# Neural dynamics in the rodent motor cortex enables flexible control of vocal timing

**Arkarup Banerjee** [1, 2, 3, 4, *], **Feng Chen** [5, *], **Shaul Druckmann** [6], **and Michael A. Long** [1, 2, 3, 4, ✉]

[1] NYU Neuroscience Institute, New York University Langone Health, New York, NY 10016, USA.
[2] Department of Otolaryngology, New York University Langone Health, New York, NY 10016, USA.
[3] Center for Neural Science, New York University, New York, NY 10003, USA.
[4] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
[5] Department of Applied Physics, Stanford University, Stanford, CA 94305, USA.
[6] Department of Neuroscience, Stanford University, Stanford, CA 94304, USA.
[*] Co-first author

**ABSTRACT: Neocortical activity is thought to mediate voluntary control over vocal production, but the underlying neural mechanisms remain unclear. In a highly vocal rodent, the Alston's singing mouse, we investigate neural dynamics in the orofacial motor cortex (OMC), a structure critical for vocal behavior. We first describe neural activity that is modulated by component notes (approx. 100 ms), likely representing sensory feedback. At longer timescales, however, OMC neurons exhibit diverse and often persistent premotor firing patterns that stretch or compress with song duration (approx. 10 s). Using computational modeling, we demonstrate that such temporal scaling, acting via downstream motor production circuits, can enable vocal flexibility. These results provide a framework for studying hierarchical control circuits, a common design principle across many natural and artificial systems.**

**Correspondence:** ***mlong@med.nyu.edu***

## Introduction

Many species exert voluntary control over vocal production, allowing rapid flexibility in response to conspecific partners or other environmental cues [1, 2]. Neocortical activity observed across a range of species [3-8] has been proposed to be important for executive control of vocalization [9-12]. For instance, cortical neurons are preferentially active when non-human primates vocalize in response to a conditioned cue [6]. In contrast, the primary vocal motor network consisting of evolutionarily conserved brain areas in the midbrain and brainstem [10-14] is sufficient to generate species-typical sounds. Pioneering work in squirrel monkeys [15] and cats [16] as well as recent studies in laboratory rodents [17-19] have identified many such areas, including the periaqueductal grey and specific pattern generator nuclei in the reticular formation. While these subcortical vocal production mechanisms have been well-characterized, much less is known about how cortical activity contributes to vocal production and flexibility.

To address this issue, we focus our attention on the highly tractable vocalizations of a Costa Rican rodent [20]: the Alston's singing mouse (Scotinomys teguina, **Fig. 1a**). Singing mice produce a temporally patterned sequence of notes (approx. 20 to 200 ms) that become progressively longer over many seconds (e.g., **Fig. 2a**), henceforth referred to as a song. Singing mice can flexibly adjust their song duration in response to many internal [21] and external [22] factors, including social context [20]. Recently, we discovered that a specific forebrain region, the orofacial motor cortex (OMC), is crucial for vocal behavior in this species [20]. Electrical stimulation of OMC disrupted or paused ongoing singing, and its pharmacological inactivation abolished vocal interactions and significantly reduced variability in song duration [20]. A major gap in understanding, however, concerns the nature of the cortical activity that drives this ethologically relevant vocalization. We therefore performed the first electrophysiology recordings in singing mice to assess the impact of OMC dynamics on vocal production and flexibility.

## Results

### High-density silicon probe recordings in freely behaving singing mice

We recorded OMC neural activity during vocal production in four adult male S. teguina using high-density silicon probes (Cambridge NeuroTech or Diagnostic Biochips) (**Fig. 1b, c**). Electrodes were inserted to a final depth of 600-1000 µm, such that most recording sites were in the ventral portion (i.e., motor output layers) of OMC. We used this approach to monitor neural activity continuously over 3 to 20 days, and 13 sessions with robust vocal behavior (duration: 10.4 ± 5.7 hours, mean ± SD) were analyzed further. During these recording sessions, singing mice produced songs both spontaneously (n = 226) and in response to the playback of a conspecific vocalization (n = 79). For this study, which focuses on vocal production, we combined data across these conditions, yielding a total of 23 ± 17 (mean ± SD) songs per session (range: 8 to 72). In total, we recorded from 375 neurons (29 ± 11 per session, mean ± SD) whose spiking was stably monitored throughout those recording sessions (see Methods).

### OMC spiking is modulated during vocal production

We began by examining whether OMC neural activity was related to singing behavior. Although song-related spiking patterns often differed across neurons (e.g., **Fig. 1c**), we found that the ensemble activity of simultaneously recorded OMC neurons was similar across song epochs compared to non-singing periods (**Fig. 1d, e**). Since each session consisted of

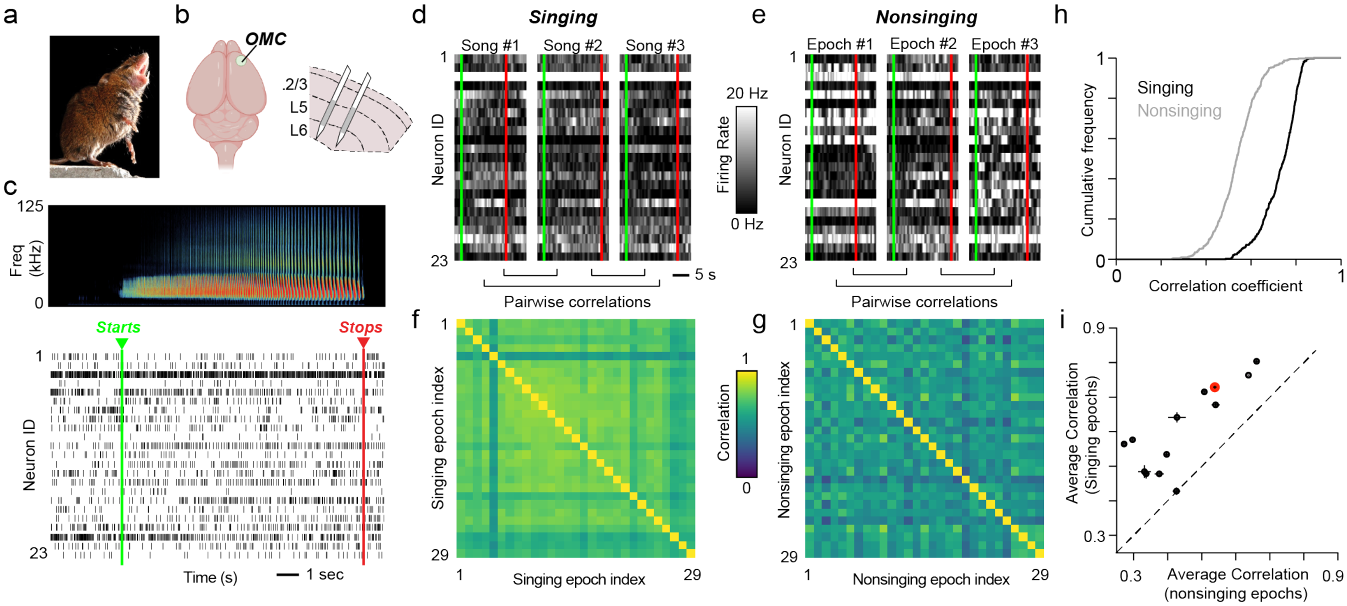


**Fig. 1. Reliable cortical population activity during singing in S. teguina. (a)** S. teguina singing (Photo credit: Christopher Auger-Dominguez). **(b)** Schematic of S. teguina brain highlighting the recording site (i.e., orofacial motor cortex, or OMC) as well as the positioning of electrodes (gray shaded region). **(c)** Spiking activity from 23 simultaneously recorded OMC neurons during song production. The sonogram at top depicts S. teguina song. Neurons with mean firing rates less than 1 spikes/s are excluded for visualization purposes. **(d** and **e)** Firing rates of OMC neural ensemble from (c) during three singing epochs (d) compared with equally timed epochs recorded outside of song (e). For plots (c) through (e), green and red dashed lines mark the beginning and end of the song, respectively. **(f)** and **(g)** For the example session, pairwise correlations of the joint activity of the OMC ensemble recorded across all singing (f) and nonsinging (g) epochs. Dimensions of this matrix reflect the total number of songs in this session (n = 29). **(h)** Correlation values across all songs are significantly higher during singing compared with nonsinging (one-sided Welch's t-test, $p = 3.0 \times 10^{-139}$). (i) Average correlation values for each recording session (mean ± S.E.M., n = 13 sessions, 4 mice). Red point refers to example session shown in (c)-(h).

multiple songs, we calculated the correlation values of OMC ensemble activity across all pairs of songs and found them to be significantly greater compared to nonsinging epochs in the example session (**Fig. 1f-h**) as well as across all recording sessions ($\text{Corr}_{\text{singing}} = 0.61 \pm 0.11$, $\text{Corr}_{\text{nonsinging}} = 0.44 \pm 0.12$, $p = 2.76 \times 10^{-6}$, paired t-test) (**Fig. 1i**). Taken together, we find that OMC population activity is consistently modulated during song production.

Since OMC ensemble activity displayed reliable neural dynamics during singing, we next proceeded to characterize song-related spiking in individual OMC neurons. Each song is composed of a series of notes (**Fig. 2a, b**); therefore, neural activity could a priori be related to the production of each note at a fast timescale (approx. 100 ms), or it could follow slower dynamics at timescales comparable to the entire song (approx. 10 s). By statistically comparing neural activity during vocal production (versus nonsinging epochs), we found that 29.6 percent of neurons (111 out of 375) were correlated with notes (**Extended Data Fig. 1a-c, Fig. 2c**) while 35.5 percent (133 out of 375) of neurons displayed dynamics spanning the entire song (**Extended Data Fig. 1d-f**, see Methods), and 13.1 percent were active at both timescales. Therefore, more than half of individual OMC neurons were significantly modulated with some aspect of singing behavior.

