neurons) from 8 mice; Fig. 2E) of L5 neurons following rhythmic stimuli with an ISI of 2 s. Consistent with the results in L2/3 neurons, only 0.8% of imaged L5 neurons show echo response without any response to the sound sequence. In addition, most of the neurons (68.2%) showed single echo responses (Fig. 2D), with a latency approximately equal to the ISI and an averaged amplitude (0.033 ± 0.001 Δf/f) significantly greater than that of spontaneous activity (0.017 ± 0.001 Δf/f; $n = 1500$ imaged neurons, $P < 0.001$; Fig. 2F,G).

We next imaged the neurons within the main thalamoreceptive layer, L4, to rhythmic sound stimulation. Of note, 1934 imaged neurons from 13 mice were obtained at depths in the range of 400 to 500 µm. We observed that the fraction of responsive cells in L4 during sound stimulation was similar to that of L2/3 but much higher than L5 (Supplementary Fig. 10E). In addition, the response rate to sound stimulation in L4 was significantly lower than that of both L2/3 and L5 (Supplementary Fig. 10F). As
illustrated in the averaged trace of all the imaged L4 neurons (Supplementary Fig. 10A), we found a significant but weak response at around the echo time point following rhythmic stimuli. Overall, we observed echo responses in only 8% of imaged neurons, with a latency approximately equal to the ISI (Supplementary Fig. 10B–D). All of the echo neurons showed responses to the preceding sound stimulus.

**Strong Correlation Between the Au1 Neuronal and Behavioral Echo Responses**

To explore potential behavioral consequence of the neuronal echo responses in Au1, we developed a simple sound-triggered licking task, in which mice were trained (see Materials and Methods) to detect auditory stimulation (broadband noise, duration 100 ms, 76 dB SPL) and report sound by licking a spout for water reward (Fig. 3A), similar to the whisker stimulation-evoked licking task that has been described previously (Sachidhamanandam et al. 2013). After 2–3 days of training, mice reached stable performance with a high success rate (80%) when we delivered sound stimulus paired with a drop of water at random times (Fig. 3B). These mice were considered as trained. The sound-evoked licking response was completely abolished by local injection of the GABA<sub>A</sub> (γ-aminobutyric acid)-receptor agonist muscimol bilaterally into Au1 with small glass pipettes in trained mice (Supplementary Fig. 11). In contrast, in a control group that bilaterally received ACSF injection into the Au1, the sound-evoked licking responses remained normal (Supplementary Fig. 11). These results indicate a causal role for Au1 in controlling this sound stimulation-associated behavior.

We then tested behavioral echo responses in trained mice with a train of 10 rhythmically repeated sound stimuli at an ISI of 3 or 4 s. The example in Figure 3C illustrates that the mouse showed one more licking response at the time point that closely corresponded to the ISI of the preceding sound sequence (see another example in Supplementary Movie 1). Figure 3D shows 4 more examples of such experiments at an ISI of 3 s (top) or 4 s (bottom). Overall, these echo licking responses following rhythmic stimulation were observed in all the mice tested (n = 7 mice for the ISI of 3 s, and 9 mice for the ISI of 4 s). The averaged latency of echo licking response to the last sound matched well the ISI (Fig. 3E). Similar to the cellular echo responses in Au1 (Supplementary Fig. 4), 5 repeated sound stimuli were insufficient for producing echo licking responses (Supplementary Fig. 12A). It should be noted that the duration of licking responses was in the range of 1–5 s (2.4 ± 0.1 s, n = 144 licking events from 8 mice; Supplementary Fig. 12B), which does not allow for testing the licking behavior with stimulus intervals shorter than 1 s in our experiments.

To explore the correlation between the echo neuronal response in Au1 and the echo licking response, we simultaneously monitored licking behavior and neuronal activity in Au1 using fiber photometry that allows for the fluorometric detection of local action potential activity (Adelsberger et al. 2005; Stroh et al. 2013). For control, we placed the second optical fiber onto the secondary somatosensory cortex (S2) (Fig. 3F, left). The cortical neurons were bulk loaded with the Ca<sup>2+</sup> indicator Oregon Green BAPTA-1 AM (OGB-1 AM) using the multicolor bolus loading technique (Stosiek et al. 2003). In mice trained in the sound-licking task, sound stimulation reliably evoked licking behavior as well as Ca<sup>2+</sup> responses in Au1 (Fig. 3F, right). In 8 trained mice, we found that 98% ± 1% of the sound-evoked licking responses and 84% ± 7% of the echo licking responses were associated with Ca<sup>2+</sup> responses in Au1 (Fig. 3G). Similarly, 94% ± 4% and 88% ± 2% of the Ca<sup>2+</sup> responses in Au1 were associated with licking responses during repeated sound stimuli and at the first echo time point, respectively. According to a signal detection theory analysis (Supplementary Table 1), we can see that, in ~80% of all echo licking trials, the echo licking responses can be predicted by the echo Ca<sup>2+</sup> responses measured by photometry.

Sound stimulation produced Ca<sup>2+</sup> responses also in caudal S2 (+0.02 mm from Bregma, n = 4 mice; Supplementary Fig. 13A,B). The response in caudal S2 could be due to activation of the tongue representation during licking behavior (Sakamoto et al. 2008) or even to evoked auditory activity during sound stimulation (Carvell and Simons 1986). On the other hand, sound stimulation was not associated with Ca<sup>2+</sup> responses in rostral S2 (+0.5 mm from Bregma, n = 4 mice; Fig. 3F and Supplementary Fig. 13C,D). Rostral S2 did not show any response associated with licking at the first echo time point either (n = 4 mice; Fig. 3F,G). These results suggest a strong correlation between behavioral echo responses and neuronal activity in Au1.

**Optogenetic Inhibition of Au1 Impairs Behavioral Echo Responses**

To determine the necessity of Au1 neuronal activity in producing behavioral echo responses, we performed 2 optogenetic silencing experiments (Fig. 4A). First, we suppressed the activity of excitatory neurons in Au1 by activating the light-driven outward proton pump archaerhodopsin (ArchT), causing rapid and effective inhibition of neuronal activity (Chow et al. 2010). We bilaterally injected rAAV-CaMKII<sup>lox</sup>-ArchT-EGFP or rAAV-CaMKII<sup>lox</sup>-EGFP (control mice) into Au1. The expression of ArchT was confirmed by histology (Fig. 4B) and its functionality was verified by in vivo cell-attached recordings showing that light stimulation at a wavelength of 593 nm completely suppressed spiking activity of Au1 neurons (5–8 trials from each neuron; n = 7 neurons; Fig. 4C). We then measured the licking responses during and after rhythmic sound stimulation. We found that light presentation produced a significant reduction in the rate of echo licking responses as compared with the control mice (Fig. 4D,E), reaching a level slightly higher than chance level (probability of observing licking activity within a time window of the same duration preceding sound stimulation). The spared behavioral echo responses in the presence of the optogenetic activation of ArchT could be due to the limited volume of light illumination, which may have been insufficient to cover the entire Au1 area.

