nature neuroscience

Article

https://doi.org/10.1038/s41593-023-01556-5

Temporal scaling of motor cortical dynamics reveals hierarchical control of vocal production

Received: 13 January 2023

Accepted: 13 December 2023

Published online: 30 January 2024

Check for updates

Arkarup Banerjee [1,2,3,4,7] ✉, Feng Chen [5,7], Shaul Druckmann [6] & Michael A. Long [1,2,3] ✉

Neocortical activity is thought to mediate voluntary control over vocal production, but the underlying neural mechanisms remain unclear. In a highly vocal rodent, the male 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 (~100 ms), probably representing sensory feedback. At longer timescales, however, OMC neurons exhibit diverse and often persistent premotor firing patterns that stretch or compress with song duration (~10 s). Using computational modeling, we demonstrate that such temporal scaling, acting through 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.

Many species exert voluntary control over vocal production 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 nonhuman primates vocalize in response to a conditioned cue[6]. By 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 works 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 gray and specific pattern generator nuclei in the reticular formation. Although these subcortical vocal production mechanisms have been well characterized, much less is known about how cortical activity contributes to vocal production.

To address this issue, we focus our attention on the highly structured 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 (~20–200 ms) that become progressively longer over many seconds, henceforth referred to as a song. Moreover, song duration can vary substantially 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]. Stimulation of the OMC revealed a functional connection to vocally relevant musculature[20] through probable brainstem targets including the reticular formation, the ventrolateral periaqueductal gray and the parabrachial nucleus[23]. Directed perturbations of OMC during singing support a hierarchical control organization in which OMC and subcortical structures mediate song timing and note production, respectively[20].

A major gap in understanding, however, concerns the nature of the cortical activity that drives the moment-by-moment control of this ethologically relevant vocalization. We therefore performed electrophysiology recordings in singing mice to assess the impact of OMC dynamics on vocal production. We found that the population activity within OMC was highly stereotyped during singing compared to nonsinging epochs. The firing rates of individual neurons were strongly

[1]NYU Neuroscience Institute, New York University Langone Health, New York, NY, USA. [2]Department of Otolaryngology, New York University Langone Health, New York, NY, USA. [3]Center for Neural Science, New York University, New York, NY, USA. [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [5]Department of Applied Physics, Stanford University, Stanford, CA, USA. [6]Department of Neurobiology, Stanford University, Stanford, CA, USA. [7]These authors contributed equally: Arkarup Banerjee, Feng Chen. ✉e-mail: abanerjee@cshl.edu; mlong@med.nyu.edu

https://doi.org/10.1038/s41593-023-01556-5

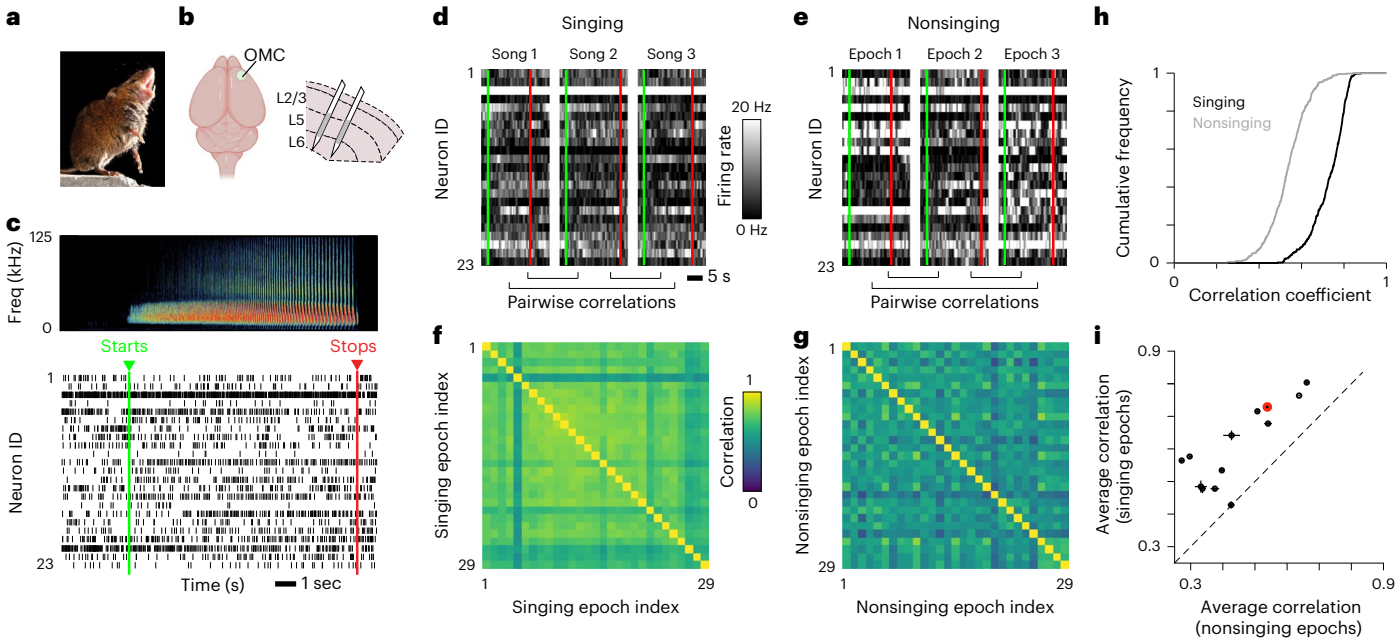


Fig. 1 | Reliable cortical population activity during singing in *S. teguina*. **a**, *S. teguina* singing (photo credit, Christopher Auger-Dominguez). **b**, Schematic of *S. teguina* brain highlighting the recording site (that is, the OMC) as well as the positioning of electrodes (gray-shaded region). **c**, Spiking activity from 23 simultaneously recorded OMC neurons during song production (bottom panel). The sonogram (top panel) depicts *S. teguina* song. Neurons with mean firing rates less than 1 Hz are excluded for visualization purposes. **d**, Firing rates of OMC neural ensemble from **c** during three singing epochs. For **c** and **d**, green and red lines mark the beginning and end of the song, respectively. **e**, Three epochs recorded outside of the song, with green and red lines reflecting the timing of the song from **d**. **f**,**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 than during 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, four mice). Red point refers to the example session shown in **c**–**h**.

modulated by song features at both short (~100 ms) and long (~10 s) timescales, which corresponded to notes and song epochs, respectively. Additionally, trial-to-trial changes in the timing of song performance were reflected in the underlying cortical activity, consistent with the notion that OMC has a central role in the temporal progression of that behavior in conjunction with downstream motor targets. The vocal production pathway of the singing mouse thus provides a compelling example of hierarchical motor control that is likely to be relevant for other behaviors for which cortical involvement is required.

Results

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 600–1,000 µm, such that most recording sites were in the ventral portion (that is, motor output layers) of the OMC. We used this approach to monitor neural activity continuously over 3–20 days, and 13 sessions with robust vocal behavior (mean ± s.d. duration, 10.4 ± 5.7 h) 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 ± s.d.) songs per session (range, 8–72). In total, we recorded data from 375 neurons (mean ± s.d., 29 ± 11 per session) from which spiking was stably monitored throughout the recording sessions (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 (for example, Fig. 1c), we found that the ensemble activity of simultaneously recorded OMC neurons was similar across song epochs and nonsinging periods (Fig. 1d,e). As each session consisted of multiple songs, we calculated the correlation values of OMC ensemble activity across all pairs of songs and found them to be significantly greater than in nonsinging epochs in the example session (Fig. 1f–h) as well as across all recording sessions ($Corr_{singing}$, 0.61 ± 0.11, $Corr_{nonsinging}$, 0.44 ± 0.12; paired t-test, $P = 2.76 \times 10^{-6}$) (Fig. 1i). Taken together, we find that OMC population activity is consistently modulated during song production.

As 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 (~100 ms) or it could follow slower dynamics at timescales comparable to the entire song (~10 s). By statistically comparing neural activity during vocal production (versus nonsinging epochs), we found that 29.6% of neurons (111 out of 375) were correlated with notes (Extended Data Fig. 1a–c and Fig. 2c) while 35.5% of neurons (133 out of 375) displayed dynamics spanning the entire song (Extended Data Fig. 1d–f; Methods) and 13.1% 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 (for example, velocity and force of effector muscles) for many movements[24]. Applying this framework to vocal production, we would expect OMC neurons to show phasic activity patterns preceding each note. To determine the relationship between OMC firing and note production, we linearly warped spiking activity to both the onset and offset