## Note-related responses of OMC neurons

Cortical activity has been shown to represent relevant kinematic features (e.g., velocity and force of effector muscles) for many movements [23]. Applying this framework to vocal production, we would expect OMC neurons to show phasic activity patterns preceding each note. To determine the relationship of OMC firing and note production, we linearly warped spiking activity to both the onset and offset of notes (**Fig. 2d**). A close inspection of note-related neurons revealed a diverse relationship between spike timing and note duration. For instance, in some cases, there appeared to be a systematic shift in the spike timing as note durations increased (e.g., **Fig. 2di**), which may arise from systematic offsets between neural activity and note production. Specifically, if this shift were due to a motor delay, or the timing needed for premotor signals to result in a behavioral change, activity would precede the production of notes [24]. Conversely, if the timing shift were due to sensory feedback, spiking activity would lag note production [25].

To explore these possibilities, we systematically varied the timing of spikes with respect to the audio recordings (**Fig. 2d, Extended Data Fig. 2a, b**) and determined the time lag that resulted in the most consistent alignment with notes (**Fig. 2e**, see Methods). Among the population of note modulated neurons, shifts resulted in significantly better alignment between neural activity and note phase in 25 cases (**Fig. 2e, f**, bootstrap p < 0.01, see Methods). Of these, 23 were consistent with sensory shifts and only 2 with motor offsets (**Fig. 2f, Extended Data Fig. 2e**). Based on the relative timing of neural activity and behavior, less than 1 percent (2 out of 375) of all recorded OMC neurons have a response profile consistent with a motor command for note production. Therefore, while we find phasic note-related activity in OMC, it is unlikely to be directly involved in the production of individual notes.

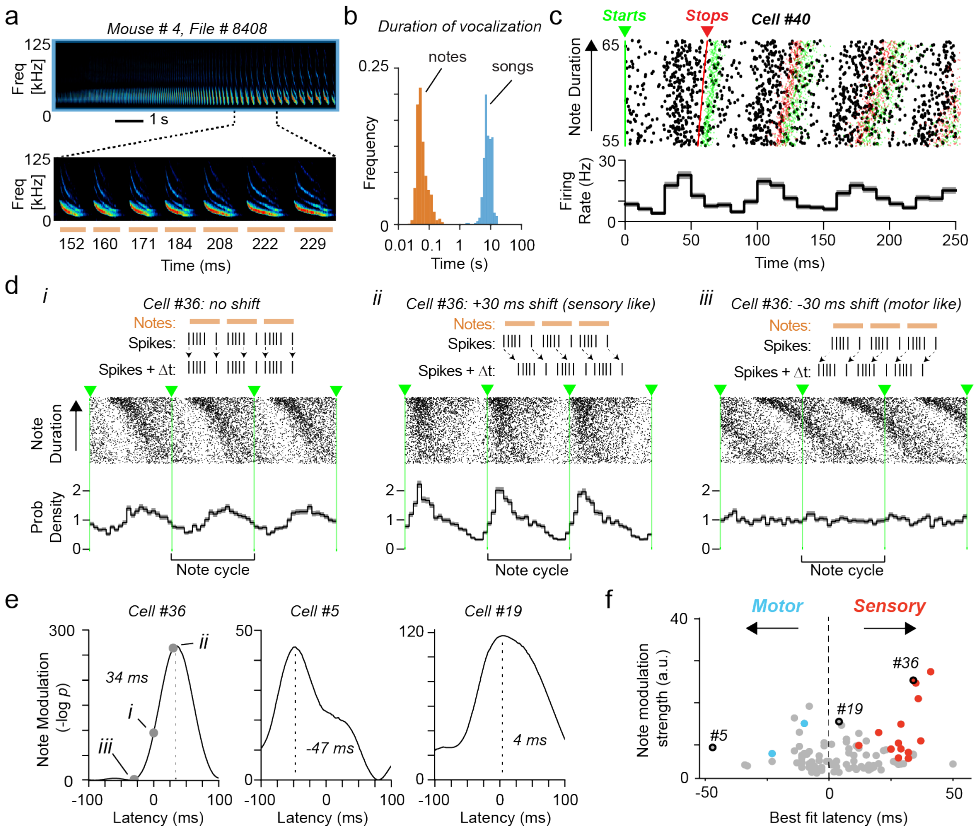


**Fig. 2. Note-related activity of OMC neurons. (a)** At top, singing behavior in a single S. teguina example song. At bottom, an expanded view of 7 notes from the above example. Horizontal lines represent the timing of notes, and the durations for each note (in ms) are provided below. **(b)** Histogram of note (n = 30,540) and song (n = 305) durations plotted on a logarithmic axis across all recorded mice in this study (n = 4). **(c)** Spiking activity corresponding to note timing for an example neuron. For visualization, the spike raster plot was restricted to notes within a range of 55 to 65 ms (full range: 31.4 to 175.9 ms). Green and red ticks indicate the onset and offset of notes, respectively. **(d)** Spiking activity of an example neuron linearly warped to a common note duration (onsets indicated by dashed green lines). Rasters (top) and spike probability density plots (bottom) are provided for the recorded spike trains (i) and after imposing a 'sensory' (- 30 ms) (ii) or 'motor' (+ 30 ms) (iii) offset. **(e)** Modulation strength and offset values for three example neurons. Gray circles and roman numerals in plot for Cell 36 refer to the corresponding panels depicted in (d) (see Methods). **(f)** Summary plot showing the best-fit latency (restricted to ± 50 ms) corresponding to the maximum note modulation strength for 96 neurons. Gray symbols represent cases that are not significantly different from zero, and red (n = 15) and blue (n = 2) symbols represent points with sensory and motor offsets, respectively. The three example cells depicted in (e) are indicated.

## Precise temporal scaling of OMC neural dynamics with song duration

We next explored an alternative schema based on hierarchical control in which OMC population dynamics is dominated by a set of motor primitives (i.e., distinct patterns of neural activity) which do not directly represent movement kinematics [26]. In this view, motor commands for note production are determined by downstream vocal pattern generators driven by time-varying OMC activity spanning the duration of the song, a dynamical systems framework that has been proposed in other motor control studies [27, 28]. Therefore, we broadened our view to examine the extent to which neural activity relates to the structure of the produced song at timescales comprising the entire song duration (approx. 10 s).

We tested how OMC neural dynamics covaries with song duration, which can substantially differ across renditions (**Fig. 3a**). The activity of individual neurons may evolve with identical timing regardless of song duration and therefore be correlated with 'Absolute Time' (**Fig. 3b**). Consequently, dynamics associated with shorter songs would simply look like truncated versions of those observed during longer songs. Alternatively, OMC neurons could reflect 'Relative Time' (**Fig. 3c**), in which neural activity expands and contracts to track the progression through longer and shorter songs, respectively. To test these models, we analyzed trial-to-trial differences in song duration across renditions (average variation: 139.9 percent, n = 13 sessions, e.g., **Fig. 3a**) and used a similarity analysis to compare the firing patterns of each modulated neuron after the timing of activity had been linearly warped to align the onset and offset of song (**Fig. 3d, e, Extended Data Fig. 3**). The Absolute Time model would predict a higher degree of correlation when maintaining original timing and comparing initial portions of longer songs to shorter songs, while the Relative Time model suggests the opposite (i.e., higher correlation after warping). We therefore directly compared these two scenarios and found that the explained variance of single trial firing rates was significantly greater in the warped condition compared with the unwarped condition ($p = 7.5 \times 10^{-7}$, one-sided paired t-test) (**Fig. 3f**), supporting the Relative Time model of OMC neural dynamics.

To further quantify the magnitude of time scaling for each neuron, we generated a consensus neural activity profile for songs with similar durations (**Fig. 3g, Extended Data Fig. 3a-c**, see Methods). For each pair of blocks, we compared the neural activity profiles to determine the scaling factor that maximized the pairwise correlation (e.g., **Fig. 3h**), which we call the neural scaling factor ($S_{\text{neural}}$). If the optimal neural scaling (i.e., the ratio of activity profiles leading to the highest correlation value) matched the relative ratio of associated song durations ($S_{\text{behavioral}}$), then the $S_{\text{neural}}/S_{\text{behavioral}}$ slope is expected to be 1 (equivalent to the Relative Time model). When $S_{\text{neural}}$ was plotted against the behavioral scaling factor (i.e., ratio of the associated song durations, $S_{\text{behavioral}}$), we found them to be linearly proportional (**Fig. 3i, j**). Across all the neurons, the neural scaling/behavioral