In the second optogenetic silencing experiment, we used an alternative approach for inactivating Au1. We took advantage of Cre-loxP recombination to express channelrhodopsin-2 (ChR2) in parvalbumin (PV)-expressing inhibitory neurons (Li et al. 2013). We bilaterally injected rAAV-EF1α-DIO-ChR2-mCherry to the Au1 of Pvlb-re transgenic mice. Post hoc histology (Supplementary Fig. 14A) and in vivo electrophysiological recordings confirmed the specific expression of ChR2 in PV<sup>+</sup> neurons in Au1 (Supplementary Fig. 14B). In these virally injected mice, blue light reduced the success rate of echo licking responses to chance level (Fig. 4F,G). In control mice, blue light presentation did not affect echo licking responses (Fig. 4F,G). Together, these 2 experiments suggest that Au1 is a critical hub in the circuit that generates anticipatory motor responses following rhythmic auditory stimulation.

**Behavioral Echo Responses Driven by Optogenetic Activation of Au1 Neurons**

The essential role of Au1 activity in generating anticipatory motor responses suggest that echo licking behavior may be
triggered by echo neuronal responses in Au1. To test this possibility, we first analyzed the temporal relationships between echo licking and echo Ca\textsuperscript{2+} responses based on the optical fiber recording. As shown in Supplementary Figure 15, we found that, on average, echo neuronal responses occurred at 387 ± 128 ms (n = 32 trials from 8 mice) before the onset of echo licking.

In addition, we tested whether a direct activation of principal neurons in Au1 using ChR2 can drive mouse echo licking behavior in our auditory associative learning task. Since L5

**Figure 3.** Mice show echo licking responses in the sound-licking task following rhythmic sound stimulation. (A) Left, behavioral diagram. Right, example of sound-evoked licking response (a.u., arbitrary unit). (B) Success rate of licking response across 6 training sessions. (C) Example of licking responses during and after 10 repetitive sound stimuli in a well-trained mouse. The echo licking response is marked in red. (D) Raster plot of onsets of licking responses in 4 representative mice tested with 2 different ISIs. Each tick represents one licking onset on one trial. (E) Latency of echo response to the last sound stimulus or to the echo time point tested with an ISI of either 3 s (n = 7 mice) or 4 s (n = 9 mice). (F) Left, schematic of simultaneous recording of licking behavior and population activity in Au1 and S2. Right, simultaneously recorded neuronal activity in S2 (the rostral area, upper; average of 4 mice), in Au1 (middle, average of 8 mice), and licking response (lower, average of 8 mice). (G) Summary of percentage of Ca\textsuperscript{2+} responses that correlated with licking responses during sound presentations (black) or at the first echo points (red). n = 8 mice for Au1, 4 mice for S2 (the rostral area), and 8 mice for licking behavior. Wilcoxon rank sum test, ***P < 0.001 (df = 46, z = 5.3227) and **P < 0.01 (df = 46, z = 3.2547).
neurons provide the major source of outputs from Au1 and also show echo responses following rhythmic sound stimulation, we used a previously reported mouse line, in which high levels of ChR2 and YFP are expressed mainly in L5 neurons under the Thy1 promoter (Supplementary Fig. 16) (Arenkiel et al. 2007). Delivering pulses of blue light through a fiber placed above the somata of L5 neurons reliably evoked their activation, as revealed by extracellularly recorded electrical responses (Supplementary Fig. 17A), and also by population Ca\(^{2+}\) signals recorded through the same fiber used for optogenetic stimulation (Supplementary Fig. 17B) (Stroh et al. 2013). In mice that were already trained for the licking task, optogenetic activation of L5 neurons in Au1 bilaterally, substituting for sound stimulation, evoked reliable licking response in all animals tested (n = 6 mice; Fig. 5A,B). For control, neither naïve mice that expressed ChR2 (ChR2 naïve mice; n = 6 mice) nor trained mice that expressed no ChR2 (WT-trained mice; n = 5 mice) showed licking in response to the same optogenetic stimulation (Fig. 5A,B). Interestingly, although light stimulation was indiscriminately applied to a group of L5Au1 neurons, during light presentations we did not detect any motor behaviors other than the licking responses in the trained mice (n = 6). Also, we did not observe any motor response to light stimulation in naïve mice (n = 6, Supplementary Movie 2). This suggests that activating L5 neurons in Au1 may be predominantly meaningful for triggering the learned motor behavior. Importantly, in ChR2-expressing mice that were trained to lick following auditory stimulation, rhythmically repeated light stimuli readily produced rhythmic licking behavior as well as echo licking responses after the end of the stimulation sequence (see an example in Fig. 5C and Supplementary Movie 2, and plot of onsets of licking in 6 example mice in Fig. 5D). The latency of echo licking responses following the last light stimulus closely matched the ISI (n = 11 mice; Fig. 5E). In 5 trained mice, we also combined the recordings of licking behavior with optical fiber-based measurements of population neuronal activity in Au1 L5 while delivering rhythmic light stimuli (Fig. 5F, left). We found that both immediate and echo licking responses induced by optogenetic stimulation were tightly associated with the neuronal responses in Au1 (Fig. 5F, right, and G).

It is unlikely that optogenetically induced anticipatory motor responses were caused by the light-evoked activation of ChR2 expressing in the subcortical region in Thy1-ChR2 mice, for example, hippocampal area CA1. Indeed, the distance from the tip of the optical fiber placed above L5 to the hippocampal CA1 pyramidal cell layer was 0.8–1.1 mm. At that distance, the estimated light intensity after penetration through the highly scattering tissue (~1–2 mW/mm\(^2\); Supplementary Fig. 16B) would be insufficient for activating ChR2 (at least 5 mW/mm\(^2\) needed to drive action potential firing) (Boyden et al. 2005; Zhang et al. 2006).