https://doi.org/10.1038/s41593-023-01556-5

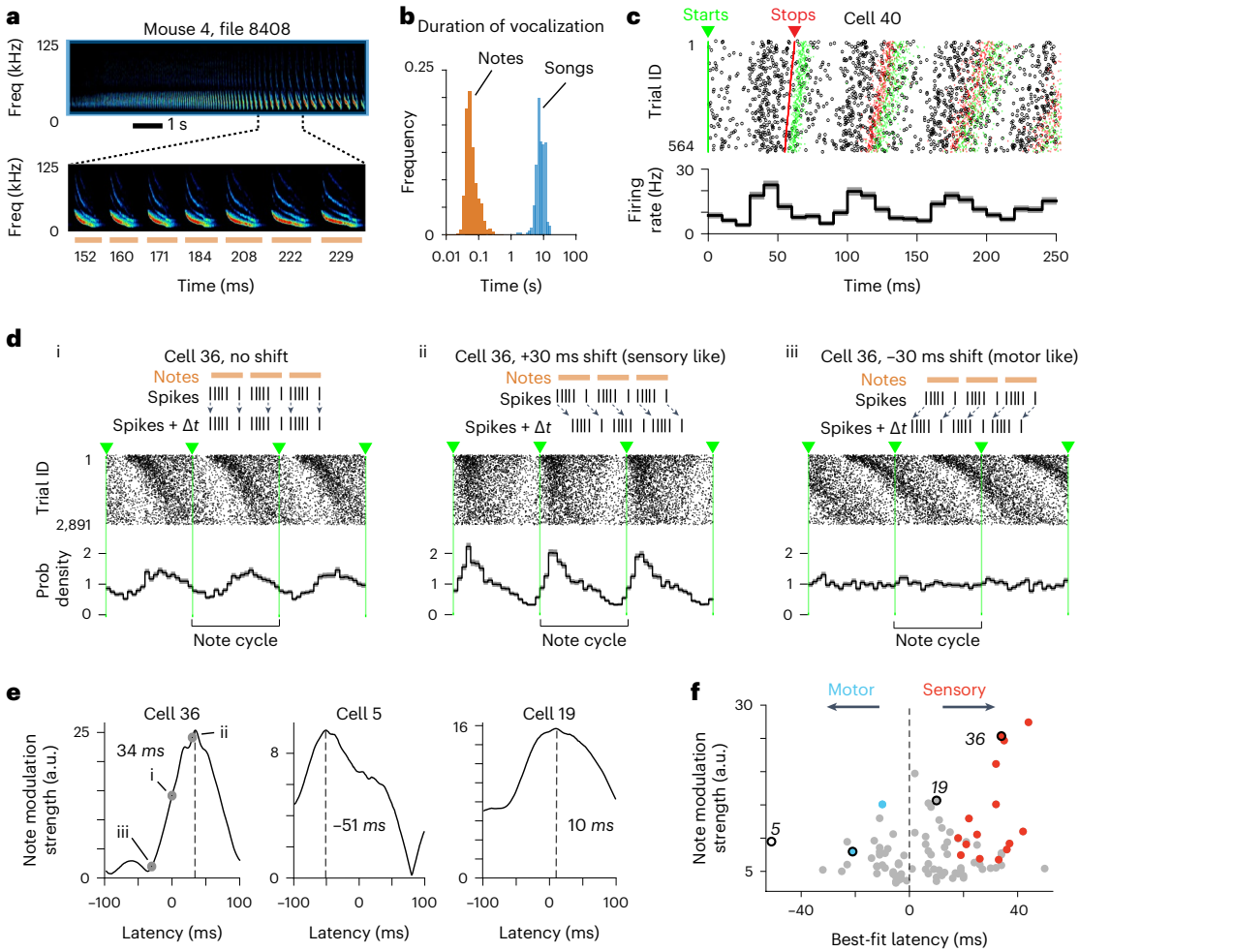


Fig. 2 | Note-related activity of OMC neurons. a, Singing behavior in a single *S. teguina* example song (top) and expanded view of seven notes within the above example (bottom). Horizontal lines represent the timing of notes, and the durations for each note (in ms) are provided below. **b**, Histogram 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, 31.4–175.9 ms), and spiking responses are arranged by accompanying note length, with the longest on the top and the shortest on the bottom. Subsequent notes are included in each row to highlight the periodic firing of this neuron. Green and red ticks indicate note onsets and offsets, respectively. **d**, Spiking activity of an example neuron linearly warped to a common note duration (onsets indicated by dashed green lines). Each row represents the firing of a neuron aligned to the beginning of a sequence of three notes; responses are sorted based on the original duration of the first note produced in this sequence from longest (top) to shortest (bottom). Rasters and spike probability density plots 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 lowercase Roman numerals in the plot for Cell 36 refer to the corresponding panels depicted in **d** (Methods). **f**, Summary plot showing the best-fit latency (restricted to x axis, ±53 ms, and y axis, 0–30 a.u.) 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.

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 (for example, Fig. 2d(i)), which may arise from systematic offsets between neural activity and note production. Specifically, if this shift were caused by a motor delay or the timing needed for premotor signals to result in a behavioral change, activity would precede the production of notes[25]. Conversely, if the timing shift were caused by sensory feedback, spiking activity would lag note production[26].

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

Precise temporal scaling of OMC activity 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 (that is, distinct patterns of neural activity) that do not directly represent movement kinematics[27]. 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

https://doi.org/10.1038/s41593-023-01556-5

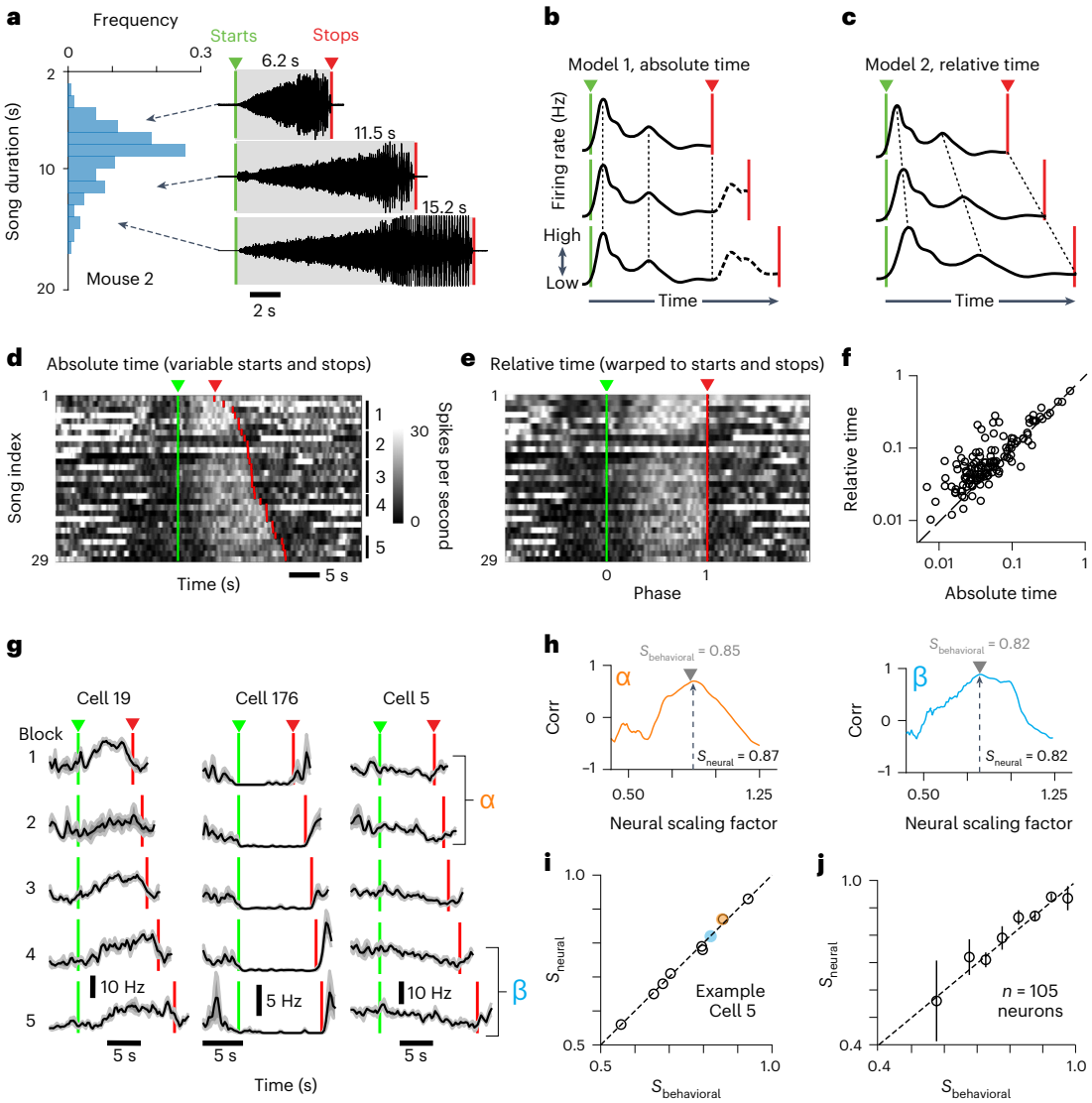


Fig. 3 | Scaling of neural activity with song duration. a, Left, duration of all songs ($n = 143$) produced from one example mouse. Right, raw waveforms for three example songs of different durations. **b**,**c**, Hypothetical time-varying neural activity from a single neuron as predicted by the absolute time (**b**) and relative time (**c**) models for three songs of varying durations. **d**,**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 in Model 2 (one-sided paired *t*-test, $P = 3.95 \times 10^{-7}$), and 101 neurons are to the left of the unity line. **g**, 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 arrowhead shows the ratio of song times ($S_{behavioral}$). **i**,**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 from four mice; *x*-axis bin size, 0.05) (**j**). Error bars, standard error of the median estimated by bootstrapping.

song, which is a configuration that has been proposed in other motor control studies[28,29]. 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 (~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 thus 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%, $n = 13$ sessions; for example, see 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 and Extended Data Fig. 3). The absolute time model predicted 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 (that is, higher correlation after warping). We directly compared these two scenarios and found that the explained variance of single trial firing rates was significantly greater in the warped condition than in 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.