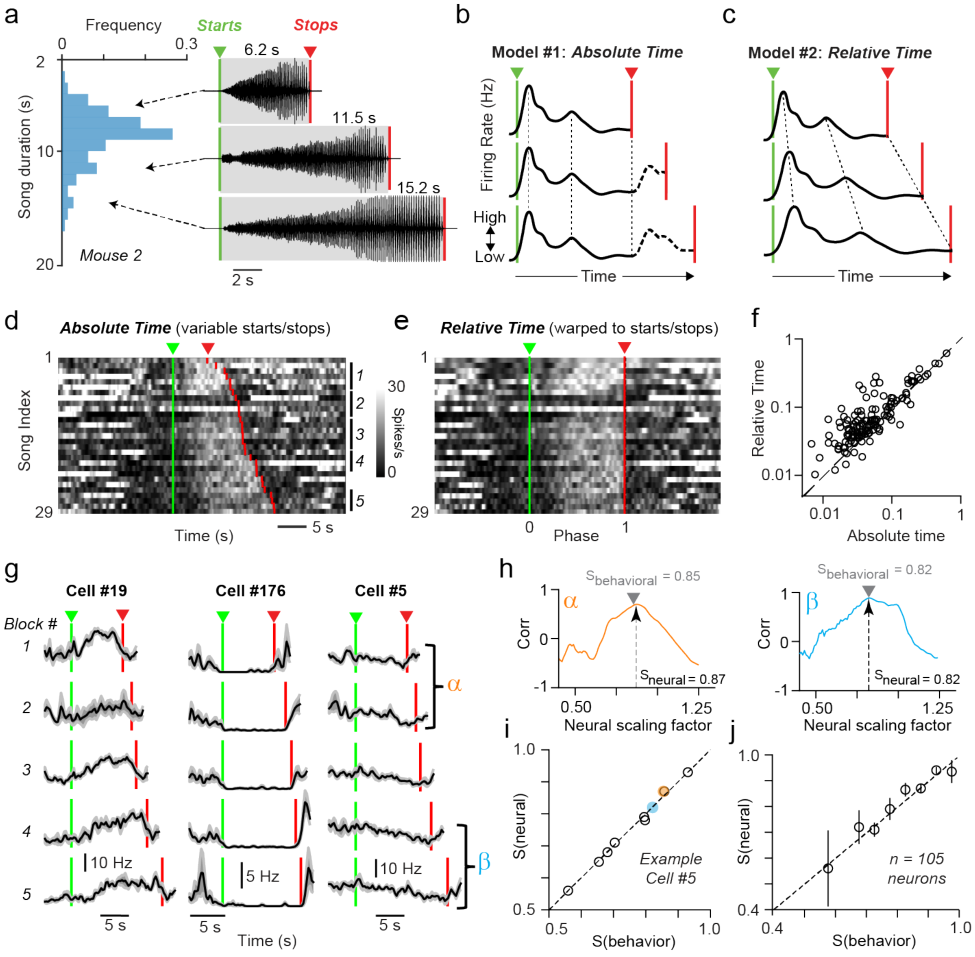


**Fig. 3. Scaling of neural activity with song duration. (a)** Duration of all songs (n = 143) produced from one example mouse (left). Raw waveforms for three example songs of different durations (right). **(b and c)** Hypothetical time-varying neural activity from a single neuron as predicted by the Absolute Time (b) and Relative Time (c) models for three songs of varying durations. **(d)** and **(e)** Spiking responses for a single neuron across 29 songs aligned to the start of the song (d) or temporally warped to the beginning and end of the song (e). **(f)** Comparison of explained variance for 133 song-modulated neurons across trials using recorded song times (Model 1, x-axis) and following temporal warping (Model 2, y-axis). Data are better fit by Model 2 (one-sided paired t-test, p = 3.95 x $10^{-7}$). **(g)** Peri-song time histograms (PSTHs) for three example neurons. Each trace represents an average of 4-21 similarly timed trials. Cell 19 is the same neuron shown in (d) and (e). Song blocks used to calculate consensus firing rate profiles are indicated by numbers and vertical lines. **(h)** Two example pairwise comparisons of the instantaneous firing plots from (g). For each pair, the black arrow indicates scaling factor with maximum correlation ($S_{neural}$), and the gray arrow shows the ratio of song times ($S_{behavioral}$). **(i)** and **(j)** All pairwise comparisons (n = 10) of $S_{neural}$ and $S_{behavioral}$ for the example neuron (i) (colored circles refer to panels in (h)) and for the entire population (n = 105 neurons, x-axis bin size: 0.05) **(j)** The error bars refer to the standard error of the median estimated by bootstrapping.

slope was 1.01 ± 0.01 (n = 659 pairs, 105 neurons, **Fig. 3j**, see Methods). For comparison, the Absolute Time model would predict a slope of 0. This result demonstrates that activity of individual OMC neurons linearly stretches or compresses by a magnitude determined by the ratio of the song durations, enabling OMC activity to precisely track the proportion of elapsed song.

## Diverse individual neuron dynamics in OMC

What are the motor primitives observed in OMC during vocalization? Since OMC circuit activity precisely scales with song duration, we linearly warped the firing rates of song-modulated neurons to both the onset and offset of song. Using this strategy, we observed diverse firing patterns within the OMC during vocalization (**Fig. 4**). To quantify this heterogeneity, we performed hierarchical clustering (**Fig. 4a**, see Methods) and found that 28.6 percent of neurons increased firing during song production while the remainder were suppressed. Further analyses of their response profiles revealed 8 distinct clusters of neurons (**Fig. 4a, b**). We observed that some neurons exhibited transient responses coincident with song onset (Cluster 7), song offset (Cluster 6), or both (Cluster 2), and other neurons showed more persistent increases (Clusters 1, 3) or decreases (Clusters 4, 5, 8) in neural activity during singing. Overall, neurons were responsive throughout the duration of the song and not just at song initiation and termination, consistent with moment-by-moment control of ongoing song production. We conclude that the population of OMC neurons that keep track of Relative Time (i.e., phase) shows diverse firing patterns during song production.

## Computational model of vocal motor control

To understand how motor commands for note timing can be generated from the motor primitives described above (**Fig. 4**), we next constructed a data-driven hierarchical model that makes experimentally testable behavioral predictions. In this model, OMC does not determine note timing directly (consistent with a lack of 'premotor' timing in **Fig. 2**), but vocal motor control is instead shared by cortical and downstream circuits. Inspired by our data, we posit that cortex dictates the moment-by-moment song phase and overall duration (**Fig. 3**), while the motor command for individual notes is generated by midbrain/brainstem areas comprising the primary vocal motor network (**Fig. 5a, Extended Data Fig. 4**). In the model, OMC activity provides descending synaptic drive, which influences the rate of note production in the subcortical song pattern generator (**Fig. 5b**). To account for the decreasing rate of note production with time, the synaptic drive onto the downstream note pattern generator may decrease throughout the song. We accomplish this in our model through linear weighting of OMC activity profiles directly

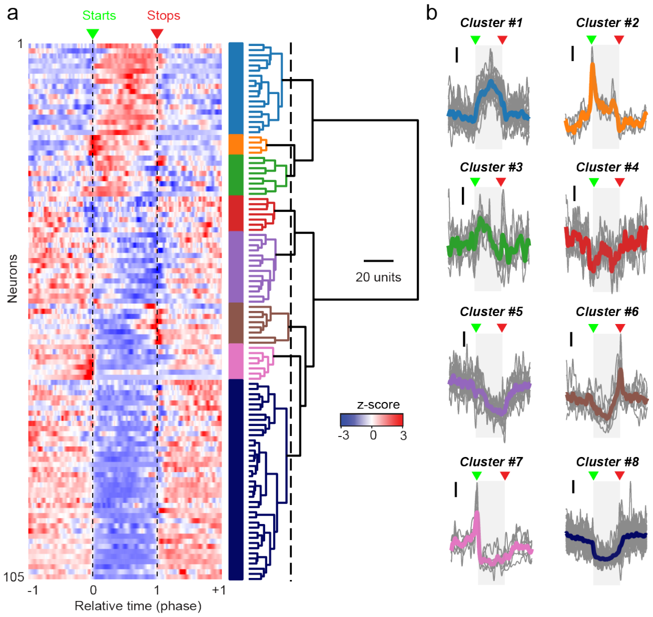


**Fig. 4. Diverse categories of OMC firing patterns during singing. (a)** A hierarchical clustering plot describing the response profiles of OMC neurons whose activity was modulated during singing (see Methods, n = 105 neurons). Individual clusters are indicated by colored bars on the right. **(b)** Spiking responses for each cluster displayed as average firing rate plots. The mean activity profile of each neuron is represented with gray lines, and colored lines are average waveforms for each cluster corresponding to categories from (a). Black vertical bars indicate a normalized firing rate (z-score) of 1. Gray shaded blocks denote song epochs, with green and red arrows marking song starts and stops respectively.

measured in our recordings (**Extended Data Fig. 4a**) which sum up to produce synaptic drives with varying slopes (**Fig. 5b**). We model the workings of the note pattern generator such that individual notes are produced upon reaching a fixed firing rate threshold (see Methods), akin to an integrate-and-fire module. Appropriate time-scaling of cortical activity will thus result in songs of different durations without the need for modifying the note-generating mechanism (**Fig. 5b**). Importantly, this role of OMC is robust to the choice of the precise means by which note generation is implemented in the note pattern generator, either via postsynaptic adaptation mechanisms or synaptic drive from another brain region (**Extended Data Fig. 4b, c**).

We next test a specific behavioral prediction of our hierarchical model to assess its validity. Our model predicts that songs become longer by incorporating more notes and not by increasing the duration of individual notes (**Fig. 5b, c**). Alternately, if note timing were directly triggered by note-modulated OMC activity (**Fig. 2**), longer songs would have the same number of notes with their durations proportionately stretched, as observed in the songbird [29, 30]. We tested these predictions by examining the structure of songs produced with different durations and found that the number of notes systematically increased as a function of song duration (n = 13 animals, 4 from this study and an additional 9 from a published data set [20]) (**Fig. 5d**), a finding that strongly agrees with our hierarchical model.

We considered a directed circuit perturbation to assess whether the relationship between notes and song duration relies upon activity within OMC. We reanalyzed a data set in which OMC was focally cooled in 9 mice [20]. Previous experimental [29, 31-33] and theoretical [34] work predicts that mild focal cooling should dilate the temporal profile of OMC neural activity thereby slowing the progression of subcortically controlled note production. For each animal, OMC cooling resulted in an increase in both song duration (control: 8.0 ± 0.3 s, cooled: 9.3 ± 0.4 s, p = 0.002, paired t-test) as well as the number of notes (control: 92.8 ± 3.2, cooled: 103.9 ± 3.5, p = 0.004, paired t-test) (**Fig. 5e**). Therefore, OMC-cooled songs became longer by incorporating more notes, further supporting the role of OMC activity in our hierarchical model. In sum, these results suggest that cortical activity can generate the necessary vocal motor commands to account for natural variability in behavior.

## Discussion

In this study, we observed robust modulation of motor cortical activity during vocalization corresponding to two behaviorally relevant timescales: (1) phasic responses during note production (approx. 100 ms) and (2) persistent song-related dynamics (approx. 10 sec). We found that many neurons modulated at the faster timescale exhibited a delay between note timing and spiking that could represent either sensory feedback or efference copy signals (**Fig. 5f**). Sensory feedback is known to be important in animal and human vocal motor control [35-38], and a systematic perturbation of sensory streams (e.g., auditory, proprioceptive) [39] could test whether these signals are important in similar control processes in the singing mouse. Nevertheless, our time-shift analysis, modeling, and perturbation results confirm that these fast-varying responses in OMC do not reflect vocal motor commands to produce individual notes. At the slow timescale, responses were heterogeneous (e.g., transient at song onsets, ramping responses, etc.) and appear to reflect a set of motor primitives related to the control of song duration and the rate of note production. Future work will determine whether these spiking profiles map onto specific neuronal cell types in the OMC defined by critical circuit features, such as their output targets, as seen in motor cortical circuits in the laboratory mouse [40-42].

