Title: **Motor corollary discharge activates layer six circuits in the auditory cortex**

Author names: Kameron K. Clayton[1*], Ross S. Williamson†[1,2], Yurika Watanabe[1], Kenneth E. Hancock[1,2], Gen-ichi Tasaka[3], Adi Mizrahi[3], Troy Hackett[4] and Daniel B Polley*[1,2]

1- Eaton-Peabody Laboratories, Massachusetts Eye and Ear Infirmary, Boston MA 02114 USA
2- Dept. Otolaryngology, Harvard Medical School, Boston MA 02114 USA
3- The Edmond and Lily Safra Center for Brain Sciences, Dept. Neurobiology, Hebrew University of Jerusalem, Jerusalem ISR
4- Dept. Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville TN, USA
* Corresponding author

*Correspondence: daniel_polley@meei.harvard.edu
†Current address: Dept. Otolaryngology, University of Pittsburgh, rsw@pitt.edu

Keywords: motor preparatory, corticothalamic, sensorimotor integration, globus pallidus, FoxP2

SUMMARY

During active sensing, neural responses to sensory inputs directly generated by our own movements are suppressed. In the auditory cortex (ACtx), self-initiated movements elicit corollary discharge from secondary motor cortex (M2) that suppresses pyramidal neuron (PyrN) spiking via recruitment of local inhibitory neurons. Here, we observed that ACtx layer (L)6 PyrNs were also activated hundreds of milliseconds prior to movement onset, at approximately the same time as fast spiking inhibitory neurons. Most L6 PyrNs were corticothalamic (CT) cells, which all expressed FoxP2, a protein marker enriched in brain areas that integrate sensory inputs to control vocal motor behaviors. L6 CTs were strongly activated prior to orofacial movements, but not locomotion, and received ten times more direct inputs from the basal ganglia than M2. These findings identify new pathways and local circuits for motor modulation of sound processing and suggest a new role for CT neurons in active sensing.

INTRODUCTION

When scanning a crowded room or fumbling with a button on a shirt cuff, reafferent sensory inputs created directly from the movements of the eyes or hands are canceled out to accentuate the environmental stimuli needed to guide behavior. Canceling out reafferent sensory input is no less important in the auditory system, but can only be precisely accomplished in special circumstances where the acoustic consequences of movement are nearly immediate and highly stereotyped (Eliades and Wang, 2003, 2008; Keller and Hahnloser, 2009; Poulet and Hedwig, 2002; Schneider et al., 2018; Singla et al., 2017). In most cases, the timing and acoustics of sounds generated by our movements are not fully predictable, and perhaps for this reason self-initiated movement is associated with a generalized suppression of sound-evoked activity throughout the auditory pathway (Schneider and Mooney, 2018).

Neural suppression during movement has been studied most extensively at the level of the auditory cortex (ACtx) and reflects at least three separate components: *i*) a subcortical pathway, *ii*) a cholinergic basal forebrain pathway, *iii*) a motor cortex pathway. Sub-cortical attenuation of sound-evoked activity during movement has been observed in the middle ear (Carmel and Starr, 1963), cochlear nucleus (Singla et al., 2017), inferior colliculus (Chen and Song, 2019; Yang et al., 2020) and medial geniculate body (McGinley et al., 2015; Rummell et al., 2016; Williamson et al., 2015), but is relatively subtle and can account for less than 40% of the total suppression observed in cortical neurons (Rummell et al., 2016; Schneider et al., 2014). Movement-related suppression of cortical PyrN activity is greatly enhanced through feedforward inhibition from local GABAergic interneurons (Schneider et al., 2014; Zhou et al., 2014). Cholinergic projections from the caudal substantia innominata (SIn) make monosynaptic connections onto ACtx inhibitory neurons to suppress cortical sound processing during states of movement or active listening, though SIn inputs tend to lag rather than lead movement initiation (Abs et al., 2018; Kuchibhotla et al., 2017; Nelson and Mooney, 2016). The best-defined pathway comes from neurons in secondary motor cortex (M2) that make monosynaptic connections onto local inhibitory neurons to suppress PyrN

responses prior to the initiation of self-generated sounds (Nelson et al., 2013; Schneider et al., 2014, 2018).

In addition to their monosynaptic inputs onto local GABA cells, M2 afferents also innervate deep layers of ACtx and reliably elicit excitatory synaptic currents in L6 PyrNs (Nelson et al., 2013). In mice, most L6 PyrNs are corticothalamic (CT) neurons that powerfully regulate sensory gain across all layers of the cortical column, mostly through their intracolumnar synaptic connections rather than modulation of thalamic sound processing (Bortone et al., 2014; Olsen et al., 2012). In ACtx, optogenetic activation of L6 CTs recruit transcolumnar activation of local PyrNs as well as intense activation of fast-spiking (FS) interneurons (Guo et al., 2017; Williamson and Polley, 2019). Interestingly, the strongest modulation of cortical sound processing and sound perception occurs after optogenetically induced L6 CT spiking has ended, not during activation (Guo et al., 2017). This raised the possibility that L6 CT spiking might naturally precede the onset of upcoming sounds, as in the case of sounds produced by movement. However, any role for L6 CTs in active sensing is purely speculative, as targeted recordings of this cell type have never been made in behaving animals. Here, we found that L6 PyrNs – including L6 CTs – are activated by motor corollary discharge hundreds of milliseconds prior to movement. L6 CT activation was elicited by motor corollary inputs that precede licking, but not bouts of running. Further, we found only very sparse monosynaptic inputs from M2 onto L6 CT cells, but instead observed numerous monosynaptic inputs from the globus pallidus. These findings suggest new pathways and local circuits for modulating cortical sound processing during self-initiated movements.

RESULTS

Movement impairs sound frequency discrimination

As a first step, we wanted to better establish the association between movement and psychophysical measures of sound perception. A few studies have reported that movement-induced ACtx suppression elevated behavioral detection thresholds for unpredictable sounds at low signal-to-noise ratios (SNR) (McGinley et al., 2015; Schneider et al., 2018). While ACtx suppression clearly interferes with detection of faint sounds, in some cases cortical suppression supports more accurate discrimination of similar sound frequencies (Blake et al., 2006; Christensen et al., 2019; Guo et al., 2017; Maor et al., 2020). To assess whether the deleterious effects of movement on sound perception extended beyond sound detection paradigms, we trained head-fixed mice in a Go/NoGo frequency recognition task (**Fig. 1A**). Mice licked the initiation spout to start a trial, which triggered the delivery of the target (12 kHz) or non-target frequency tone 1.5s later. Mice licked the decision spout following the target frequency to receive a water reward (**Fig. 1A**). Mice were reliably under stimulus control within 1-2 weeks of shaping, and produced psychometric functions comparable to other rodents tested with this paradigm (Polley et al., 2006) (**Fig. 1B**).

Mice exhibited substantial session-to-session variation in lickspout contact during the 1.5s foreperiod separating the initiation and decision periods (N=12 mice, n=96 test

sessions; **Fig. 1C**). Examining the trial lick rasters revealed that behavioral sessions with poor discrimination thresholds were associated with persistent licking of the initiation spout throughout the foreperiod separating initiation and tone presentation (**Fig. 1D, left).** In sessions with better discrimination thresholds, mice briefly licked the initiation spout and then paused before the target or non-target tone was presented (**Fig. 1D, right**). Across all behavioral sessions, initiation spout lick rate during the foreperiod was negatively correlated with the discrimination threshold, even after controlling for other mitigating variables including the individual mouse and the number of prior testing sessions (partial correlation in the multi-variate linear regression model; $r = -0.33$, $p = 0.001$; **Fig. 1E**).

Suppressed cortical processing of lick-triggered sounds

Our behavioral findings suggest that auditory discrimination performance is impaired by motor activity preceding and impinging upon the sound delivery period. However, it was difficult to relate poor performance to foreperiod movement *per se,* as exuberant licking might have been the byproduct of heightened thirst, arousal, impatience, procedural confusion, or any number of other cognitive states associated with excess false positive rates. To focus only on movement-related modulation of neural sound processing, we switched to a minimalistic behavior in which delivery of a broadband sound was either triggered by initiation of a lick bout or was presented in the absence of licking (**Fig. 2A**). Lick bouts had a 50% probability of generating a sound and a 50% probability of resulting in reward (each independently determined). Continuous background masking noise obscured sound arising from the lick itself. From prior work, we anticipated that motor corollary discharge from M2 would recruit A1 inhibitory neurons (e.g., FS neurons) that would suppress PyrN activity, thereby reducing the magnitude of lick-triggered sound responses (**Fig. 2B**) (Rummell et al., 2016; Schneider et al., 2014, 2018). Two-photon calcium imaging of L2/3 PyrNs confirmed that lick-triggered sound responses were significantly reduced compared to sounds presented at rest (paired t-test, $p < 1 \times 10^{-6}$; **Fig. 2C-D**).

Suppression of sound-evoked cortical responses during movement has now been replicated by many different laboratories, yet the particular form of the suppression across layers and between regular spiking (RS) PyrNs versus FS units is unresolved (Bigelow et al., 2019; McGinley et al., 2015; Rummell et al., 2016; Schneider et al., 2018; Zhou et al., 2014). To compare laminar differences and to address uncertainties about FS responsiveness during movement, we recorded from FS and RS putative PyrNs from all layers of A1 with a 64-channel high-density linear probe during the lickspout sampling task (**Fig. 2E**). Significant movement-related suppression was observed both in RS and FS units (one-sample t-test against a population mean of 0, $p < 0.0005$ for both RS and FS, $n = 125$ and 52, respectively from 7 mice) with no difference in the degree of suppression between cell types (unpaired t-test for RS vs FS, $p = 0.33$; **Fig. 2F**). Laminar differences went in the opposite direction of prior work,

in that we observed weaker and more variable spiking suppression to self-generated sounds in deeper cortical layers than more superficial layers (**Fig. 2G**) (Rummell et al., 2016; Schneider et al., 2018).

Orofacial movements activate deep layer A1 neurons

Because firing rates vary from cell to cell, sensory-evoked responses are generally measured by subtracting each unit's pre-stimulus baseline activity from the post-stimulus response. Sensory-evoked suppression could therefore reflect elevated pre-stimulus activity or reduced post-stimulus activity. A closer inspection of units recorded in our study revealed examples of both (**Fig. 3A**). We computed the movement modulation index separately on both pre-stimulus and post-stimulus firing rates and observed a significant suppression of lick-triggered sound responses (**Fig. 3B, top**) but also a significant elevation of spiking during the baseline period before the lick-triggered sound (one-sample t-tests against a population mean of 0, $p = 0.00005$ and 0.003, respectively, **Fig. 3B, bottom**). This was puzzling, as M2 inputs suppress PyrNs prior to bouts of locomotion through recruitment of local inhibitory neurons (Schneider et al., 2014). Referring to the schematic in Fig. 2B, elevated firing rates during and preceding the stimulus are expected in inhibitory neurons, while baseline activity in RS PyrNs (which comprise the majority of our sample), should be suppressed, not enhanced.

To characterize the laminar and cell-type specificity of motor-related signals on A1 firing rates, we focused on lick-only trials in which lick bout initiation did not trigger a sound. Significantly suppressed spiking during licking was observed in roughly 25% of units and was expressed similarly across layers (n = 28, 43, 132 and 135 units in L2/3, L4, L5 and L6 respectively; ANOVA, main effect for layer $F = 1.19$, $p = 0.7$; **Fig. 3C**). Firing rates were significantly increased during licking in approximately 40% of single units, but particularly in L6, where spiking ramped up hundreds of milliseconds before contact with the lickspout and was significantly greater than other layers (ANOVA, main effect for layer, $F = 8.55$, $p < 0.00003$; post-hoc pairwise comparisons with Holm-Bonferroni correction for multiple comparisons, $p = 0.02$, 0.00009 and 0.02 for L6 vs L5, L4 and L2/3, respectively). Averaging over all functional response types, we confirmed prior reports that firing rates of FS units increased over a 1s period prior to movement onset without any systematic differences across layers (n = 79, Mixed design ANOVA, main effect for time and layer, $F = 4.03$ and 2.06, $p < 1 \times 10^{-6}$ and 0.11, respectively; time × layer interaction term, $F = 0.93$, $p = 0.65$; **Fig. 3D, top**). Although a minority of RS units were strongly suppressed prior to lick onset, the overall effect was an increase in firing rate (main effect for time, $F = 4.29$, $p < 1 \times 10^{-6}$), that was mainly attributable to strong motor-related activation of L6 RS units (n = 236, main effect for layer, $F = 2.66$, $p = 0.04$; time × layer interaction term, $F = 1.58$, $p < 0.001$; **Fig. 3D, bottom**).

