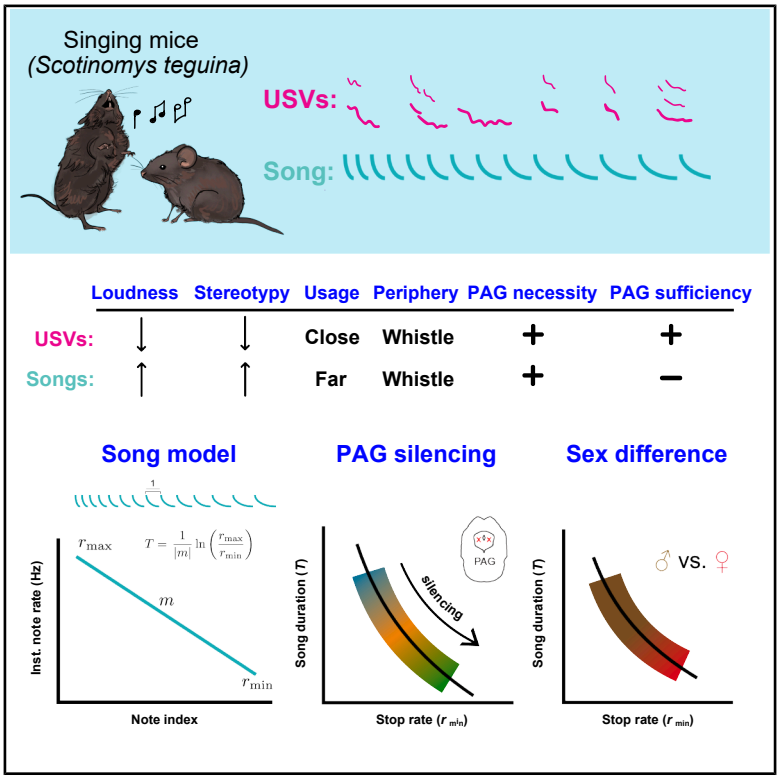
Article

Current Biology

Vocal repertoire expansion in singing mice by co-opting a conserved midbrain circuit node

Graphical abstract



Authors

Xiaoyue Mike Zheng, Clifford E. Harpole, Martin B. Davis, Arkarup Banerjee

Correspondence

abanerjee@cshl.edu

In brief

Zheng and Harpole et al. show that singing mice produce two major vocal modes: ancestral USVs and novel songs that follow a linearly decelerating rhythm. Although acoustically distinct, songs and USVs share conserved phonatory mechanisms. Synaptic silencing of clPAG progressively truncates song, recapitulating natural behavioral variation, including sexual dimorphism.

Highlights

- Singing mice produce two context-dependent vocal modes: soft USVs and loud songs
- Although acoustically distinct, songs and USVs share conserved phonatory mechanisms
- Song progression follows a decelerating rhythm governed by a linear equation
- Synaptic silencing of clPAG progressively truncates song, mimicking sex differences

Zheng et al., 2025, Current Biology *35*, 5762–5778
December 1, 2025
https://doi.org/10.1016/j.cub.2025.10.036

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

Vocal repertoire expansion in singing mice by co-opting a conserved midbrain circuit node

Xiaoyue Mike Zheng,[1,2,3] Clifford E. Harpole,[1,3] Martin B. Davis,[1] and Arkarup Banerjee[1,2,4,*]
[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Cold Spring Harbor Laboratory School of Biological Sciences, Cold Spring Harbor, NY 11724, USA
[3]These authors contributed equally
[4]Lead contact
*Correspondence: abanerjee@cshl.edu
https://doi.org/10.1016/j.cub.2025.10.036

SUMMARY

How neural circuits generate diverse behaviors is a fundamental question in neuroscience. Distinct behavioral outputs may arise from either dedicated motor circuits or shared circuits operating in different functional states. Although multifunctional circuits offer an efficient solution for behavioral flexibility and may drive rapid evolutionary adaptations, their neural mechanisms remain poorly understood, especially in mammals. Here, we leverage the rich vocal repertoire of the singing mouse (*Scotinomys teguina*) to investigate the organizational logic of multifunctional motor circuits. We developed a behavioral assay (partial acoustic isolation reveals identity [PAIRId]) that enables precise attribution of vocalizations during social interactions. This paradigm revealed two distinct vocal modes: soft, variable, ultrasonic vocalizations (USVs) ancestral to rodents, used for short-range communication, and loud, rhythmic, human-audible songs unique to the singing mouse lineage, used for long-range communication. Despite their substantial acoustic and contextual differences, we found that USVs and songs do not arise from parallel pathways. Instead, they share the same sound production mechanism, phonatory-respiratory coupling, and vocal gating from the midbrain caudolateral periaqueductal gray (clPAG). To understand the mechanism governing song production, we combined mathematical modeling of song rhythm with synaptic silencing of clPAG, which progressively reduced song amplitude and duration. We demonstrate that song duration decreases via a single parameter controlling its termination. Notably, this mechanism also accounts for sexual dimorphism in songs, identifying clPAG as a key locus for driving natural behavioral variability. Our findings reveal how parametric tuning of a central circuit node produces distinct vocal modes, providing a mechanistic basis for rapid behavioral evolution in mammals.

INTRODUCTION

Behavioral flexibility—the ability of animals to produce different actions depending upon context—is essential for survival. How the brain generates this flexibility remains a fundamental question in neuroscience. Distinct behavioral modes can be produced by dedicated motor circuits[1–3] or instantiated through shared circuits capable of operating at different functional regimes.[4–7] Determining such organizational logic is critical for understanding not only the function but also the evolution of neural circuits. For instance, repurposing of shared, multifunctional circuits may drive rapid behavioral evolution, but the specific neural mechanisms underlying this process remain obscure, especially in mammalian brains.

Vocal communication, which requires coordinated respiratory, laryngeal, and orofacial motor control in different behavioral contexts, provides an ideal framework to investigate the rules of multifunctional circuit function.[8–12] Singing mice (*Scotinomys teguina*), known for their distinctive songs used for turn-taking, are an attractive model system to explore this issue[13–17] (Video S1). We have previously shown that this behavior is dependent upon motor cortical function.[18,19] However, the downstream subcortical mechanisms that drive vocalizations remain unexplored. The periaqueductal gray (PAG), which projects to brainstem phonatory and respiratory networks, has been established as a necessary node for vocal production across vertebrate evolution.[11,20,21] As we will show, the rich vocal behaviors of the singing mice and their control by the midbrain PAG provide an ideal testbed for addressing whether distinct vocal behaviors are generated by separate dedicated pathways or parametric modulation of a common circuit.

In this study, we address three specific questions: do singing mice employ distinct vocal modes across different social contexts? If so, what are the acoustic characteristics and usage patterns of these modes? Do these distinct vocal behaviors emerge from separate neural pathways or from shared circuits operating in different regimes? Although songs have been well studied, it remains unclear whether they represent their only major form of vocal communication. Additional vocalizations have been observed during social encounters,[15] yet their acoustic features, prevalence, and neural mechanisms remain unknown. By combining a novel behavioral paradigm with acoustic analysis,