These results provide a striking example of how motor cortical dynamics can modulate song production, perhaps reflecting a voluntary mechanism of generating adaptive vocal flexibility. To accomplish this moment-to-moment control, our cortical recordings support a model in which OMC acts hierarchically via downstream song pattern-generator circuits (**Fig. 5f, Extended Data Fig. 4b, c**), likely corresponding to regions that have been recently characterized in the laboratory mouse [17-19] and appear to be highly conserved across vocalizing species [10, 11]. The hierarchical model proposed here is consistent with our previous work, where we found that OMC inactivation did not abolish singing but significantly reduced the variability in song durations [20], suggesting that activity in OMC is providing necessary input to the brainstem to generate socially appropriate vocalizations (**Extended Data Fig. 4b, c**). Future work is needed to determine the full song circuit in the singing mouse and

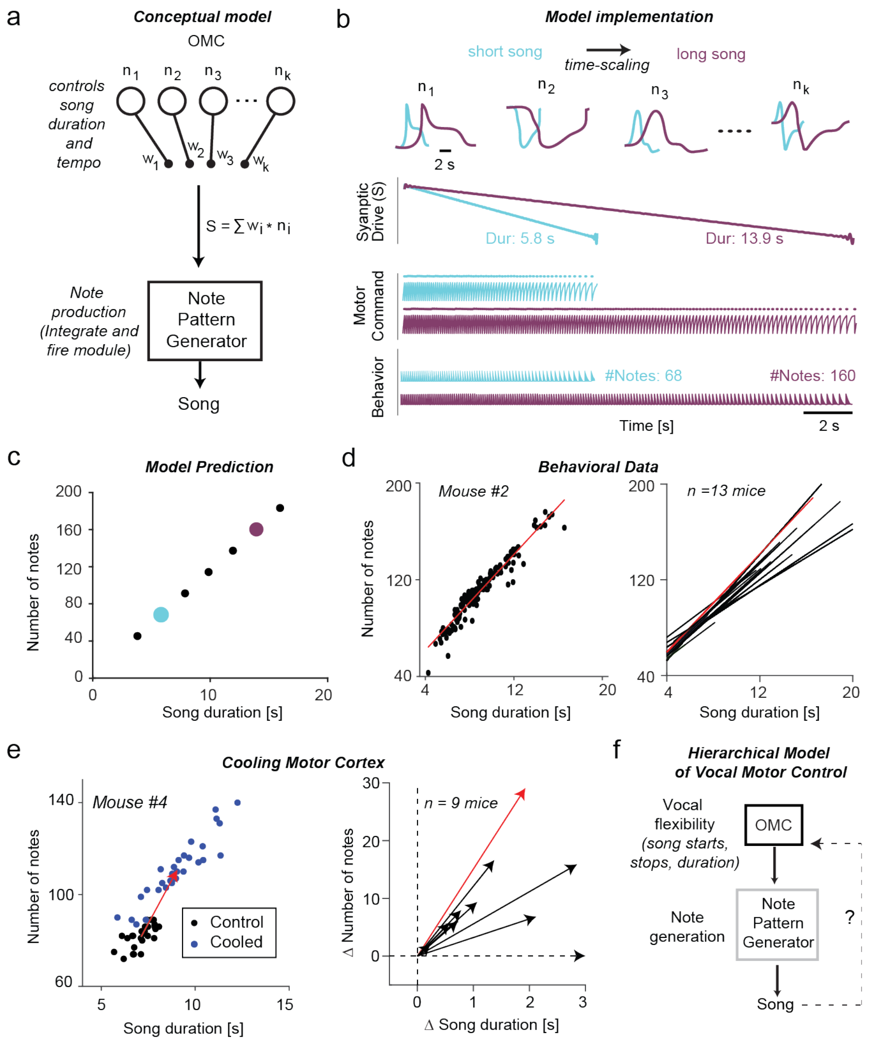


**Fig. 5. Hierarchical model of vocal motor control. (a)** Schematic depicting shared control of vocal production, where OMC controls song duration and rate of progression while individual notes are produced by a downstream note pattern generator. The synaptic drive to the note pattern generator is derived from OMC neural activity (see **Extended Data Fig. 4**). **(b)** Activity profiles of four model OMC neurons for a long song (purple) compared to a short song (cyan). Linear summation of neural activity creates the synaptic drive to the note pattern generator. The note pattern generator is modeled as an integrate-and-fire module, such that the rate of note production depends upon the strength of the OMC synaptic input. **(c)** Model output using seven different values of time-scaling, leading to a prediction in which the number of notes linearly covaries with song duration. Cyan and purple indicate examples from (b). **(d)** The number of notes scales with song duration in an example mouse (n = 144 songs, left) as well as across the population (n = 13 mice, right). Diagonal lines at right represent linear regression fits for each individual animal. Red line indicates data from the example at left. **(e)** Cooling OMC results in a shift in both song duration and number of notes in one example animal (Mouse 4, left). The average change in song duration and number of notes as the result of cooling for each animal (n = 9 mice, right). Across all animals, OMC cooling significantly increased average song durations (control: 8.0 ± 0.3 s, cooled: 9.3 ± 0.4 s, p = 0.002, paired t-test) as well as average number of notes (control: 92.8 ± 3.2, cooled: 103.9 ± 3.5, p = 0.004, paired t-test). Red line indicates data from Mouse 4 (left). **(f)** Hierarchical model of vocal motor control, wherein OMC confers flexibility to a downstream song pattern generator.

elucidate the synaptic mechanisms by which OMC influences downstream vocal production circuits.

The singing mouse vocal control network appears to operate in a partially autonomous hierarchical configuration – a successful design principle for biological and artificial systems – wherein a higher order modulator (i.e., OMC) extends the capabilities of lower-level motor controllers (i.e., note production circuitry) without being necessary for generating the basic motor program [43-45]. Such an arrangement enables behavioral flexibility without relying upon synaptic plasticity in downstream motor patterning circuits. Similar mechanisms have been observed when animals are trained to keep track of time [46-51] or in primate cortex during cycling tasks at different speeds [52]. Our results extend the scope of this temporal scaling algorithm over an expanded time window (approx. 10 s) and to a new domain: controlling vocal flexibility in mammals. Despite its ubiquity, the neural mechanisms contributing to temporal scaling are not well-understood, though several ideas have been proposed, including feedback loops [46, 51] and neuromodulatory gain control [53]. The OMC circuit in the singing mouse offers a valuable opportunity to examine these and other circuit features for generating motor flexibility in the context of an ethologically-relevant behavior.

**Acknowledgements** We thank Steve Shea, Florin Albeanu, Walter Bast, Joseph del Rosario, Hadas Sloin, and members of the Long and Banerjee laboratories for comments on earlier versions of this manuscript. Abby Paulson provided technical assistance.

## Funding

National Institutes of Health grant R01 NS113071 (MAL, SD)
Simons Collaboration on the Global Brain (MAL, SD)
Searle Scholars Program (AB)
Simons Foundation Junior Fellows Program (AB)

## Author contribution

Conceptualization: AB, MAL
Methodology: AB, FC, SD, MAL
Investigation: AB, MAL
Visualization: AB, FC, SD, MAL
Funding acquisition: AB, SD, MAL
Project administration: SD, MAL
Supervision: SD, MAL

Writing – original draft: AB, MAL
Writing – review and editing: AB, FC, SD, MAL

### Competing interests

Authors declare that they have no competing interests.

### Data and materials availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Michael Long (mlong@med.nyu.edu). This study did not generate new unique reagents. The data sets generated during this study are available upon request from the Lead Contact.

### Bibliography

1. Banerjee, A., and Vallentin, D. (2022). Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. Curr Opin Neurobiol 73, 102529.
2. Pika, S., Wilkinson, R., Kendrick, K.H., and Vernes, S.C. (2018). Taking turns: bridging the gap between human and animal communication. Proceedings. Biological sciences 285.
3. Castellucci, G.A., Guenther, F.H., and Long, M.A. (2022). A Theoretical Framework for Human and Nonhuman Vocal Interaction. Annu Rev Neurosci.
4. Miller, C.T., Thomas, A.W., Nummela, S.U., and de la Mothe, L.A. (2015). Responses of primate frontal cortex neurons during natural vocal communication. J Neurophysiol 114, 1158-1171.
5. Roy, S., Zhao, L., and Wang, X. (2016). Distinct Neural Activities in Premotor Cortex during Natural Vocal Behaviors in a New World Primate, the Common Marmoset (Callithrix jacchus). J Neurosci 36, 12168-12179.
6. Hage, S.R., Gavrilov, N., and Nieder, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. J Cogn Neurosci 25, 1692-1701.
7. Hage, S.R., and Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. Nat Commun 4, 2409.
8. Castellucci, G.A., Kovach, C.K., Howard, M.A., 3rd, Greenlee, J.D.W., and Long, M.A. (2022). A speech planning network for interactive language use. Nature 602, 117-122.
9. Hage, S.R., and Nieder, A. (2016). Dual Neural Network Model for the Evolution of Speech and Language. Trends Neurosci 39, 813-829.
10. Jurgens, U. (2009). The neural control of vocalization in mammals: a review. Journal of voice : official journal of the Voice Foundation 23, 1-10.
11. Nieder, A., and Mooney, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. Philos Trans R Soc Lond B Biol Sci 375, 20190054.
12. Zhang, Y.S., and Ghazanfar, A.A. (2020). A Hierarchy of Autonomous Systems for Vocal Production. Trends Neurosci 43, 115-126.
13. Kittelberger, J.M., Land, B.R., and Bass, A.H. (2006). Midbrain periaqueductal gray and vocal patterning in a teleost fish. Journal of Neurophysiology 96, 71-85.
14. Bass, A.H. (2014). Central pattern generator for vocalization: is there a vertebrate morphotype? Curr Opin Neurobiol 28, 94-100.
15. Jurgens, U. (1994). The role of the periaqueductal grey in vocal behaviour. Behav Brain Res 62, 107-117.
16. Zhang, S.P., Davis, P.J., Bandler, R., and Carrive, P. (1994). Brain stem integration of vocalization: role of the midbrain periaqueductal gray. J Neurophysiol 72, 1337-1356.
17. Tschida, K., Michael, V., Takatoh, J., Han, B.X., Zhao, S., Sakurai, K., Mooney, R., and Wang, F. (2019). A Specialized Neural Circuit Gates Social Vocalizations in the Mouse. Neuron 103, 459-472 e454.
18. Michael, V., Goffinet, J., Pearson, J., Wang, F., Tschida, K., and Mooney, R. (2020). Circuit and synaptic organization of forebrain-to-midbrain pathways that promote and suppress vocalization. Elife 9.
19. Chen, J., Markowitz, J.E., Lilascharoen, V., Taylor, S., Sheurpukdi, P., Keller, J.A., Jensen, J.R., Lim, B.K., Datta, S.R., and Stowers, L. (2021). Flexible scaling and persistence of social vocal communication. Nature 593, 108-113.
20. Okobi, D.E., Jr., 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 363, 983-988.
21. Burkhard, T.T., Westwick, R.R., and Phelps, S.M. (2018). Adiposity signals predict vocal effort in Alston's singing mice. Proceedings of the Royal Society B: Biological Sciences 285.
22. Banerjee, A., Phelps, S.M., and Long, M.A. (2019). Singing mice. Current biology : CB 29, R190-R191.
23. Evarts, E.V. (1968). Relation of Pyramidal Tract Activity to Force Exerted during Voluntary Movement. Journal of Neurophysiology 31, 14-+.
24. Fee, M.S., Kozhevnikov, A.A., and Hahnloser, R.H.R. (2004). Neural mechanisms of vocal sequence generation in the songbird. Ann Ny Acad Sci 1016, 153-170.
25. Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. J Neurosci 3, 1039-1057.
26. Fetz, E.E. (1992). Are Movement Parameters Recognizably Coded in the Activity of Single Neurons. Behav Brain Sci 15, 679-690.
27. Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature 487, 51-+.
28. Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical Control of Arm Movements: A Dynamical Systems Perspective. Annual Review of Neuroscience, Vol 36 36, 337-359.
29. Long, M.A., and Fee, M.S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature 456, 189-194.