To determine the latency of motor-related changes in spiking, we followed the analysis approach of prior reports and replicated their finding that FS units recruited by movement first begin to increase their spiking hundreds of milliseconds prior to movement onset (mean latency for L2/3, L5, L6 FS units = -292ms, -400ms, -396ms; **Fig. 3E**) (Schneider et al., 2014). Motor-related changes in RS unit spiking tended to lag behind FS units with one notable exception: L6 RS units began to increase their spiking at approximately the same latency as FS units (-349ms prior to lick spout contact; unpaired t-test for enhanced L6 RS vs. FS latencies, $p = 0.63$). In sum, our findings confirm that upcoming movements increase the firing rate of FS units and that the first wave of increased spiking begins 300-400ms prior to movement. To this, our experiments add something new: a subclass of L6 RS neurons that are also activated by upcoming movements with similar timing as local inhibitory neurons.

Overt orofacial movements lag behind motor-related changes in A1 spiking

Although A1 firing rates preceded lickspout contact by hundreds of milliseconds, it was unclear whether these changes reflect internalized motor corollary inputs or sensory reafference arising from orofacial movements that occur before the tongue first contacted the lick spout. We measured the timing of orofacial movements and pupil diameter with DeepLabCut, a video analysis method for markerless tracking of body movements based on deep neural networks (n = 12 videography sessions in 4 mice; **Fig. 4A**) (Mathis et al., 2018; Nath et al., 2019). We observed a reproducible sequence of orofacial movements associated with licking both within and between mice. Movement of the whisker pad, nose and lower jaw reliably preceded lickspout contact, whereas changes in pupil diameter and pinna movement lagged behind lick onset (**Fig. 4B**). Importantly, the earliest facial movements preceded lickspout contact by only ~30ms (**Fig. 4C**). As the earliest changes in A1 firing rates begin more than an order of magnitude earlier than the beginning of orofacial movements, they very likely reflect motor corollary discharge rather than sensory feedback arising from ongoing movements.

Functional and histological markers of motor processing in A1 L6 corticothalamic neurons

The majority of L6 cells in mouse A1 are corticothalamic (Guo et al., 2017). Neurotensin receptor 1-Cre (Ntsr1-Cre) expression is highly selective for L6 CT neurons (Bortone et al., 2014; Chevée et al., 2018; Gong et al., 2007; Guo et al., 2017; Kim et al., 2014; Olsen et al., 2012). Crossing Ntsr1-Cre mice to a Cre-dependent GCaMP reporter line (Ai148) labeled densely packed CT cell bodies in L6a and sparse, vertically oriented neuropil predominantly terminating in layer 5a (**Fig. 5A**). We immunolabeled A1 of Ntsr1-Cre x Ai148 mice with a neuron-specific marker (NeuN) and for the forkhead box protein P2 (FoxP2), and quantified co-expression between all three

markers in L6. FoxP2 is highly expressed in motor control nuclei, particularly in brain regions subserving audiomotor integration for speech production and vocal motor learning (French and Fisher, 2014). *FOXP2* mutations are linked to vocal communication disorders in humans (Lai et al., 2001), birds (Haesler et al., 2007) and to more generally disrupted audiomotor learning and sensorimotor integration in mice (Castellucci et al., 2016; Kurt et al., 2012). FoxP2 is highly expressed in A1, but only in layer 6 (**Fig. 5B**). We found that 74% of L6 FoxP2 neurons in A1 (1381/1875) are Ntsr1-Cre+ and – importantly – every Ntsr1-Cre+ neuron expressed FoxP2 (**Fig. 5C**). Given the strong recruitment of L6 RS units by orofacial movements and the well-established role for FoxP2+ cells in coordinating sensory inputs with orofacial movements, we next made targeted recordings from Ntsr1+ L6 CT neurons to characterize their recruitment by orofacial movements associated with licking.

We expressed channelrhodopsin-2 in A1 of Ntsr1-Cre mice and implanted an optic fiber such that the fiber tip terminated just above the dorsal surface of the medial geniculate body (MGB) (**Fig. 5D**). We optogenetically activated thalamic axon terminals of L6 CT neurons with a brief (1ms) flash of blue light and quantified the temporal jitter of antidromically activated spikes recorded from A1 cell bodies (**Fig. 5E**). As confirmed in our recent study, units with minimal trial-to-trial jitter in first spike latency were Ntsr1+ CT cells directly activated by the laser pulse, whereas higher jitter (>0.35 ms) were activated by polysynaptic pathways and were not necessarily CT units (**Fig. 5F**) (Williamson and Polley, 2019). Firing rates of optogenetically tagged L6 CT units were strongly modulated both before and during orofacial movements (n = 32, N = 6 mice; **Fig. 5G**). Latencies for motor related enhancement or suppression were not significantly different than unidentified L6 RS units, which is unsurprising given that the majority of L6 RS units should be CT neurons (two-sample K-S test on L6 CT vs L6 RS latency distributions, $p > 0.68$ for both enhanced and suppressed units; **Fig. 5H**).

Motor-related inputs from orofacial movements – but not locomotion – recruit A1 L6 corticothalamic neurons

Single unit recordings measure spiking at millisecond resolution, which is useful when calculating onset latencies for neural changes relative to overt movements. However, the yield in single unit recording experiments is much lower than 2-photon imaging, particularly with antidromic phototagging, where we optogenetically tagged less than one CT unit on average in any given recording session. Although the temporal dynamics and relationship to spike coupling with calcium indicators preclude any strong claims about the relative timing of neural activity and movement, we extended our motor characterization experiments with 2-photon imaging to simultaneously visualize larger neural ensembles (**Fig. 6A**).

Consistent with our small sample of L2/3 single unit recordings (Fig. 3C), orofacial movement elicited weak responses in L2/3 PyrNs, with 68% of cells not

significantly affected by movement, and the remainder of cells mostly activated or suppressed after the lick bout had begun (n = 624; N=5; **Fig. 6B**). Two-photon imaging of L6 cell bodies is impossible under most circumstances on account of fluorescence from more superficial layers that greatly reduces the SNR of somatic GCaMP signals. Because GCaMP expression is limited to L6a in Ntsr1-Cre x Ai148 mice and because these CT cells have sparse, slender, and short apical dendrites (Fig. 5A), there is very little out of plane fluorescence, permitting 2-photon calcium imaging at depths of ~0.7mm without using high excitation laser power (Daigle et al., 2018; Liang et al., 2019; Voigts et al., 2019) (**Fig. 6C**). In agreement with our phototagged recordings, we found that L6 CT neurons were strongly recruited by orofacial movements, where the activity rates of 74% of cells was significantly changed from baseline, mostly reflecting enhanced activity beginning prior to lickspout contact (n = 739; N = 3, **Fig. 6D**).

The most obvious difference between our paradigm and previous studies was that we focused on orofacial movements, whereas previous studies have used locomotion on a treadmill or other body movements. To determine if the type of movement accounted for the differences in our findings, we performed an additional 2-photon imaging experiment on ensembles of Ntsr1+ L6 CT neurons when mice initiated bouts of running on a treadmill. To our surprise, L6 CT activation prior to movement onset was virtually absent during treadmill running. Whereas 53% of L6 CT neurons showed significantly enhanced activity rates related to orofacial movement, only 15% showed significantly elevated activity around bouts of running (n = 606; N = 3, **Fig. 6E**). In sum, L6 CT PyrNs showed similar patterns of motor-related activity as their L2/3 PyrN counterparts when the source of movement changed from the face to the body (**Fig. 6F**).

Basal ganglia projections target A1 L6 corticothalamic neurons

These diverse forms of L6 modulation by upcoming movements are hard to reconcile with a single circuit involving excitatory inputs from M2 onto A1 inhibitory interneurons. One idea is that distinct anatomical pathways subserving different movement classes might converge on distinct local circuits to promote different forms of motor-related modulation. The globus pallidus (GP) is an obvious candidate based on many reports that firing rates of GP neurons are modulated before self-initiated movements (Dodson et al., 2015; Lee and Assad, 2003). The external GP borders the SIn and contains a sub-compartment that projects to neocortex, including targeting of deep layer neurons in primary somatosensory cortex (Abecassis et al., 2020; Saunders et al., 2015). As a next step, we determined whether GP or other regions of the basal ganglia make monosynaptic connections onto A1 L6 CT neurons.

We characterized long-range monosynaptic inputs onto A1 L6 CT neurons using RVdG-based trans-synaptic tracing (Wall et al., 2010; Wickersham et al., 2007). Ntsr1-Cre+ L6 neurons were transduced with Cre-dependent viral constructs to become

“starter cells” that express mCherry, the EnvA receptor TVA, and the rabies glycoprotein, G (AAV2-CAG-FLEx-TVA and AAV2-CAG-FLEx-oG, respectively; **Fig. 7A**) (Kim et al., 2016; Miyamichi et al., 2013). A second injection of an RVdG pseudotyped with an avian EnvA envelope protein (SADΔG-GFP) was made into the same A1 location to selectively transduce L6 CT starter neurons and drive GFP expression in their monosynaptic inputs. We confirmed that A1 L6 CTs are innervated by neighboring cortical regions, in addition to reciprocal inputs from ventral and medial divisions of the MGB, the same subdivisions targeted by their subcortical axons (Andersen et al., 1980; Winer et al., 2001) (**Fig. 7B, left**). As expected, we observed scattered inputs from SIn, but we also found projection neurons in more central regions of GP, even as far dorsolateral as the caudate/putamen (**Fig. 7B, right**). We quantified the brain-wide distribution of presynaptic inputs to A1 L6 CTs by sectioning at 40μm resolution and reconstructing the location of all starter and input neurons (n = 1,628 starters and 22,625 input neurons in five mice; **Fig. 7C**). As expected, the greatest number of inputs came from neighboring regions of ACtx, with a smaller fraction of inputs from a secondary visual field (**Fig. 7D**). Of motor-related brain regions, A1 L6 CTs received approximately ten times more input from GP than M2, suggesting an alternate pathway for routing motor corollary discharge into A1.

DISCUSSION

Here, we used a combination of psychophysics, 2-photon imaging, quantitative videography, optogenetically guided single unit recordings and RVdG input tracing to advance our understanding of how L6 PyrNs are recruited by motor corollary discharge to modulate processing of self-generated sounds. To a first approximation, our findings agree with the consensus view that sound-evoked activity is suppressed during movement (Fig. 2) and extends upon previous work in sound detection tasks by showing that frequency discrimination is also impaired by movement (Fig. 1) (McGinley et al., 2015; Schneider et al., 2018). Most paradigms have focused on central auditory suppression that occurs in the general context of ongoing locomotion, heightened arousal or task engagement (Bigelow et al., 2019; Buran et al., 2014; Carcea et al., 2017; Chen and Song, 2019; Jaramillo and Zador, 2011; Kuchibhotla et al., 2017; McGinley et al., 2015; Otazu et al., 2009; Schneider et al., 2014; Williamson et al., 2015; Yang et al., 2020; Zhou et al., 2014). Our findings are most directly comparable to studies that have characterized suppression of self-generated communication sounds (Eliades and Wang, 2003, 2008; Houde et al., 2002; Keller and Hahnloser, 2009; Poulet and Hedwig, 2002; Schneider et al., 2014) or – even more specifically – suppression of artificial sounds (e.g., tones or frequency modulated sweeps) generated by an arbitrary operant behavior (e.g., steps on a treadmill, bar presses, or licking) (Rummell et al., 2016; Schneider et al., 2018). In contrasting ACtx response to self-generated sounds

versus sounds with unpredictable timing or acoustics, these studies have collectively demonstrated that *i)* spike rates in local inhibitory neurons are increased hundreds of milliseconds prior to movement onset, *ii)* spike rates of PyrNs are suppressed prior to movement onset, *iii*) suppression of self-generated sounds is similar between RS/PyrNs and FS units; *iv*) suppression of self-generated sounds is greatest in deeper layers of ACtx.