Nevertheless, to directly rule out a role for optogenetic activation of neurons located outside Au1 in eliciting the motor responses, we bilaterally injected rAAV-CaMKIIa-ChR2-mCherry to Au1. The ChR2 expression in Au1 was confirmed by post hoc histology and in vivo electrophysiological recordings (Fig. 5F). As in the Thy1-ChR2-YFP transgenic mice, delivering blue light bilaterally through 2 fibers placed above the virally transduced Au1 neurons (~500 µm deep from the cortical surface) evoked reliable licking responses in all trained mice (n = 7 mice) (Fig. 5I,J). In contrast, in trained mice that received a control viral injection of AAV-CaMKIIa-mCherry, light illumination did not evoke any licking responses (Fig. 5I,J).
Figure 5. Optogenetic activation of Au1 excitatory neurons induces echo licking response in trained mice. (A) Examples of licking response to light illumination (473 nm wavelength) of a Thy1-ChR2 trained (top), a Thy1-ChR2-naïve (middle), or a WT-trained mouse (bottom). (B) Comparison of the success rates of light-induced licking in these 3 groups, n = 6, 6, and 5 mice, respectively. Wilcoxon rank sum test, ChR2 trained versus naïve: ***P < 0.001 (df = 115, z = 9.5388) and ChR2 naïve versus WT: ns (df = 78, z = −0.2163); ns denotes no significance in this figure. (C) Example of echo licking after repetitive light stimuli at an ISI of 4 s in a trained mouse. The echo licking is marked in red in this figure. The last 5 stimuli are shown. (D) Raster plot of onsets of licking responses from 6 representative trained mice. Each row represents a mouse. Each tick represents one licking onset on one trial. (E) Latency of the first echo licking to the last light stimulus (left) or to the echo time point (right), n = 11 mice. (F) Schematic of simultaneous optical stimulation and recording in L5 through the same fiber. Right, licking response (lower, average of 11 mice) and the corresponding population neuronal activity in the L5 (upper, average of 5 mice) during and after repetitive light stimulation. The light stimulation artifacts were removed. (G) Summary of percentage of Ca²⁺ events that correlated with licking responses during light presentations (blue) or at the first echo points (red) or during spontaneous licking before light stimulation (Spont. in black), n = 5 mice for each group. Wilcoxon signed rank test, ns (df = 9, z = −0.9567) and *P < 0.05 (df = 9, z = 2.3204). (H) Left, image showing the expression of AAV-CaMKIIα-ChR2-mCherry in Au1. Right, local field potential response to a 50 ms blue light stimulation. (I) Examples of licking response to light illumination of an AAV-CaMKIIα-ChR2-injected (upper) or a control mouse (lower). Both mice were trained for the licking task. (J) Comparison of the success rates of light-induced licking in CaMKIIα-ChR2-expressing (n = 7 mice) and control mice (n = 6 mice). Wilcoxon rank sum test, **P < 0.01 (df = 11, z = 2.9326). (K) Example of echo licking after repetitive light stimuli in a CaMKIIα-ChR2-mCherry-expressing mouse. The echo licking is marked in red. The last 5 stimuli are shown. (L) Raster plot of onsets of licking responses from 7 CaMKIIα-ChR2-mCherry-expressing mice. Each row represents a mouse. (M) Latency of the first echo licking to the last light stimulus (left) or to the echo time point (right) in 7 CaMKIIα-ChR2-mCherry-expressing mice.
As expected, in CaMKIIα-ChR2-mCherry-expressing mice that were trained in the behavioral task, rhythmically repeated light stimuli evoked clear echo licking responses following the end of the light sequence (Fig. 5K,L). The latency of the echo licking responses following the last light stimulus closely matched the ISI (n = 7 mice; Fig. 5M). These results show that anticipatory motor response can be effectively triggered by optogenetic activation of a subgroup of Au1 neurons.

The Auditory Thalamus is Not Necessary for Echo Responses

The echo responses observed here in Au1 neurons could result from activity generated in subcortical regions, for example, auditory thalamus. Previous work has observed 2 types of expected responses in the auditory thalamus of guinea pigs following rhythmic sound stimulation at an interval of a few seconds (Gao et al. 2009): first, intracellular recordings revealed up-to-down transitions of the membrane potentials in the neurons of the dorsal and caudomedial divisions (nonlemniscal subdivisions) of the MGB. This transition was associated with a decrease in spike activity and therefore cannot explain our (excitatory) cortical echo response. Second, there may be an increase in spike activity in the ventral division (the lemniscal subdivision) of the MGB, as revealed by multiunit extracellular recordings. Such effects occurred in the auditory thalamus of guinea pigs during slow-wave sleep but not in the awake state, and may represent entrainment of the slow-wave activity by the sensory stimulus. In contradistinction, our cortical echo responses occur in the awake state. Therefore, thalamic responses are unlikely to be the origin of echo responses in mouse Au1. Nevertheless, given the differences in species and in experimental conditions, we tested whether mouse auditory thalamus could exhibit echo responses under the same experimental conditions we used to elicit them in auditory cortex.

As a first step, we examined the anatomical connections between Au1 and the MGB using a fluorescent retrograde tracer, cholera toxin β subunit conjugated to Alexa Fluor-488 (CTB-green). We injected small amounts of CTB-green into the cortical region that has been imaged (Fig. 6A, top). After a 5-day waiting period for retrograde transport, we observed that retrogradely labeled neurons were always located in the lateral part of the ventral division of the MGB (MGBv) (Fig. 6A, bottom; n = 5 mice). This is consistent with MGB-Au1 connectivity, as reported in previous studies (Llano and Sherman 2008; Horie et al. 2013; Tsukano et al. 2015) and thus verified that our 2-photon imaging was performed in Au1. In addition, a CTB-green injection that spanned both MGBv and MGBd (dorsal division) resulted in dense labeling in the deep layers of Au1 and secondary auditory cortex (in both dorsal and ventral areas, AuD and AuV) (Fig. 6B), while an injection into MGBd without involvement of MGBv caused labeling mainly in the deep layers of secondary auditory cortex (AuD and AuV) (Fig. 6C). These results confirmed the locations of MGBv and MGBd in mice (Llano and Sherman 2008; Hackett et al. 2011) and were used for the following experiments.

We next used a fiber photometry approach to record neuronal responses in MGB. The thalamic neurons were labeled with OGB-1 AM. The tip of the optical fiber was inserted into the stained thalamic area (Fig. 6D,E: left panels). In both MGBv (Fig. 6D, right) and MGBd (Fig. 6E, right), we reliably observed sound-evoked responses to a train of 10 broadband noise stimuli with an interval of 4 s. However, unlike our 2-photon imaging results in Au1 neurons, in these 2 MGB divisions, we did not detect any echo response following the end of sound stimulus sequence (Fig. 6D–F). Therefore, under the same experimental conditions, neuronal echo responses could be detected in Au1 but not in auditory thalamus. These results suggest that in mice, the auditory thalamus has at most a minimal contribution to the generation of the cortical echo responses following rhythmic auditory stimulation.

The lack of cellular echo responses in thalamic neurons suggests that these neurons may have little role in the generation of behavioral echo responses. To directly test this possibility, we conducted optogenetic silencing experiments by inactivating ArchT-expressing MGB neurons during the licking task. We bilaterally expressed ArchT in MGB by local injection of AAV-CaMKIIα-ArchT-EYFP. In vitro electrophysiological recordings in slice preparations verified the blocking effect of yellow light (Fig. 7A,B). ArchT expression was restricted to MGB, as verified by post hoc histology (Fig. 7C). We then measured the echo licking responses following rhythmic auditory stimulation in the trained mice. We found that delivering light stimulation at a wavelength of 593 nm through 2 fibers (400 µm diameter) bilaterally placed above the transected MGB regions had no significant effect on the echo licking responses (Fig. 7D,E). Therefore, these results suggest that auditory thalamus plays no or minor (if any) role in the generation of behavioral echo responses.