30. Glaze, C.M., and Troyer, T.W. (2006). Temporal structure in zebra finch song: implications for motor coding. J Neurosci 26, 991-1005.

31. Tang, L.S., Goeritz, M.L., Caplan, J.S., Taylor, A.L., Fisek, M., and Marder, E. (2010). Precise temperature compensation of phase in a rhythmic motor pattern. PLoS Biology 8, 21-22.

32. Elmaleh, M., Kranz, D., Asensio, A.C., Moll, F.W., and Long, M.A. (2021). Sleep replay reveals premotor circuit structure for a skilled behavior. Neuron 109, 3851-3861 e3854.

33. Yamaguchi, A., Gooler, D., Herrold, A., Patel, S., and Pong, W.W. (2008). Temperature-dependent regulation of vocal pattern generator. J Neurophysiol 100, 3134-3143.

34. Banerjee, A., Egger, R., and Long, M.A. (2021). Using focal cooling to link neural dynamics and behavior. Neuron 109, 2508-2518.

35. Crapse, T.B., and Sommer, M.A. (2008). Corollary discharge across the animal kingdom. Nat Rev Neurosci 9, 587-600.

36. Houde, J.F., and Chang, E.F. (2015). The cortical computations underlying feedback control in vocal production. Curr Opin Neurobiol 33, 174-181.

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

38. Eliades, S.J., and Miller, C.T. (2017). Marmoset vocal communication: Behavior and neurobiology. Dev Neurobiol 77, 286-299.

39. Vallentin, D., and Long, M.A. (2015). Motor origin of precise synaptic inputs onto forebrain neurons driving a skilled behavior. J Neurosci 35, 299-307.

40. Economo, M.N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L.T., Nguyen, T.N., Smith, K.A., Yao, Z., et al. (2018). Distinct descending motor cortex pathways and their roles in movement. Nature 563, 79-84.

41. Network, B.I.C.C. (2021). A multimodal cell census and atlas of the mammalian primary motor cortex. Nature 598, 86-102.

42. Warriner, C.L., Fageiry, S.K., Carmona, L.M., and Miri, A. (2020). Towards Cell and Subtype Resolved Functional Organization: Mouse as a Model for the Cortical Control of Movement. Neuroscience 450, 151-160.

43. Merel, J., Botvinick, M., and Wayne, G. (2019). Hierarchical motor control in mammals and machines. Nat Commun 10, 5489.

44. Lopes, G., Nogueira, J., Paton, J.J., and Kampff, A.R. (2016). A robust role for motor cortex. bioRxiv, 058917.

45. Ebbesen, C.L., and Brecht, M. (2017). Motor cortex - to act or not to act? Nat Rev Neurosci 18, 694-705.

46. Wang, J., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. Nat Neurosci 21, 102-110.

47. Remington, E.D., Egger, S.W., Narain, D., Wang, J., and Jazayeri, M. (2018). A Dynamical Systems Perspective on Flexible Motor Timing. Trends in cognitive sciences 22, 938-952.

48. Mello, G.B., Soares, S., and Paton, J.J. (2015). A scalable population code for time in the striatum. Current biology : CB 25, 1113-1122.

49. Paton, J.J., and Buonomano, D.V. (2018). The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions. Neuron 98, 687-705.

50. Xu, M., Zhang, S.Y., Dan, Y., and Poo, M.M. (2014). Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex. Proc Natl Acad Sci U S A 111, 480-485.

51. Remington, E.D., Narain, D., Hosseini, E.A., and Jazayeri, M. (2018). Flexible Sensorimotor Computations through Rapid Reconfiguration of Cortical Dynamics. Neuron 98, 1005-1019 e1005.

52. Saxena, S., Russo, A.A., Cunningham, J., and Churchland, M.M. (2022). Motor cortex activity across movement speeds is predicted by network-level strategies for generating muscle activity. Elife 11.

53. Stroud, J.P., Porter, M.A., Hennequin, G., and Vogels, T.P. (2018). Motor primitives in space and time via targeted gain modulation in cortical networks. Nat Neurosci 21, 1774-1783.

Supplementary Materials for

# Neural dynamics in the rodent motor cortex enables flexible control of vocal timing

**Authors:** Arkarup Banerjee[1,2,3,4*], Feng Chen[5*], Shaul Druckmann[6], Michael A. Long[1,2,3#]

*Equal contribution

# Corresponding author. Email: Michael A. Long (mlong@med.nyu.edu)

**This file includes:**

Materials and Methods
Extended Data Figs. 1 to 4

## MATERIALS AND METHODS

### Animals

All procedures were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee of NYU Langone Medical Center. Animals used in the study were adult (> 3 months) male laboratory-reared offspring of wild-captured *Scotinomys teguina* from La Carpintera and San Gerardo de Dota, Costa Rica. Mice were maintained at 22 ± 3 °C with a 12:12 L:D cycle.

### Behavioral recordings

*S. teguina* were housed in individual recording chambers (Med Associates) lined with sound insulation foam (Soundproof Cow). Vocalizations were recorded using a condenser microphone (Avisoft Bioacoustics CM16/CMPA) placed within home cages. Acoustic signals were sampled at 250 kHz and digitized with Avisoft UltraSoundGate 116Hb. For playback experiments, we used an ultrasonic tweeter (Vifa), as described previously [1]. To precisely align the audio and electrophysiology signals, each data stream was additionally recorded continuously into an INTAN recording system at a fixed sampling rate between 20-30 KHz.

### Silicon-probe recordings

Chronic recordings were performed using either 64-channel (Cambridge Neurotech, E-1) or integrated 128-channel high-density silicon probes (Diagnostic Biochips, 128-5). Prior to surgery, probes were mounted to a plastic microdrive (NeuroNexus, d-XL) and a stainless-steel ground wire (0.001”, A-M systems) was soldered to the reference of the headstage, which was held in place by a custom-made 3D printed enclosure (Formlabs). For all surgical procedures, mice were anesthetized with 1-2% isoflurane in oxygen and placed in a stereotaxic apparatus. Neural activity of freely moving singing mice was recorded using an electrically assisted commutator (Doric Lenses) and the RHD USB Interface Board or RHD Recording Controller (Intan Technologies). For all chronic recordings, silicon probes were implanted directly into the OMC using the following stereotaxic coordinates: +2.25 mm anterior to bregma, +2.25 mm lateral to the midline. This location represents the center of the OMC region identified by electrical microstimulation [1]. The ground wire was inserted between the skull and the dura above the visual cortex or cerebellum contralateral to the probe implantation. Silicon elastomer (Kwik-Cast, WPI) was applied to the craniotomy once the probe was inserted to the desired depth (1 mm for OMC). The microdrive and the enclosure were secured to the skull with dental acrylic and Metabond cement (Parkell). Animals were monitored and allowed to recover for three to seven days prior to the start of electrophysiology experiments. Spike detection and clustering were performed using KiloSort software [2] and manual post-processing (merging/ splitting of clusters) was performed using phy [3]. Clusters that drifted during the recording session were not included in further analyses. Spike times of all clusters were aligned to onsets and offsets of individual notes or songs as specified below.

### Behavioral annotation of acoustic parameters

We analyzed song structure using custom software (MATLAB) as described previously [1]. Briefly, we first smoothed the sound waveform with a 4-ms sliding window. We then identified individual notes, which typically exhibited an absolute intensity threshold corresponding to 25-40 dB below the mouse's loudest note. Exact note start times and stop times were calculated based on the maximum intensity of each note, such that onsets and offsets were first and last crossings of 1% (20 dB quieter) of each note's maximum intensity. Note duration was calculated as the difference between the offset and the onset for each note. Song duration was defined as the difference between the offset of the last note and the beginning of the first note. For each song, the number of notes was plotted against the overall song duration. For each animal, linear regression (MATLAB function: polyfit) was used to describe how the number of notes vary as a function of song duration. For reanalysis of the previously published cooling data set [1], the number of notes for each song was plotted against the song duration for both control and cooled conditions. A small minority of songs shorter than five seconds, which often had breaks, were ignored. To summarize the effect of cooling, for each animal, the difference between the average number of notes before and after cooling was plotted against the difference of song durations before and after cooling. Since the average song duration varies for individual animals, the difference between cooled and control conditions (Δ notes and Δ song duration) were plotted as opposed to the absolute values.