Multiple pathways for audiomotor monitoring

Our interpretation is that we did not find anything contradictory to these findings, but rather that we observed something additional that had not been seen in prior studies either because they did not study orofacial or vocal behaviors or – if they did – they did not make a point of focusing their analysis on deep layer PyrN spiking (Eliades and Wang, 2003, 2008; Schneider et al., 2014). To this point, we replicated each of the above findings by reporting that: *i)* FS firing rates increased hundreds of milliseconds prior to lick onset (Fig. 3E), *ii)* many cases of suppressed PyrN firing prior to lick onset, particularly outside of L6 (Fig. 3C), *iii)* similar suppression of sound-evoked responses in FS and RS units (Fig 2F), and *iv)* many examples of deep layer units with suppressed sound-evoked responses (Fig. 2G). The major discovery here was the finding that a sub-population of deep layer PyrNs that essentially acted like motor neurons in the ACtx, in that they increased the firing rates hundreds of milliseconds prior to movement onset.

Our interpretation and overall hypothesis to be tested in future studies is that the net-suppressive influence of motor-related inputs from the cholinergic basal forebrain and M2 shape cortical responses during a broad class of self-initiated movements, but that orofacial movements (including, perhaps, vocalizations) recruit an additional pathway involving pallidocortical projections onto FoxP2+ deep layer PyrNs, which are mostly L6 CT neurons. A rigorous test of this hypothesis would involve targeted recordings from L6 CT ensembles across a much wider array of natural movements to distill the features that elicit activation versus suppression of L6 CTs. In addition, the molecular and chemical anatomy of the caudal basal forebrain needs to be better defined, as well as their cell type-specific projections to primary and non-primary areas of the ACtx. In rats, neurons in the external GP and caudate/putamen project to secondary auditory fields, but not A1, although the cell type-specific targets within the cortex have yet to be defined (Chavez and Zaborszky, 2017). The vast majority of cells that project from the caudal basal forebrain to A1 are cholinergic and are primarily localized to SIn, although there are a fringe of cholinergic cells extending dorsally and laterally that are not easily assigned to GP versus SIn (Abs et al., 2018; Chavez and Zaborszky, 2017; Guo et al., 2019; Hangya et al., 2015; Saunders et al., 2015). Here, we describe numerous pallidocortical inputs from the central GP that have not been described when injections are strictly limited to A1. Whereas cholinergic SIn neurons

that project to the ACtx predominantly receive inputs from arousal- and motor-related brainstem nuclei (Nelson and Mooney, 2016), we would expect many of these pallidocortical neurons to be GABAergic and receive both direct and indirect pathway input (Saunders et al., 2015). Future work can put this presumption (and others like it) to the test by carefully positioning tracers or viral reagents within primary or non-primary fields of the auditory cortex (Romero et al., 2019) in combination with transgenic approaches to restrict input mapping onto cell types of interest and high-throughput *in situ* sequencing to provide a more complete transcriptional profile of labeled input neurons (Chen et al., 2019).

Corticothalamic neurons and active sensing

L6 CTs are the largest component of the corticofugal projection system and one of the largest single class of projection neurons in the brain. The advent of optogenetic approaches to activate and silence L6 CT neurons in Ntsr1-Cre transgenic mice have breathed new life into research on CT circuits, inspiring new hypotheses about their role in sensory gain control and predictive coding (Crandall et al., 2015; Guo et al., 2017; Olsen et al., 2012; Vélez-Fort et al., 2014; Voigts et al., 2019; Williamson and Polley, 2019). Our prior work noted that optogenetic activation of L6 CTs reset the phase of low-frequency oscillations, modified sensory tuning in the thalamus and cortex, and switched perception into modes of enhanced detection or enhanced discrimination (Guo et al., 2017). Because the largest and clearest effects of L6 CT activation on thalamocortical tuning and perception occurred after photoactivated volleys of L6 spiking had ended, we speculated that L6 CT neurons might be naturally recruited to spike prior to anticipated sounds during active sensing. Here, we discovered that motor corollary inputs may fulfill this role by activating L6 CT neurons prior to sounds generated by orofacial movements. Preparatory facial movements provide a critical reference frame for anticipating the timing and acoustics of forthcoming vocalizations, provided there are neurons in the auditory system that can both register these motor triggers and also coordinate local circuits and ongoing rhythms to process predicted sound features (Morillon et al., 2015). L6 CTs fulfill both of these requirements, suggesting that they might be integrated into distributed motor control systems that support active sensing during vocal motor communication in mice (Okobi et al., 2019). Given the direct linkage between *FOXP2* mutations and monogenic speech and language disorders, one could predict that disruptions of ACtx L6 CT processing could have devastating consequences on integrating the timing of vocal motor signals with auditory feedback.

Acknowledgments: We thank Eyal Kimchi for guidance on using DeepLabCut, Huixin Xi for assistance with cell plotting, Evan Foss and Ishmael Stefanov for behavioral hardware development and Christine Liu for contributing to behavioral training; We thank D. Kim and the

Genetically Encoded Neuronal Indicator and Effector Project at the HHMI's Janelia Farm Research Campus for making the Thy1-GCaMP6s mouse publicly available. AAV5-Ef1a-DIO hChR2(E123T/T159C)-EYFP was developed by Karl Deisseroth. **Funding:** This work was supported by NIH grants DC017178 (DBP) DC015388 (TAH), ERC grant 616063 (AM) and DC018327 (RSW), NSF fellowship DGE1745303 (KKC), NIH fellowship DC015376 (RSW), [AM] [Troy?]); **Author contributions:** KCC and YW performed all experiments and analyses on live animals with supervisory input from RSW and DBP. TH performed imaging and analysis of fixed tissue. KEH developed neurobehavioral software control; GT and AM provided reagents and protocols for pseudotyped rabies viral tracing. DBP and KCC wrote the manuscript with feedback from all authors; **Competing interests:** Authors declare no competing interests.

REFERENCES

Abecassis, Z.A., Berceau, B.L., Win, P.H., García, D., Xenias, H.S., Cui, Q., Pamukcu, A., Cherian, S., Hernández, V.M., Chon, U., et al. (2020). Npas1+-Nkx2.1+ Neurons Are an Integral Part of the Cortico-pallido-cortical Loop. J. Neurosci. *40*, 743–768.

Abs, E., Poorthuis, R.B., Apelblat, D., Muhammad, K., Pardi, M.B., Enke, L., Kushinsky, D., Pu, D.L., Eizinger, M.F., Conzelmann, K.K., et al. (2018). Learning-related plasticity in dendrite-targeting layer 1 interneurons. Neuron *100*, 684-699.e6.

Andersen, R.A., Knight, P.L., and Merzenich, M.M. (1980). The thalamocortical and corticothalamic connections of AI, AII, and the anterior auditory field (AAF) in the cat: evidence for two largely segregated systems of connections. J Comp Neurol *194*, 663–701.

Bigelow, J., Morrill, R.J., Dekloe, J., and Hasenstaub, A.R. (2019). Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. ENeuro *6*.

Blake, D.T., Heiser, M.A., Caywood, M., and Merzenich, M.M. (2006). Experience-dependent adult cortical plasticity requires cognitive association between sensation and reward. Neuron *52*, 371–381.

Bortone, D.S., Olsen, S.R., and Scanziani, M. (2014). Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. Neuron *82*, 474–485.

Buran, B.N., von Trapp, G., and Sanes, D.H. (2014). Behaviorally gated reduction of spontaneous discharge can improve detection thresholds in auditory cortex. J. Neurosci. *34*, 4076–4081.

Carcea, I., Insanally, M.N., and Froemke, R.C. (2017). Dynamics of auditory cortical activity during behavioural engagement and auditory perception. Nat. Commun. *8*, 14412.

Carmel, P.W., and Starr, A. (1963). Acoustic and nonacoustic factors modifying middle-ear muscle activity in waking cats. J. Neurophysiol. *26*, 598–616.

Castellucci, G.A., McGinley, M.J., and McCormick, D.A. (2016). Knockout of Foxp2 disrupts vocal development in mice. Sci. Rep. *6*.

Chavez, C., and Zaborszky, L. (2017). Basal forebrain cholinergic-auditory cortical network: primary versus nonprimary auditory cortical areas. Cereb. Cortex *27*, 2335–2347.

Chen, C., and Song, S. (2019). Differential cell-type dependent brain state modulations of sensory representations in the non-lemniscal mouse inferior colliculus. Commun. Biol. *2*.

Chen, X., Sun, Y.C., Zhan, H., Kebschull, J.M., Fischer, S., Matho, K., Huang, Z.J., Gillis, J., and Zador, A.M. (2019). High-throughput mapping of long-range neuronal projection using in situ sequencing. Cell *179*, 772-786.e19.

Chevée, M., Robertson, J.D.J., Cannon, G.H., Brown, S.P., and Goff, L.A. (2018). Variation in Activity State, Axonal Projection, and Position Define the Transcriptional Identity of Individual Neocortical Projection Neurons. Cell Rep. *22*, 441–455.

Christensen, R.K., Lindén, H., Nakamura, M., and Barkat, T.R. (2019). White noise background improves tone discrimination by suppressing cortical tuning curves. Cell Rep. *29*, 2041-2053.e4.

Crandall, S.R., Cruikshank, S.J., Connors, B.W., Crandall, S.R., Cruikshank, S.J., and Connors, B.W. (2015). A corticothalamic switch: controlling the thalamus with dynamic synapses. Neuron *86*, 768–782.

Daigle, T.L., Madisen, L., Hage, T.A., Valley, M.T., Knoblich, U., Larsen, R.S., Takeno, M.M., Huang, L., Gu, H., Larsen, R., et al. (2018). A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. Cell *174*, 465-480.e22.

Dodson, P.D., Larvin, J.T., Duffell, J.M., Garas, F.N., Doig, N.M., Kessaris, N., Duguid, I.C., Bogacz, R., Butt, S.J.B., and Magill, P.J. (2015). Distinct developmental origins manifest in the specialized encoding of movement by adult neurons of the external globus pallidus. Neuron *86*, 501–513.

Eliades, S.J., and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. J. Neurophysiol. *89*, 2194–2207.

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

French, C.A., and Fisher, S.E. (2014). What can mice tell us about Foxp2 function? Curr. Opin. Neurobiol. *28*, 72–79.

Friedrich, J., Zhou, P., and Paninski, L. (2017). Fast online deconvolution of calcium imaging data. PLoS Comput. Biol. *13*, p.e1005423.

Gong, S., Doughty, M., Harbaugh, C.R., Cummins, A., Hatten, M.E., Heintz, N., and Gerfen, C.R. (2007). Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. J. Neurosci. *27*, 9817–9823.

Guo, W., Clause, A.R., Barth-maron, A., Polley, D.B., Guo, W., Clause, A.R., Barth-maron, A., and Polley, D.B. (2017). A corticothalamic circuit for dynamic switching between feature detection and discrimination. Neuron *95*, 180–194.

Guo, W., Robert, B., and Polley, D.B. (2019). The cholinergic basal forebrain links auditory stimuli with delayed reinforcement to support learning. Neuron *103*, 1164-1177.e6.

Guo, Z. V., Hires, S.A., Li, N., O'Connor, D.H., Komiyama, T., Ophir, E., Huber, D., Bonardi, C., Morandell, K., Gutnisky, D., et al. (2014). Procedures for behavioral experiments in head-

fixed mice. PLoS One *2*.

Hackett, T.A., Rinaldi Barkat, T., O'Brien, B.M.J., Hensch, T.K., and Polley, D.B. (2011). Linking topography to tonotopy in the mouse auditory thalamocortical circuit. J. Neurosci. *31*, 2983–2995.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. PLoS Biol. *5*, 2885–2897.

Hangya, B., Ranade, S.P., Lorenc, M., and Kepecs, A. (2015). Central cholinergic neurons are rapidly recruited by reinforcement feedback. Cell *162*, 1155–1168.

Houde, J.F., Nagarajan, S.S., Sekihara, K., and Merzenich, M.M. (2002). Modulation of the Auditory Cortex during Speech: An MEG Study. J. Cogn. Neurosci. *14*, 1125–1138.