CellPress
OPEN ACCESS

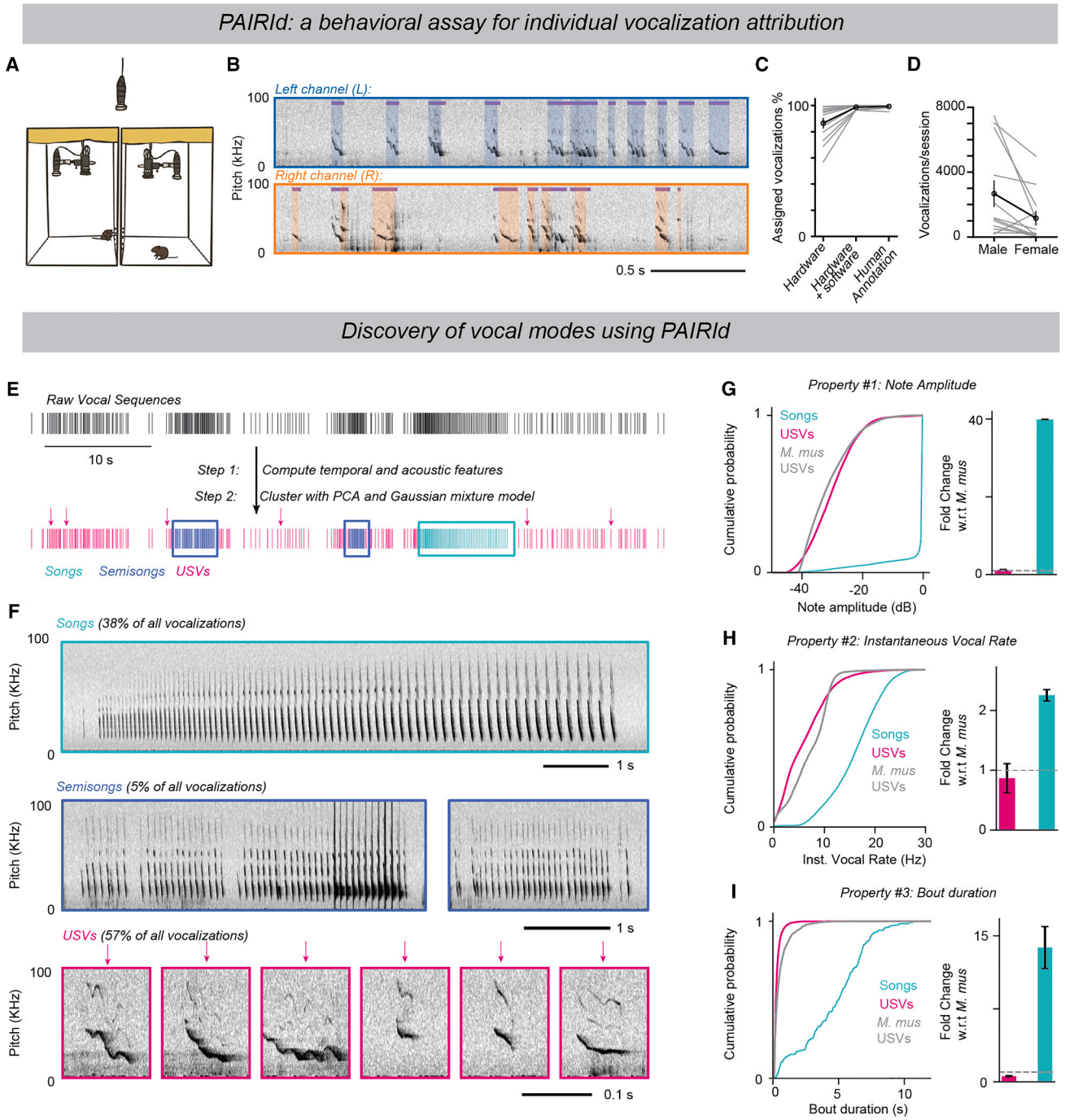


Figure 1. Discovery of distinct vocal modes of singing mice using PAIRId

(A) Schematic of the PAIRId behavioral paradigm, where two mice housed in separate acoustically dampened enclosures—each equipped with its own microphone and camera—interact across a perforated plane.

(B) Example audio data from two singing mice interacting in PAIRId, where purple lines indicate vocalizations detected and blue and orange boxes denote vocalizations attributed to individual mice.

(C) Proportion of vocalizations attributed to individual mice using only hardware acoustic isolation (86.6% ± 4.0%) and the full analysis pipeline (98.8% ± 0.5%), compared with human annotations (99.3% ± 0.4%, n = 12 sessions of 5 h each).

(D) Number of vocalizations in PAIRId by the male (2,673.3 ± 818.0, n = 12 sessions of 5 h each, 6 males) and the female (1,162.3 ± 451.0, n = 12 sessions of 5 h each, 4 females).

(E) Schematic for identifying vocal modes from temporal sequence of vocalizations. Top: raster plot illustrating vocalizations of an animal during an example 60-s period, with each line marking the onset of a vocalization. Bottom: the same raster plot, with lines color-coded by the vocal modes (songs, cyan; semisongs, blue; USVs, magenta) identified by combining temporal stereotypy and loudness features.

(legend continued on next page)

mathematical modeling, peripheral sound production experiments, and targeted manipulations of the caudolateral PAG (clPAG), we report that singing mice use shared neural control of phonation operating under different amplitude modulation (AM) and frequency modulation (FM) regimes to generate categorically distinct vocal behaviors.

RESULTS

Discovery of distinct vocal modes using PAIRId

An accurate and quantitative description of behavior is essential for investigating neural circuit mechanisms. We begin by measuring the vocal repertoire of individual singing mice during close-range social interactions. One major challenge in bioacoustics has been the assignment of quiet vocalizations produced in close proximity to one other. This has led to the development of methods such as wearable microphones or microphone arrays, with subsequent sound localization using neural networks.[22–26] Complementary approaches have created physical separations between rodents to facilitate the attribution of vocalizations to individual animals during social exchanges.[27–30] Using this general strategy, we developed a behavioral paradigm—PAIRId (partial acoustic isolation reveals identity)—in which two animals placed in separate acoustically dampened enclosures, equipped with private microphones and cameras, can interact across a perforated plane (Figure 1A; Video S2).

To identify vocalizations of individual animals, we leverage both hardware-level acoustic isolation and software-level signal processing (Figure 1B). Briefly, vocalizations detected exclusively in one channel due to acoustic dampening are assigned directly to that animal (86.6% ± 4.0%). For the remaining detections with temporal overlap across channels, we compare spectrotemporal properties to distinguish coincident vocalizations and acoustic bleed-through (Figures 1B and S1A; STAR Methods). This approach allows for the attribution of almost all vocalizations with high accuracy, in experiments with two intact mice (assignment rate 98.8% ± 0.5%, 99.3% ± 0.4% with human annotation, 12 male-female pairs; F1 accuracy 0.87 ± 0.02; Figures 1C and S1B–S1D) and experiments where one mouse is muted (assignment rate 86.0% ± 4.7%, 96.9% ± 1.8% with human annotation, F1 accuracy 0.89 ± 0.03; Figures S1E–S1G). This performance is comparable with other published methods that report attribution rates of more than 90%.[23,24,27] Together, PAIRId allows us to quantify the vocalizations of individual singing mice during social interactions, providing the necessary foundation for exploring underlying neural mechanisms.

Using PAIRId, we found that the singing mouse vocal repertoire is much richer than previously anticipated. Both male and female singing mice vocalize robustly in these close-range social encounters (Figure 1D). This paradigm further revealed the complexity of temporally sequenced vocal streams produced by individual singing mice (Figure 1E). To quantitatively describe their vocal repertoire, we analyzed the temporal rhythm of notes within vocal bouts (Figure S2A; STAR Methods). We identified rhythmic patterns—where mice produce notes at steady or smoothly varying rates—by first detecting short, contiguous sequences with high temporal stereotypy. These vocal segments were then hierarchically merged into longer sequences, which we clustered using a Gaussian mixture model to reveal distinct rhythms (Figures S2B–S2E). In addition to these rhythmic vocal sequences, we also observed substantial differences in loudness of individual notes (Figure S2G bottom). Combining temporal characteristics and loudness information, we define two major vocal modes—songs (38% of vocalizations) and ultrasonic vocalizations (USVs) (57% of vocalizations) (Figures 1E, 1F, and S2F–S2H). We also observed a rarer vocal mode (semisongs, 5% of vocalizations) with acoustic properties intermediate between USVs and songs (Figures S2I and S3A–S3D).

To place the two vocal modes of singing mouse in a phylogenetic context, we compare them with vocalizations of other rodents. Laboratory mouse (*Mus musculus*) USVs serve as an ideal comparative reference due to their thorough characterization in the literature.[31–33] We found that singing mouse USVs have overlapping acoustic properties with laboratory mouse USVs (Figures 1G–1I, gray)—including lower amplitude, lower temporal stereotypy, shorter bouts (Figures 1G–1I, magenta), and variable note shapes (Figure 1F, magenta). In contrast, songs are much louder, have faster tempos, and contain many more stereotyped notes over much longer durations (Figures 1G–1I, cyan). We conclude that songs have acoustic characteristics in both amplitude and temporal domains that are categorically distinct from the USVs of both species.

To test whether the acoustic distinction is reflected in context-dependent usage, we examined vocal usage during close-range social interactions in PAIRId—when mice were within a few body lengths of one another—and compared it with usage when mice were housed alone without social cues (Figures 2A, 2B, and 2I). In close-range encounters, singing mice produce predominantly USVs, which are much softer than songs (Figures 2B–2E). Both males and females make USVs in these interactions (Figure S3A), and they are produced mainly near the interaction plane in temporally correlated bouts (Figures 2F–2H, S3D, and S3E). By contrast, when mice are housed alone, USVs are rarely

(F) Example spectrograms for each of the three identified vocal modes—songs, semisongs, and USVs—highlighted in (E). To account for the large loudness difference among vocalizations, the song spectrogram was computed from low-gain audio recording, whereas the semisong and USV spectrograms were computed from high-gain audio recording.

(G) Loudness of notes in different vocal modes. Left: cumulative distribution of note loudness for each vocal mode (singing mouse USV, $n = 26{,}246$ notes from 10 mice, −29.78 ± 0.05 dB; singing mouse song, 17,681 notes from 10 mice, −2.27 ± 0.05 dB; laboratory mouse USV, $n = 7{,}295$ notes from 3 mice, −30.82 ± 0.09 dB). Right: fold change of median note loudness per mouse relative to the median of laboratory mouse USVs (singing mouse USV, 1.1 ± 0.2, $p > 0.05$; singing mouse song, 39.9 ± 0.0, $p = 0.007$).

(H) Same as (G) but for instantaneous note rate (singing mouse USV, 6.26 ± 0.03 Hz, fold change 0.9 ± 0.2, $p > 0.05$; singing mouse song, 16.16 ± 0.04 Hz, fold change 2.3 ± 0.1, $p = 0.007$; laboratory mouse USV, 7.24 ± 0.05 Hz).

(I) Same as (G) but for vocal bout duration (singing mouse USV, 0.31 ± 0.00 s, fold change 0.6 ± 0.1, $p = 0.007$; singing mouse song, 4.64 ± 0.15 s, fold change 13.8 ± 2.2, $p = 0.007$; laboratory mouse USV, 0.56 ± 0.02 s). Values reported are mean ± SEM, and hypothesis testing was performed using Mann-Whitney U test. See also Figures S1–S3 and S7 and Videos S1 and S2.