### Correlation analysis of neuronal ensembles during singing

We performed a correlation analysis for each session individually. We estimated the firing rates from the spike trains using a Gaussian kernel ($\sigma = 0.2$ s). For correlation analyses, we chose the window size based on the longest song duration $T_{\max}$ in that session. To better capture the modulation at the onset and offset, we included 2 s before the song onset and 2 s after the song offset, so the total window size is $T_{\max} + 4\ s$. In this time window, for each song in the session, we sampled every 200 ms from the estimated firing rates to construct the peri-song time histograms (PSTHs). We concatenated the PSTHs from all the neurons for each song into a single vector. The correlation matrix was then constructed by taking the correlation between all pairs of songs. For nonsinging epochs, we performed the same analysis but with song timing (onsets) replaced by control epochs set to be 30 seconds after the song offsets. For each session we averaged the off-diagonal elements in the correlation matrix and performed a one-sided paired t-test to determine the significance.

### Selection of note and song modulated neurons

*Note modulated neurons*: Within a song, consecutive notes usually have short gaps between them (~1/3 of note duration, e.g., **Fig. 2a**). We define a note cycle ($T_{\text{cycle}}$) as the time between the subsequent note onsets. Some songs may have short pauses. To distinguish actual note cycles from pauses, we required the note cycle duration to be less than 3 times the note onset-offset duration. All the analyses on notes shown in this paper were performed with note cycles that passed the above criterion. We verified that our results are robust if we change note cycles to be the time from note onset to offset or the time between the offsets of successive notes. Because notes have variable durations, our analyses were carried out after warping spiking activity to align onsets and offsets, which enabled the calculation of phase tuning. We defined note phase as the relative time within a note cycle, $\phi(t) \equiv \frac{t-t_{\text{onset}}}{T_{\text{cycle}}}$. To select note-modulated neurons, we summarized the spike phases for all the notes and used the Rayleigh $z$-test ($\alpha = 0.01$) to test against the null hypothesis that the spikes within each note cycle were uniformly distributed.

*Song modulated neurons:* We selected the song-modulated neurons initially without warping, i.e., in Absolute Time. Because each song within a session has a different duration, and the different durations could affect estimations of variance, we used the same window size for all songs. Specifically, we chose the window size based on the shortest song duration $T_{\text{min}}$ in that session. We performed statistical tests twice: once for song onset alignment and once for song offset alignment (**Extended Data Fig. 1d**). To better capture the modulations at song onsets or offsets, we include 2 s before the song onsets or 2s after the song offsets. For song onset alignment, we calculated the averaged firing rates within the time window by counting the spikes between 2 s before the song onsets and $T_{\text{min}}$ after the song onsets. For song offset alignment, we calculated the averaged firing rates within the time window by counting the spikes between $T_{\text{min}}$ before the song offsets and 2 s after the song offsets. As a control, we created a baseline nonsinging epoch for each song by counting the spikes from 10 s to 70 s after the song offset. In rare cases when another song appeared in this time window, we excluded the song period and extended the time window to include a total of 60 s of baseline activity. We then performed a two-sided paired $t$-test ($\alpha = 0.01$) to test the null hypothesis that the firing rates within a song were the same as baseline firing rates.

Analysis of note-related neural activity

We found that for many neurons the time course of modulation by notes had a peak that shifted with note duration (e.g., **Fig. 2di**). One possible explanation is that there exists a latency in absolute time between the behavioral recordings and neural activity. To quantify this offset, we reasoned that the optimal latency should give the strongest modulation that is characterized by the $p$-value of the Rayleigh $z$-test. We defined the modified phase as $\tilde{\phi}(t, T_{\text{cycle}}, \Delta T) \equiv \frac{t-t_{\text{onset}}+\Delta T}{T_{\text{cycle}}}$, where $\Delta T$ is the fixed latency in absolute time, $T_{\text{cycle}}$ is the note cycle duration, and $t_{\text{onset}}$ is the note onset. We performed the Rayleigh $z$-test in this modified phase frame and obtained the $p$-value as a function of the latency $\Delta T$. The optimal latency was determined from $\Delta T_{\text{op}} \equiv \underset{\Delta T}{\operatorname{argmin}}\, p(\Delta T)$. To calculate the modulation strength, we first defined the modulation vector as $\vec{m} \equiv \left(\frac{1}{n}\sum_i \sin 2\pi\tilde{\phi}_i, \frac{1}{n}\sum_i \cos 2\pi\tilde{\phi}_i\right)$, where $n$ is the total number of spikes in all the note cycles. We estimated the standard error of the $L_2$ norm of the modulation vector and denoted it as $\Delta\|\vec{m}\|_2$. The modulation strength is then defined as $\frac{\|\vec{m}\|_2}{\Delta\|\vec{m}\|_2}$.

To check whether the latency was sensory- or motor-like, we first selected neurons that had a latency that significantly differed from 0 based on bootstrapping; we randomly sampled the note cycles 1000 times to get the distribution for inferred optimal latency. We then selected neurons which had an optimal latency distribution significantly different from 0 (two sides, $\alpha = 0.01$). We define song modulation strength by the larger absolute $t$-value of the two $t$-tests (i.e., performed on onset- and offset-aligned data).

Analysis of song-related neural activity

To differentiate between the absolute time and relative time models, we constructed a mean template and compared the variance explained by each model. We estimated the firing rates from the spike trains using a Gaussian kernel ($\sigma = 0.2$ s) and denoted this continuous function as $r_\sigma(t)$. For the absolute time model, we set the time window to be the shortest song duration $T_{\min}$ in that session and sampled every 200 ms in this window from $r_\sigma(t)$ to construct the PSTHs. This gave a matrix $\boldsymbol{R}^{\text{abs}}$, which is of size $(n_{\text{song}},\ 5 * T_{\min})$. For each neuron, the mean template was then constructed by taking averages across the rows (i.e., song dimension). We then computed the explained variance ($\rho^2_{\text{abs}}$of the PSTHs about the mean template. For the relative time model, we sampled the same number ($5T_{\min}$) of points evenly from the firing rate function $r_\sigma(t)$ of each neuron after linear warping of time between song onset and song offset. Explicitly stated, $\boldsymbol{R}^{\text{rel}}_{ij} = r_\sigma\left(t^i_{\text{onset}} + \frac{t^i_{\text{offset}} - t^i_{\text{onset}}}{5T_{\min}} j\right)$, where $t^i_{\text{onset}}$ and $t^i_{\text{offset}}$ denote the onset and offset for the i$^{\text{th}}$ song in the session. Following this, identical to above, we computed the mean template and the explained variance using $\boldsymbol{R}^{\text{rel}}$ in place of $\boldsymbol{R}^{\text{abs}}$.

To further quantify the degree of stretching and compression in the relative time model, we performed the following scaling analysis. For each session, we first grouped songs of similar durations using the Jenks Natural Breaks method [4]. We required a valid cluster to have at least four songs and chose the number of clusters to maximize the total number of valid clusters in each session. We then averaged the neural firing rates within each song cluster, $r_\sigma^{(c)}(t) = \frac{1}{|\mathcal{S}_c|}\sum_{j\in\mathcal{S}_c} r_\sigma(t^j_{onset} + t)$, where the superscript $(c)$ denotes the cluster, and $\mathcal{S}_c$ denotes the set of song indices in cluster c. For any two clusters (e.g., $c_1$ and $c_2$), the goal was to find the scaling factor $s_{neural}$ that gave the largest correlation between the two cluster-averaged firing rates $r_\sigma^{(c_1)}(t)$ and $r_\sigma^{(c_2)}(t)$. Formally, for a given scaling factor $s$, we first chose the window size $T_w(s) = \max\left(\frac{T^{(c_1)}}{s},\ T^{(c_2)}\right)$, where $T^{(c_i)}$ is the average song duration in that cluster. We then computed the correlation between the two cluster-averaged firing rates along the time dimension using 31 sampling points from 0 to $T_w(s)$. The optimal neural scaling is defined by $s_{neural} = \underset{0.4\leq s\leq 2.5}{\arg\max}\, \rho\left(r_\sigma^{(c_1)}, r_\sigma^{(c_2)}, s\right)$. We get the behavior scaling factor from $s_{behavioral} \equiv \frac{T^{(c_1)}}{T^{(c_2)}}$. If the neural firing rates can be explained by relative time, we would get $s_{neural} \approx s_{behavioral}$. Depending on whether $T^{(c_2)}$ is longer or shorter than $T^{(c1)}$, the behavior scaling factor $s_{behavioral}$ would be either larger or smaller than 1. To eliminate the ambiguity of these two choices of orders, we required $s_{behavioral} \leq 1$, i.e., we chose the order such that $T^{(c_1)}$ is smaller than $T^{(c_2)}$. For a single neuron, we performed the scaling analysis on all possible combinations of the cluster pairs. To perform this analysis, two valid clusters per session were required (12/13 sessions met this criterion). Scaling analyses were only performed on song modulated neurons whose firing rates exceeded 1 Hz either within the song or within the control. To summarize the results, we binned $s_{behavioral}$ (bin size = 0.05) and plotted the median of $s_{neural}$ within each bin. The best fit line was estimated using quantile regression without intercept.

## Hierarchical Clustering

We estimated firing rates from spike trains using a Gaussian kernel ($\sigma = 0.2$ s) and denoted this continuous function as $r_\sigma(t)$. For the song-modulated neurons, we linearly warped their absolute time firing rates to the relative time firing rates and take the mean across songs, $\bar{r}_\sigma(\theta) = \frac{1}{n_{songs}} \sum_i r_\sigma\left(\left(t^i_{offset} - t^i_{onset}\right)\theta + t^i_{onset}\right)$. We then transformed $\bar{r}_\sigma(\theta)$ to its $z$-score. For each neuron, we sampled $z(\theta)$ from -0.2 to 1.2 with an interval of 0.01, which composes the vector representation of the neural modulation with the song. Agglomerative clustering was carried out on those vector representations. We used Euclidean distance as the affinity function. We chose the distance threshold to be 25. An average template was computed for each cluster by averaging across the neurons within the cluster.