Jaramillo, S., and Zador, A.M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat. Neurosci. *14*, 246–251.

Keller, G.B., and Hahnloser, R.H.R. (2009). Neural processing of auditory feedback during vocal practice in a songbird. Nature *457*, 187–190.

Kim, E.J., Jacobs, M.W., Ito-cole, T., Callaway, E.M., Kim, E.J., Jacobs, M.W., Ito-cole, T., and Callaway, E.M. (2016). Improved monosynaptic neural circuit tracing using engineered rabies virus glycoproteins. Cell Rep. *15*, 692–699.

Kim, J., Matney, C.J., Blankenship, A., Hestrin, S., and Brown, S.P. (2014). Layer 6 corticothalamic neurons activate a cortical output layer, layer 5a. J. Neurosci. *34*, 9656–9664.

Kuchibhotla, K. V, Gill, J. V, Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., and Froemke, R.C. (2017). Parallel processing by cortical inhibition enables

context-dependent behavior. Nat. Neurosci. *20*, 62–71.

Kurt, S., Fisher, S.E., and Ehret, G. (2012). Foxp2 mutations impair auditory-motor association learning. PLoS One *7*, 3–7.

Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519–523.

Lee, I.H., and Assad, J.A. (2003). Putaminal activity for simple reactions or self-timed movements. J. Neurophysiol. *89*, 2528–2537.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Liang, Y., Sun, W., Lu, R., Chen, M., and Ji, N. (2019). A distinct population of L6 neurons in mouse V1 mediate cross-callosal communication. BioRxiv 778019.

Maor, I., Shwartz-Ziv, R., Feigin, L., Elyada, Y., Sompolinsky, H., and Mizrahi, A. (2020). Neural correlates of learning pure tones or natural sounds in the auditory cortex. Front. Neural Circuits *13*, 1–21.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. *21*, 1281–1289.

McGinley, M.J., David, S. V., and McCormick, D.A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. Neuron *87*, 179–192.

Metherate, R., Kaur, S., Kawai, H., Lazar, R., Liang, K., and Rose, H.J. (2005). Spectral integration in auditory cortex: Mechanisms and modulation. Hear. Res. *206*, 146–158.

Miyamichi, K., Shlomai-Fuchs, Y., Shu, M., Weissbourd, B.C., Luo, L., and Mizrahi, A. (2013). Dissecting local circuits: Parvalbumin interneurons underlie broad feedback control of olfactory bulb output. Neuron *80*, 1232–1245.

Morillon, B., Hackett, T.A., Kajikawa, Y., and Schroeder, C.E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. Curr. Opin. Neurobiol. *31C*, 230–238.

Muller-Preuss, P., and Mitzdorf, U. (1984). Functional anatomy of the inferior colliculus and the auditory cortex: current source density analyses of click-evoked potentials. Hear. Res. *16*, 133–142.

Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., and Mathis, M.W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. Nat. Protoc. *14*, 2152–2176.

Nelson, A., and Mooney, R. (2016). The basal rorebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. Neuron *90*, 635–648.

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

Okobi, D.E., Banerjee, A., Matheson, A.M.M., Phelps, S.M., and Long, M.A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. Science (80-.). *363*, 983–988.

Olsen, S.R., Bortone, D.S., Adesnik, H., and Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. Nature *482*, 47–52.

Otazu, G.H., Tai, L.-H., Yang, Y., and Zador, A.M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. Nat. Neurosci. *12*, 646–654.

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K.D. (2016). Kilosort:

realtime spike-sorting for extracellular electrophysiology with hundreds of channels. BioRxiv *preprint*, 061481.

Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. *26*, 4970–4982.

Poulet, J.F.A., and Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity during sound production. Nature *418*, 872–876.

Romero, S., Hight, A.E., Clayton, K.K., Resnik, J., Williamson, R.S., Hancock, K.E., and Polley, D.B. (2019). Cellular and widefield imaging of sound frequency organization in primary and higher order fields of the mouse auditory cortex. Cereb. Cortex *10.1093/ce*.

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

Saunders, A., Oldenburg, I.A., Berezovskii, V.K., Johnson, C.A., Kingery, N.D., Elliott, H.L., Xie, T., Gerfen, C.R., and Sabatini, B.L. (2015). A direct GABAergic output from the basal ganglia to frontal cortex. Nature *521*, 85–89.

Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A.D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience *131*, 1–11.

Schneider, D.M., and Mooney, R. (2018). How movement modulates hearing. Annu. Rev. Neurosci. *41*, 553–572.

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

Schneider, D.M., Sundararajan, J., and Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. Nature *561*, 391–395.

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

Vélez-Fort, M., Rousseau, C. V., Niedworok, C.J., Wickersham, I.R., Rancz, E.A., Brown, A.P.Y., Strom, M., and Margrie, T.W. (2014). The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing. Neuron *83*, 1431–1443.

Voigts, J., Deister, C.A., and Moore, C.I. (2019). Layer 6 ensembles can selectively regulate the behavioral impact and layer-specific representation of sensory deviants. BioRxiv.

Wall, N.R., Wickersham, I.R., Cetin, A., De La Parra, M., and Callaway, E.M. (2010). Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus. Proc. Natl. Acad. Sci. U. S. A. *107*, 21848–21853.

Wickersham, I.R., Lyon, D.C., Barnard, R.J.O., Mori, T., Finke, S., Conzelmann, K.K., Young, J.A.T., and Callaway, E.M. (2007). Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. Neuron *53*, 639–647.

Williamson, R.S., and Polley, D.B. (2019). Parallel pathways for sound processing and functional connectivity among layer 5 and 6 auditory corticofugal neurons. ELife *8*, e42974.

Williamson, R.S., Hancock, K.E., Shinn-Cunningham, B.G., and Polley, D.B. (2015). Locomotion and task demands differentially modulate thalamic audiovisual processing during active search. Curr. Biol. *25*, 1885–1891.

Winer, J.A., Diehl, J., and Larue, D.T. (2001). Projections of auditory cortex to the medial geniculate body of the cat. J. Comp. Neurol. *430*, 27–55.

Yang, Y., Lee, J., and Kim, G. (2020). Integration of locomotion and auditory signals in the mouse inferior colliculus. ELife *9*, 1–17.

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

FIGURES AND LEGENDS

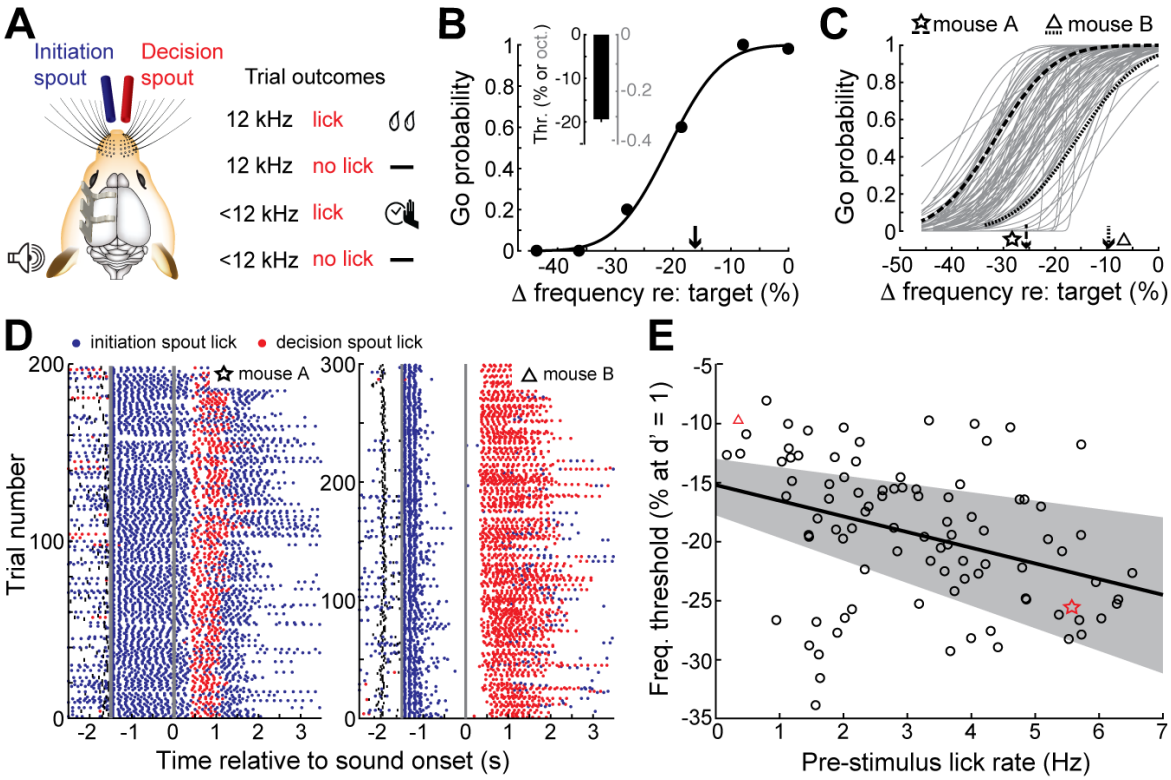

Figure 1. Movement impairs frequency discrimination performance.
(**A**) Schematic of a self-initiated Go/NoGo frequency recognition task for head-fixed mice.
(**B**) Psychometric function from an example mouse. *Inset*: mean ± SEM frequency recognition threshold measured from 93 sessions in 12 mice. Threshold is defined as the frequency difference where d-prime = 1.0 as expressed in units of percent difference or octaves from the target frequency.
(**C**) Large variations in task performance between sessions. Arrows denote frequency threshold for two example testing sessions
(**D**) Lick rasters for the two example sessions shown in *C*. Blue and red dots denote lick timing on the initiation and decision spouts, respectively. Thick gray vertical lines indicate the timing of trial initiation and sound onset. Small vertical black lines indicate the timing of visual ready cue.
(**E**) Frequency thresholds are improved in sessions with minimal licking prior to tone onset. Thick black line and gray shading represent the linear regression line and the bootstrapped 95% confidence interval of the linear regression, respectively. Each individual test session is represented by a symbol, red symbols indicate example behavior sessions shown in *C and D*.