Up to now, we have shown that the L2/3, L4, and L5 cortical neurons, but probably not those in MGB, displayed echo responses, suggesting that the echo responses originate in the cortex. We investigated the possible cortical origin of the echo responses in acute cortical slices of adult mice. We performed whole-cell patch-clamp recordings from L5 and from L2/3 neurons in Au1. In current-clamp mode, rhythmic stimulation by positive current injections via the patch electrode at an ISI of 2 s reliably resulted in depolarizations with AP firing in all the recorded L5 and L2/3 neurons (Supplementary Fig. 18A,B). Interestingly, 8 of 29 (28%) L5 neurons exhibited small depolarization responses with amplitudes in the range of −0.8–4.7 mV (1.3 ± 0.1 mV, n = 33 events) approximately at the expected time point (Supplementary Fig. 18C–E). On the other hand, none of 19 neurons in L2/3 had such echo depolarization responses (Supplementary Fig. 18H). These results suggest that the echo depolarization responses observed in L5 neurons in vitro may underlie the echo spike activity in cortical neurons in vivo. Interestingly, we found that the rhythmic current injection-evoked echo depolarizations were completely abolished by application of the ionotropic glutamate receptor antagonists APV and CNQX (Supplementary Fig. 18F,G), suggesting that a network mechanism (e.g., spontaneous synaptic events (Timofeev et al. 2000; Chauvette et al. 2010)) but not intrinsic properties of the cortical neurons may account for the generation of echo responses.

The lack of spike activity associated with the echo depolarizations in brain slices may be explained, at least in part, by 3 possibilities. First, the neurons in slice preparations have lower background synaptic activity than those obtained from intact brains (Pare et al. 1998), increasing the size of the depolarization required for reaching action potential threshold. Second, the experiments were conducted in brain slices, in which many local connections have been cut, possibly reducing the ability of the remaining local circuit to support the echo spike potential. Third, these small echo depolarizations were elicited by single-cell stimulation, which could be far below action potential threshold, unlike the in vivo conditions in which the auditory stimulation can recruit many neurons in the cortex. Taken together, these slice experiments suggest a cortical
origin of echo responses. L5 neurons in Au1 may have the ability to initiate the echo responses, and local feedback connectivity from L5 probably generates the echo responses in L2/3 neurons and even weaker responses in L4. Consistent with this hypothesis, in vivo neurons in L5 have somewhat higher probability to show echo responses than neurons in L2/3 or L4.
These results are also consistent with the dominant role of cortical L5 neurons in the generation of spontaneous slow waves (Sanchez-Vives and McCormick 2000; Sakata and Harris 2009; Chauvette et al. 2010; Stroh et al. 2013).

Discussion

Two-photon Ca\(^{2+}\) imaging in combination with improved fluorescent Ca\(^{2+}\) indicators enabled us to precisely and simultaneously monitor spiking activity of many neurons in the intact brain (Stosiek et al. 2003) and therefore provided a suitable tool for searching for neurons that encode expected sensory events. Using this approach, we demonstrate here the existence of neuronal activity representing predictive signals in mouse Au1. We found that following a train of rhythmic auditory stimulation, a subgroup of L2/3, L4, and L5 Au1 neurons exhibited one or a few repetitions of “echo” responses, in the absence of sound, at the expected time interval. After learning to perform a sound-triggered licking task, mice showed echo licking responses following rhythmic sound stimulation that were highly correlated with Au1 neuronal echo responses, as indicated by fiber photometry-based recording of neuronal activity (Adelsberger et al. 2005; Stroh et al. 2013). Optogenetic silencing (Chow et al. 2010; Li et al. 2013) of Au1 neurons significantly impaired echo licking responses. Furthermore, ChR2-mediated optogenetic activation (Nagel et al. 2003; Boyden et al. 2005) of a group of Au1 principal neurons could substitute for sound stimulation and reliably produce echo behavioral responses. Finally, we showed that the auditory thalamus did not show neuronal echo responses and was not necessary for the generation of echo licking responses, suggesting that they are generated by cortical mechanisms in Au1. Thus, these results identify an essential role of Au1 circuits for coding expected sensory information, which is transmitted to downstream targets for driving motor responses to expected events.

Coding of Predictive Information in Sensory Cortical Circuits

Past work suggests that cortical representations of predictive sensory information can occur in at least 3 different manners. First, temporal expectation can enhance sensory-evoked neuronal responses in primary auditory cortex (Jaramillo and Zador 2011). This enhanced neuronal representation of an expected sensory stimulus may have a causal role in improving speed and accuracy of associated behavioral responses (Jaramillo and Zador 2011). This perceptual enhancement can also explain the improved performance caused by temporal expectation in humans (Stefanics et al. 2010). Second, the most intensively studied single-neuron correlate for implicit predictive coding in auditory cortex is stimulus-specific adaptation (SSA) (Ulanovsky et al. 2003), the reduction in the responses of a neuron to a repeating common sound which does not fully generalize to other, rare, stimuli (see review in Nelken 2012). This expected stimulation-induced reduction in neural responses has been also observed in visual cortex (Muller et al. 1999; Summerfield and de Lange 2014) as well as in somatosensory cortex (Musall et al. 2015).

Third, and most important for this work, omitted stimuli following rhythmic sensory stimulation have been reported to...
evoke neural responses. Several studies using low resolution methods have reported the presence of responses to omitted sound stimuli in the vicinity of human and monkey auditory cortex (Busse and Woldorff 2003; Lakatos et al. 2013). To understand the cortical representations of these expected stimuli, we studied omission responses and the associated network dynamics in mouse primary auditory cortex at a cellular resolution.

Using in vivo 2-photon imaging-based mapping experiments, we demonstrate that a subpopulation of L2/3, L4, and L5 neurons in mouse Au1 responds to expected stimulation following rhythmic sound stimulation that we termed “echo responses.” A higher fraction of cells in L5 than in L2/3 or L4 showed echo responses, and in vitro electrophysiological recordings in brain slice preparations suggested a dominant role of local circuits in L5 in generating them. Moreover, neuronal echo responses were exclusively evoked by stimuli at repetition rates of 0.25–0.5 Hz and occurred even under anesthesia. This observation suggests that entrainment of slow oscillations, predominately being initiated through cortical L5 neuronal networks (see Supplementary Fig. 19) (Sanchez-Vives and McCormick 2000; Sakata and Harris 2009; Chauvette et al. 2010; Stroh et al. 2013), probably plays an important role in producing the echo responses. In addition, using GAD67-GFP knock-in mice, we found that both excitatory and inhibitory neurons showed echo responses, suggesting that different neuronal types jointly encode predictive information.

It is interesting to note that omission neuronal responses have not been reported in the SSA literature (Farley et al. 2010; Hershenhoren et al. 2014). A number of reasons may account for this difference. First, we used long intervals between stimuli (2–4 s), while most SSA studies used faster presentation rates (ISIs < 1 s, typically 300–500 ms). Omitted stimulus-evoked responses may be preferentially elicited at slow repetition rates because such rates fit the time scales of motor responses (Supplementary Fig. 12B). Second, we looked for a single omitted stimulus-evoked response, while SSA studies often averaged the responses of tens of stimuli. It may well be that omitted stimulus-evoked responses disappear after 1 or 2 repetitions of the omission. Finally, we imaged simultaneously the responses of tens of hundreds of neurons, making it possible to pick up the relatively small population that did emit the echo response.