https://doi.org/10.1038/s41593-023-01556-5

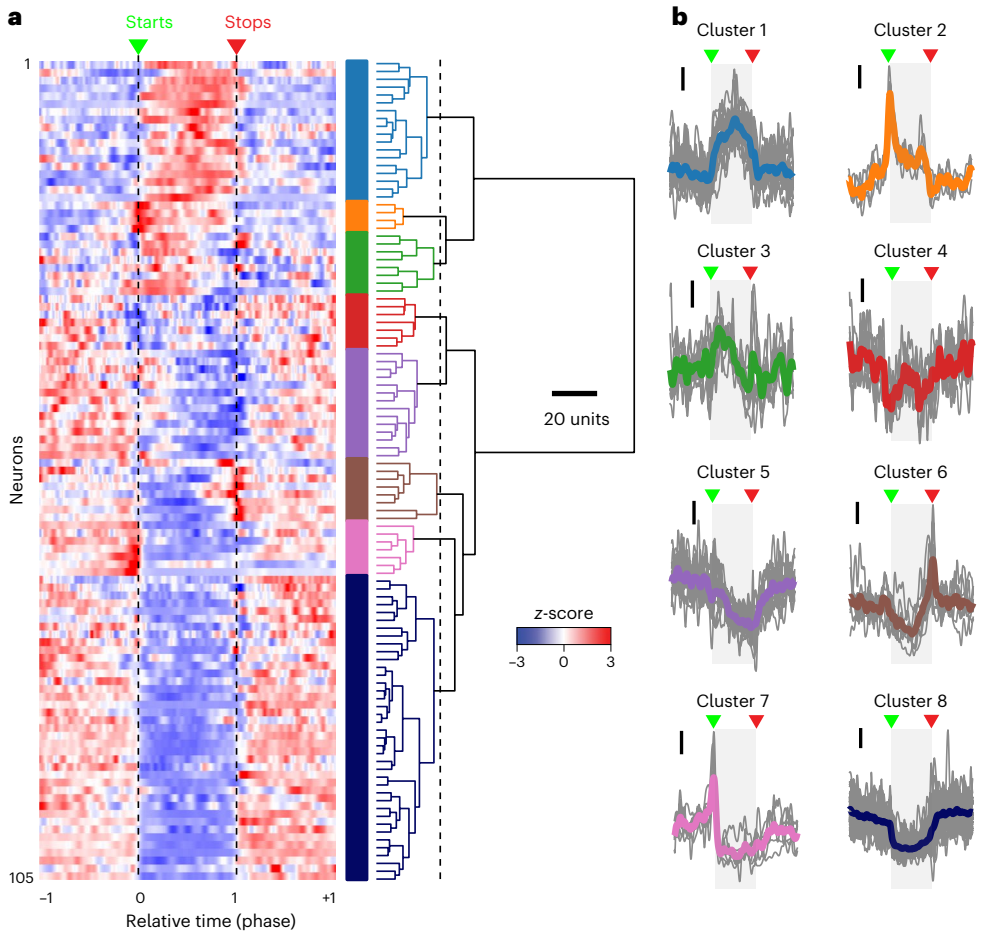


Fig. 4 | Diverse categories of OMC firing patterns during singing. **a**, A hierarchical clustering plot describing the response profiles of OMC neurons whose activity was modulated during singing (Methods; $n = 105$ neurons from four mice). Individual clusters are indicated by colored bars on the right. **b**, Spiking responses for each cluster displayed as average firing rate plots. The mean activity profile of each neuron is represented with gray lines; 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 arrowheads marking song starts and stops, respectively.

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 and Extended Data Fig. 3; Methods). For each pair of blocks, we compared the neural activity profiles to determine the scaling factor that maximized the pairwise correlation (for example, Fig. 3h), which we call the neural scaling factor (S_{neural}). If the optimal neural scaling (that is, the ratio of activity profiles leading to the highest correlation value) matched the relative ratio of associated song durations ($S_{behavioral}$), then the $S_{neural}/S_{behavioral}$ slope is expected to be 1 (equivalent to the relative time model). When S_{neural} was plotted against the behavioral scaling factor (that is, ratio of the associated song durations, $S_{behavioral}$), we found them to be linearly proportional (Fig. 3i,j). Across all the neurons, the neural scaling/behavioral slope was 1.01 ± 0.01 (quantile regression without intercept, $n = 659$ pairs, 105 neurons; Fig. 3j; Methods). These results are consistent across individual animals as well as trial type (that is, solo songs or counter-songs) (Extended Data Fig. 4). For comparison, the absolute time model predicts a slope of 0. This result demonstrates that the 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 the elapsed song.

Diverse individual neuron dynamics in OMC

We next asked what are the motor primitives observed in OMC during vocalization. Given that 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 the 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; Methods) and found that 28.6% of neurons increased firing during song production while the remainder were suppressed. Further analyses of their response profiles—presumably reflecting the sensorimotor processes occurring during vocal production—revealed eight distinct clusters of neurons (Fig. 4a,b), which were confirmed through cross-validation (Extended Data Fig. 5a,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 and 3) or decreases (Clusters 4, 5 and 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. Each cluster exhibited a ratio of neural and behavioral scaling values that did not differ from 1 (Extended Data Fig. 5c), confirming the relative time model. We conclude that the population of OMC neurons that keep track of relative time (that is, 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

https://doi.org/10.1038/s41593-023-01556-5

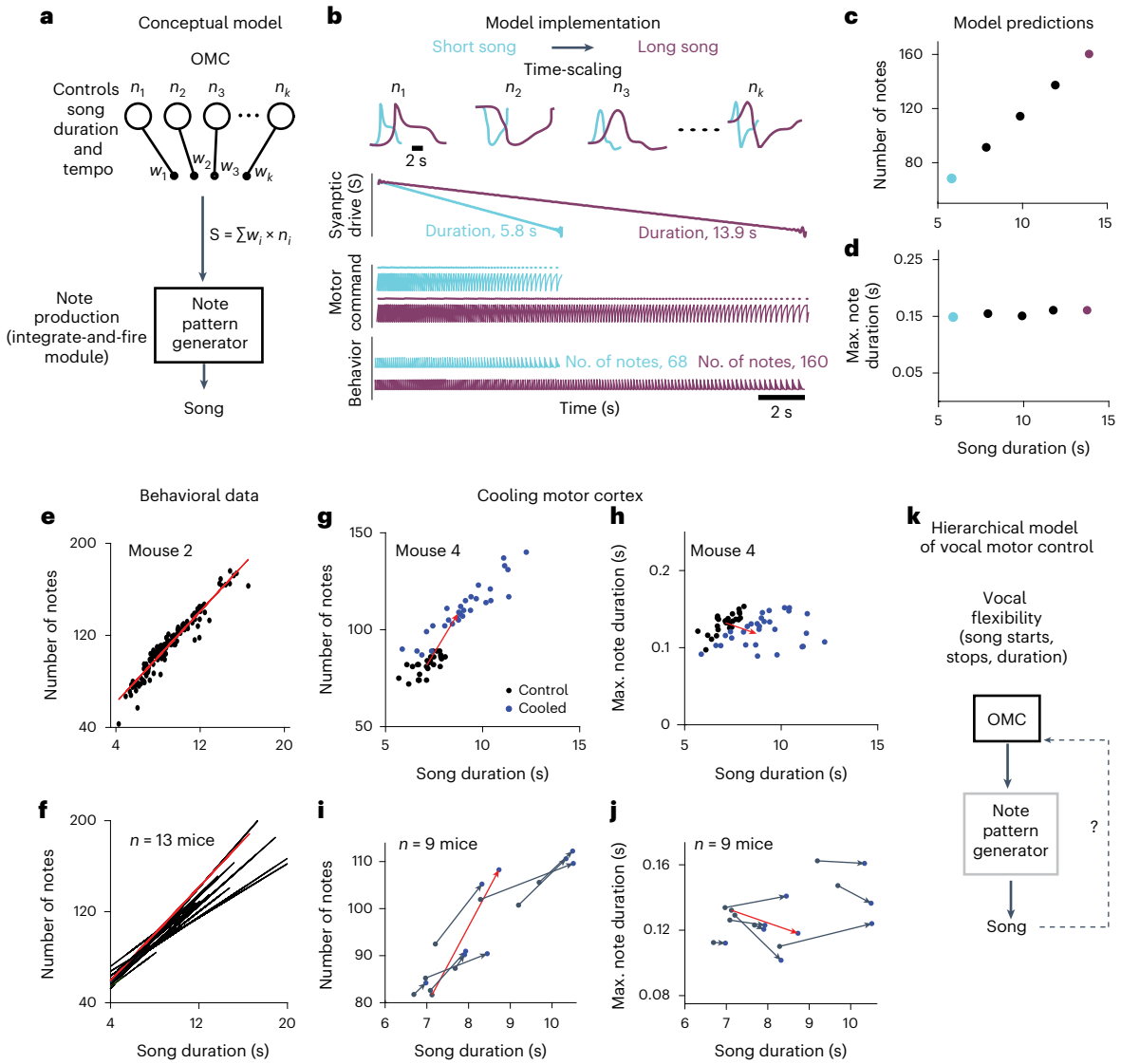


Fig. 5 | Hierarchical model of vocal motor control. a, Schematic depicting shared control of vocal production, in which OMC controls song duration and rate of progression while individual notes are produced by a downstream note pattern generator. The synaptic drive to the note pattern generator is derived from OMC neural activity (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**,**d**, Model output using five different values of time-scaling predicts that the number of notes linearly covaries with song duration (**c**) without an appreciable increase in the duration of the longest note (**d**). Cyan and purple indicate examples from **b**. **e**,**f**, The note number scales with song duration in an example mouse (**e**; $n = 144$ songs, left) as well as across the population (**f**; $n = 13$ mice). Lines represent linear regression fits for each individual animal. Red line indicates data from the example on the left. **g**,**h**, In an example animal (mouse 4), cooling-related expansion of song duration is associated with an increase in the number of produced notes (**g**) without changing the maximum note duration (**h**). **i**,**j**, The average change number of notes (**i**) and maximum note length (**j**) as the result of cooling across the population ($n = 9$ mice). OMC cooling significantly increased average song durations (control, 7.8 ± 0.4 s; cooled, 8.8 ± 0.4 s; paired two-sided t-test, $P = 0.0009$) as well as average number of notes (control, 91.0 ± 3.2; cooled, 100.2 ± 3.7; paired two-sided t-test, $P = 0.005$) without significantly changing the maximum note length (control, 130.7 ± 5.4 ms; cooled, 126.4 ± 5.8 ms; paired t-test, $P = 0.31$). Red lines in **i** and **j** indicate data from mouse 4. **k**, Hierarchical model of vocal motor control, wherein OMC confers flexibility to a downstream song pattern generator.