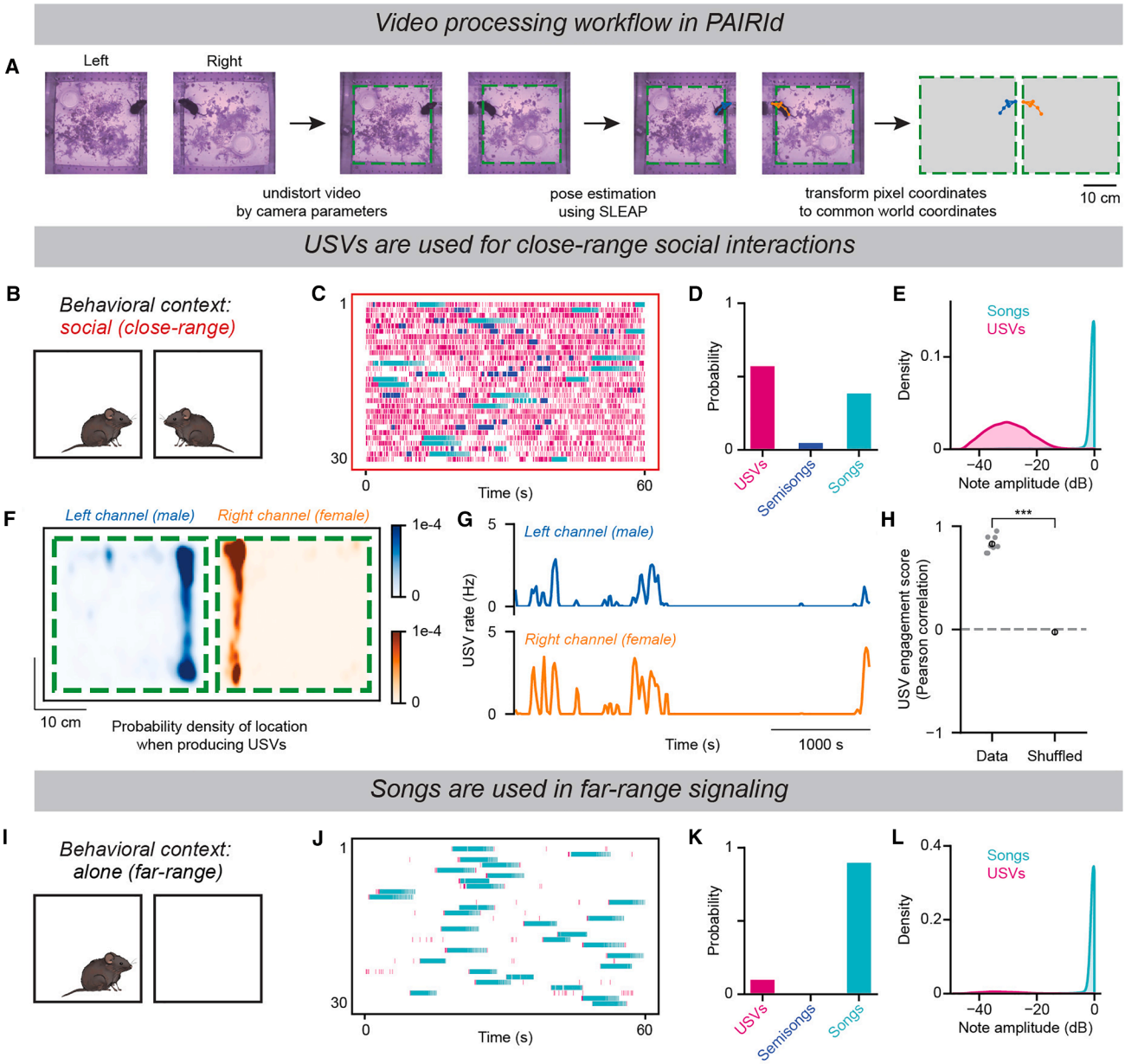


Figure 2. Context-dependent usage of USVs and songs

(A) The PAIRId video processing pipeline aligns two mice in common world coordinates. In steps, the distortion produced by the lens is removed relative to the enclosure floor, followed by pose estimation using SLEAP. The output positions of the 6-node skeleton are then transformed from pixel values to common world coordinates.

(B) Schematic illustrating that (C)–(H) refer to data collected during close-range social interactions in PAIRId.

(C) Raster plots of 30 segments, with the highest vocal activity during close-range social interactions in PAIRId. Cyan represents songs, magenta USVs, and blue semisongs.

(D) During close-range social interactions in PAIRId, proportion of notes classified as USVs (57.0%, 26,246 notes), semisong notes (4.6%, 2,100 notes), and song notes (38.4%, 17,681 notes) across 12 sessions of 5 h each, 6 males and 4 females.

(E) Distribution of loudness of vocal modes when mice interact socially in PAIRId (cyan, n = 17,681 song notes; magenta, n = 26,246 USVs).

(F) Density of mouse locations (centroid) during USVs (n = 26,246 notes across 12 sessions of 5 h each, 6 males and 4 females).

(G) Rates of USV production of two interacting mice (top, male; bottom, female) in an example hour with high vocal engagement. Epochs of USV production were correlated between the two animals (Pearson correlation = 0.74).

(H) Pearson correlation of USV production across 8 h of engaged social interaction between males and females, including (G) (mean 0.83 ± 0.03). Correlations were computed between the two animals' vocalization rates within the same hour and compared with control correlations obtained by pairing rates from different hours (−0.03 ± 0.02, p = 4.5e−10). Gray dots are individual hours and black circles the mean ± SEM.

(I) Schematic illustrating that (J)–(L) refer to data collected in a second behavioral context when the mouse is alone in a similar enclosure to that in PAIRId.

(J) Raster plots of 30 segments, with the highest vocal activity when the mouse is alone.

(legend continued on next page)

produced (Figures 2I–2L). Together, these observations indicate that proximity to a conspecific strongly promotes USV production, consistent with a role as short-range social signals, also observed in other species.[34] Songs, by comparison, are louder and capable of propagating over much greater distances, making them well-suited to long-range communication. Indeed, when housed alone—a proxy for far or effectively infinite distance from conspecifics—singing mice produced songs almost exclusively (Figures 2I–2L and S3B). This finding, along with prior evidence that mice engage in turn-taking with songs to interact with remote partners,[17,19] supports the idea that songs function primarily as a far-range signal. Notably, singing mice also produce songs, and the acoustically intermediate semisongs, during close-range interactions (Figures 2B–2E and S3A–S3C) and can switch rapidly between different vocal modes in succession (Figures 2C, S3F, and S3G), suggesting their functions may be more nuanced. In sum, by developing the PAIRId paradigm, we found that singing mice produce two major, categorically distinct vocal modes—songs and USVs—that differ not only in their acoustic structure but also in the social contexts in which they are used.

A linear model of the song rhythm

Singing mouse songs are strikingly divergent from the USVs in their temporal stereotypy. To capture this stereotyped rhythmic structure, we sought to develop a mathematical model describing the temporal progression of individual notes using a compact set of parameters. Songs are composed of a series of increasingly longer notes produced over several seconds (Figure 3A). We serendipitously observed that, during song progression, instantaneous note rate $r(i)$ —defined as the inverse of inter-note interval—varies linearly with note index i. A song always begins with notes emitted at a high rate, which steadily decreases throughout the song, finally ending at a much lower rate (Figure 3B). A linear model with just three parameters (start rate r_{max}, slope m, and stop rate r_{min}) adequately describes the instantaneous note rate throughout the song:

$$r(i) = m(i-1) + r_{max}$$

for $i = 1,2,\ldots,N-1$, where N is the number of notes in a song (Figure 3B; STAR Methods). This simple generative model reliably captures the temporal progression of songs across our entire dataset (Figure 3C). Moreover, song duration T estimated by integrating the instantaneous rate in the time domain matches the measured song durations accurately across all animals (Figure 3D):

$$T = \int_1^N \frac{1}{r(i)}di = \frac{1}{|m|}\ln\left(\frac{r_{min}}{r_{max}}\right).$$

Thus, this linear model captures the stereotyped rhythm of singing mouse songs and explains the observed variability through just three patterning parameters.

Peripheral mechanisms for distinct vocal modes

As we have shown above, singing mouse songs are rhythmic and loud, but their USVs are much softer with less temporal stereotypy. What might be the neural mechanisms driving these two distinct vocal modes? Given their categorical differences, one possibility is that these two vocal modes are largely driven by parallel motor pathways. Alternatively, the presence of the intermediate semisongs and the smooth transitions between songs and USVs might imply that the two distinct vocal modes share a common motor pathway operating in two different regimes (Figures 2C, S2I, S3F, and S3G). Building on the current understanding of the mammalian vocal motor hierarchy,[11,12] we tested these alternative models at three different levels: peripheral mode of sound production, phonation-respiration coupling, and mechanisms of vocal gating by the PAG in the midbrain.