### Cortical Representation of Predictive Information Does Not Reflect Afferent Thalamic Input

In mammals, thalamic neurons provide the principal afferent inputs to primary sensory cortex during sensory information processing. Previous work has revealed that neurons in the auditory thalamus of guinea pigs exhibit responses to expected stimulation following stimulation by rhythmic sound sequences (Gao et al. 2009). We argue that the echo responses we observed in the mouse are different in nature from those observed by Gao et al. (2009) in guinea pig thalamus. Indeed, they mainly observed reductions of the spiking responses in nonlemniscal thalamus; and while increased responses were observed in lemniscal thalamus, they occurred only during slow-wave sleep but not in the awake state. In marked contrast, we documented cortical echo responses in both anesthetized and awake mice.

We also addressed the role of the auditory thalamus directly. We used optical fiber-based recordings, a simple but powerful optical technique that enables the study of population neuronal activity in deep brain areas such as the thalamus (Stroh et al. 2013). Using this approach, we observed in the mouse MGB reliable responses to sound stimulation but did not detect any echo responses in the mouse MGB, either in the lemniscal (MGBv) or in the nonlemniscal (MGDb) subdivision.

It could be argued that the fiber photometry method could miss the activity from a potentially small fraction of thalamic neurons, as it only detects population neuronal activity. This is unlikely to be the case, because when placing the same fiber onto the cortical surface we were able to detect robust echo responses in L2/3 of Au1 (Fig. 3F) that contains only 15–20% echo neurons (Fig. 1F). Since the Ca^{2+} fluorescence level during the echo window in MGB was more than 10 times smaller than the echo responses in Au1 (Fig. 1G vs. Fig. 6F), this suggests that only a negligible fraction (<1%) of MGB neurons could have echo responses.

This finding raises a question: why were echo responses undetectable in the thalamus, given that they are present in cortex and given the extensive projections from Au1 to MGB in mice (Ilano and Sherman 2008; Hackett et al. 2011; Hackett 2015)? Activation of auditory cortex has been reported to mostly suppress MGDb neurons (He 2003). This is consistent with the reductions of the spiking responses observed in guinea pig thalamus (Gao et al. 2009) and the absence of echo responses in mouse thalamus from our work. While this argument does not necessarily hold for MGBv, the descending connections from cortex to MGBv are weak (Lee and Sherman 2010a, 2010b). In fact, these connections are classified as “modulators” rather than “drivers.” When there is no afferent activity in the MGBv, there may be nothing to modulate, consistent with our failure to detect echo responses in MGBv.

Therefore, our results suggest a cortical origin for the echo responses. This suggestion is further supported by in vitro slice experiments indicating that single-cell stimulation can produce significant (although, admittedly, week) echo depolarizations in a subgroup of cortical L5 neurons. Although more experimental work is needed to figure out how echo response is generated in cortical circuits, a recent model proposed by Mi et al. (2013) provides a possible mechanism that matches well our experimental results. In their model, the network has scale-free topology (containing hub neurons that are difficult to be excited) and consists of many low-degree loops with different sizes. These loops endow the system with the ability to follow the trained rhythm, while hub neurons can synchronize the entire network. Based on this model and our entire experimental data, we propose that L5 serves as a reservoir that contains low-degree loops, while L2/3 neurons is most likely to be hub cells that fire sparsely and connect to the network extensively (Supplementary Fig. 19).

Through learning by rhythmic sensory stimulation, a certain loop is selected from L5 networks, whose size matches the trained rhythm. In addition, the connectivity from the neurons in this L5 loop to L2/3 hub neurons is strongly potentiated during learning. Such strengthened circuits allow the signals to flow from a few L5 neurons to L2/3 and then across the entire network. In addition, this network-wide activity and the potential need for exerting behavioral actions may determine that echo responses only occur following a particular rhythm, for example, 0.25–0.5 Hz (slow oscillation range) in our study.

### Cortical Processing of Predictive Information is Required for Anticipatory Motor Output

Although neural correlates of predictive information have been reported broadly in mammalian sensory systems (Bullock et al. 1994; Schwartz et al. 2007; Gao et al. 2009), the necessity of any of these sites for generating anticipatory behavioral consequence
remained open. The only available evidence relating sensory stimulation and anticipatory motor action came from work performed in the optic tectum of zebrafish larvae, in which the entrained neuronal activity following rhythmic stimulation at an interval of a few seconds was correlated with initiation of visuo-motor escape responses (Sumbre et al. 2008). However, the causal relationship between neuronal activity and behavioral response has not been established. In the present study, we applied fiber photometry in combination of optogenetic loss- and gain-of-function experiments and demonstrated that activating neurons in mouse auditory cortex is both necessary and sufficient for eliciting motor responses to predicted auditory stimuli in trained (but not in naive) animals.

Au1 sends projections to a number of downstream targets, including several motor-related regions, such as the striatum, secondary motor cortex, and the pontine nuclei (Perales et al. 2006; Hackett 2015). Among these, recent work has revealed that the neural pathway from auditory cortex to the striatum plays a critical role in action selection during an auditory discrimination task (Znamenskiy and Zador 2013). This projection or any of the other ones may be responsible for transforming the expected information from primary auditory cortex into motor commands.

Altogether, we identify a circuit in Au1 that includes L2/3, L4, and L5 neurons and that encodes both actual and expected sensory stimulation. Following a conditioning training protocol involving rhythmic auditory stimulation, a subgroup of neurons in this circuit was activated at the expected time interval in the absence of sensory stimulation. This echo response may reflect a transient storage of rhythmic sensory information in Au1. A local recurrent network model with scale-free topology matches well these results and offers a possible mechanism for holding information transiently by activating a specific loop in L5 and L2/3 neurons (Supplementary Fig. 19). Still, we cannot exclude the possibility of the involvement of long-range recurrent interactions between higher order cortices (e.g., prefrontal cortex) and Au1. Further studies on the cellular and circuit mechanisms of this echo signal are needed. Moreover, the group of neurons encoding expected sensory stimulation was essential for transmitting the autonomously generated activity in the Au1 to other regions and for driving the associated motor output. Therefore, primary sensory cortex may provide a general pathway for the transient storage of rhythmic sensory information, and potentially for the selection of which predictive activities would be translated into a motor action.

**Supplementary Material**

Supplementary material are available at Cerebral Cortex online.

**Funding**

The 1000 Talents Program for Young Scholars; the Natural Science Foundation of China (No. 81671106, 31400933, 31371113); the National Basic Research Program of China ("973 Program": 2015CB759500, 2014CB541600). X.C. is a junior fellow of the CAS Center for Excellence in Brain Science and Intelligence Technology. I.N. holds the Milton z"l and Brindell Gottlieb Chair in Brain Sciences.

**Notes**

We thank J. Lou for assistance in composing figures, X.Q. Wang and Z. Varga for valuable discussions, Y. Zhou for the help of sound level calibration. Conflict of Interest: None declared.

**References**


Lee CC, Sherman SM. 2010a. Drivers and modulators in the central auditory pathways. Front Neurosci. 4:79–86.