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 and/or brainstem areas comprising the primary vocal motor network (Fig. 5a and Extended Data Fig. 6). 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 measured in our recordings (Extended Data Fig. 6a), 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 (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 the OMC is robust to the choice of the precise means by which note generation is implemented in the note pattern generator, either by postsynaptic adaptation mechanisms or synaptic drive from another brain region (Extended Data Fig. 6b,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 (Fig. 5c) and not by increasing the duration of individual notes (Fig. 5d). 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 songbirds[30,31]. 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 ($n = 13$ mice; four from this study and an additional nine from a published dataset[20]) (Fig. 5e,f), a finding that strongly agrees with our hierarchical model.

We further considered a directed circuit perturbation to assess whether the relationship between notes and song duration relies on activity within OMC. We reanalyzed a dataset in which OMC was focally cooled in nine mice[20]. Previous experimental[30,32–34] and theoretical[35] 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, 7.8 ± 0.4 s; cooled, 8.8 ± 0.4 s; paired t-test, $P = 0.0009$) and the number of notes (control, 91.0 ± 3.2; cooled, 100.2 ± 3.7; paired t-test, $P = 0.005$), without significantly changing the maximum note length (control, 130.7 ± 5.4 ms; cooled, 126.4 ± 5.8 ms; paired t-test, $P = 0.31$) (Fig. 5g–j). These experimental results concerning the relationship between song length and note production match the predictions of our hierarchical model (Fig. 5c,d). We conclude that cortical activity can generate the necessary vocal motor commands to account for natural variability in behavior.

Discussion

In this study, we used chronic silicon probe recordings to observe neural population activity in *S. teguina* during vocal production. We found that neurons within the orofacial motor cortex exhibited reliable activity across songs, which reflected the highly structured nature of *S. teguina* vocalizations. Specifically, we observed neurons whose activity reflected two behaviorally relevant timescales related to the song: phasic responses during note production (~100 ms) and persistent song-related dynamics (~10 s). 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. 5k). Although the impact of sensory and motor processing on OMC activity during song production remains difficult to disentangle, sensory feedback is known to be important in animal and human vocal motor control[36–39], and a systematic perturbation of sensory streams (for example, auditory, proprioceptive)[40] 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 (for example, transient at song onsets, ramping responses and so on) 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[41–43].

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 through downstream song pattern-generator circuits (Fig. 5k and Extended Data Fig. 6b,c), probably 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, in which we found that OMC inactivation did not abolish singing but significantly reduced the variability in song durations[20], suggesting that activity in OMC provides necessary input to the brainstem to generate socially appropriate vocalizations (Extended Data Fig. 6b,c). As instantaneous note frequency is tightly correlated with song progression (phase), it is difficult to disambiguate whether the OMC neural activity tracks relative time versus a more motor-centric signal that directly influences note frequency. Social, more variable songs that transiently decouple note frequency from song progression (for example, long pauses) can potentially resolve this dichotomy in the future. More broadly, subsequent studies are needed to determine the full song circuit in the singing mouse and elucidate the synaptic mechanisms by which OMC influences downstream vocal production circuits.

Our results support the notion that the singing mouse vocal control network contains a higher-order modulator (that is, the OMC) that extends the capabilities of lower-level motor controllers (that is, note production circuitry) without being necessary for generating the basic motor program. This arrangement—referred to as a partially autonomous hierarchical configuration[44–46]—is a successful design principle for both biological and artificial systems, and it enables behavioral flexibility without relying upon synaptic plasticity in downstream motor patterning circuits. Similar mechanisms have been observed in experiments in which animals are trained to keep track of time[47–55] and in the primate cortex during motor tasks performed at different speeds[56]. Our results extend the scope of such temporal scaling algorithms over an expanded time window (~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, although several ideas have been proposed, including feedback loops[47,52] and neuromodulatory gain control[57]. 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.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-023-01556-5.

References

1. Banerjee, A. & Vallentin, D. Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. *Curr. Opin. Neurobiol.* **73**, 102529 (2022).
2. Pika, S., Wilkinson, R., Kendrick, K. H. & Vernes, S. C. Taking turns: bridging the gap between human and animal communication. *Proc. Biol. Sci.* **285**, 20180598 (2018).
3. Castellucci, G. A., Guenther, F. H. & Long, M. A. A theoretical framework for human and nonhuman vocal interaction. *Annu. Rev. Neurosci.* **45**, 295–316 (2022).
4. Miller, C. T., Thomas, A. W., Nummela, S. U. & de la Mothe, L. A. Responses of primate frontal cortex neurons during natural vocal communication. *J. Neurophysiol.* **114**, 1158–1171 (2015).
5. Roy, S., Zhao, L. & Wang, X. 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 (2016).

6. Hage, S. R., Gavrilov, N. & Nieder, A. Cognitive control of distinct vocalizations in rhesus monkeys. *J. Cogn. Neurosci.* **25**, 1692–1701 (2013).
7. Hage, S. R. & Nieder, A. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409 (2013).
8. Castellucci, G. A., Kovach, C. K., Howard, M. A. 3rd, Greenlee, J. D. W. & Long, M. A. A speech planning network for interactive language use. *Nature* **602**, 117–122 (2022).
9. Hage, S. R. & Nieder, A. Dual neural network model for the evolution of speech and language. *Trends Neurosci.* **39**, 813–829 (2016).
10. Jürgens, U. The neural control of vocalization in mammals: a review. *J. Voice* **23**, 1–10 (2009).
11. Nieder, A. & Mooney, R. The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Phil. Trans. R. Soc. B* **375**, 20190054 (2020).
12. Zhang, Y. S. & Ghazanfar, A. A. A hierarchy of autonomous systems for vocal production. *Trends Neurosci.* **43**, 115–126 (2020).
13. Kittelberger, J. M., Land, B. R. & Bass, A. H. Midbrain periaqueductal gray and vocal patterning in a teleost fish. *J. Neurophysiol.* **96**, 71–85 (2006).
14. Bass, A. H. Central pattern generator for vocalization: is there a vertebrate morphotype? *Curr. Opin. Neurobiol.* **28**, 94–100 (2014).
15. Jurgens, U. The role of the periaqueductal grey in vocal behaviour. *Behav. Brain Res.* **62**, 107–117 (1994).
16. Zhang, S. P., Davis, P. J., Bandler, R. & Carrive, P. Brain stem integration of vocalization: role of the midbrain periaqueductal gray. *J. Neurophysiol.* **72**, 1337–1356 (1994).
17. Tschida, K. et al. A specialized neural circuit gates social vocalizations in the mouse. *Neuron* **103**, 459–472.e4 (2019).
18. Michael, V. et al. Circuit and synaptic organization of forebrain-to-midbrain pathways that promote and suppress vocalization. *eLife* **9**, e63493 (2020).
19. Chen, J. et al. Flexible scaling and persistence of social vocal communication. *Nature* **593**, 108–113 (2021).
20. Okobi, D. E. Jr, Banerjee, A., Matheson, A. M. M., Phelps, S. M. & Long, M. A. Motor cortical control of vocal interaction in neotropical singing mice. *Science* **363**, 983–988 (2019).
21. Burkhard, T. T., Westwick, R. R. & Phelps, S. M. Adiposity signals predict vocal effort in Alston's singing mice. *Proc. R. Soc. B* **285**, 20180090 (2018).
22. Banerjee, A., Phelps, S. M. & Long, M. A. Singing mice. *Curr. Biol.* **29**, R190–R191 (2019).
23. Zheng, D. J. et al. Mapping the vocal circuitry of Alston's singing mouse with pseudorabies virus. *J. Comp. Neurol.* **530**, 2075–2099 (2022).
24. Evarts, E. V. Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* **31**, 14–27 (1968).
25. Fee, M. S., Kozhevnikov, A. A. & Hahnloser, R. H. R. Neural mechanisms of vocal sequence generation in the songbird. *Ann. N. Y. Acad. Sci.* **1016**, 153–170 (2004).
26. Margoliash, D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* **3**, 1039–1057 (1983).
27. Fetz, E. E. Are movement parameters recognizably coded in the activity of single neurons? *Behav. Brain Sci.* **15**, 679–690 (1992).
28. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
29. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
30. Long, M. A. & Fee, M. S. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).
31. Glaze, C. M. & Troyer, T. W. Temporal structure in zebra finch song: implications for motor coding. *J. Neurosci.* **26**, 991–1005 (2006).
32. Tang, L. S. et al. Precise temperature compensation of phase in a rhythmic motor pattern. *PLoS Biol.* **8**, e1000469 (2010).
33. Elmaleh, M., Kranz, D., Asensio, A. C., Moll, F. W. & Long, M. A. Sleep replay reveals premotor circuit structure for a skilled behavior. *Neuron* **109**, 3851–3861.e4 (2021).
34. Yamaguchi, A., Gooler, D., Herrold, A., Patel, S. & Pong, W. W. Temperature-dependent regulation of vocal pattern generator. *J. Neurophysiol.* **100**, 3134–3143 (2008).
35. Banerjee, A., Egger, R. & Long, M. A. Using focal cooling to link neural dynamics and behavior. *Neuron* **109**, 2508–2518 (2021).
36. Crapse, T. B. & Sommer, M. A. Corollary discharge across the animal kingdom. *Nat. Rev. Neurosci.* **9**, 587–600 (2008).
37. Houde, J. F. & Chang, E. F. The cortical computations underlying feedback control in vocal production. *Curr. Opin. Neurobiol.* **33**, 174–181 (2015).
38. Eliades, S. J. & Wang, X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* **453**, 1102–1106 (2008).
39. Eliades, S. J. & Miller, C. T. Marmoset vocal communication: behavior and neurobiology. *Dev. Neurobiol.* **77**, 286–299 (2017).
40. Vallentin, D. & Long, M. A. Motor origin of precise synaptic inputs onto forebrain neurons driving a skilled behavior. *J. Neurosci.* **35**, 299–307 (2015).
41. Economo, M. N. et al. Distinct descending motor cortex pathways and their roles in movement. *Nature* **563**, 79–84 (2018).
42. Network, B. I. C. C. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
43. Warriner, C. L., Fageiry, S. K., Carmona, L. M. & Miri, A. Towards cell and subtype resolved functional organization: mouse as a model for the cortical control of movement. *Neuroscience* **450**, 151–160 (2020).
44. Merel, J., Botvinick, M. & Wayne, G. Hierarchical motor control in mammals and machines. *Nat. Commun.* **10**, 5489 (2019).
45. Lopes, G. et al. A robust role for motor cortex. *Front. Neurosci.* **17**, 971980 (2023).
46. Ebbesen, C. L. & Brecht, M. Motor cortex—to act or not to act? *Nat. Rev. Neurosci.* **18**, 694–705 (2017).
47. Wang, J., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* **21**, 102–110 (2018).
48. Remington, E. D., Egger, S. W., Narain, D., Wang, J. & Jazayeri, M. A dynamical systems perspective on flexible motor timing. *Trends Cogn. Sci.* **22**, 938–952 (2018).
49. Mello, G. B., Soares, S. & Paton, J. J. A scalable population code for time in the striatum. *Curr. Biol.* **25**, 1113–1122 (2015).
50. Paton, J. J. & Buonomano, D. V. The neural basis of timing: distributed mechanisms for diverse functions. *Neuron* **98**, 687–705 (2018).
51. Xu, M., Zhang, S. Y., Dan, Y. & Poo, M. M. Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex. *Proc. Natl Acad. Sci. USA* **111**, 480–485 (2014).
52. Remington, E. D., Narain, D., Hosseini, E. A. & Jazayeri, M. Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron* **98**, 1005–1019.e5 (2018).
53. De Corte, B. J., Akdogan, B. & Balsam, P. D. Temporal scaling and computing time in neural circuits: should we stop watching the clock and look for its gears? *Front. Behav. Neurosci.* **16**, 1022713 (2022).
54. Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y. & Tanji, J. Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nat. Neurosci.* **12**, 502–507 (2009).