We first determined whether the biophysical mechanism for sound production differs between songs and USVs (Figures 4A–4G). Rodents produce sounds using their larynx in two ways: by vocal fold vibrations (e.g., squeaks in laboratory mouse) or by generating aerodynamic whistles within the larynx (e.g., USVs in laboratory mouse).[35] Although the precise details of the laryngeal and aerodynamic mechanisms remain under investigation,[36–38] the two broad categories can be distinguished by simply changing the density of the surrounding air.[35,39,40] If sounds are produced by a whistle mechanism, the fundamental frequency (F0) will depend on air density and increase in a less dense medium. In contrast, if sounds are produced by a vibrational mechanism, F0 is expected to remain unchanged. To test these alternatives, we replaced air in the behavioral enclosure of two freely interacting singing mice with a helium-oxygen mixture (heliox; 80% helium, 20% oxygen). Under heliox, F0 of both USVs and song notes increased significantly—and by similar ratios—consistent with an aerodynamic whistle mechanism (Figures 4B–4E). Crucially, we verified that singing mice were also able to vocalize via a vibrational mechanism, as the F0 of their squeaks recorded concurrently remained stable (Figures 4F and 4G). These findings demonstrate that both USVs and songs are produced by a similar whistle mechanism at the vocal periphery.

Moving up the vocal motor hierarchy, we next examined the coupling between phonation and respiration during songs and USVs (Figures 4H–4O). We implanted a thermistor in the nasal cavity to continuously monitor temperature changes as a proxy for respiration while the male mouse interacted with a mute female stimulus animal (Figures 4I–4L). Consistent with rodent respiratory patterns, singing mice exhibit a bimodal distribution of breathing rates during non-vocal periods (Figure 4M, black), likely reflecting slower baseline respiration and faster rates associated with sniffing and whisking.[41,42] We found that each song note was associated with a single respiration cycle, with phonations occurring exclusively during exhalation (Figures 4K and 4N), consistent with previous findings using a different method.[19] Therefore, the temporal patterning of song notes arises from the

(K) While alone, proportion of notes classified as USVs (9.8%, 2,259 notes), semisong notes (0.5%, 120 notes), and song notes (89.7%, 20,649 notes), across 12 sessions of 5 h each, 6 males and 4 females.
(L) Distribution of loudness of vocal modes when mice are alone (cyan, n = 20,649 song notes; magenta, n = 2,259 USVs). Values reported are mean ± SEM and hypothesis testing was performed using Mann-Whitney U test.
See also Figures S3 and S7 and Videos S1 and S2.

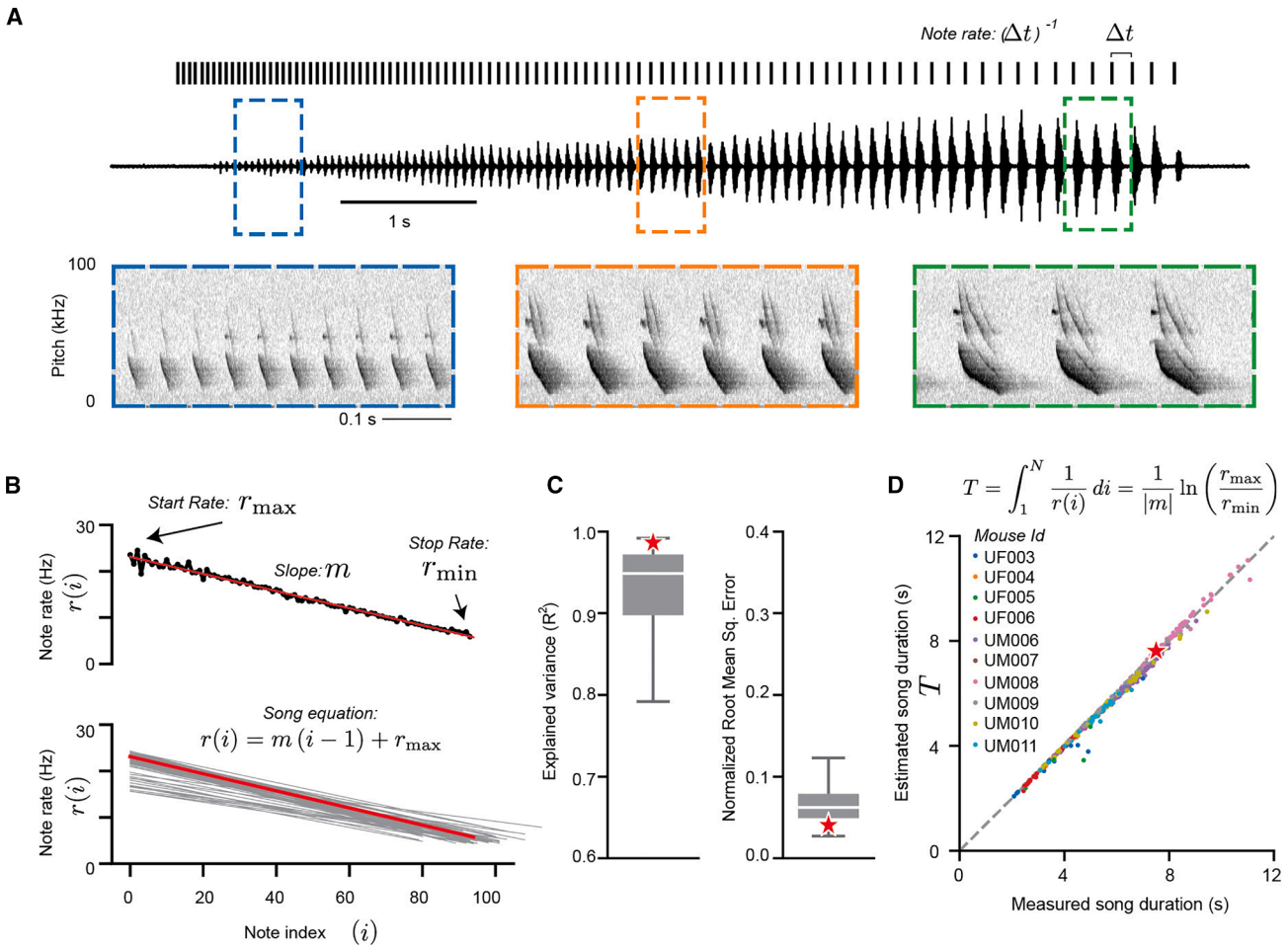


Figure 3. A linear model of the song rhythm

(A) An example singing mouse song. Top: raster plot of the song, with each line marking the onset of a song note. The annotation indicates that the instantaneous note rate is calculated as the inverse of the interval between note onsets. Middle: audio waveform of the song. Bottom: spectrograms of the highlighted snippets of the song.

(B) Rhythm trajectories of example songs. Top: the trajectory of the example song from (A) is plotted with each note's instantaneous rate against its sequential index; the red line shows the linear model fit. Bottom: model fits of rhythm trajectories for all 60 songs produced by a single mouse, with the red line indicating the example song.

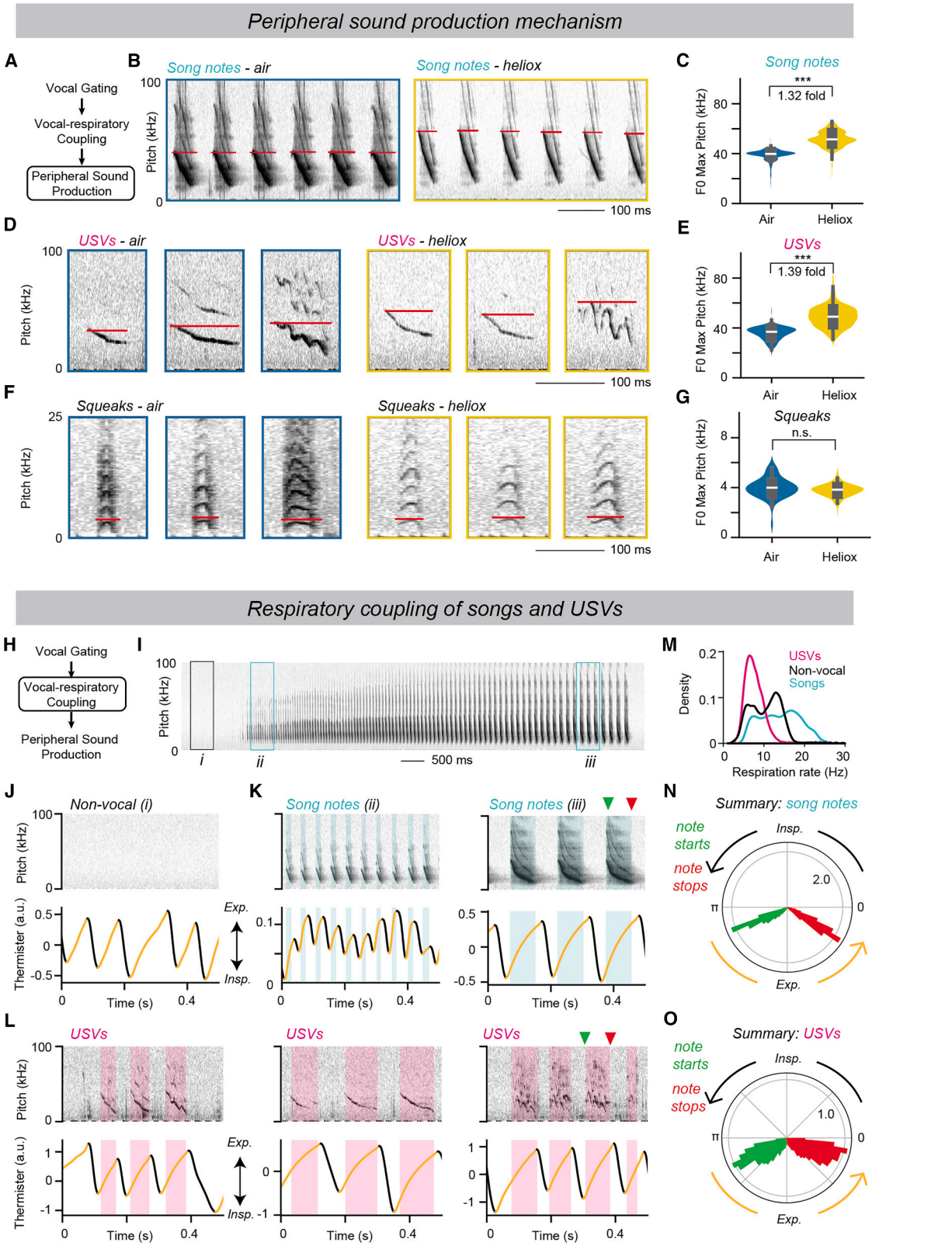
(C) Performance of the linear model fit (n = 456 songs from 10 mice across alone and social contexts, R^2 = 0.90 ± 0.01, normalized root mean squared error [NRMSE] = 0.07 ± 0.00). Star represents the example song from (A).