Primary auditory cortex is required for anticipatory motor response

Jingcheng Li¹,², Xiang Liao¹, Jianxiong Zhang¹, Meng Wang¹,³, Nian Yang², Jun Zhang², Guanghui Lv³, Haohong Li³, Jian Lu¹, Ran Ding¹, , Xingyi Li¹, Yu Guang¹, Zhiqi Yang¹, Han Qin¹, Wenjun Jin¹, Kuan Zhang¹, Chao He², Hongbo Jia⁴, Shaoqun Zeng³, Zhian Hu², Israel Nelken⁵ & Xiaowei Chen¹,⁶

¹Brain Research Center and State Key Laboratory of Trauma, Burns, and Combined Injury, Third Military Medical University, Chongqing 400038, China.
²Department of Physiology, Third Military Medical University, Chongqing 400038, China.
³Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China.
⁴Brain Research Instrument Innovation Center, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, Jiangsu, China
⁵Department of Neurobiology, Silberman Institute of Life Sciences and the Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem 91904, Israel.
⁶CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

Address correspondence to Xiaowei Chen or Xiang Liao, Brain Research Center, Third Military Medical University, Chongqing 400038, China. Email: xiaowei_chen@tmmu.edu.cn or xiang.liao@aliyun.com

Running title: Anticipatory response in auditory cortex
Supplementary Materials

Supplementary Figure 1. Head post and its fixation device for two-photon imaging and behavior/fiber recording/optogenetic experiments in head-fixed awake mice.

(A) Head post and its fixation device for two-photon imaging in awake mice. From left to right: plastic fixation device used for fixing the head post to the imaging setup, head post used for fixing the mouse head to the fixation device, head post skewed to the fixation device, an awake mouse fixed to the imaging setup using this head post. (B) Head post and its fixation device for behavior/fiber recording/optogenetic experiments in awake mice. From left to right: fixation device used for fixing the head post, head post (the same as that in panel A), head post skewed to the fixation device, an awake mouse fixed with this device during behavior tests.
Supplementary Figure 2. Two-photon imaging of spontaneous and sound-evoked responses of L2/3 neurons in Au1.

(A) A merged image showing neurons stained with Cal-520 AM in green and glial cells stained with both Cal-520 AM and sulforhodamine 101 (SR101, red) in yellow in vivo. (B) Simultaneous two-photon Ca\(^{2+}\) imaging and cell-attached recording in vivo. The glass electrode is indicated by an arrowhead. (C) Simultaneously-recorded spontaneous Ca\(^{2+}\) transients (lower) and their corresponding action potentials (upper) from the neuron in vivo indicated in panel B. The numbers of action potentials are indicated above the electrical trace. (D) Detection rate of spontaneous Ca\(^{2+}\) transients in vivo for different numbers of action potentials (n = 7 neurons). The time window of detection was 200 ms. (E) Two-photon image of Cal-520 AM-labeled cells at a depth of 249 µm from the pial surface. (F) Spontaneous (left) and sound-evoked responses (right) of seven neurons and the neuropil as indicated in panel E. The vertical gray bars denote broadband noise stimuli. (G) No contamination of somatic Ca\(^{2+}\) signals of neurons by nearby neuropil. Data were from 15 neurons and their corresponding 15 adjacent regions of neuropil (3 mice). Upper panel, 109 traces from both neurons and neuropil were superimposed respectively. Lower panel, 67 traces from both neurons and neuropil were superimposed respectively. Each Ca\(^{2+}\) trace was selected from the period when either neuropil showed signals (upper) or neurons showed signals (lower). (H) Distribution of success rate for evoked responses to broadband noise (n = 486 neurons from 4 mice).
Supplementary Figure 3. No significant fraction of cells in Au1 L2/3 that follows the ISI change.

(A) Upper: raster plot of onsets of Ca$^{2+}$ transients before, during and after sound stimulus sequence at ISI 2/3 s. Lower, averaged Ca$^{2+}$ responses of all imaged neurons. The number of mice or neurons is indicated above the raster plot. (B) Fraction of cells showing echo responses at the first echo time point (3 s after sound offset) and the second echo time point (5 s after sound offset) versus its baseline level calculated from the time window before stimulus sequence (10 time windows that were selected for this calculation). Permutation test, ns, ns denote no significance (n = 669 neurons).
Supplementary Figure 4. Echo responses in Au1 neurons induced by different numbers of sound stimuli.

(A) Averaged Ca$^{2+}$ responses from all the imaged neurons by 5, 10, or 30 sound stimuli. (B) Comparison of response amplitude between echo response and the corresponding baseline in conditions of 5, 10, 30 sound stimuli. Wilcoxon signed rank test, 5 stimuli: ns (df = 479, z = 0.5028), 10 stimuli: ***P < 0.001 (df = 469, z = -12.5038), 30 stimuli: ***P < 0.001 (df = 543, z = -11.3509). Error bars represent SEM. (C) Fraction of neurons showing echo response after 10, 20, 30 sound stimuli. 95% confidence intervals of the fraction of cells were calculated by bootstrapping. (D) Success rate of sound response in 'echo neurons' and 'non-echo neurons'. Non-echo neurons are the neurons that did not respond during the echo time point, including both sound responding cells and silent cells. Wilcoxon rank sum test, ***P < 0.001 (df = 1261, z = 3.8526). Error bars represent SEM.
Supplementary Figure 5. Echo responses in Au1 neurons with different ISIs.

(A) Summary of Ca^{2+} transients in all imaged 496 neurons in 4 mice at an ISI of 3 s. Upper, plot of onsets of Ca^{2+} transients. Lower, averaged Ca^{2+} responses of all neurons. (B) Comparison of response amplitude between echo response and the corresponding baseline in the condition of 3 s ISI. Wilcoxon signed rank test, ***P < 0.001 (df = 495, z = -13.1676). (C) comparison of echo Ca^{2+} transient latency to last sound at ISIs of 2 s, 3 s, and 4 s. Wilcoxon rank sum test, ISI 2s vs 3s: ***P < 0.001 (df = 278, z = -13.7079), ISI 3s vs 4s: ***P < 0.001 (df
(D) Plot of echo response latency per animal. The dark green, red and blue circuits donate neurons imaged with 2 s, 3 s and 4 s ISI respectively in anesthetized mice, the light green circuits represent neurons imaged with 2 s ISI in awake mice. The average latency of all 536 neurons is 14 ms. (E) Fraction of neurons showing echo response with 2 s, 3 s and 4 s ISI respectively. The 95% confidence intervals of the fraction of cells were calculated by bootstrapping. (F) Comparison of Ca²⁺ transient amplitude at echo time point and that of the corresponding baseline with 1 s, 6 s, 10 s ISI respectively. Wilcoxon signed rank test, ISI 1s: ns (df = 483, z = 0.7315), ISI 6s: ns (df=221, z = 1.6966), ISI 10s: ns (df = 251, z = -0.5707), ns denote no significance. n = 484 neurons from 4 mice for 1 s, 222 neurons from 2 mice for 6 s, 252 neurons from 4 mice for 10 s.
Supplementary Figure 6. No dependence of echo response on stimulus type and intensity.