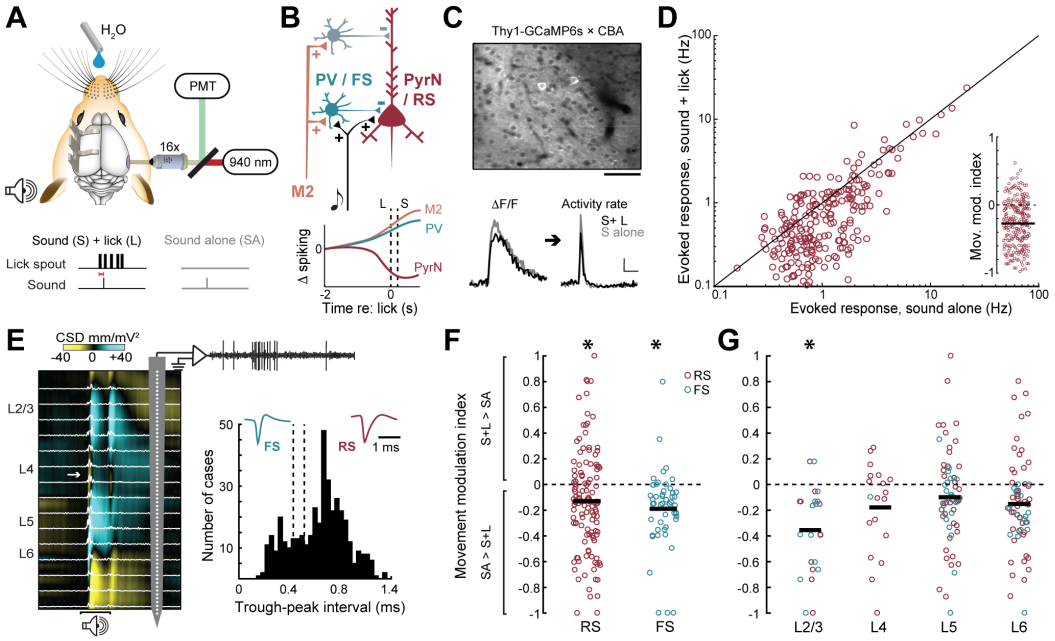

Figure 2. Suppression of lick-triggered sounds in A1.
(**A**) 2-photon imaging of A1 L2/3 PyrN activity during a lick spout sampling behavior in which a frequency modulated sweep is either presented 0.1s following lick initiation or during periods of rest.
(**B**) *Top:* cartoon of ACtx circuitry for suppression of sound-evoked activity during self-initiated movement. M2 = secondary motor cortex; PV = parvalbumin; FS = fast spiking; PyrN = pyramidal neuron; RS = regular spiking. *Bottom:* anticipated changes in PyrN, PV and M2 spike rate changes relative to lick (L) and sound (S) onset.
(**C**) *Top:* GCaMP6s fluorescence from L2/3 of a *Thy1*GCaMP6s × CBA mouse. Scale bar = 50 μm. *Bottom*: Sound-evoked fractional change in fluorescence (left) and deconvolved activity rate (right) from a L2/3 PyrN during licking or quiescence. Horizontal scale bar = 0.5s, Vertical scale bar = 20% dF/F (left) and 5 deconvolved events per second (right).
(**D**) Suppression of lick-triggered sound responses in L2/3 PyrNs. *Inset:* The same data presented as a movement modulation index where values > 0 reflect facilitated responses during licking and < 0 reflect suppressed responses during licking.
(**E**) Extracellular recordings were made from all layers of A1 with a 64-channel silicon probe during the lick spout sampling behavior. Electrophysiological responses are filtered offline to separately analyze unit spiking (white trace) and the current source density (CSD). White arrow in CSD trace identifies the early current sink in layer L4 elicited by a 50ms white noise burst used to assign units to layers. Isolated single units were classified as FS or RS (trough-to-peak delay < 0.5 ms or > 0.6, gray dashed lines). Spike waveforms reflect mean FS and RS waveforms.
(**F-G**) Scatterplots present movement modulation index for every sound-responsive single unit (circle) according to RS or FS spike shape (*F*) or cortical layer (*G*). Horizontal bars denote sample means. Asterisks indicate $p < 0.05$ with a one-sample t-test against a population mean of 0 after correcting for multiple comparisons.

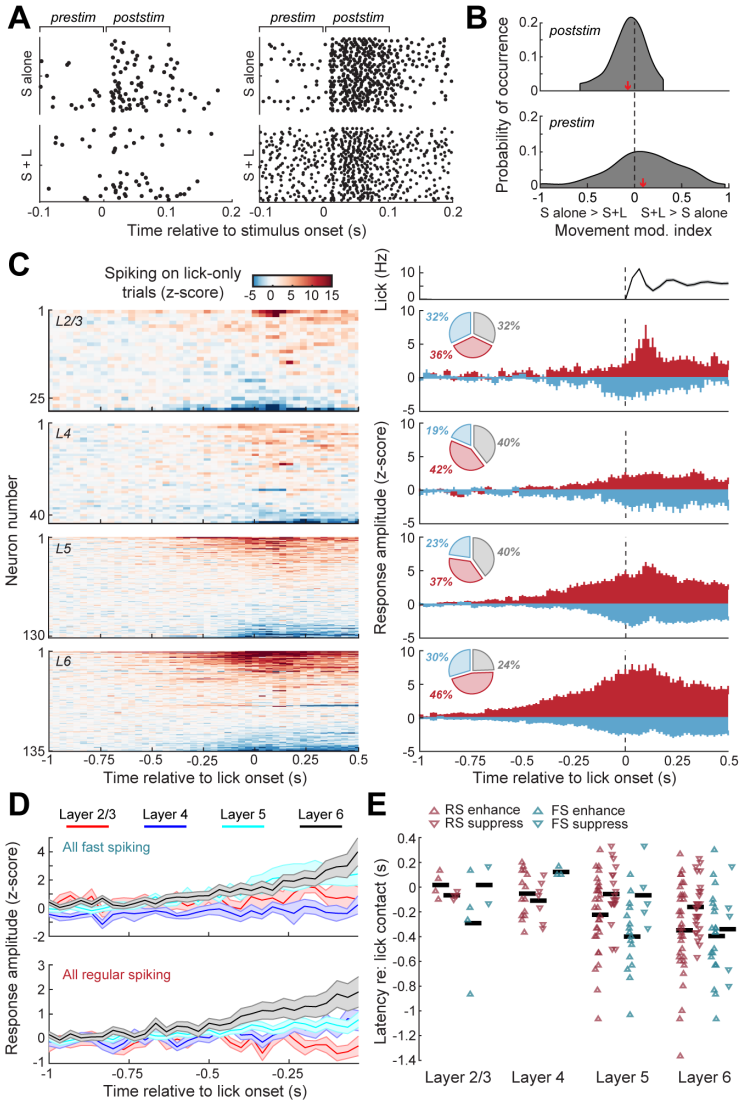

Figure 3. Motor corollary inputs recruit deep layer units in A1

(**A**) Example single unit rasters illustrate cells with suppressed spiking during sound stimulation or enhanced spiking during the pre-stimulus baseline, either of which would produce negative movement modulation index.

(**B**) Histogram of movement modulation index (SA – SL / SA + SL) based only on spiking during the post- or pre-stimulus periods (top and bottom, respectively). Red arrows = sample means.

(**C**) *Left:* Neurograms are grouped by layer and present normalized spike rates for all single units on lick-only trials. Units are sorted by mean normalized activity. *Right*: Pie charts present percentage of units significantly enhanced (red) or suppressed (blue) during the peri-lick period. PSTHs present mean ± SEM normalized firing rates for enhanced and suppressed units from the corresponding layer. Top panel presents the mean ± SEM lick rate during the recording sessions.

(**D**) Mean ± SEM changes in activity for all FS (top) and RS (bottom) units.

(**E**) Motor-related response latencies for each significantly enhanced or suppressed single unit. Horizontal bar = sample mean.

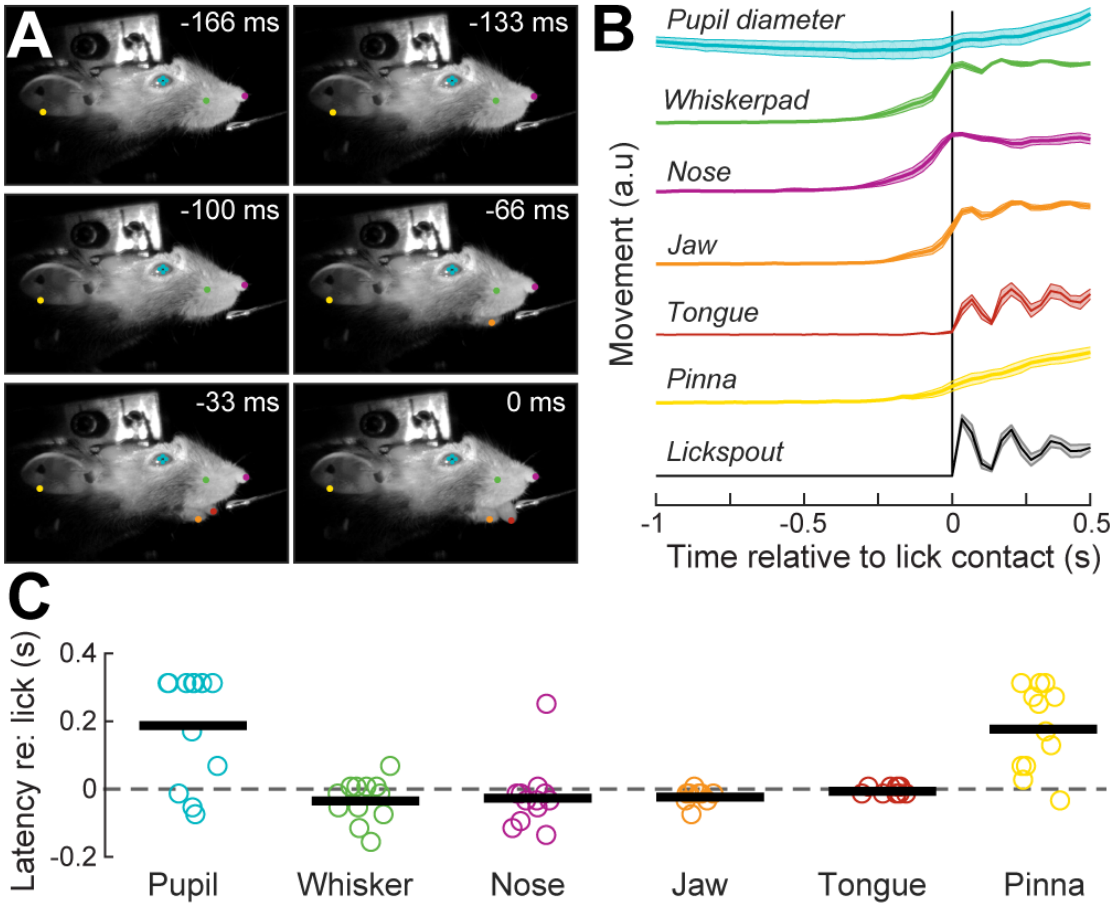

Figure 4. Overt facial movements follow motor-related activation of A1 units
(**A**) Six consecutive video frames culminating in the moment where the tongue first contacted the lick spout. DeepLabCut marker positions are indicated on each frame for the pinna, whisker pad, nose, tongue, jaw and pupil.
(**B**) Mean ± SEM movement amplitudes and pupil diameter are shown relative to electrical contact with the lickspout (n = 12 video sessions from 4 mice).
(**C**) Onset latencies for initial changes in facial markers and pupil diameter relative lickspout contact. Each data point represents a single imaging session. Horizontal bar = sample mean.

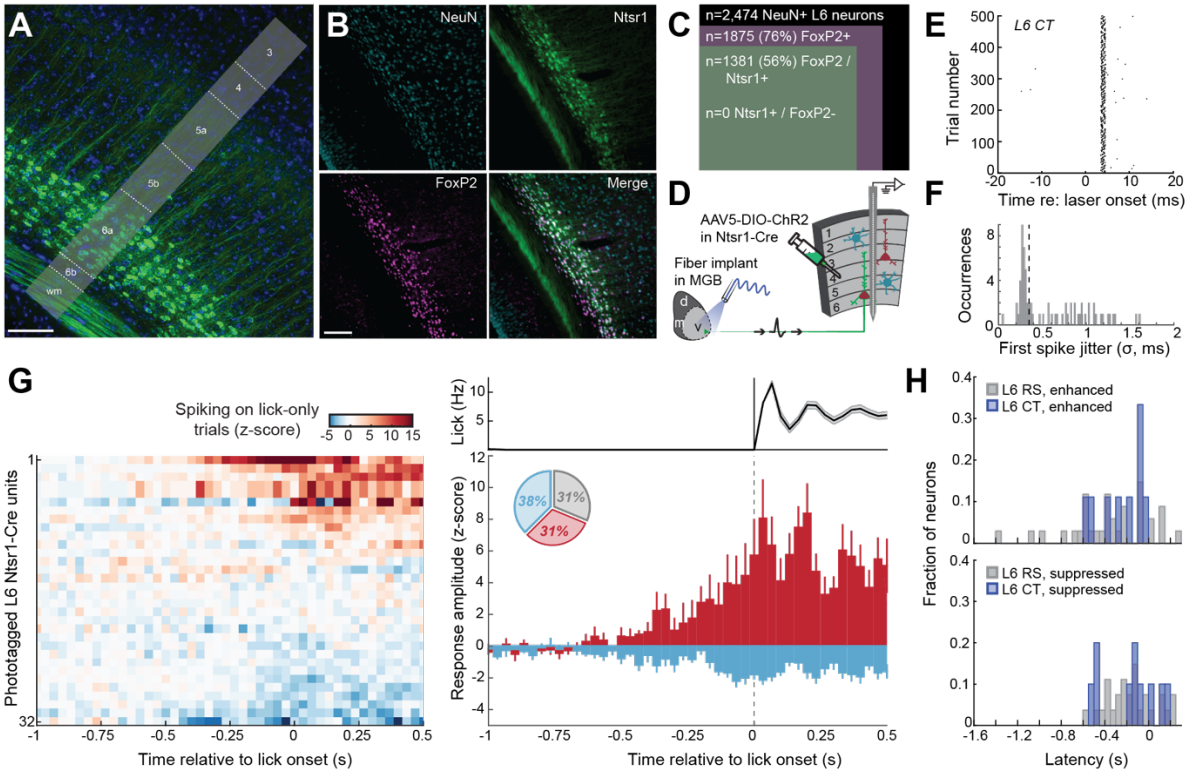

Figure 5. Histological and functional markers for motor-related activation of L6 corticothalamic cells.