(D) Model-predicted vs. measured song duration (Pearson correlation coefficient = 0.9970; n = 456 songs from 10 mice). Values reported are mean ± SEM. See also Figure S7 and Video S1.

rhythmic structure of respiration itself. Similar to song notes, each USV, regardless of duration, is produced during exhalation with a similar phase relationship (Figures 4L and 4O). This is in line with previous literature demonstrating that laboratory mouse USVs are tightly coupled to underlying respiratory cycles and are produced exclusively during exhalations.[27,33,43] Although the respiration rates during USVs fall within the range observed during silent, non-vocal periods, respiratory rates needed for singing are substantially higher and can easily exceed 20 cycles per second (Figure 4M). Our results demonstrate that the two strikingly distinct vocal modes—songs and USVs—share peripheral sound production mechanism as well as vocal-respiratory coupling. This suggests that, in the singing mouse, the newly evolved song mode has co-opted the phonatory mechanism of USV production, while operating it at substantially higher amplitude and frequency regimes for song motor control.

Vocal gating by the caudolateral PAG

Given that USVs in singing mice and laboratory mice share similar acoustic properties, behavioral contexts, and peripheral sound production mechanisms, we hypothesized that their vocal gating mechanisms might also be conserved. The midbrain PAG has been established as a critical locus for innate behaviors, including vocalizations, across vertebrates.[21] Specifically, neurons in the caudolateral PAG (clPAG) project to hindbrain phonation circuits (e.g., the nucleus retroambiguus [RAm], the intermediate reticular oscillator [iRO], and the pre-Bötzinger complex [preBötC]) and are necessary and sufficient



(legend on next page)

for USV production in laboratory mice.[44–49] Therefore, we decided to examine the role of the clPAG region for vocal gating in the singing mouse.

We relied on gross morphological similarities to identify the homologous PAG subregion (Figure S4). Optogenetic activation of channelrhodopsin-2 (ChR2)-expressing CAMKII-positive clPAG neurons was sufficient to reliably elicit vocalizations in the singing mouse (Figures 5A and 5B; Video S3). Vocalizations began immediately after light onset and scaled with the duration of photostimulation (Figures 5B–5D, n = 4 male mice). Acoustic analyses revealed that these vocalizations resembled USVs in both amplitude and instantaneous vocal rates, and no optogenetically evoked vocalizations were song-like (Figure 5E). Therefore, clPAG stimulation is sufficient to evoke USVs, but not songs, in singing mice. Conversely, we expressed tetanus toxin light chain (TeLC) in CAMKII-positive neurons of the clPAG in male mice and monitored their vocalizations during interactions with intact females in PAIRId (Figure 5A). TeLC cleaves synaptobrevin, blocking neurotransmitter release without affecting neuronal excitability or viability.[50] Synaptic silencing of clPAG neurons with TeLC caused a severe reduction in USVs (Figures 5B and 5C, n = 5 male mice). We conclude that the clPAG region is both necessary and sufficient to generate species-typical USVs in the singing mouse. We also found that synaptic silencing of the clPAG eliminated songs; in effect, the animals were rendered mute (Figure 6C). This observation rules out a separate, parallel pathway for song production independent of the clPAG. Instead, it points to a shared vocal motor control circuit for both USVs and songs.

Based on similarities in acoustic properties, behavioral context, peripheral production mechanisms, and sufficiency of eliciting USVs from the clPAG, we posit that USVs in singing mice and laboratory mice are homologous behaviors. This is despite superficial differences in pitch (USVs in laboratory mice have a much higher pitch), which presumably reflects differences in airway and laryngeal morphology, such as the size of the ventral pouch.[51–53] In contrast, songs represent a drastically divergent vocal behavior unique to the singing mice lineage, characterized by two key innovations in motor control: significantly higher amplitude and highly stereotyped temporal patterning over long bouts.

Parametric control of song progression by the clPAG

How does the neural circuitry governing the elaborate songs differ from that of USVs? So far, we have shown that the mechanisms for phonatory control is shared between songs and USVs. However, given the drastic differences in loudness and rhythm during songs (compared with USVs), we hypothesized that song production involves driving the same circuit node (clPAG) in a higher amplitude and frequency regime. This would predict that progressive silencing of the clPAG would lead to graded disruptions in song quality. To test this possibility, we summarized data across all TeLC-injected animals by binning each animal's post-injection period into two equal halves: "post-early" (containing the first half of all analyzed vocalizations) and "post-late" (containing the second half). Consistent with our hypothesis, we observed a progressive shift in rhythmicity away from songs toward semisongs and USVs (Figures 6D and S5A), resulting in an increasing proportion of those shorter vocalizations (Figure 6E). At the same time, the loudness of both songs and USVs declined, with the reduction especially pronounced for songs (Figure 6F). Therefore, synaptic silencing of clPAG was associated with a progressive dialing down of both loudness and rhythmicity—the two acoustic features that distinguish the novel song mode compared with the ancestral USVs.

To better understand how clPAG influences song production, we used our analytical framework to examine the changes in the temporal progression of songs following TeLC-injection. Compared with pre-injection control songs, we observed a progressive reduction in song duration in both example songs (Figures 7A and 7B) and across animals (Figure 7C). To analyze these intermediates, we used our linear model of song rhythm to decompose each song into its three underlying patterning parameters. We found that clPAG silencing led to a large increase in the stop rate (r_{min}), while the other two parameters were only modestly affected (Figures 7D and S5B). In fact, the relationship between song durations, T, and stop rates, r_{min}, for all songs

Figure 4. Shared peripheral vocal production mechanisms between songs and USVs

(A) Schematic of the vocal motor hierarchy, highlighting the level addressed by the laryngeal phonation mechanism experiment.

(B) Example spectrograms of song notes produced in air and in helium-oxygen mixture (heliox), with the red lines illustrating the maximum of the fundamental frequency (F0).

(C) The maximum F0 of song notes is significantly higher in heliox (51.7 ± 0.1 kHz, n = 1,648 notes) than in air (39.0 ± 0.1 kHz, n = 2,262 notes; p = 0.0), consistent with a whistle mechanism.

(D and E) Same as (B) and (C), but for USVs instead of song notes, showing an increase in F0 in heliox (35.8 ± 0.2 kHz, n = 400 notes in air; 49.1 ± 0.4 kHz, n = 400 notes in heliox; p = 1.38e−94), also consistent with a whistle mechanism.

(F and G) Same as (B) and (C), but for squeaks instead of song notes, showing that F0 in heliox is not significantly higher than in air (4.0 ± 0.1 kHz, n = 75 notes in air; 3.8 ± 0.0 kHz, n = 208 notes in heliox; $p > 0.05$), consistent with a vibration mechanism.

(H) Schematic of the vocal motor hierarchy, highlighting the level addressed by the vocal-respiratory coordination experiment.

(I) Spectrogram of a song preceded by a non-vocal period, recorded with simultaneous thermistor-based respiration monitoring.

(J) Example respiration dynamics during silence (box *i* in I), with black indicating inhalation and orange indicating exhalation.

(K) Same as (J) but during song production (boxes *ii* and *iii* in I). Song notes, shown in cyan, occur during exhalation.

(L) Same as (J) but for when the mouse makes USVs. USVs, shown in magenta, also occur during the exhalation phase.

(M) Distribution of respiration rate when the mouse is non-vocal (black), making songs (cyan), or making USVs (magenta).

(N) Polar plots showing distribution of note onsets (green) and offsets (red) relative to the respiratory cycle (inspiration: 0–π; exhalation: π–2π) for song notes (onset 3.71 ± 0.01 rad, offset 5.74 ± 0.00 rad, n = 2,384 notes).

(O) Same as (N) but for USVs (onset 3.93 ± 0.01 rad, offset 5.70 ± 0.01 rad, n = 1,910 notes). Values reported are mean ± SEM and hypothesis testing was performed using Mann-Whitney U test.

See also Figure S7.