(A) Upper, raster plot of onsets of Ca\(^{2+}\) transients before, during and after pure tone (12 kHz) stimulus sequence at ISI 2 s. Lower, averaged Ca\(^{2+}\) responses of all imaged neurons. Left, 0 dB SPL attenuation (0 dB). Right, 40 dB SPL attenuation (-40 dB). The number of mice or neurons is indicated above the raster plot. (B) Comparison of Ca\(^{2+}\) transient amplitude at echo time point and that of the corresponding baseline for 12 kHz at three sound intensities (0, 20, 40 dB SPL attenuation). Wilcoxon signed rank test, 0 dB: ***P < 0.001 (df = 360, z = -5.4603), -20 dB: ***P < 0.001 (df = 360, z = -7.9016), -40 dB: ***P < 0.001 (df = 360, z = -9.0458), n = 361 imaged neurons. (C) The summary of Ca\(^{2+}\) transient amplitude for pure tones at five frequencies. Wilcoxon signed rank test, 0 dB: ***P < 0.001 (df = 360, z = -7.7739), -20 dB: ***P < 0.001 (df = 360, z = -5.2675), -40 dB: **P < 0.01 (df = 360, z = -2.9769), n = 361 imaged neurons. (D) Fraction of neurons showing echo response for 12 kHz at three sound intensities. The 95% confidence intervals of the fraction of cells were calculated by bootstrapping. Permutation test, 0 dB vs -20 dB: ns, -20 dB vs -40 dB: ns, n = 361 imaged neurons. (E) The summary of fraction of neurons for pure tones at five frequencies. Permutation test, 0 dB vs -20 dB: ***P < 0.001, -20 dB vs -40 dB: ns, n = 361 imaged neurons.
Supplementary Figure 7. Echo responses in Au1 neurons in awake mice.

(A) Infrared image of a head-fixed mouse. (B) Left panel, whole-cell patch-clamp recording in a neuron in the awake mouse. Right panel, distribution of membrane potential (binned at 1 ms) from the same neuron in the left panel. Inset: summary of peak membrane potentials from 5 neurons in 3 awake mice. (C) Summary of Ca$^{2+}$ response of L2/3 neurons before, during, and after 20 repetitive sound stimuli (n = 855 imaged neurons from 8 awake mice). Top, raster plot of Ca$^{2+}$ signals. Middle, plot of onsets of Ca$^{2+}$ transients. Bottom, number of Ca$^{2+}$ transients, binned with 250 ms. (D) Autocorrelation of the Ca$^{2+}$ trace before (left), during (middle), and after sound stimuli (right) from the neurons shown in panel C. The autocorrelation functions were constructed by calculating the auto-correlation for each single cell and then averaging across cells.
Supplementary Figure 8. Echo responses in both pyramidal and GABAergic neurons.

(A) Two-photon image of L2/3 neurons labeled with Cal-520 AM in a GAD67-GFP mouse. Neurons labeled with both GFP and Cal-520 AM are shown in yellow. Cells in green are GFP− cells. (B) Example of both GFP+ and GFP− neurons (outlined in panel A) showing echo responses. (C) Fraction of GFP+ and GFP− neurons with or without echo response. n = 430 neurons from 4 mice.
Supplementary Figure 9. Two-photon imaging of spontaneous and sound-evoked responses in L5 neurons of Au1.

(A) Schematic of dye loading or virus injection into L5 of Au1. (B) Confocal image of expression of AAV-syn-EGFP and DAPI staining in L5. (C) Two-photon image of Cal-520 AM-labeled L5 neurons in L5 at a depth of 568 μm from pial surface. (D) Spontaneous activity (left) and sound-evoked activity (right) of 6 neurons and neuropil as indicated in panel C. (E) Distribution of success rate of sound-evoked response in L5 neurons to broadband noise stimulation. n = 1500 neurons from 8 mice.
Supplementary Figure 10. Echo responses in Au1 L4 neurons.

(A) Summary of Ca$^{2+}$ responses of L4 neurons before, during, and after 20 repetitive sound stimuli (n = 1934 imaged neurons from 13 mice). First row, raster plot of Ca$^{2+}$ signals. Second row, averaged Ca$^{2+}$ responses of all the imaged neurons. Third row, plot of onsets of Ca$^{2+}$ transients. Fourth row, number of Ca$^{2+}$ transients, binned with 250 ms. (B) Fractions of echo neurons in L4. Error bar represents 95% confidence interval of the proportion. (C) Comparison of response amplitude between echo response and the corresponding baseline. n = 1934 imaged neurons. Wilcoxon signed rank test, ***$P < 0.001$ (df = 1933, z = -10.1978) (D) Latency of the first echo response to the last auditory stimulus (upper) or to the echo time point (lower). Red circles represent individual data points; black circles represent the mean values. n = 167 echo neurons. (E) Fraction of non-responder to sound stimulation in L2/3, L4 and L5 respectively. (F) Comparison of responsive success rate to sound of Au1 L2/3, L4 and L5 responsive neurons. Wilcoxon rank sum test, L2/3 vs L4: **$P < 0.01$ (df = 1951, z = 2.8789), L4 vs L5: ***$P < 0.001$: (df = 3029, z = -20.9763), L2/3 vs L5: ***$P < 0.001$: (df = 1874, z = -10.3394).
Supplementary Figure 11. Pharmacological inactivation of Au1 blocked sound-evoked licking responses.

(A) Schematic of drug injection (ACSF or ACSF+muscimol) into Au1 (upper) and post-hoc histology for verification of the injection site after injection of muscimol mixed with OGB-1 AM (lower). The ACSF without or with muscimol (total volume: 0.5 µl) was applied through a glass electrode with a tip resistance of ~1 MΩ. (B) Example of licking responses to sound stimulation in the presence of ACSF (upper) or muscimol (lower). (C) Summary of success rate of sound-evoked licking responses in two groups. n = 8 mice for ACSF group and muscimol group respectively. Wilcoxon signed rank test, **P < 0.01 (df = 7, z = 2.5846).
Supplementary Figure 12. No echo licking response following five broadband noise stimuli.

(A) Licking response to 5 broadband noise stimuli. The upper one is a single trial, and the lower trace is an average trace of 5 mice. Note that no echo licking response was induced. (B) Distribution of licking duration induced by broadband noise (n = 144 events from 8 mice).
**Supplementary Figure 13.** Licking-related neuronal activity in rostral and caudal parts of the second somatosensory cortex (S2).

**A,C** Coronal slice showing the caudal (A) or the rostral (C) S2 labeled with OGB-1AM. **B,D** Simultaneously recorded neuronal activity in S2 (upper), in Au1 (middle), and licking response (lower) in trained mice, corresponding to A and C respectively.
Supplementary Figure 14. Histology and *in vivo* electrophysiological recordings of PV-ChR2 expressing neurons.

(A) Left, a coronal slice showing the expression of AAV-EF1α-DIO-ChR2-mCherry in Au1 of a Pvalb-cre transgenic mouse. Right, confocal image showing the PV-ChR2 positive neurons in an expanded scale. (B) Raster plot of onsets of spikes during delivering pulses of blue light. These experiments were performed in four PV-ChR2 positive cells with cell-attached recordings *in vivo*. 
Supplementary Figure 15. Temporal relationship between echo Ca^{2+} signal and echo licking.