## Computational Model

We constructed a two-step model for hierarchical vocal motor control in the singing mouse. We assumed that a note pattern generator integrates synaptic input and fires upon reaching a fixed threshold using the leaky integrate-and-fire equation:

$$\tau \frac{dV}{dt} = -V + S * r$$

where $V$ = Instantaneous voltage of the note pattern generator and $S$ = synaptic drive onto the pattern generator. $r$ and $\tau$ are the membrane resistance and the membrane time-constant respectively, with units chosen appropriately. $V$ was initialized and reset to 0 mV whenever it reached a particular threshold voltage ($V_{th}$ = 50 mV). This constituted the motor command for producing each note.

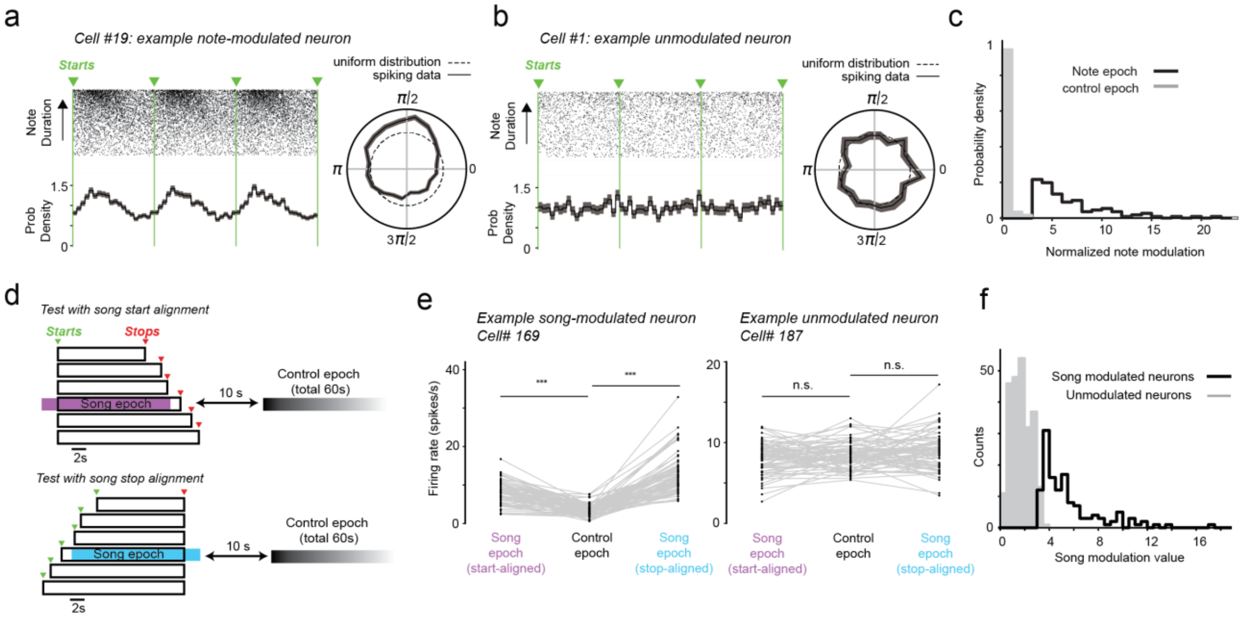
Since the rate of note production per unit time steadily decreases as the song progresses, the overall synaptic drive was required to have a negative slope. In the simplest version of the model, we assumed that the synaptic drive is entirely derived from OMC population activity. The synaptic drive was estimated using a linear combination of synaptic weights from the empirical neural data. The synaptic weights were calculated for one standard song duration (~8 s) close to the average song duration in this species. Notice that the shape of the synaptic drive (sloping down) does not require individual OMC neurons to do so. This should be interpreted as the effective influence of OMC on the note pattern generator. To generate songs of different durations (e.g., T = 4 to 16 s), OMC neural activity was time-scaled by the exact ratio of the song durations (i.e., T/8) based on our empirical result without modifying the synaptic weights. This generates steeper slopes for songs shorter than 8 s and shallower slopes for songs longer than 8 s. This model predicts that the total number of notes corresponding to each song duration increases linearly, which is recapitulated by the behavioral and cooling data. We find that this key result holds for large ranges of the values of the model parameters ($V$, $V_{th}$, $S$, $r$, $\tau$)

Currently, mechanistic details of the pattern generator circuit are unknown. Thus, we explore an alternative scenario by relaxing the assumption that the synaptic drive is entirely driven by OMC without any loss of generality. Its origin can be either entirely driven by OMC, or a combination of OMC and other brain areas. Moreover, the downward sloping synaptic drive can in practice result from a combination of a time-scaled duration signal and spike-frequency adaption (**Extended Data Fig. 4**).

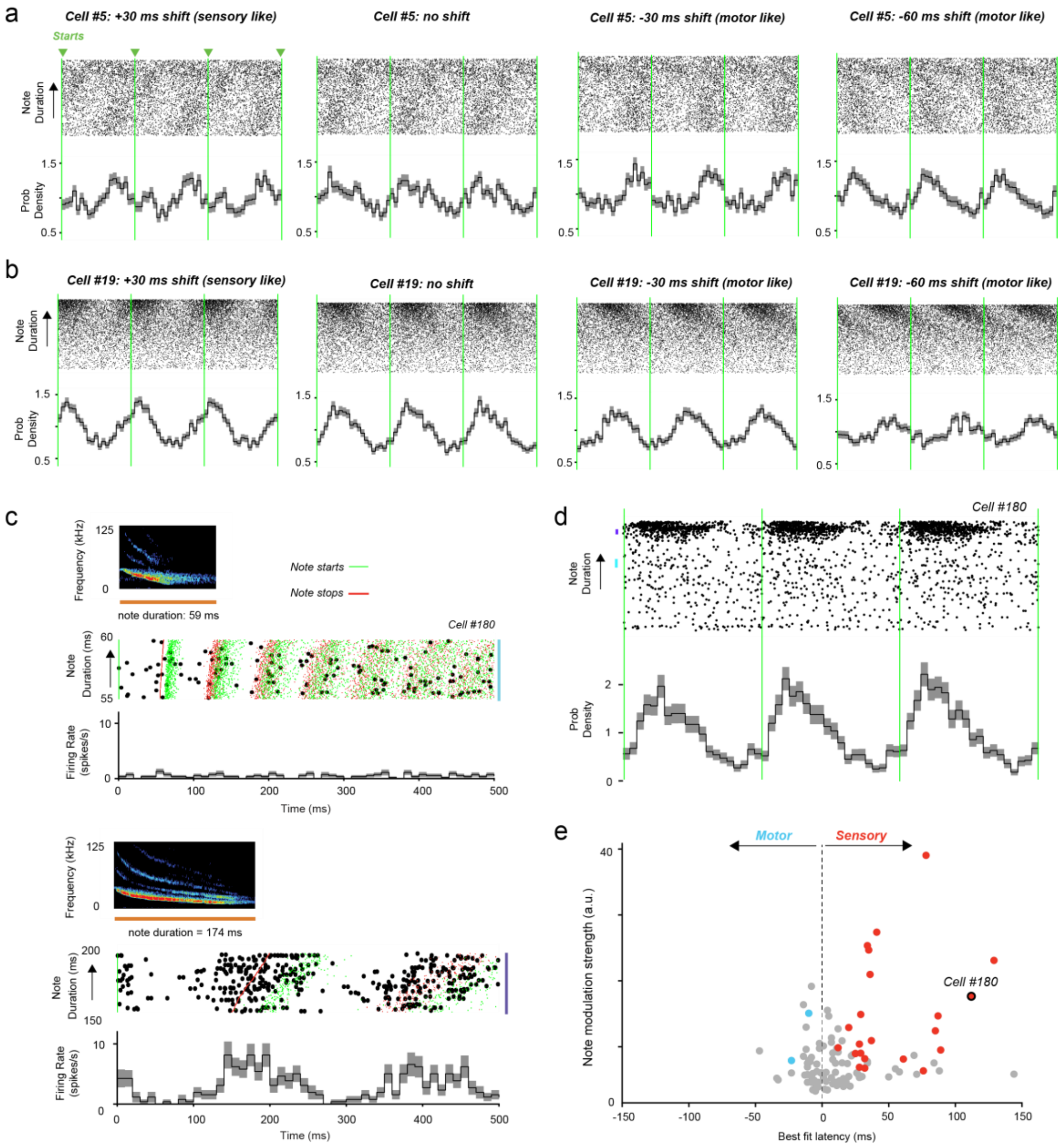
## BIBLIOGRAPHY

1. Okobi, D.E., Jr., 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 *363*, 983-988.
2. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Kenneth D, H. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv, 061481.
3. Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., et al. (2016). Spike sorting for large, dense electrode arrays. Nat Neurosci *19*, 634-641.
4. Jenks, G.F. (1967). The data model concept in statistical mapping. International yearbook of cartography *7*, 186-190.