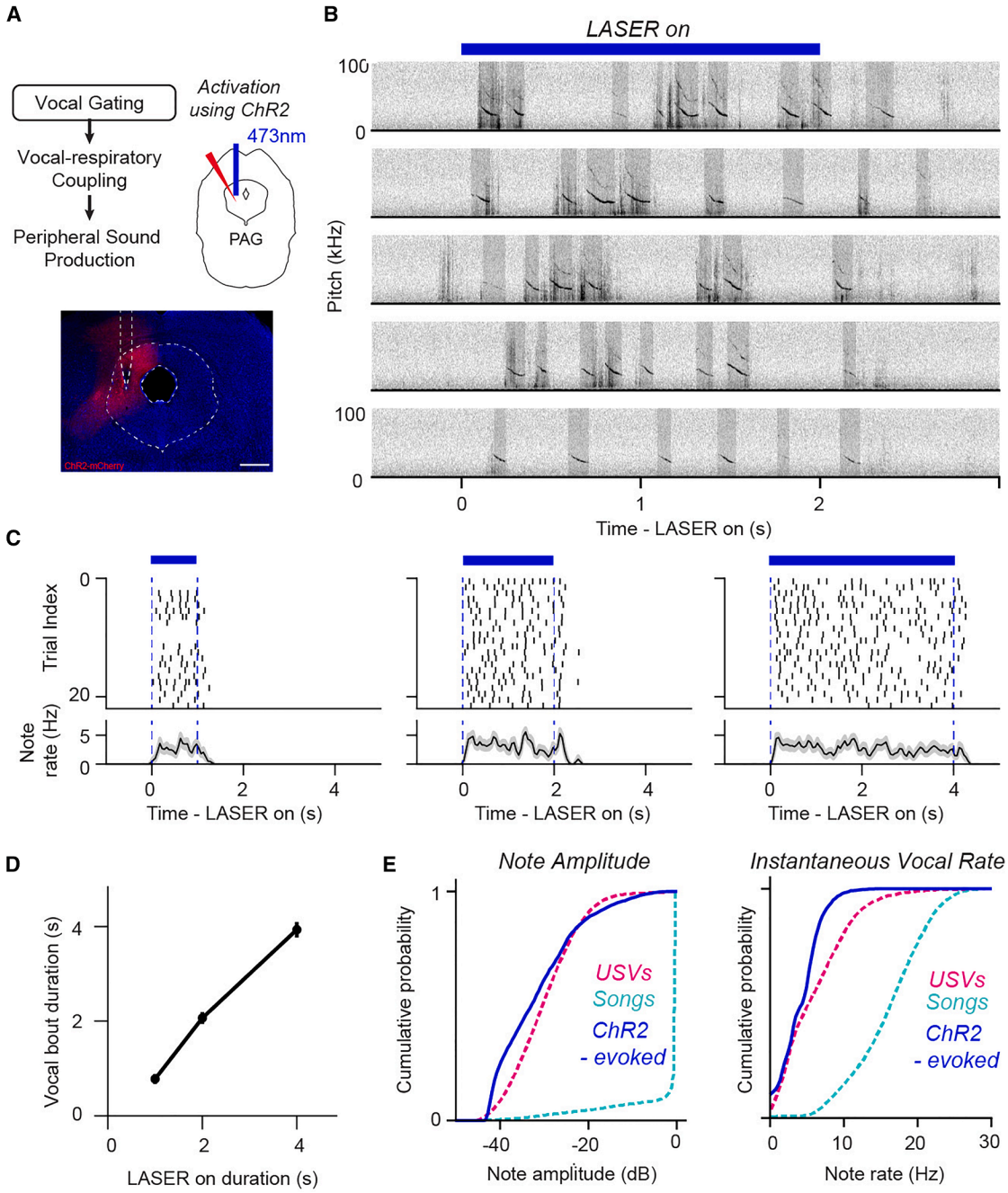


Figure 5. Activating midbrain cIPAG is sufficient to elicit USVs in singing mice

(A) Top left: schematic of the vocal motor hierarchy, highlighting the level addressed by the cIPAG optogenetic activation experiment. Top right: schematic of the experiment, showing the unilateral virus injection in the cIPAG to express ChR2-mCherry, with the fiber implanted above. Bottom: example image of the cIPAG displaying virus expression and the fiber implant.

(B) Spectrograms from five example trials of 2-s tonic optogenetic activation of the cIPAG in an example singing mouse, showing evoked vocalizations during the stimulation period.

(C) Vocalization raster and rate (mean ± SEM) for optogenetic activations of 1 s (left), 2 s (middle), and 4 s (right) in an example mouse, demonstrating that vocalizations are elicited throughout the stimulation period in each condition.

(D) Duration of the evoked vocalization bout compared with the duration of the optogenetic stimulation for all mice (1 s stimulation, 0.78 ± 0.07 s, n = 86 trials; 2 s stimulation, 2.07 ± 0.08 s, n = 85 trials; 4 s stimulation, 3.93 ± 0.12 s, n = 85 trials; n = 4 male mice).

(E) Cumulative distributions of note amplitude (left) and instantaneous note rate (right) of optogenetically evoked vocalizations (amplitude, −31.4 ± 0.2 dB; rate, 4.27 ± 0.06 Hz; n = 2,063 notes) are similar to those of natural USVs (amplitude, −29.8 ± 0.0 dB; rate, 6.26 ± 0.03 Hz; n = 26,246 notes) and distinct from songs (amplitude, −2.3 ± 0.1 dB; rate, 16.16 ± 0.04 Hz; n = 17,681 notes). Values reported are mean ± SEM.

See also Figures S4 and S7 and Video S3.

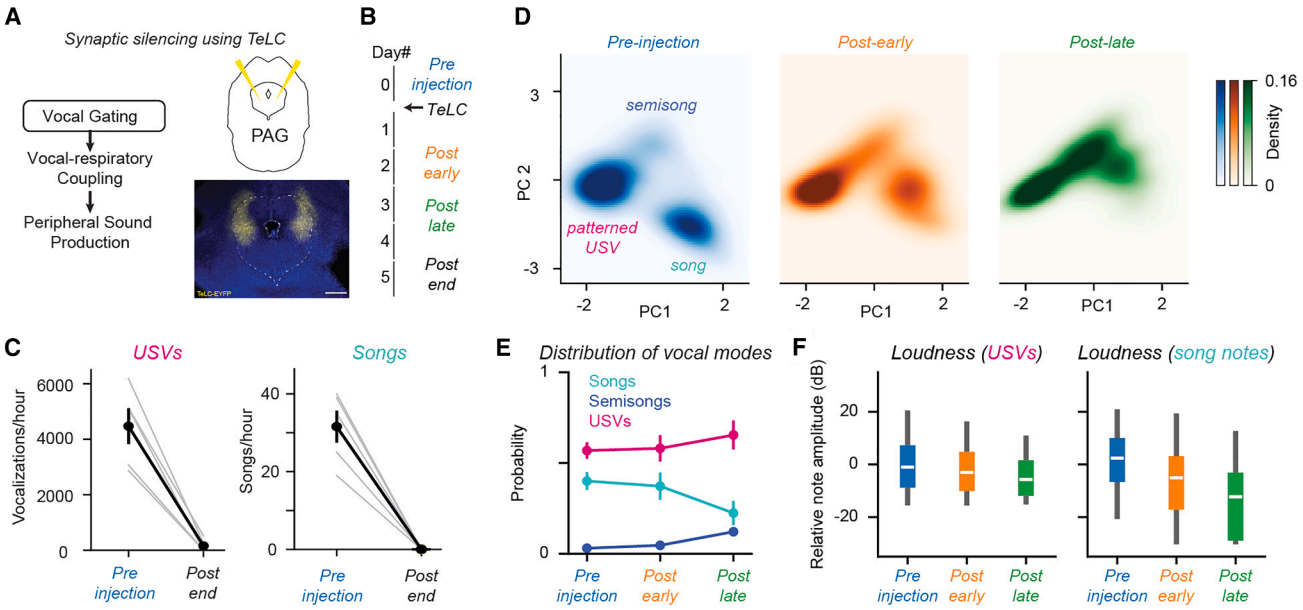


Figure 6. Synaptic silencing of the clPAG progressively degrades the vocal repertoire of singing mice

(A) Left: schematic of the vocal motor hierarchy, highlighting the level addressed by the clPAG synaptic-silencing experiment. Top right: schematic of the experiment, showing the bilateral virus injection in the clPAG to express TeLC-EYFP. Bottom right: example image of the clPAG displaying virus expression.

(B) Experimental timeline of the clPAG-silencing experiment with tetanus toxin light chain (TeLC). To summarize data across slightly different timelines across animals, the post-injection period for each animal was binned into two halves, with "post-early" covering the hours containing the first half and "post-late" the second half of all curated vocalizations.

(C) Number of USVs (left; pre-injection, 4,469 ± 645 notes/h; post-injection, 157 ± 102 notes/h, $p = 0.008$) and songs (right; pre-injection, 32 ± 4 songs; post-injection, 0 ± 0 songs, $p = 0.008$) before and after the virus injection ($n = 5$ male mice), quantified as the number of vocalizations in a highly vocal hour before injection and 5–6 days after injection.

(D) Distribution of vocal segments in the temporal feature space (see also Figure S2) for curated hours before virus injection (left; 428 segments from 39,596 notes), during the post-early period (middle; $n = 408$ segments from 34,826 notes), and during the post-late period (right; $n = 252$ segments from 21,084 notes), showing a shift in density away from songs.

(E) Proportion of USVs, semisong notes, and song notes before virus injection, during the post-early period, and during post-late period (pre-injection vs. post-late: song notes, U = 21, $p = 0.0476$; semisong notes, U = 1, $p = 0.008$; USVs, U = 8, $p = 0.210$; $n = 5$ male mice).