55. Renoult, L., Roux, S. & Riehle, A. Time is a rubberband: neuronal activity in monkey motor cortex in relation to time estimation. *Eur. J. Neurosci.* **23**, 3098–3108 (2006).
56. Saxena, S., Russo, A. A., Cunningham, J. & Churchland, M. M. Motor cortex activity across movement speeds is predicted by network-level strategies for generating muscle activity. *eLife* **11**, e67620 (2022).
57. Stroud, J. P., Porter, M. A., Hennequin, G. & Vogels, T. P. Motor primitives in space and time via targeted gain modulation in cortical networks. *Nat. Neurosci.* **21**, 1774–1783 (2018).

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 old) male laboratory-reared offspring of wild-captured *Scotinomys teguina* from La Carpintera and San Gerardo de Dota, Costa Rica. Mice were maintained in a temperature-controlled (22 ± 3 °C) and humidity-controlled environment with a 12:12 h light/dark cycle. Animals were housed in large rat-style cages with a running wheel and dietary enrichment. Careful veterinary oversight was provided, with regular inspections of both breeding facilities and behavioral arenas.

Behavioral recordings

S. teguina adults 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[20]. 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 and 30 kHz. Natural variability in vocal behavior in this species resulted in sessions during which a variable number of songs were recorded. For statistical reasons, sessions with at least eight songs were analyzed further (13 sessions from four animals).

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). Before 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. The 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[20]. 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 3–7 days before the start of electrophysiology experiments. Spike detection and clustering were performed using KiloSort1 software[58] and manual post-processing (merging and/or splitting of clusters) was performed using phy1 (ref. 59). Clusters that drifted during the recording session were not included in further analyses. Neuron counts reflect the number of 'session-neurons', which does not rule out the possibility that some individual neurons may be recorded across multiple sessions. 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 R2016b) as described previously[20]. In brief, 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 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 varies as a function of song duration. For reanalysis of the previously published cooling dataset[20], the number of notes as well as the note duration of the longest note (maximum note duration) for each song was plotted against the song duration for both control and cooled conditions. A small minority of songs (~3% of total attempts) that were shorter than 3 s were ignored. To summarize the cooling result, for each animal, we calculated the averages of note number, maximum note duration and song durations.

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). The window size for correlation analyses was determined according to the longest song duration (T_{max}) in that session. To better capture the modulation at the onset and offset, an additional 2 s was included before the song onset and after the song offset, leading to a total window size of $T_{max} + 4$ s. Within this time window, we sampled at 200 ms intervals from the estimated firing rates to construct the peri-song time histograms (PSTHs) for each song in the session. We then concatenated the PSTHs from all the neurons for each song into a single vector. A correlation matrix was then constructed by taking the correlation between all pairs of songs. For the nonsinging epochs, we followed a similar procedure but with song timing (onsets) replaced by control epochs, which were set to be 30 s 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 assess the significance.

Selection of note-modulated and song-modulated neurons

Note-modulated neurons. Within a song, consecutive notes usually possess short gaps between them (~1/3 of the note duration, as shown in Fig. 2a). We define a note cycle (T_{cycle}) as the time between the onsets of subsequent notes. Some songs may have short pauses. To distinguish actual T_{cycles} from these pauses, we required that the T_{cycle} duration should be less than three times the note onset–offset duration. All the analyses on notes in this paper were performed with T_{cycles} that met the above criterion. We verified that our results remain consistent if we change T_{cycles} to be the time from note onset to offset or the time between the offsets of successive notes. Given that 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 T_{cycle} as $\phi(t) \equiv \frac{t - t_{onset}}{T_{cycle}}$. To identify 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, positing uniform distribution of spikes within each note cycle.

Song-modulated neurons. We selected the song-modulated neurons initially without warping, that is, in absolute time. As 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 determined the window size based on the shortest song duration (T_{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 included an additional 2 s before the song onsets or following the song offsets. For song onset alignment, we calculated the averaged firing rates within the time window by counting the spikes from 2 s before the song onsets to $T_{\min}$ after the song onsets. For song offset alignment, we calculated the averaged firing rates within the time window by counting the spikes from $T_{\min}$ before the song offsets to 2 s after the song offsets. As a control, we created a baseline nonsinging epoch for each song by counting the spikes from 10 to 70 s after the song offsets. 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 ($\alpha = 0.01$, unless stated otherwise) 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 (for example, Fig. 2d(i)). One possible explanation is that there is a latency in absolute time between the behavioral recordings and neural activity. To quantify this offset, we postulated that the optimal latency should give the strongest modulation. 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 ΔT is the fixed latency in absolute time, T_{cycle} is the note cycle duration and t_{onset} is the note onset. We can obtain the modulation vector by summarizing the modified phases: $\vec{m} \equiv \left(\frac{1}{n}\sum_i \sin 2\pi\tilde{\phi}_i, \frac{1}{n}\sum_i \cos 2\pi\tilde{\phi}_i\right)$, where n is the total number of spikes in all T_{cycles} and the summation is overall spike times indexed by i. We also 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 $M(\Delta T) \equiv \frac{\|\vec{m}\|_2}{\Delta\|\vec{m}\|_2}$, which is a function of the absolute latency ΔT applied to obtain the modified phases. The optimal latency was determined from $\Delta T_{\text{op}} \equiv \underset{\Delta T}{\operatorname{argmin}} M(\Delta T)$.

To determine whether the latency was sensory-like or motor-like, we selected neurons that had a latency significantly different from zero based on bootstrapping. We randomly sampled the note cycles 1,000 times to obtain the distribution for inferred optimal latency. We then selected neurons that had an optimal latency distribution significantly different from zero (two sides, $\alpha = 0.01$). In total, 25 neurons were found to have latencies that were significantly different from zero.

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. To do so, 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 $T_{\min}$ in that session and sampled every 200 ms in this window from $r_\sigma(t)$ to construct the PSTHs. This gave a matrix $\boldsymbol{R}^{\text{abs}}$ with dimension (n_{song}, $5 \times T_{\min}$). For each neuron, we then constructed the mean template by taking averages across the rows (that is, song dimension) and computed the explained variance of the PSTHs about the mean template. For the relative time model, we sampled the same number ($5 \times T_{\min}$) of points evenly from the firing rate function $r_\sigma(t)$ after linearly warping 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_{onset} and t^i_{offset} denote the onset and offset for the ith song in the session. Following this calculation, identical to the above process, 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[60]. We averaged the neural firing rates within each song cluster, $r^{(c)}_\sigma(t) = \frac{1}{|\mathcal{S}_c|}\sum_{j\in\mathcal{S}_c} r_\sigma(t^j_{\text{onset}} + t)$, in which the superscript (c) denotes the cluster and $\mathcal{S}_c$ denotes the set of song indices in cluster c. For any two clusters (for example, 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 $r^{(c_1)}_\sigma(t)$ and $r^{(c_2)}_\sigma(t)$. 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 $\rho\left(r^{(c_1)}_\sigma, r^{(c_2)}_\sigma, s\right)$ between the two cluster-averaged firing rates along the time dimension using 31 points sampled equidistantly from 0 to $T_w(s)$. The optimal neural scaling was obtained from $s_{\text{neural}} = \underset{0.4\le s\le 2.5}{\operatorname{argmax}} \rho\left(r^{(c_1)}_\sigma, r^{(c_2)}_\sigma, s\right)$. We obtained the behavior scaling factor from $s_{\text{behavioral}} \equiv \frac{T^{(c_1)}}{T^{(c_2)}}$. If the neural firing rates can be explained by relative time, we would obtain $s_{\text{neural}} \approx s_{\text{behavioral}}$. Depending on whether $T^{(c_2)}$ is longer or shorter than $T^{(c1)}$, the behavior scaling factor $s_{\text{behavioral}}$ would be either larger or smaller than 1. To eliminate the ambiguity of these two choices of orders, we required $s_{\text{behavioral}} \le 1$; that is, we chose the order such that $T^{(c_1)}$ is smaller than $T^{(c_2)}$. We performed the scaling analysis on all possible combinations of the cluster pairs for each neuron. To perform this analysis, two valid clusters per session were required (12 out of 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 $s_{\text{behavioral}}$ (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. Note that for this scaling analysis sorted by individual animals (Extended Data Fig. 4b), we were limited by the total number of songs from mouse 1. Therefore, the significance threshold for including a neuron was relaxed to $P = 0.05$.