(**A**) A1 corticothalamic cells labeled in NTSR1-Cre x Ai148 mice have somata in L6a and sparse vertically oriented neuropil labeling up to L5a. Scale bar = 0.1mm.

(**B**) A1 cells in NTSR1-Cre x Ai148 mice with were immunolabeled for NeuN and FoxP2. Scale bar = 0.1mm.

(**C**) Venn diagram depicts the co-localization of FoxP2 in A1 L6 neurons and NTSR1-Cre in FoxP2+ and FoxP2- L6 neurons.

(**D**) Schematic for antidromic optogenetic phototagging in NTSR1-Cre+ L6 corticothalamic neurons that express ChR2.

(**E**) Rastergram of antidromic spikes in an example unit elicited by photoactivation of thalamic axon terminals of a L6 corticothalamic unit.

(**F**) Histogram of first spike latency jitter in 77 units with optogenetically evoked responses. Directly activated L6 corticothalamic cells were distinguished from indirectly activated units as having a jitter in first spike latency less than 0.35ms (dashed horizontal line).

(**G**) *Left:* Neurograms present normalized spike rates for optogenetically tagged L6 corticothalamic units sorted by mean normalized activity. *Right*: Pie chart presents percentage of units significantly enhanced (red) or suppressed (blue) during the peri-lick period. PSTHs present mean ± SEM normalized firing rates for enhanced and suppressed units. Top panel presents the mean ± SEM lick rate for the corresponding recordings sessions.

(**H**) Histograms of motor-related response latencies for all significantly enhanced (top) or suppressed (bottom) L6 CT units relative to all other unidentified RS L6 units.

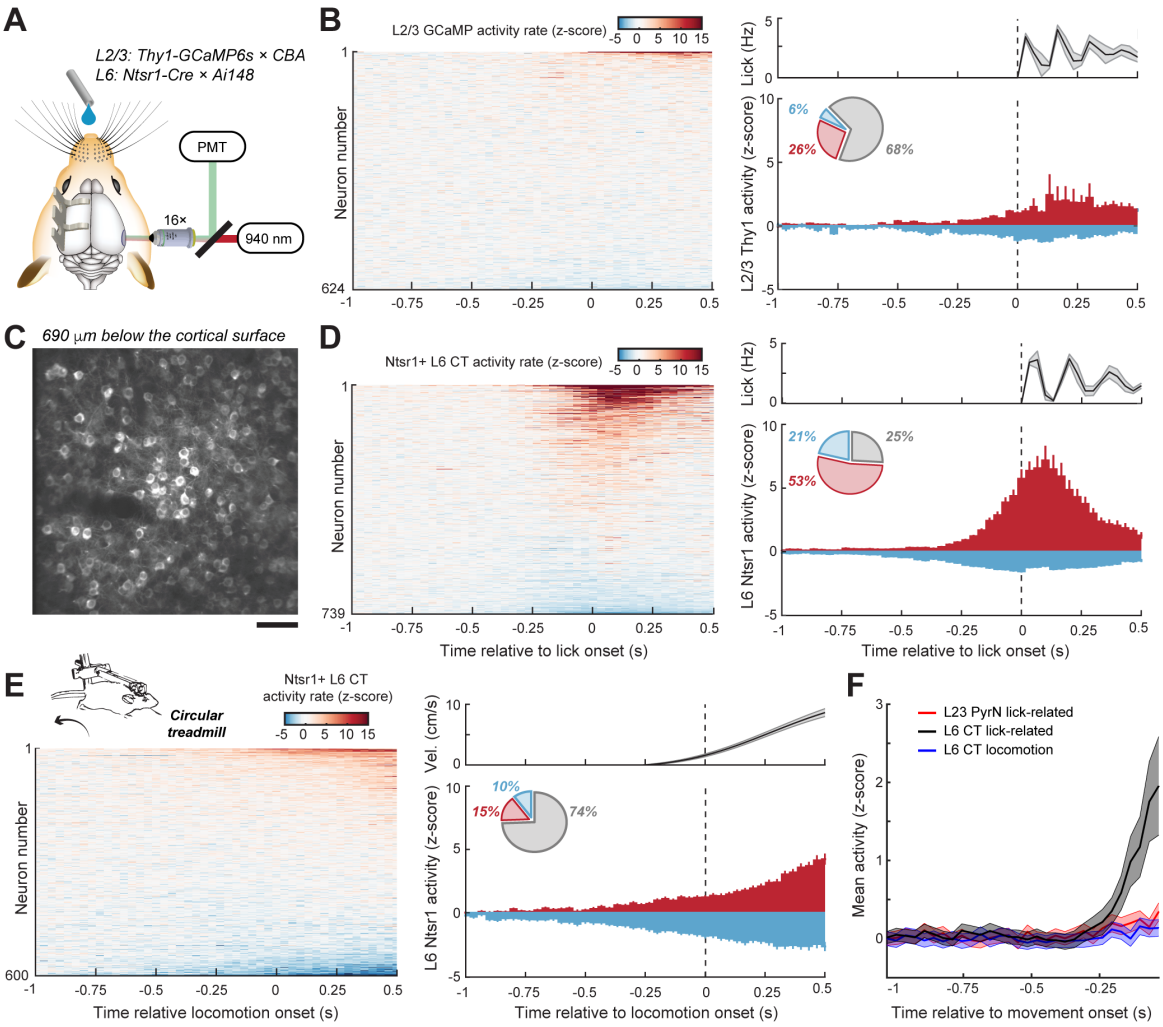

Figure 6. Ensembles of L6 CT neurons are activated by upcoming orofacial movements, but not bouts of running.

(**A**) Schematic for 2-photon imaging of motor-related activation in L2/3 PyrNs and L6 corticothalamic neurons.

(**B**) Lick-related calcium activity in L2/3 PyrNs. *Left:* Neurograms present normalized activity rates for all L2/3 PyrNs on lick-only trials. Neurons are sorted by mean normalized activity. *Right*: Pie charts present percentage of units significantly enhanced (red) or suppressed (blue) during the peri-lick period. PSTHs present mean ± SEM normalized activity rates. Top panel presents the mean ± SEM lick rate for the corresponding recordings sessions.

(**C**) GCaMP6s fluorescence in L6 corticothalamic cells in a NTSR1-Cre × Ai148 mouse. Scale bar = 50 μm.

(**D**) Lick-related calcium activity in L6 corticothalamic neurons. Graphing conventions match *B.*

(**E**) Locomotion-related calcium activity in L6 corticothalamic neurons. Graphing conventions match *B.* Top panel presents the mean ± SEM treadmill velocity from the corresponding recordings sessions.

(**F**) Mean ± SEM motor-related changes in PyrN activity for all 2-p imaging experiments.

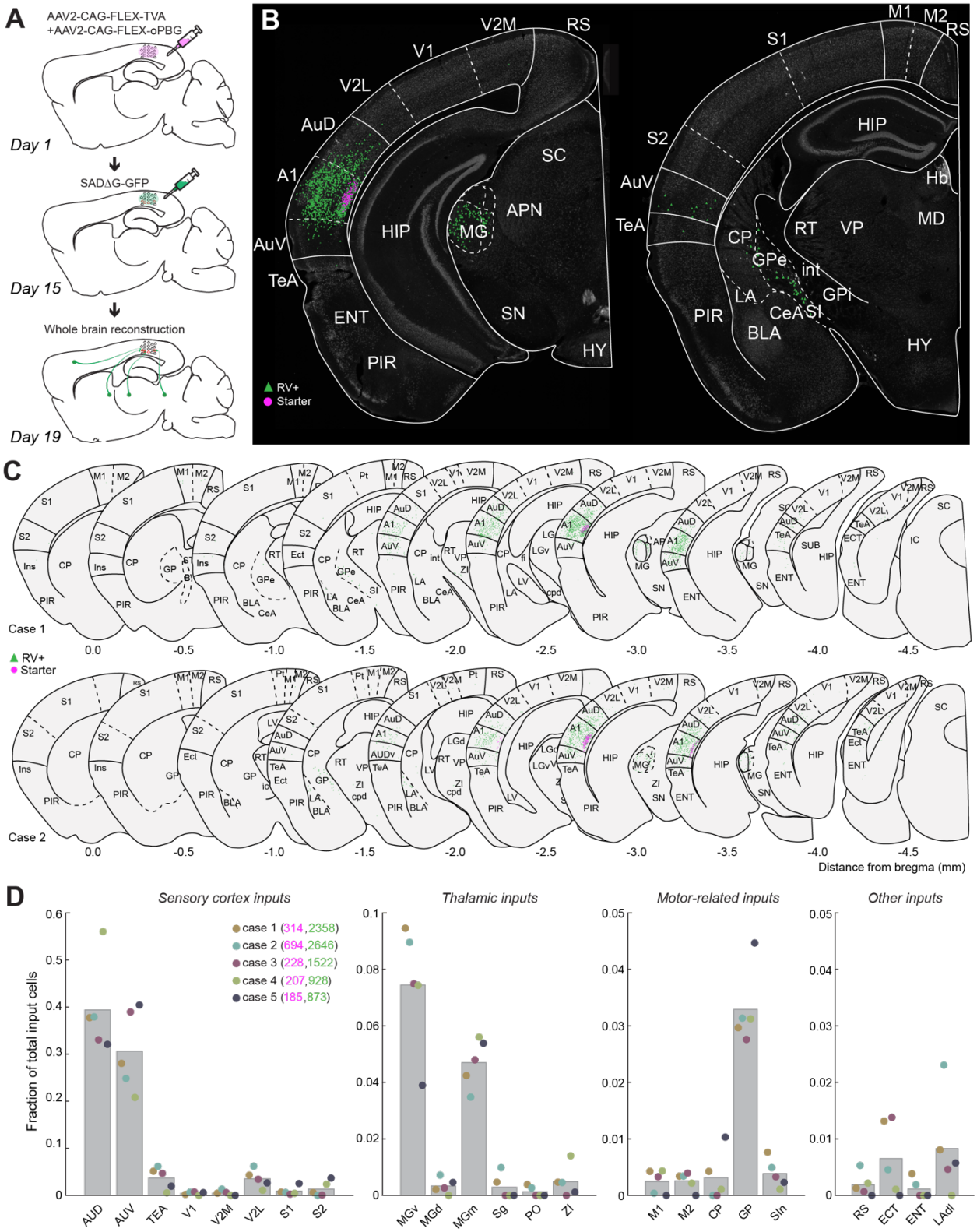

Figure 7. Brain-wide sources of monosynaptic inputs onto A1 L6 corticothalamic neurons.

(**A**) Schematic of viral labeling strategy to identify monosynaptic inputs onto Ntsr1-Cre L6 corticothalamic neurons.
(**B**) Two coronal sections showing L6 corticothalamic starter cells in A1 and RV-labeled monosynaptic inputs in auditory forebrain (left) and a more rostral section containing the motor cortex and portions of the basal ganglia and basal forebrain (right).
(**C**) Coronal series from two mice showing reconstructions of all labeled cells
(**D**) Proportion of RV-labeled input cells from the top-23 regions were independently calculated for each brain (N=5). Gray bars = mean. *Sensory cortical inputs*: AUD = dorsal auditory area, AUV = ventral auditory area, TEA = temporal association cortex, V1 = primary visual cortex, V2M/L = medial and lateral secondary visual area, S1/S2 = primary and secondary somatosensory cortex. *Thalamic inputs*: MGv/d/m = ventral, dorsal and medial subdivisions of the medial geniculate body, Sg = suprageniuclate thalamic nucleus, PO = posterior thalamic nucleus, ZI = zona incerta. *Motor-related inputs*: M1/2 = primary and secondary motor cortex, CP = caudate putamen, GPe = external globus pallidus, SI = substantia innominata. Other: RS = retrosplenial cortex, ECT = ectorhinal cortex, ENT = entorhinal cortex, LAdl = dorsolateral subdivision of the lateral amygdala.