(F) Distribution of the changes in note amplitude relative to the mean loudness in the pre-injection epoch for USVs (left; pre-injection, 0.00 ± 0.05 dB, $n = 22,663$ notes; post-early, −1.78 ± 0.05 dB, $n = 19,599$ notes; post-late, −4.42 ± 0.05 dB, $n = 14,955$ notes) and song notes (right; pre-injection, 0.00 ± 0.09 dB, $n = 15,583$ notes; post-early, −8.07 ± 0.10 dB, $n = 13,490$ notes; post-late, −13.90 ± 0.17 dB, $n = 3,774$ notes) across experimental epochs. Values reported are mean ± SEM and hypothesis testing was performed using Mann-Whitney U test.

See also Figures S4–S7.

closely followed our model's prediction: T would vary inversely with the negative logarithm of the r_{min} (Figure 7E). Thus, clPAG silencing shortened songs by causing earlier termination. In other words, male mice were no longer able to sustain their song motor pattern and append long, loud notes at the end of songs (Figures 7F and S5D). As songs progressively deteriorated over days, they eventually broke into multiple short, quiet, barely recognizable fragments before disappearing entirely—consistent with cumulative degradation of phonation, loudness, and rhythmic structure (Figure S6). Taken together with the necessity of the clPAG for producing both songs and USVs, we conclude that the song mode is produced not via a separate pathway but through AM of individual notes and FM of motor patterning in the clPAG.

Given that the clPAG modulates song rhythm via the stop rate r_{min}, we wondered whether this mechanism might also account for natural variability in song duration. Consistent with previous work,[15,54] we observed that songs are longer in males than in females (Figure 7G). Decomposing each song into its constituent parameters, we found that neither the start rate nor the slope significantly differed between the sexes (Figure S5C). However, stop rates were significantly lower for males compared with females (Figure 7H). Again, across all songs, durations closely followed the theoretical relationship with r_{min} predicted by our model (Figure 7I). Thus, the longer songs in males reflect their ability to maintain the vocal pattern (FM) at lower stop rates—effectively extending the songs by adding more long notes at the end (Figure 7J). Therefore, the parameter most affected by clPAG silencing—the stop rate—also accounts for natural sexual dimorphism in song production. We conclude that clPAG is a key locus not only for orchestrating distinct vocal modes (songs and USVs) but also for driving the graded, sexually dimorphic natural variability in song production.

DISCUSSION

In this study, we leveraged the rich vocal behavior of the singing mouse (*Scotinomys teguina*) to investigate the organizational

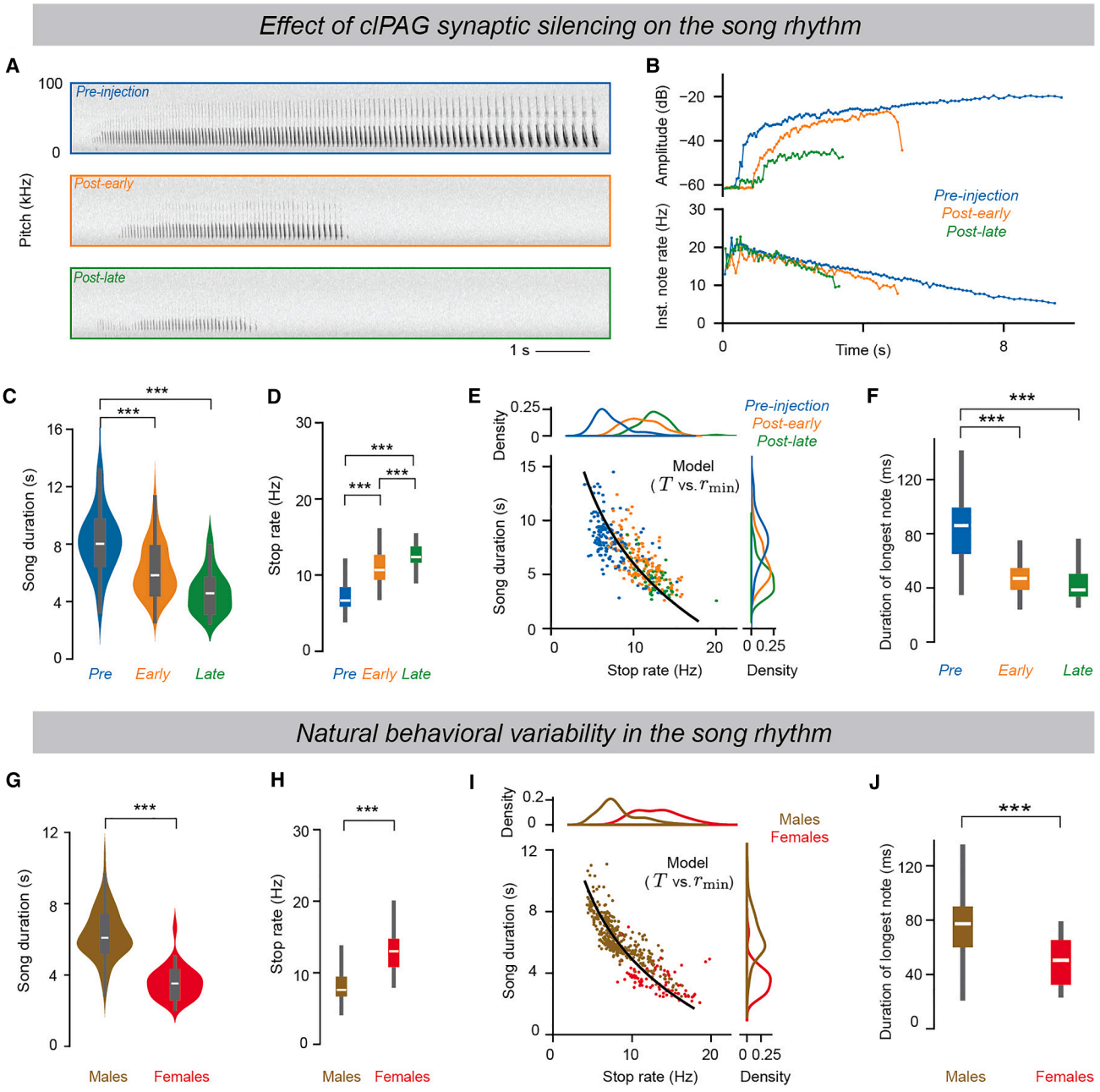


Figure 7. Parametric modulation of song patterning by the clPAG

(A) Spectrograms of example songs from one mouse before the PAG TeLC virus injection (top), during the post-early period (middle), and during post-late period (bottom).

(B) Trajectories of note amplitude (top) and instantaneous note rate (bottom) for the example songs shown in (A).

(C) Song duration distributions for songs across different PAG TeLC experimental time points (pre-injection, n = 157 songs, 8.37 ± 0.20 s; post-early, n = 153 songs, 6.22 ± 0.15 s; post-late, n = 54 songs, 4.64 ± 0.20 s; pre-injection vs. post-early, p = 5.73e−16; post-early vs. post-late, p = 4.81e−8; pre-injection vs. post-late, p = 1.53e−19).

(D) Distribution of the stop rate (r_{min})—one of the patterning parameters in our song rhythm model—for songs in (C) (pre-injection, 7.27 ± 0.16 Hz; post-early, 10.95 ± 0.17 Hz; post-late, 12.57 ± 0.25 Hz; pre-injection vs. post-early, p = 2.25e−33; post-early vs. post-late, p = 2.04e−6; pre-injection vs. post-late, p = 6.52e−24).

(E) Song duration vs. the stop rate for songs in (C). The black line represents the model prediction, based on a mean slope of −0.11 and a mean start rate of 19.15 Hz.

(F) Distribution of the duration of the longest note for songs in (C) (pre-injection, 84.1 ± 1.8 ms; post-early, 48.3 ± 1.0 ms; post-late, 43.2 ± 2.0 ms; pre-injection vs. post-early, p = 4.18e−35; post-early vs. post-late, p = 0.001; post-early vs. post-late, p = 2.76e−2).

(legend continued on next page)

logic of multifunctional motor circuits. We developed a behavioral assay (PAIRId) that enables precise attribution of vocalizations to individual animals during social interactions, solving a major technical bottleneck in bioacoustics (Figure 1). Using PAIRId, we found that singing mice employ two major, categorically distinct vocal modes: soft, unstructured USVs and loud, temporally patterned songs, often in quick succession during social encounters (Figures 1 and 2). We derived a simple linear model that captures the rhythmic structure of the songs with just three parameters (Figure 3). Despite their dramatic differences in acoustic properties, temporal organization, and social context, both songs and USVs share peripheral sound production mechanisms, vocal-respiratory coupling, and central neural control by the cIPAG (Figures 4, 5, and 6). Using our mathematical model, we demonstrated how progressive silencing of cIPAG neurons systematically alters specific aspects of song production before eliminating all vocalizations (Figure 7). Collectively, our findings suggest a potential principle of multifunctional circuit reuse: a shared phonatory control pathway supports multiple vocal modes through AM of individual notes and FM of global motor patterning (Figure S7). This work also offers a window into how neural circuits could be modified during evolution to diversify vocal repertoire in mammals, providing insights into the mechanistic basis of behavioral evolution.