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 determined the mean across songs, $\bar{r}_\sigma(\theta) = \frac{1}{n_{\text{songs}}}\sum_i r_\sigma\left(\left(t^i_{\text{offset}} - t^i_{\text{onset}}\right)\theta + t^i_{\text{onset}}\right)$. We then transformed $\bar{r}_\sigma(\theta)$ to its z-score $z(\theta)$. For each neuron, we sampled the z-score from $\theta = -0.2$ to $\theta = 1.2$ 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.

Cross-validation for hierarchical clustering

To cross-validate the hierarchical clustering results, we randomly split the songs into two halves: one for training and another for testing. During training, we took the mean across the training songs, $\bar{r}^{(\text{train})}_\sigma(\theta) = \frac{1}{|\mathcal{S}_{\text{train}}|}\sum_{i\in\mathcal{S}_{\text{train}}} r_\sigma\left(\left(t^i_{\text{offset}} - t^i_{\text{onset}}\right)\theta + t^i_{\text{onset}}\right)$ and transformed it to its z-score $z^{(\text{train})}(\theta)$. For each neuron, we sampled the z-score from $\theta = -0.2$ to $\theta = 1.2$ with an interval of 0.01, which composes the vector representation of the neural modulation with the song. Agglomerative clustering was then carried out on those vector representations. We used Euclidean distance as the affinity function and chose the threshold so that there were eight clusters. During testing, we recomputed the z-score $z^{(\text{test})}(\theta)$ from the test songs. We showed the test matrix in the same order as the training matrix. We also computed the cluster-averaged traces of the test z-scores against those of the training z-scores.

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:

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

where V is the instantaneous voltage of the note pattern generator and S is the synaptic drive onto the pattern generator; r and τ 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.

Given that the rate of note production per unit of 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), which is close to the average song duration in this species. Note 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 (for example, T = 4 to 16 s), OMC neural activity was time-scaled by the exact ratio of the song durations (that is, 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, τ).

Currently, the mechanistic details of the pattern generator circuit are unknown. Therefore, 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 driven either entirely by OMC or by 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 adaptation (Extended Data Fig. 6).

Statistics and reproducibility

No statistical methods were used to pre-determine sample sizes but they were constrained by the amount of spontaneous singing that occurred within a single session (for example, 1 out of 13 sessions was excluded owing to an insufficient number of songs). Additionally, neurons with firing rates less than one spike per second were excluded to ensure statistical rigor. Using cross-validation, we have demonstrated sufficient statistical power to support our claims (for example, Extended Data Fig. 5). The main results reported in the paper were replicated by two authors (A.B. and F.C.) with analysis codes written independently using two different software packages (MATLAB R2016b and Python v.3.8.10). Randomization and blinding were not performed because our study does not include experimental and control conditions.

Reporting summary

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

Data availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, M.A.L. (mlong@med.nyu.edu). This study did not generate new unique reagents. The datasets generated during this study are available upon request from the lead contact. Source code and documentation required for running all analyses are available.

Code availability

Analysis code is available in a GitHub repository (https://github.com/ccffccffcc/NNSingingMouse).

References

58. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Kenneth D. H. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. Preprint at *bioRxiv* https://doi.org/10.1101/061481 (2016).
59. Rossant, C. et al. Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).
60. Jenks, G. F. The data model concept in statistical mapping. *Int. Yearb. Cartogr.* **7**, 186–190 (1967).

Acknowledgements

We thank S. Shea, F. Albeanu, W. Bast, J. del Rosario, H. Sloin and members of the Long and Banerjee laboratories for comments on earlier versions of the manuscript. A. Paulson provided technical assistance. Funding was provided by the National Institutes of Health grant R01 NS113071 (M.A.L., S.D.), Simons Collaboration on the Global Brain (M.A.L., S.D.), Searle Scholars Program (A.B.), Klingenstein–Simons fellowship (A.B.) and the Simons Foundation Junior Fellows Program (A.B.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.B. and M.A.L. conceived the project; A.B., F.C., S.D. and M.A.L. designed the methodology. A.B. and M.A.L. performed the investigation. A.B., F.C., S.D. and M.A.L. visualized the project. A.B., S.D. and M.A.L. acquired funding. S.D. and M.A.L. administered and supervised the project. A.B. and M.A.L. wrote the original draft of the manuscript; A.B., F.C., S.D. and M.A.L. contributed to writing, review and editing.

Competing interests

The authors declare no competing interests.

Additional information

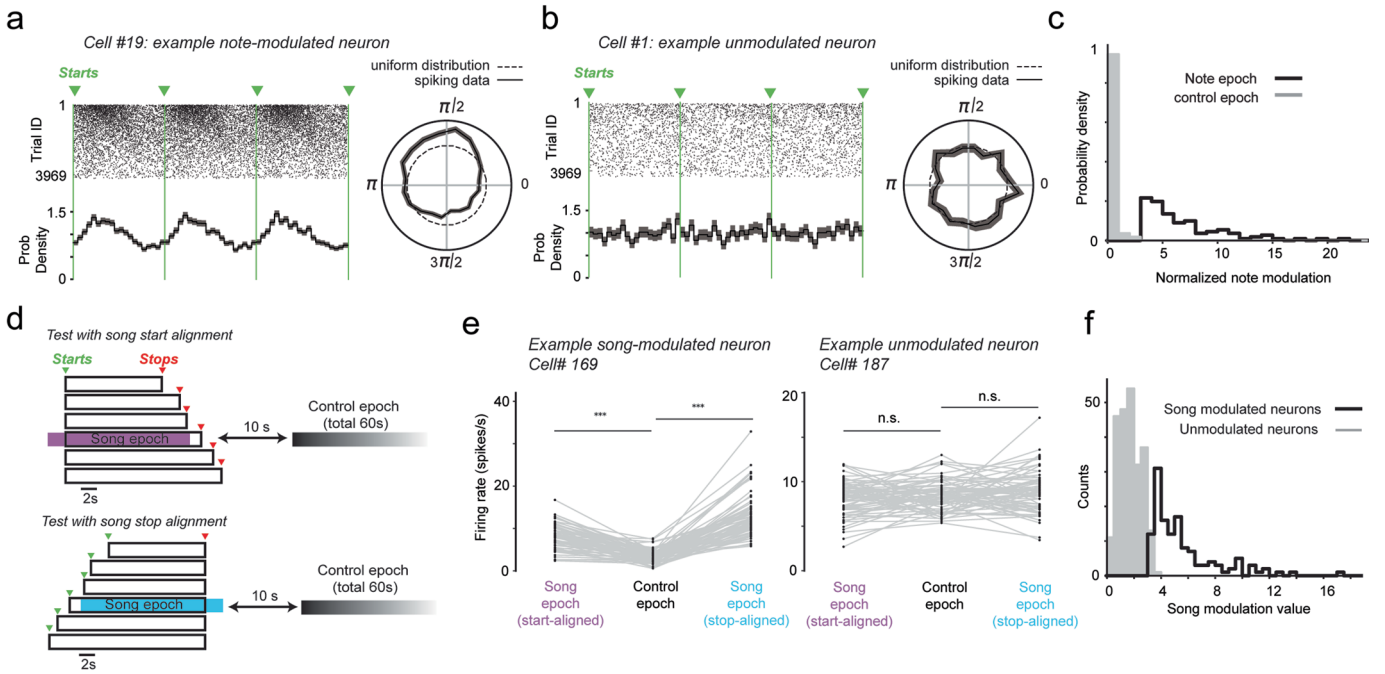
Extended data is available for this paper at https://doi.org/10.1038/s41593-023-01556-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-023-01556-5.

Correspondence and requests for materials should be addressed to Arkarup Banerjee or Michael A. Long.