(A) Two examples from two different mice showing the temporal relationship between echo Ca^{2+} signal and echo licking response. (B) Plot of latency from the onset of echo Ca^{2+} signal to the onset of echo licking response. Inset shows mean ± SEM of latency. Optical fiber-based Ca^{2+} signals were recorded simultaneously with licking behavior. n = 32 trials from 8 mice, the same data set as shown in Figure 3G.
Supplementary Figure 16. Transgenic mouse line and the profile of light intensity for optogenetic experiments.

(A) Sagittal brain section of a Thy1-ChR2-YFP mouse. (B) Normalized light intensity (mW mm\(^{-2}\)) as a function of cortex depth. Inset shows a schematic of optogenetic stimulation in Au1 L5. We estimate that the distance from the fiber tip to the CA1 pyramidal layer of hippocampus (HP) is 0.8-1.0 mm and the light intensity that reaches the CA1 pyramidal layer is less than 2 mW mm\(^{-2}\). We made this curve according to the procedure described here: https://web.stanford.edu/group/dlab/cgi-bin/graph/chart.php. We used the following parameters: light wavelength: 473 nm; fiber numerical aperture (NA):0.48; light intensity from fiber tip: 12.5 mW; fiber core radius: 0.2 mm. The principle ideas for the calculation are as follows: When light propagating in the brain tissue, light intensity attenuates mainly because of tissue scattering. According to Kubelka-Munk model the light intensity in different depth can be calculated.

\[
P = \frac{P_0}{Sz + 1}
\]

Where \(P_0\) is the light intensity at the tip of optical fiber. \(S\) is the scattering coefficient per unit thickness and \(z\) is the depth of the sample. Here, we define a constant which is decided by the parameters of the optical fiber.

\[
\rho = r \sqrt{\left(\frac{n}{M}\right)^2 - 1}
\]

Where \(n\) is the refraction index of gray matter. NA is the numerical aperture and \(r\) is the radius of optical fiber. So the light surface density in different depth is decided by both light scattering and geometrical propagation.

\[
\frac{I(z)}{I_0} = \frac{\rho^2}{(Sz + 1)(z + \rho)}
\]

\(I(z)\) is the light surface density at the depth of \(z\) of sample. \(I_0\) is light surface density at the tip of optical fiber. Using this principle, we can calculate light surface intensity at any depth of the sample.
Supplementary Figure 17. Functional verification of ChR2 expression in Thy1-ChR2-YFP mice.

(A) Left, schematic of extracellular electrical recording in L5 while stimulating with 473 nm laser. Middle, local field electrical response to 200 ms light stimulation. Five consecutive trials are shown. Right, a single response from the dashed box in the middle panel, shown in an expanded scale. (B) Left, schematic of simultaneous optical recording and stimulation in L5 through the same fiber. Middle, Ca²⁺ signals from a local region by 10 ms light stimulation. Five consecutive trials are shown. Right, a single response from the dashed box in the middle panel, shown in an expanded scale. The light stimulation artifacts are removed for clarity. These experiments were repeated in 4 mice.
Supplementary Figure 18. Echo EPSP responses in layer 5 neurons following rhythmic current-injection stimulation in vitro.
(A) One example layer 5 (L5) neuron in Au1 recorded from an acute slice of an adult mouse, showing depolarization + spike responses by current injections (+300 pA, 100 ms) at an ISI of 2 s. One to three spikes were reliably evoked in this neuron. (B) A single response expanded from that pointed out by an asterisk in panel A. (C) Left, 4 single trials from a L5 neuron following 20 rhythmic current-injection stimuli, showing small depolarization responses (EPSPs; marked in red) around the echo time point. Right, histograms summarizing the number of depolarization events from 8 consecutive trials in this neuron. (D) Summary of depolarization responses (EPSPs; black dots) and action potential responses (AP; grey dots) in all the recorded L5 neurons before, during and after rhythmic current-injection stimulation. Top, plot of onsets of EPSPs (black dots) and APs (grey dots). Three trials from one neuron are shown. Bottom, histograms showing the number of EPSPs (blue) and APs (grey). The number of APs is truncated for clarity. (E) Histograms showing the number of EPSPs around the echo time point, expanded from panel D. Wilcoxon signed rank test, one-tailed, n = 29 neurons, \( P = 0.0196 \) (df = 28, \( z = -2.0617 \)). (F) The same experimental arrangement as in panel C in the presence of APV (20 \( \mu \)M) and CNQX (40 \( \mu \)M). The histograms showing the number of depolarization events from 10 consecutive trials in this neuron. (G) Histograms summarizing the number of depolarization events from 13 neurons in L5 in the presence of APV and CNQX. Note no significant increase in the number of depolarization events within a time window from -0.5 s to 0.5 s (around the echo time point), as compared to that in the time window from -1.5 s to -0.5 s (preceding the echo time point). Wilcoxon signed rank test, one-tailed, n = 13 neurons, \( P = 0.6250 \) (df = 12, \( z = 0.1414 \)). (H) Histograms summarizing the number of depolarization events from 19 neurons in L2/3. Note no significant increase in the number of depolarization events within a time window from -0.5 s to 0.5 s, as compared to that in the time window from -1.5 s to -0.5 s. Wilcoxon signed rank test, one-tailed, n = 19 neurons, \( P = 0.9743 \) (df = 18, \( z = 1.9480 \)).
Supplementary Figure 19. Working model for generating echo response in cortical circuits.

Left, during actual stimulation, sensory input signals flow from L4 to L2/3 and then to L5 in primary sensory cortex. L5 serves as a reservoir of many low-degree loops with different sizes, as L5 neurons usually fire more but connect less extensively to the network than L2/3, while L2/3 contains hub neurons. Right, following rhythmic sensory stimulation, the system selects a loop in L5 with a proper size that matches the interval of the trained rhythm. In addition, the connectivity from L5 to L2/3 hub neurons has been potentiated, which is essential for transmitting signals from the selected L5 loop (pacemaker-like loop) to L2/3 and then to the entire network. According to the intrinsic properties of cortical circuits and the need for behavioral actions, this proposed circuit probably can only follow a certain rhythm, e.g. 0.25 to 0.5 Hz (slow oscillation) in our study.
<table>
<thead>
<tr>
<th></th>
<th>Echo licking</th>
<th>No echo licking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Echo Ca^{2+} response</td>
<td>37</td>
<td>3</td>
</tr>
<tr>
<td>No echo Ca^{2+} response</td>
<td>9</td>
<td>39</td>
</tr>
</tbody>
</table>

**Supplementary Table 1.** Two-way table for association between echo licking and echo Ca^{2+} response. The data were from fiber photometry experiments (8 mice).
SUPPLEMENTARY MOVIES

Supplementary Movie 1. An example of echo licking response following rhythmic sound stimulation in a trained mouse.

The last 5 sound stimuli and 2 echo time points are shown (ISI = 4 s). Yellow vertical lines denote sound stimuli; red vertical lines denote echo time points.

Supplementary Movie 2. An example of echo licking response following rhythmic light stimulation.

This video includes two parts. The first part shows echo licking response in a trained mouse. The second part shows no licking response in a naïve mouse.