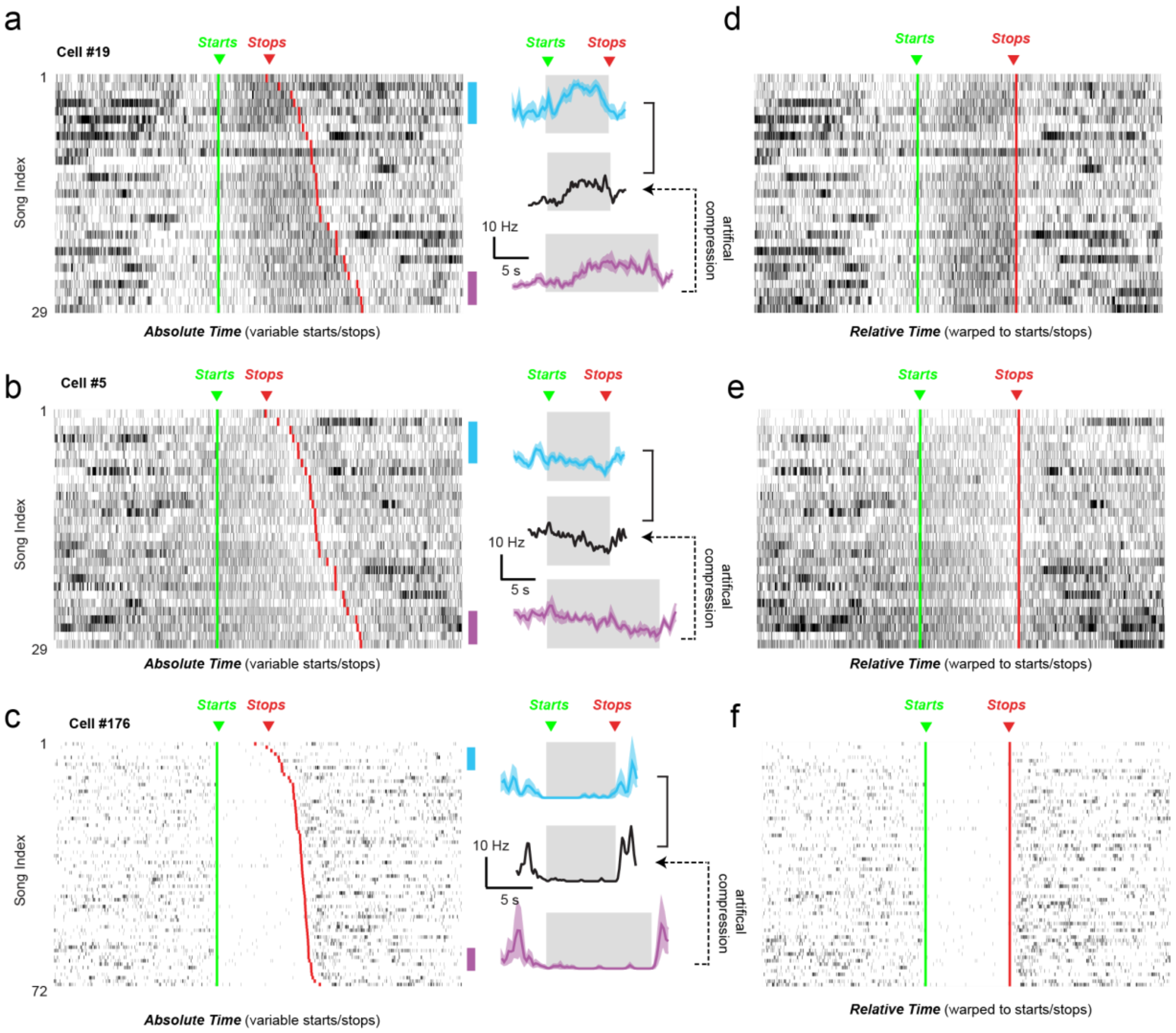
# EXTENDED DATA FIGURES



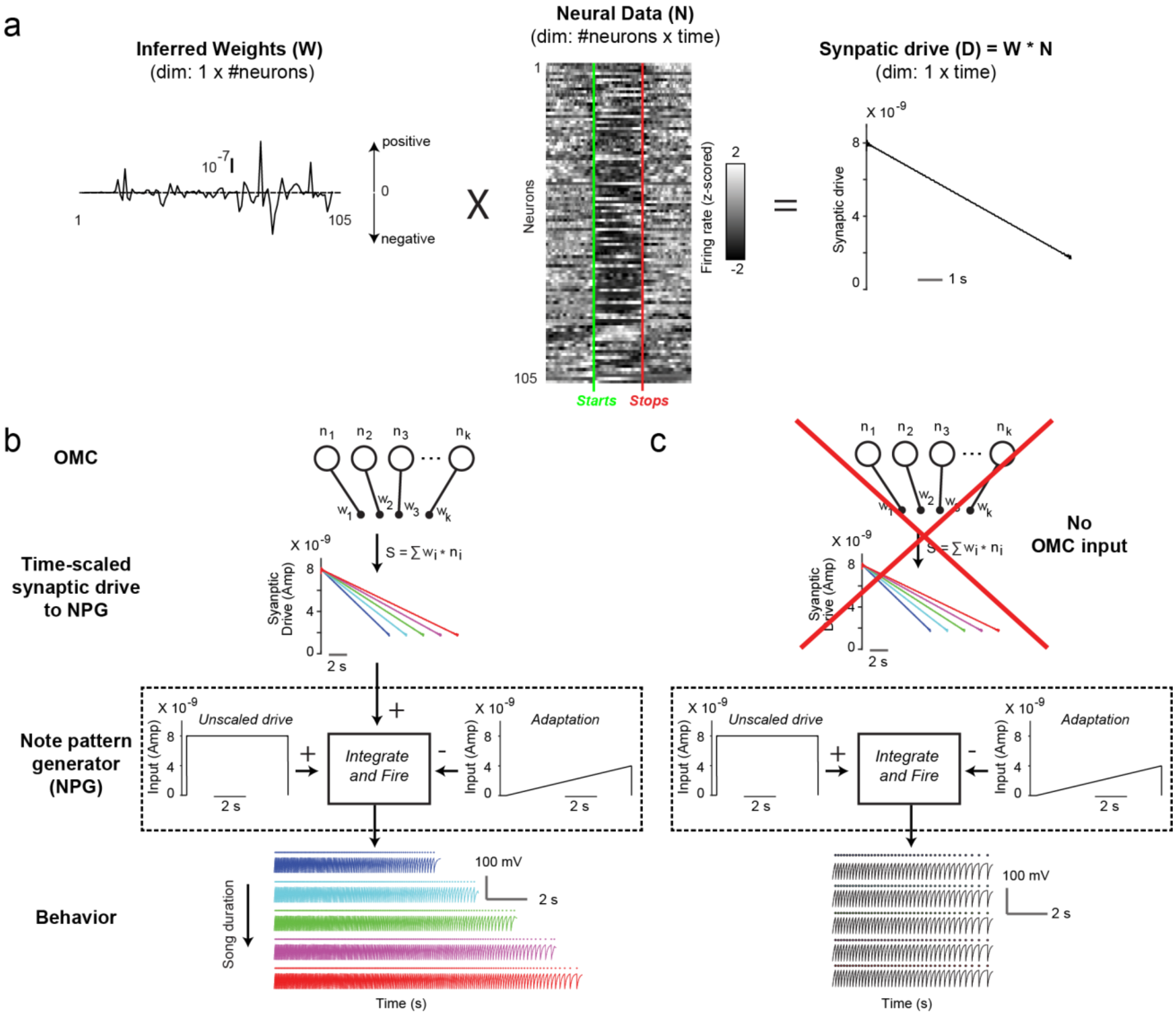
**Extended Data Fig. 1. Determination of significant note- and song-related responses.** (**a** and **b**) Example neurons with (a, Cell #19) and without (b, Cell #1) significant note modulation. Rasters (top) and spike probability density plots (bottom) for example neurons whose activity profiles have been linearly warped to a common note duration (onsets indicated by green lines). At right, polar plots describing the tuning of spike times with respect to the relative phase of note production. Dashed lines indicate a uniform distribution. (**c**) Histogram of note modulation (see Methods) for significantly note-modulated neurons (n = 111) compared with the same analysis applied to nonsinging epochs. (**d**) Song modulation analysis protocol. Neural activity for songs (black rectangles) are aligned either to their starts (top) or stops (bottom). The evaluation window (song epoch) begins and ends two seconds before and after the shortest song duration of that session. (**e**) The relative firing rate difference between the song-aligned spiking activity and a nonsinging period for a modulated (left, Cell #169) and unmodulated (right, Cell #187) neuron. 72 song trials are represented by separate lines for each neuron. Significance determined by bootstrap resampling (***: p < 0.01, n.s.: not significant). (**f**) Histogram of song modulation values (see Methods) for all song modulated neurons (n = 133) and those not modulated by song (n = 242).



**Extended Data Fig. 2. Further characterization of note-related responses.** (**a** and **b**) Spiking activity of two example neurons – Cell #5 (a) and Cell #19 (b) - linearly warped to a common note duration (onsets indicated by dashed lines). At right, the alignment of spikes under normal conditions and after imposing a 'sensory' (- 30 ms) or two different 'motor' (+ 30 ms) and (+ 60 ms) offsets. Examples in (a) and (b) relate to analyses in **Fig. 2e**. (**c**) Spiking activity corresponding to note timing for an example neuron (Cell #180 from Mouse #4). For visualization, analysis was restricted to notes of prespecified durations (top: 55 to 60 ms; bottom: 150 to 200 ms, sample note sonograms provided for each range). For long note durations, robust spiking emerges near the end of each note. Green and red ticks indicate the onset and offset of notes, respectively. (**d**) Spiking activity from Cell #180 linearly warped to a common note duration (onsets indicated by dashed lines). Timing shifted by a best fit latency of 110 ms (sensory-like shift). (**e**) Summary plot (extension from **Fig. 2f**) showing the latency resulting in the maximum note modulation strength for all note modulated neurons (n = 111). Gray symbols represent cases that are not significantly different from zero, and red (n = 23) and blue (n = 2) symbols represent points with sensory and motor offsets, respectively.



**Extended Data Fig. 3. Song-modulated neurons.** (**a-c**) Spiking raster plots for three example neurons – Cell #19 (a), Cell #5 (b), and Cell #176 (c) – across all trials. At right, a peri-song time histogram (PSTH) for song blocks representing the shortest and longest songs in the session (indicated by cyan and magenta vertical lines on right of raster plots). Black curve represents temporally compressed PSTHs from longest trials as a comparison. The magnitude of compression was chosen to match the ratio of the song durations. (**d-f**) Spike times of neurons in (a-c) after temporally warping to the beginning and end of song. Green and red lines indicate the onset and offset of songs, respectively.



**Extended Data Fig. 4. Details of the computational model.** (**a**) Inferred weights (shown at left) for each song-modulated OMC neuron (shown in middle) which leads to a descending synaptic drive (shown at right) to the downstream note pattern generator. (**b**) An alternative implementation of the hierarchical model, in which the note pattern generator produces a song by combining an unscaled step-like input with a characteristic time-dependent adaptation. These inputs could be intrinsic to the pattern generator or could be inherited from a different brain area. In both cases, time-scaled OMC activity can interface with the existing note generating mechanism to produce adaptive behavioral variability. (**c**) In the absence of the OMC input, the note pattern generator can produce notes but loses flexibility resulting in songs with higher stereotypy, consistent with a partially autonomous motor control system.