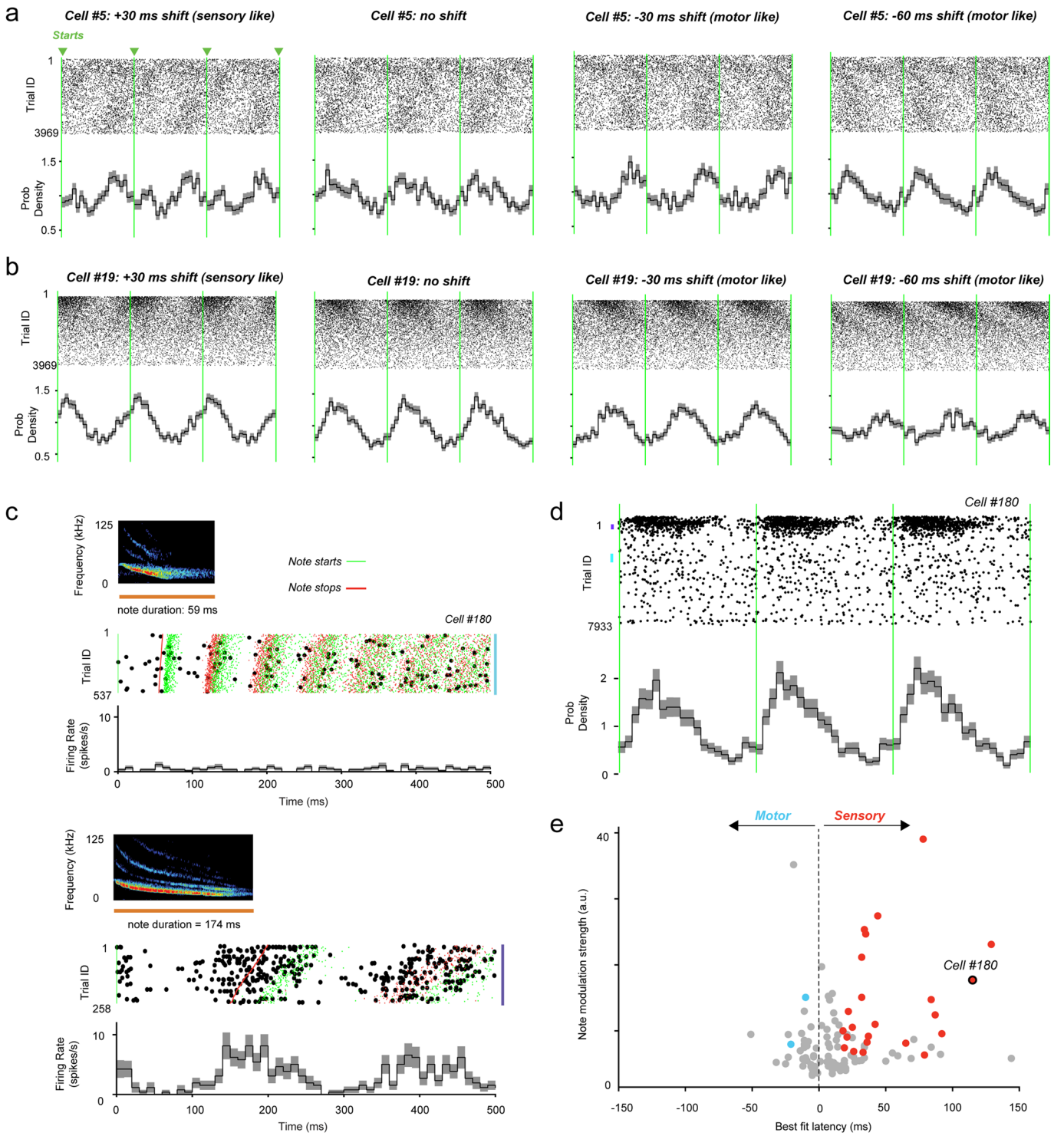
Peer review information *Nature Neuroscience* thanks Steffen Hage, Mehrdad Jazayeri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41593-023-01556-5



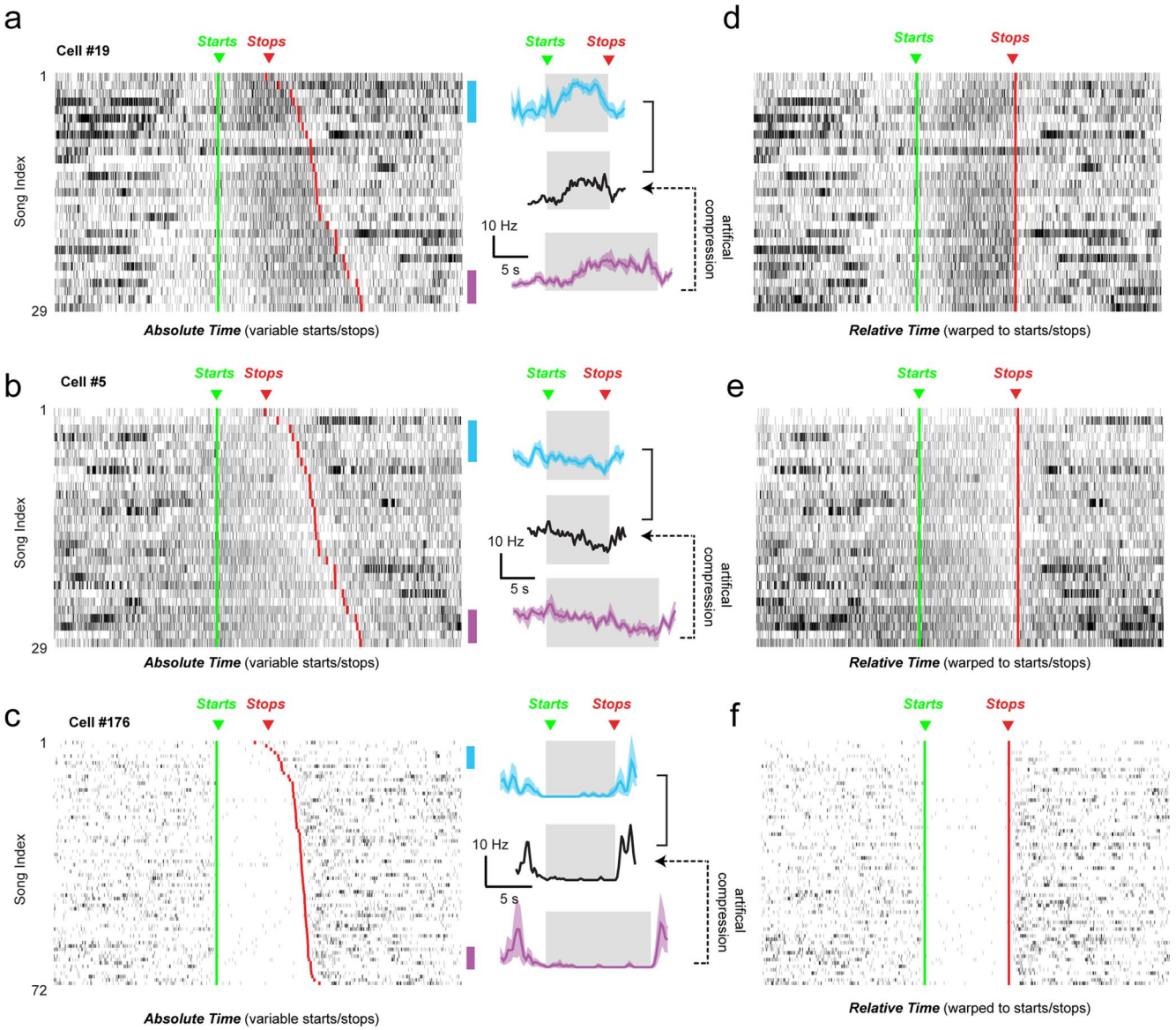
Extended Data Fig. 1 | Determination of significant note- and song-related responses. (**a**,**b**) Example neurons with (a, Cell #19) and without (b, Cell #1) significant note modulation. Spike rasters (top) and spike probability density plots (bottom) for example neurons whose activity profiles have been linearly warped to a common note duration (onsets indicated by green lines). Each row represents the warped spike raster of a neuron aligned to the beginning of a sequence of three notes; responses are sorted based on the original duration of the first note produced in this sequence from longest (top) to shortest (bottom). 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 (***: $p < 0.01$ two-sided test, 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).

https://doi.org/10.1038/s41593-023-01556-5



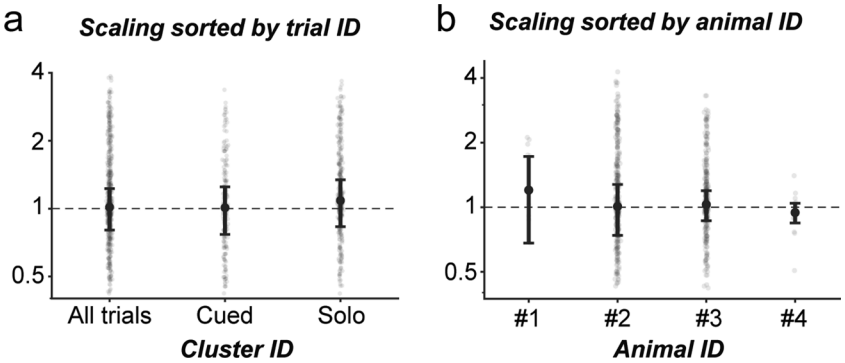
Extended Data Fig. 2 | Further characterization of note-related responses. (**a** and **b**) Spike times of two example neurons – Cell #5 (a) and Cell #19 (b) - linearly warped to a common note duration (onsets indicated by dashed lines). Each row represents the warped spike raster of a neuron aligned to the beginning of a sequence of three notes; responses are sorted based on the original duration of the first note produced in this sequence from longest (top) to shortest (bottom). 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.