STAR ★ Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests should be directed to and will be fulfilled by the lead contact, Daniel Polley (daniel_polley@meei.harvard.edu). This study did not generate new reagents.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-NeuN	Millipore	MAB377
Goat anti-ChAT	Millipore	AB144P
Goat anti FoxP2	Novus	55411
Donkey anti-mouse, Alexa 647	Lifetech	A329
Donkey anti-goat, Alexa 647	Lifetech	A32789
Bacterial and Virus Strains		
AAV5-Ef1a-DIO hChR2(E123T/T159C)-EYFP	Addgene	35509-AAV5
AAV2-CAG-Flex-TVA	Adi Mizhrahi	N/A
AAV2-CAG-Flex-oPGB	Adi Mizrahi	N/A
Sad-delta-G	Adi Mizrahi	N/A
Chemicals, Peptides, and Recombinant Proteins		
Lidocaine hydrochloride	Hospira Inc	Cat# 71-157-DK
Buprenorphine hydrochloride	Buprenex	Cat#NDC 12496-0757-5
Isoflourane	Piramal	Cat#NDC 66794-013-10
Flow-It ALC Flowable Composite	Pentron	Cat#N11B
C&B Metabond Quick Adhesive Cement System	Parkell	Cat#S380
Silicon adhesive	WPI	Cat#KWIK-SIL
Prolong Gold antifade media	Lifetech	Cat# P36930
Experimental Models: Organisms/Strains		
Mouse:Ntsr1-Cre	GENSAT	B6.FVB(Cg)-Tg(Ntsr1-Cre)GN220Gsat/Mcd
Mouse: Ai148D: B66.Cg-Igs7tm148.1(tetO-GCaMP6f, CAG-tTA2)Hze/J	The Jackson Laboratory	RRID:IMSR_JAX:030328
Mouse: Tg(Thy1-GCaMP6s)GP4.3Dkim	The Jackson Laboratory	RRID:IMSR_JAX:024275
Mouse: CBA/CaJ	The Jackson Laboratory	RRID:IMSR_JAX:000654)
Software and Algorithms		
Labview 2015	National Instruments	https://www.ni.com/en-us/shop/labview.html
Synapse	Tucker-Davis Technologies	http://www.tdt.com/component/synapse-software/

ThorImage 3.0	Thorlabs	https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=9072#ad-image-0
Kilosort	Github	https://github.com/cortex-lab/Kilosort
Suite2P	Github	https://github.com/cortex-lab/Suite2P
DeepLabCut	Github	https://github.com/AlexEMG/DeepLabCut
Matlab 2016b	Mathworks	https://www.mathworks.com/products/matlab.html
Python 3.7	Python	https://www.python.org/downloads/release/python-370/
Other		
Linear silicone recording electrode	Cambridge Neurotech	H3
Linear silicone recording electrode	NeuroNexus	NeuroNexus A1x16-50-177-5mm
Diode laser (473 nm)	Omnicon	LuxX_473-100
PZ5 Neurodigitizer amplifier	Tucker-Davis Technologies	Pz-128
BioAmp processor	Tucker-Davis Technologies	RZ5D
Power meter	Thorlabs	S120C
Solenoid driver	Eaton-Peabody Lab	https://github.com/EPL-Engineering/epl_valve
Lickometer	Eaton-Peabody Lab	https://github.com/EPL-Engineering/epl_lickometer
PXI Controller	National Instruments	PXIe-8840
Free-field speaker	Parts Express	275-010
Ti-Sapphire laser	Spectra Physics	Mai Tai HP DeepSee
16x/.8NA Objective	Nikon	CFI75 LWD 16X W
Two-Photon Microscope	Thorlabs	Bergamo
Titanium headplate	iMaterialise	Custom
IR Camera	Sony	Playstation Eye
LED driver	Thorlabs	Cat# LEDD1B

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All procedures were approved by the Massachusetts Eye and Ear Animal Care and Use Committee and followed the guidelines established by the National Institutes of Health for the care and use of laboratory animals. Mice of both sexes were used for this study. All mice were maintained on a 12 hr light/12 h dark cycle with ad libitum access to food and water, unless undergoing behavioral testing. Mice undergoing behavioral

testing were kept on a reversed 12hr light (7:00 a.m.-7:00 p.m) /12 hr dark cycle. Mice were grouped housed unless they had undergone a major survival surgery.

For two lickspout behavior experiments, we used 12 NTSR1-Cre mice, aged 6-8 weeks. For 2-photon imaging experiments, we used 5 Thy1-GCaMP6sxCBA-CaJ mice and 6 NTSR1-Cre x Ai148 mice. For single-unit electrophysiology experiments, we used 7 NTSR1-Cre mice. For anatomy experiments, we used 3 NTSR1-Cre x Ai148 mice and 5 NTSR1-Cre mice.

METHOD DETAILS

Recovery surgeries

Mice were anesthetized with isoflourane in oxygen (5% induction, 2% maintenance). A homeothermic blanket system was used to maintain body temperature at 36.5º (FHC). Lidocaine hydrochloride was administered subcutaneously to numb the scalp. The dorsal surface of the scalp was retracted and the periosteum was removed. For mice to be used in behavior and physiology experiments, the skull surface was prepped with etchant (C&B metabond) and 70% ethanol before affixing a titanium head plate (iMaterialise) to the dorsal surface with dental cement (C&B metabond). At the conclusion of all recovery procedures, Buprenex (.05 mg/kg) and meloxicam (0.1 mg/kg) were administered and the animal was transferred to a warmed recovery chamber.

Go / NoGo frequency recognition task

Three days after headplate surgery, animals were weighed and placed on a water restriction schedule (1 ml per day). During behavioral training, animals were weighed daily to ensure they remained above 80% of their initial weight and examined for signs of dehydration, such as fur tenting (Guo et al., 2014). Mice were given supplemental water if they received less than 1 mL during a training session. Mice were head-fixed in a dimly lit, single-walled sound-attenuating booth (Acoustic Systems), with their bodies resting in conductive cradle. The lickspout apparatus consisted of two spouts placed 4 cm apart, positioned 1 cm below and centered 0.5 cm to the right of the animal's snout using a 3D micromanipulator (Thorlabs). Tongue contact on the lickspout closed an electrical circuit that was digitized (at 40 Hz) and encoded to calculate lick timing. Acoustic stimuli were delivered through an inverted dome tweeter positioned 10 cm from the animal's left ear (CUI, CMS0201KLX). A broad spectrum LED (Thorlabs) was placed 20 cm in front of the head to provide diffuse, full-field illumination. Digital and analog signals controlling sound delivery, water reward, and visual stimuli were controlled by a PXI system with custom software programed in LabVIEW (National Instruments). Free-field stimuli were calibrated before recording using a wide-band ultrasonic acoustic sensor (Knowles Acoustics, model SPM0204UD5).

Most mice required 1-2 weeks of behavioral shaping before they could perform the complete frequency recognition task. After mice were habituated to head-fixation, they were conditioned to lick the trial initiation spout within 8 seconds of LED onset to receive a small quantity of water (2 μL). Once mice learned to initiate trials, they were then conditioned to lick the decision spout within 0.2-2s after the 12kHz target tone was presented, but not a 6.5 kHz non-target tone (1.5s after initiation, 0.1s tone duration,

5ms raised cosine onset/offset ramps at 70 dB SPL). Licking to the non-target tone triggered a 4-5 s time out. Licking the decision spout before tone onset ended the trial. Inter-trial intervals were drawn from an exponential distribution (3-10s). Once mice learned to withhold licking on >80% of non-target trials, additional non-target frequencies were added to measure a psychometric function. Daily sessions consisted of 200-600 trials.

Non-target tones were presented in 50% of trials and were randomly selected from five logarithmically spaced frequencies centered on the indecision point (50% false alarm rate) from the previous session. Of 96 behavioral sessions, 3 were excluded either because the hit rate was less than 80%, or fewer than 100 total trials were performed. Psychometric functions were fit using binary logistic regression. Go probability was expressed as d-prime, calculated as z(hit rate) - z(false alarm rate). The frequency recognition threshold was defined as the frequency difference from target where d-prime equaled 1.0. The mean number of licks during a 1s period prior to tone was used to calculate the foreperiod lick rate for each session.

Lick spout sampling task

A different cohort of water-deprived mice performed a simple head-fixed single lick spout sampling task (N = 15). Licks were either encoded with the electrical circuit described above or an infrared photobeam emitter/detector for electrophysiology experiments. Mice were initially conditioned to lick the spout to receive a 4-6μL bolus of water 0.5s later. Once mice reliably licked to receive reward, we varied the outcome such that 50% of lick bouts triggered a sound and 50% of lick bouts resulted in reward. Sound and reward probability were determined independently. This task design ensured that both sound presentation and reward were unpredictable, and that sound presentation did not predict reward. Sound onset was delayed by 0.1s following lick bout initiation. Rapid frequency modulated sweeps were used in these experiments because they elicited strong A1 neural responses across a wide range of best frequency tuning (50 ms sweeps presented at 70 dB SPL; 4:64 kHz or 64:4 kHz range [80 octaves/s], 5ms raised cosine onset/offset ramps). To avoid photobleaching, 2-photon imaging sessions were limited to 135 trials per day (approximately 30 minutes) by only presenting upward FM sweeps. For electrophysiology and video tracking experiments, 200-500 trials were performed each day and included both upward and downward FM sweeps.

Treadmill running

Head-fixed mice were habituated to running on a low-inertia, quiet cylindrical treadmill over 4-6d period before neural recordings began. Locomotion was recorded using an optical rotary encoder. Locomotion signals were downsampled to 30 Hz and filtered using a zero-phase digital filter. Locomotion was operationally defined as periods where running speed exceeded 2 cm/s. In order to be included for analysis, at least 3s of quiescence were required prior a running bout that lasted for a minimum of 1s. Only sessions that included at least 15 locomotion bouts were used in the analysis. Continuous broadband noise (50 dB SPL) was used to mask sounds generated by running.

Two-photon calcium imaging

Three round glass coverslips (one 4mm, two 3mm, #1 thickness, Warner Instruments) were etched with piranha solution and bonded into a vertical stack using transparent, UV-cured adhesive (Norland Products, Warner Instruments). Headplate attachment, anesthesia and analgesia follow the procedure listed above. A 3 mm craniotomy was made over right ACtx using a scalpel and the coverslip stack was cemented into the craniotomy (C&B Metabond). Animals recovered for at least 5 days before beginning water restriction for the lickspout sampling task.

An initial widefield epifluorescence imaging session was performed to visualize the tonotopic gradients of the ACtx and identify the position of A1 as described previously (Romero et al., 2019). Two-photon excitation was provided by a Mai-Tai eHP DS Ti:Sapphire-pulsed laser tuned to 940 nm (Spectra-Physics). Imaging was performed with a 16×/0.8NA water-immersion objective (Nikon) from a 512 x 512 pixel field of view at 30Hz with a Bergamo II Galvo-Resonant 8 kHz scanning microscope (Thorlabs). Scanning software (Thorlabs) was synchronized to the stimulus generation hardware (National Instruments) with digital pulse trains. Widefield and 2-photon microscopes were rotated by 50-60 degrees off the vertical axis to obtain images from the lateral aspect of the mouse cortex while the animal was maintained in an upright head position. Imaging was performed in light-tight, sound attenuating chambers. Animals were monitored throughout the experiment to confirm all imaging was performed in the awake condition using modified cameras (PlayStation Eye, Sony) coupled to infrared light sources.

For 2-photon imaging of L2/3 PyrNs, imaging was performed 175-225 μm below the pial surface. The focal plane for L6 imaging was 600-700 μm below the pial surface, which can be accomplished with relatively low excitation power (107-138 mW) because Ntsr1-Cre neurons have sparse apical dendritic fields that produce minimal out of plane excitation (Daigle et al., 2018; Liang et al., 2019; Voigts et al., 2019). The DeepSee precompensation oscillator (Spectra Physics) was adjusted for each imaging session to improve image quality and reduce laser power for L6 imaging.

Fluorescence images were captured at 2x digital zoom. Raw calcium movies were processed using Suite2P, a publicly available analysis pipeline (Pachitariu et al., 2016). Briefly, movies are registered to account for brain motion. Regions of interest are established by clustering neighboring pixels with similar time courses. Manual curation is then performed in the Suite2P GUI to eliminate for low quality or non-somatic ROIs. Spike deconvolution was performed in Suite2P, using the default method based on the OASIS algorithm (Friedrich et al., 2017).

Two-photon imaging analysis

Sound-responsive neurons were identified by measuring deconvolved activity during a pre-stimulus (-133 - 0ms prior to sound onset) and post-stimulus period (33 - 167ms following sound onset) in sound alone trials. Sound responsive neurons were operationally defined as having significantly elevated activity in the post-stimulus period based on a one-tailed paired-test ($p < 0.05$). To calculate sound-evoked activity rates for sound alone and sound + lick conditions, the baseline activity level during the pre-stimulus window was subtracted from the post-stimulus window.

The modulation index for sound-evoked activity was calculated as:

$$Modulation\ index = \frac{Sound\ alone - Sound\ and\ lick}{Sound\ alone + Sound\ and\ Lick}$$

To calculate activity rates during lick only conditions, we first defined baseline activity during a 0.5s period that was separated by at least 2s from the last lick of a preceding bout and 1.5s from the first lick of an upcoming bout. Lick-related activity was calculated during a 1.5s period beginning 1s prior to lick onset, where mean activity in each 33ms time bin was expressed in units of z-score relative to the baseline distribution. Neurons were classified as significantly enhanced or suppressed during licking by calculating the mean activity rates during a period of licking (-0.5 to 0.5s relative to lick onset) versus a period of quiescence (-2 to -1s relative to lick onset) based on a $p < 0.05$ threshold of a two-tailed Wilcoxon rank-sum test.

Virus mediated gene-delivery

For mice used in optogenetics experiments, two burr holes (<1mm diameter each) were made in the skull overlying A1, 1.75 - 2.25mm rostral to the lambdoid suture. A motorized injection system (Stoelting) was used to inject 200 nL of AAV5-Ef1a-DIO hChR2(E123T/T159C)-EYFP into each burr hole 600 μm below the pial surface. Electrophysiology experiments began 3-4 weeks after injection. For anatomical tracing experiments using pseudotyped rabies virus, 150 nL of a 1:2 mixture AAV2-CAG-FLEx-TVA-mCherry and AAV2-CAG-FLEx-oG were injected through a single burr hole, 600μm below the surface of A1. After a 14-21 day waiting period, we targeted the same location and depth for a second injection of SAD-ΔG-GFP (500 nL). Mice were perfused 4 days after the SAD-ΔG injection. Immediately following each injections, the scalp was sutured and the animal recovered in a warmed chamber. All virus solutions were injected at a rate of 15 nL/min.

Preparation for single unit-recordings in head-fixed mice

A ground wire (AgCl) was implanted over the left occipital cortex. A craniotomy was made above the MGB (centered 2.75 mm lateral to the midline and 2.75 mm caudal to bregma) and the position of the ventral subdivision was identified with a cursory mapping of pure tone receptive fields as described previously (Guo et al., 2017; Williamson and Polley, 2019; Williamson et al., 2015). An optic fiber (flat tip, 0.2mm diameter, Thorlabs) was inserted 2.4 – 2.8mm to rest 0.2mm above the physiologically identified dorsal surface of MGB. The fiber assembly was cemented in place and painted with black nail polish to prevent light leakage. Animals were allowed to recover for at least three days before water restriction began.

On the day of the first recording session, mice were briefly anaesthetized with isoflurane (2%) and a small craniotomy (0.5 medial-lateral x 1.25 mm rostral-caudal) was made over A1, centered 2mm from the lambdoid suture. Each day, a small well was made around the craniotomy with UV-cured cement and filled with lubricating ointment (Paralub Vet Ointment). At the conclusion of each recording, the chamber was flushed, filled with new ointment, and capped with UV-cured cement.

Single unit recordings in head-fixed mice

All recordings took place in a dimly lit, single-wall sound attenuating chamber (Acoustic Systems). A 64-channel silicon probe (H3 probe, Cambridge Neurotech) was slowly advanced (100 μm/min) into ACtx perpendicular to the pial surface until the tip of the electrode was 1.3-1.4mm below the cortical surface, ensuring full coverage of all layers of A1. The brain was allowed to settle for at least 15 minutes before recordings began. On the day of the first recording, multiple penetrations were made to identify the tonotopic reversal which represents the rostral border of A1 (Hackett et al., 2011).

Raw neural data was digitized at 32-bit, 24.4 kHz (Neurodigitizer and RZ5 BioAmp Processor; Tucker-Davis Technologies) and stored in binary format. To eliminate artifacts, the common mode signal (channel-averaged neural traces) was subtracted from all channels in the brain. Signals were notch filtered at 60Hz, then band-pass filtered (300-3000 Hz, second order Butterworth filters). Raw signals were notch filtered at 60 Hz and downsampled to 1000 Hz to extract local field potentials. The CSD was calculated as the second spatial derivative of the LFP signal. To eliminate potential artifacts introduced by impedance mismatching between channels, signals were spatially smoothed along the channels with a triangle filter (5-point Hanning window). Two CSD signatures were used to identify L4 in accordance with prior studies: A brief current sink first occurs approximately 10ms after the onset of broadband noise burst (50ms duration, 70 dB SPL, 100 trials), which was used to determine the lower border of L4. A tri-phasic CSD pattern (sink-source-sink from upper to lower channels) occurs between 20ms and 50ms, where the border between the upper sink and the source was used to define the upper boundary of L4 (Guo et al., 2017; Metherate et al., 2005; Muller-Preuss and Mitzdorf, 1984).

Kilosort was used to sort spikes into single unit clusters (Pachitariu et al., 2016). Single-unit isolation was based on the presence of both a refractory period within the interspike interval histogram, and an isolation distance (>10) indicating that single-unit clusters were well separated from the surrounding noise (Schmitzer-Torbert et al., 2005). RS and FS designation was based on the ratio of the mean trough to peak interval > 0.6 (RS) or < 0.5 (FS). Ntsr1-Cre L6 corticothalamic units were identified using antidromic phototagging. A 1ms pulse of 473nm light was delivered via the implanted MGB fiber to fire antidromic spikes in L6 corticothalamic units that expressed hChR2 (5-45 mW laser power, repeated at 4Hz for 1000 repetitions with a diode laser, Omicron, LuxX). Ntsr1+ L6 corticothalamic cells were distinguished from indirectly activated neurons based on laser-evoked spiking rates at least 5 sd above baseline and first spike latencies that varies by less than 0.35 SD, as confirmed in a previous study (Williamson and Polley, 2019).

Single unit recording analysis

Single unit spike times were binned at 30 Hz for direct comparison with 2-photon and behavior tracking data. Units were excluded from further analysis if spontaneous spike rates were < 0.1 Hz. Sound-responsive units, evoked-activity rates and the modulation index for sound alone and sound + lick conditions were calculated similarly to the 2-photon data except that pre- and post-stimulus spike measurement windows were -66-0ms and 33-99ms relative to sound onset, respectively. Lick related activity was calculated with the same methods as described above for 2-photon imaging.

Latencies of motor-related excitation or suppression was defined as the first time point that activity consistently exceeded the 5th or 95th percentile of the baseline spike rate distribution (as per Schneider et al, 2014). Analyses was initiated at the minimum (suppressed) or maximum (enhanced) activity time bin and preceded backwards in time until this condition was first met.

Capture and quantification of orofacial movements

Videos of the animal's face were obtained using a camera (Playstation Eye, Sony) coupled with infrared light sources (Thorlabs) and a 5-50 mm varifocal lens (Computar CS-Mount). Videos were acquired at 30 Hz (N=2) or 60 Hz (N=2) at a resolution of 512x512 pixels. A single point on the vibrissae array, nose, jaw, tongue and pinna were labeled alongside four cardinal positions on the circumference of the pupil in 60 frames for each imaging session. Manually labeled frames were split into training and test sets and the network was trained for 1.3 million iterations using default parameters in DeepLabCut (Mathis et al., 2018; Nath et al., 2019). Low-confidence outlier frames were manually re-labeled, the model re-trained and full video dataset reanalyzed. As the lower jaw and tongue markers were not visible until lick initiation, the point where they first appeared was manually entered as their position for all quiescent frames. Any remaining spurious position values where model confidence was < 0.9 was replaced with the interpolated value determined from surrounding frames.

Each tracked point was expressed as a 3-D vector as X × Y × Time. All videos recorded at 60 Hz were downsampled to 30 Hz for further analysis. Pupil diameter was calculated as the mean of the two diameter measurements. All other points were reduced to a 1-dimensional activity trace d(t) as:

$$d(t) = \sqrt{\left(\frac{x(t) - \bar{x}}{\sigma(x)}\right)^2 + \left(\frac{y(t) - \bar{y}}{\sigma(y)}\right)^2}$$

where x(t) is the x position at timepoint t and y(t) is the y position at time point t, which are normalized using the mean and the standard deviation of the entire x and y position traces respectively. A zero phase digital filter was then used to smooth activity traces without introducing temporal distortions. Movement latency was computed identically to neural latencies.

Anatomy

Mice were deeply anaesthetized with ketamine and transcardially perfused using 4% paraformaldehyde in 0.01M phosphate buffered saline. Brains were extracted and post-fixed in 4% paraformaldehyde for 12 hours, then placed in 30% sucrose. Brains were sectioned coronally at 40 μm into 0.1M phosphate buffer (PB), preserving all sections containing cerebral cortex.

To co-localize Ntsr1-Cre and FoxP2 in NeuN-labled ACtx neurons, fluorescent immunohistochemistry assays were performed on sections of NTSR1 × Ai148 brains. Sections were rinsed in 0.1M phosphate buffered saline (PBS), permeabilized for 2 hours at room temperature (RT) in a blocking solution of 0.1M PBS, 2% BSA, 5%

donkey serum, 0.1% Tween-20, then incubated for 48 hours at 4°C in blocking solution containing the primary antibodies. Sections were rinsed in PBS then incubated for 2 hours at RT in blocking solution containing secondary antibodies, counterstained in DAPI for 5 minutes, rinsed in 0.01M PB, mounted onto glass slides, then coverslipped using Prolong Gold antifade media (Lifetech, Eugene OR).

ACtx sections were imaged in z-stacks over 20 um (2 um steps) on a Nikon 90i epifluorescence microscope (40x Plan Fluor, oil DIC N2). Co-localization of NeuN, NTSR1-Cre labeled GFP, and FoxP2 were quantified in 0.25mm wide regions of layer 6 in A1. Labeled cells were counted independently for each fluorescence channel in 2 adjacent sections from each hemisphere of three brains using the Taxonomy function in Nikon Elements AR software. Analysis was restricted to NeuN-labled cells containing a DAPI-labeled nucleus.

To count and localize cells in RV-injected brains, all sections were scanned at 20x using a Leica Versa widefield epifluorescence microscope, then converted to multichannel TIFF files for identification of labeled cells and brain areas. Reconstructions and cell counts were performed in layers using Adobe Illustrator, with reference to the Allen Brain Atlas (Lein et al., 2007) and the stereotaxic mouse brain atlases.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed in MATLAB 2016b (Mathworks). Data are reported as mean ± SEM unless otherwise indicated. Non-parametric tests were used when data samples did not meet assumptions for parametric statistical tests. Inflated familywise error rates from multiple comparisons of the same sample were adjusted with the Holm-Bonferroni correction. Statistical significance was defined as $p < 0.05$.

DATA AND CODE AVAILABILITY

Data acquisition and analysis was performed with custom scripts in MATLAB, LabVIEW, and Python. Spike sorting was done in Kilosort (https://github.com/cortex-lab/Kilosort). Two photon imaging ROI extraction and spike deconvolution was done in Suite2p (https://github.com/cortex-lab/Suite2P). Markerless behavior tracking was done with DeepLabCut (https://github.com/AlexEMG/DeepLabCut).

ADDITIONAL RESOURCES

Not applicable.