https://doi.org/10.1038/s41593-023-01556-5



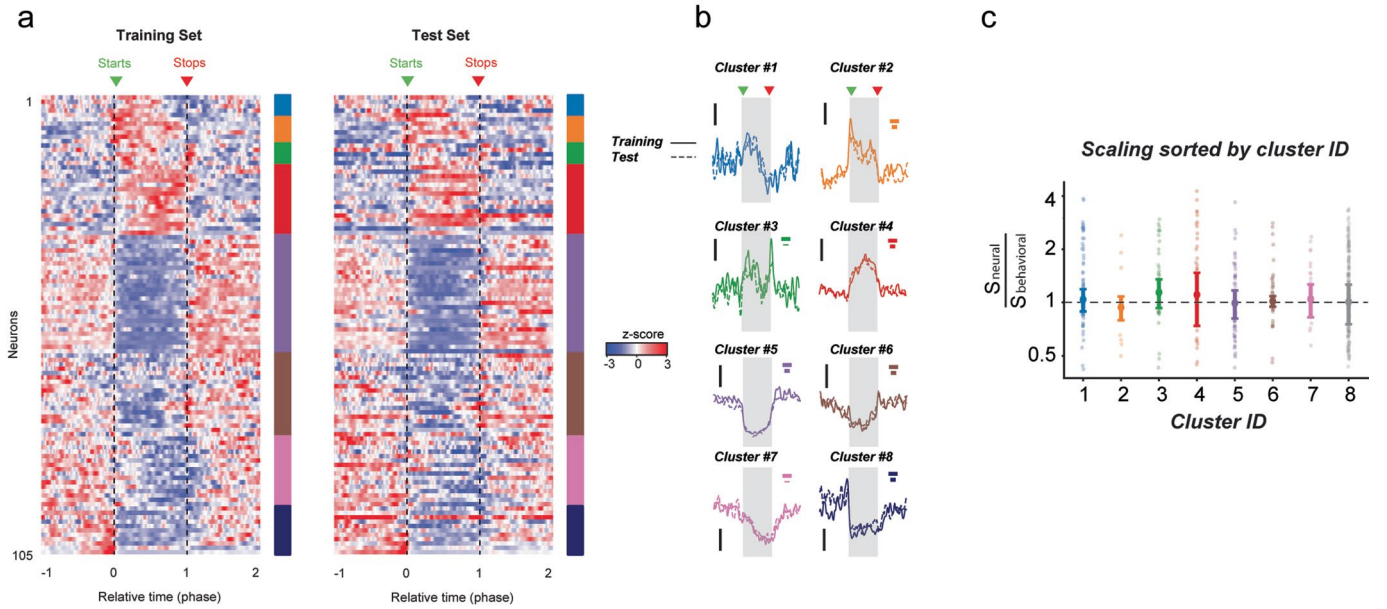
Extended Data Fig. 3 | Song-modulated neurons. (**a-c**) Spiking raster plots for three example neurons – Cell #19 (a), Cell #5 (b), and Cell #176 (c) – across all trials. At right, a peri-song time histogram (PSTH) for song blocks representing the shortest and longest songs in the session (indicated 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.

https://doi.org/10.1038/s41593-023-01556-5



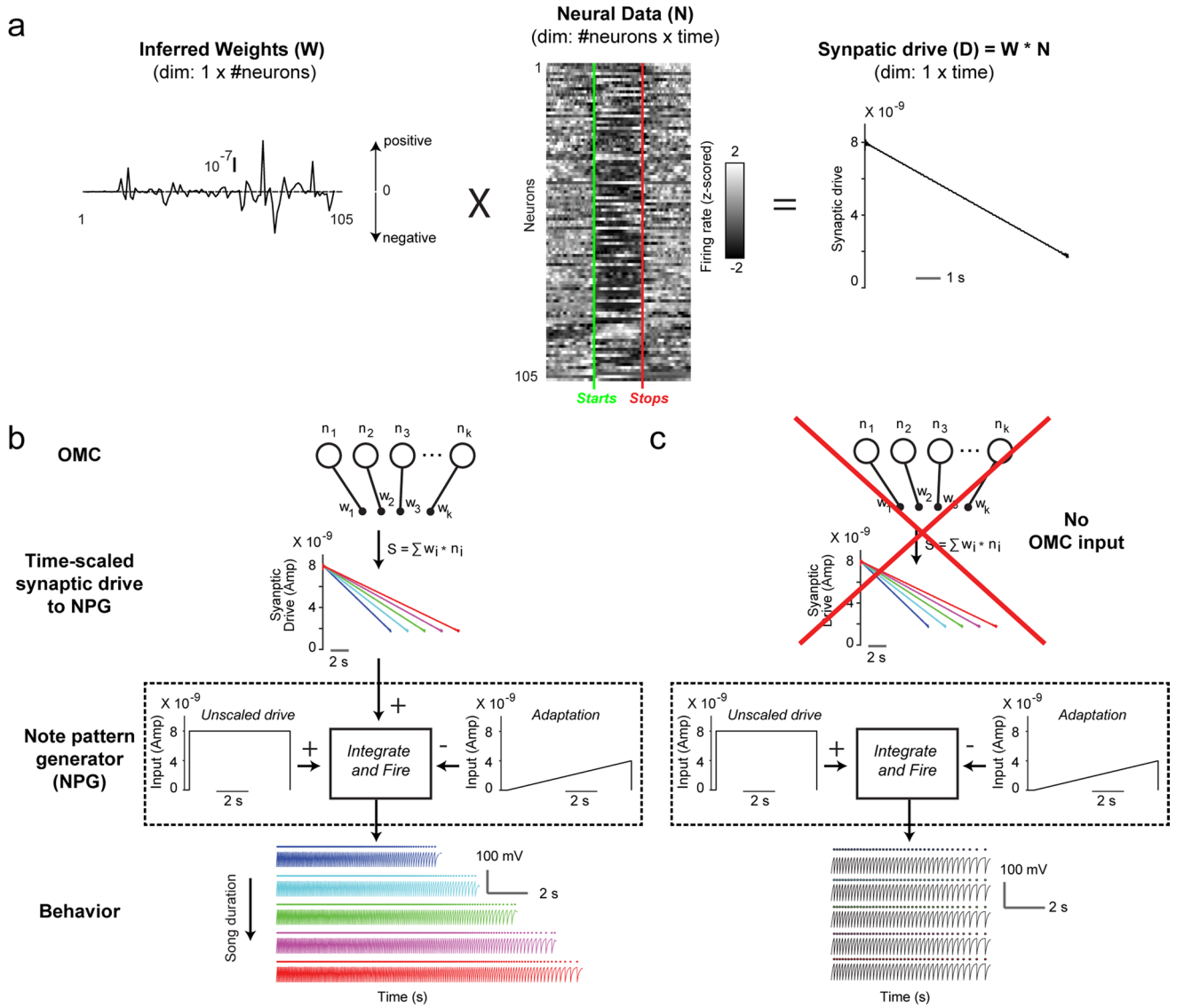
Extended Data Fig. 4 | Quantifying neural scaling as a function of behavior across categories. (**a,b**) The ratio of the neural scaling factor (S_{neural}) to the behavioral scaling factor ($S_{behavioral}$) with neurons grouped across different categories, namely reactive versus spontaneous singing (a) and animal ID (b). Each dot represents a comparison of similarly timed trial blocks (n = 4 - 21) for an individual neuron; quantifications denote median ± MAD. For the analysis shown in (b), Mouse #1 was limited in its total number of song trials per session with our original stringent criterion for significance threshold (p = 0.01), which prevented us from testing our hypothesis. We therefore relaxed this threshold across all animals to p = 0.05, which enabled a direct comparison of scaling factor. In all cases for both (a) and (b), the median neural:behavioral scaling ratio overlapped with 1, which denotes perfect co-variance between the duration of the song and the underlying OMC neural dynamics. See Methods and Fig. 3h for further information concerning how these parameters were calculated. Two-sided tests were used unless specified otherwise.

https://doi.org/10.1038/s41593-023-01556-5



Extended Data Fig. 5 | Cross-validation of hierarchical clustering. (**a**) Shown are the results of hierarchical clustering performed on the training (*left*) and test (*right*) set of trials sorted with respect to cluster affiliation (*left*). (**b**) Cross-validated firing rate profiles of the eight clusters evaluated on the training (solid) and the test (dashed) data set. (**c**) The ratio of the neural scaling factor (S_{neural}) to the behavioral scaling factor ($S_{behavioral}$) with neurons grouped across different categories. Each dot represents a comparison of similarly timed trial blocks (n = 4–21) for an individual neuron; quantifications denote median ± MAD. In all cases, the median neural: behavioral scaling ratio overlapped with 1, which denotes perfect co-variance between the duration of the song and the underlying OMC neural dynamics. See Methods and Fig. 3h for further information concerning how these parameters were calculated.

https://doi.org/10.1038/s41593-023-01556-5



Extended Data Fig. 6 | 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.

nature portfolio

Corresponding author(s): Michael Long and Arkarup Banerjee

Last updated by author(s): Dec 4 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Avisoft Recorder H116 for audio recordings and INTAN recording system for acquiring electrophysiology data.
Data analysis	Analysis code in MATLAB (R2016b) and Python (3.8.10); spike sorting carried out using Kilosort 1 and visualized with Phy

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

This study did not generate new unique reagents. Analysis code and data are available at the following GitHub repository: https://github.com/ccffccffcc/NNSingingMouse

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We have recorded from 375 neurons in the motor cortex of four singing mice. All analyses were performed on this data set. Sample sizes are constrained by the amount of spontaneous singing in a single session. Using cross-validation, however, we have demonstrated sufficient statistical power to support our claims.
Data exclusions	Data was not excluded unless specifically mentioned in the text and methods sections. For example, neurons with firing rates less than 1 spike/s were excluded to ensure statistical rigor. These are mentioned prominently in the manuscript, whenever applicable.
Replication	Main results reported in the manuscript were replicated by two authors (AB and FC) with analyses codes written independently using two different softwares (MATLAB and python).
Randomization	We did not have an 'experimental' or 'control' condition. Instead, we performed a descriptive study in which we simply measured cortical activity occurring during song production. Randomization is not an appropriate feature of such studies.
Blinding	We did not have an 'experimental' or 'control' condition. Instead, we performed a descriptive study in which we simply measured cortical activity occurring during song production. Blinding is not an appropriate feature of such studies.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	These experiments are done on adult male Alston's singing mice (Scotinomys teguina) bred and maintained in a temperature and humidity controlled facility at NYU. Large communal cages, running wheels, and dietary enrichment are provided.
Wild animals	All animals used in this study were laboratory-reared offspring of wild-captured Scotinomys teguina from La Carpintera and San Gerardo de Dota, Costa Rica. No wild-caught mice were used as part of this study.
Reporting on sex	Experiments were performed on male singing mice. They have been previously reported to sing and respond to songs of other males (Okobi, Banerjee et al, Science, 2019).
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	Institutional Animal Care and Use Committee of NYU Langone medical center

Note that full information on the approval of the study protocol must also be provided in the manuscript.