Two vocal modes: Conserved USVs and novel songs

Our behavioral results suggest that singing mice have evolved a novel song mode while retaining ancestral USVs common to other rodents. The acoustic properties of singing mouse USVs—their soft amplitude, reduced temporal stereotypy, and usage during close-range social interactions—closely parallel USVs observed in laboratory mice,[33,55] despite differences in pitch that can be attributed to species-specific variations in laryngeal morphology.[51–53] The shared biophysical mechanisms of sound production through aerodynamic whistles and vocal gating by the cIPAG further support this homology at a mechanistic level.[35,45] This conservation is particularly notable given the phylogenetic distance between singing mice (Cricetidae) and laboratory mice (Muridae), suggesting that USV production represents a deeply conserved trait across multiple rodent families. Indeed, species producing USV-like vocalizations can be found across every studied subfamily within both Cricetidae and Muridae,[31,34,39] providing strong evidence that USVs constitute an ancestral vocal mode that predates the divergence of these lineages approximately 20–25 million years ago.[56]

In contrast, songs represent a derived, novel vocal behavior unique to a small clade of rodents. Often accompanied by an upright posture, songs consist of a series of loud notes with a stereotyped rhythmic progression—fast and short at the start, slowing with each repetition with longer notes at the end. Songs are used primarily for long-range vocal signaling and exchanges, shaped by the need for communication in a diurnal cloud-forest niche,[14,17,57] physiological costs,[58,59] and sexual selection for display elaboration.[60,61] Similar, but less elaborate, songs have been reported in close relatives (*Baiomys* and *Scotinomys xerampelinus*[15,40]), and their restricted phylogenetic distribution suggests that songs arose relatively recently (~6.5 million years ago[56,62]). In accordance with an operational definition of evolutionary novelty—new traits or novel combinations that serve a new function within a lineage's ecology[63,64]—songs fulfill all of these criteria and may thus be considered a behavioral novelty. Importantly, however, a behavioral novelty does not necessarily imply a novel mechanism; they often arise via tinkering of preexisting ones.[65] In the case of singing mice, our results indicate that songs and USVs share a phonatory pathway. Notably, we observe rare, acoustically intermediate semisongs produced during close-range interactions. The presence of these semisongs and smooth transitions between the different vocal modes reinforce that songs likely represent a redeployment of this conserved motor program.

A defining feature of this redeployment is temporal stereotypy, with songs arising from the rhythmic activation of the conserved phonatory circuit over many seconds. This rhythmic progression of songs can be captured by a simple linear model defined by three interpretable parameters. This compact parameterization not only accurately describes the structure of the behavior but also enables us to decompose variability into distinct components, each of which may be linked to separable neural substrates. In this way, the model provides a principled framework for investigating the neural mechanisms underlying specific aspects of song production. More broadly, singing mouse songs may exemplify a special class of natural behaviors—like birdsong,[66] woodpecker drumming,[67] or locomotor gaits[68]—that are highly stereotyped, rhythmic, and low-dimensional, making them particularly amenable to quantitative analysis. Therefore, the song system in the singing mouse offers an exciting model for understanding not only the neural mechanisms of vocal motor control but also how changes in neural circuits may lead to behavioral innovations.

The role of the cIPAG in USVs and songs

The coexistence of ancestral USVs and derived songs in individual singing mice provides a unique opportunity to investigate how neural circuits generate behavioral novelty. To this end, we used neural perturbation experiments to examine the role of the cIPAG in the two vocal modes. The PAG plays a conserved role in the control of instinctive behaviors across vertebrates,[69–74] with particular importance for vocalization.[20,21] Stimulation of the PAG (also called the "central gray," and, in birds, the "dorsomedial nucleus of the intercollicular complex") elicits species-typical vocalizations across vertebrates, including fish,[75] birds,[76–78] bats,[79] rodents,[45,80,81] and primates.[82–84] Further, bilateral PAG

(G) Song duration distribution for intact males and females when alone or socially interacting in PAIRId (male, $n = 370$ songs from 6 mice, 6.28 ± 0.07 s; female, $n = 86$ songs from 4 mice, 3.58 ± 0.10 s; $p = 6.65e{-}38$).
(H) Same as (D) but for songs in (G) (male, 8.27 ± 0.13 Hz; female, 13.10 ± 0.30 Hz; $p = 1.21e{-}30$).
(I) Same as (E) but for songs in (G). The black line represents the model prediction, with a mean slope of −0.18 and a mean start rate of 24.23 Hz.
(J) Same as (F) but for songs in (G) (male, 75.1 ± 1.1 ms; female, 49.2 ± 1.3 ms; $p = 2.72e{-}24$). Values reported are mean ± SEM and hypothesis testing was performed using Mann-Whitney U test.
See also Figures S4–S7.

lesions cause mutism in both learned (speech) and innate vocalizations in humans,[85] as well as mutism in many other species.[75,86–90] Our results on USVs are consistent with this role: optogenetic activation of the clPAG was sufficient to evoke USVs, while synaptic silencing eliminated them.

The silencing experiment implicated the clPAG as a critical locus for song production, as TeLC expression ultimately abolished songs. The manner of this loss provides further insight: songs were progressively truncated, with later notes disappearing first and overall duration decreasing in a pattern well captured by the stop-rate parameter of our model. These later notes are normally louder and longer. Because TeLC expression weakens clPAG output[50]—either by reducing drive to downstream targets or by altering local dynamics—we speculate that they may be especially vulnerable. Together, these observations support our working model that the song motor program co-opts the clPAG's ancestral phonatory control circuit for USVs,[45,87] with individual song notes depending on graded clPAG output and stronger activity required to sustain the later, more demanding portions of the song. Interestingly, the same stop-rate parameter also accounts for natural variability in song duration observed both within and between sexes, suggesting that the clPAG may be the source of this variability.

Although our study establishes the clPAG as a key region for both vocal modes, future work is needed to clarify how its neural activity generates songs while also supporting USV production. For example, optogenetic activation of the clPAG elicited USVs, but not song-like vocalizations, and silencing produced progressive truncation rather than complete disruption of rhythmic structure. These observations suggest that the rhythm of the song does not originate within the clPAG but is instead likely imposed by upstream circuits, resulting in intrinsic clPAG dynamics that cannot be readily recruited by non-specific activation. This dissociation is reminiscent of findings in songbirds, where no artificial stimulation to date can evoke full songs: activating the avian PAG homolog elicits only innate vocalizations,[76,77,91] whereas stimulating forebrain song areas produces isolated song syllables but not full songs.[91] Our study also cannot yet resolve how songs and USVs are represented locally within the clPAG. Given the common phonatory mechanism, one possibility is that some neurons contribute to both songs and USVs by issuing shared motor commands to brainstem premotor centers. Alternatively, the stark differences in rhythm and loudness between songs and USVs may indicate that the two behaviors emerge from distinct neuronal populations. This question parallels a recent study in laboratory mice showing that PAG neurons controlling the whistle-based USVs are not required for the vibrational squeaks.[92] Future work, including *in vivo* electrophysiology, will be important for determining the extent of shared vs. segregated control between songs and USVs and for uncovering the neural dynamics underlying each mode, thereby providing crucial insights into the organization of the PAG.

The PAG as a key locus for rapid evolutionary diversification

Evolutionary modification of ancestral neural circuits provides a pathway for rapid behavioral diversification. The midbrain PAG is anatomically well positioned to receive ethologically relevant contextual inputs from the forebrain and to coordinate diverse motor output programs,[93–95] leading to a recent proposal positing the PAG as a key driver in the evolution of innate motor behaviors across species.[96] We put this idea to test and indeed identified the clPAG as a critical locus for the evolutionary diversification of vocal behavior in the singing mouse, supporting both the ancestral USVs and the derived song. Using clPAG as a critical circuit node, future work will dissect how the brain-wide pathways (e.g., cortical control) that underlie song production redeploys a conserved phonation program shared with ancestral USVs.[18,19,97,98]

Neural circuit co-option may represent a common strategy for evolving complex behaviors. Rather than evolving separate circuits *de novo*—a process requiring coordinated changes across multiple organizational levels—evolution appears to favor modifying ancestral circuits to operate in new regimes.[7,69,99–104] Beyond vocal communication in singing mice, quantitative modifications of preexisting circuit nodes is widespread at the molecular, genetic, and anatomical levels, representing a general principle for the rapid evolutionary emergence of biological innovations across scales and systems.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to, and will be fulfilled by, the lead contact, Arkarup Banerjee (abanerjee@cshl.edu).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- The dataset reported in this study has been deposited at Dryad and is available under the DOI: https://doi.org/10.5061/dryad.ht76hdrw5.
- The code to replicate the analyses is available at: https://github.com/singingmicelab/zheng-harpole-2025-vocalmodes-pag.
- Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health BRAIN Initiative via RF1-NS132046-01 (A.B.), Searle Scholars Program (A.B.), Pershing Square Foundation Innovator Fund (A.B.), the Esther A. & Joseph Klingenstein Fund (A.B.), Cold Spring Harbor Laboratory (A.B.), the International Society for Neuroethology Konishi Research Award (C.E.H.), and the George A. and Marjorie H. Anderson Fellowship (X.M.Z.). We are grateful to Priyanka Gupta and Devon Cowan for performing the thermistor implantation surgery and technical support. Luke Bemish contributed to earlier iterations of the song rhythm model. Bo Li generously provided the Flex TeLC virus. We thank Florin Albeanu, Benjamin Cowley, Priyanka Gupta, Michael Long, Stephen Shea, and members of the Banerjee lab for their helpful comments on earlier versions of the manuscript. We thank the staff of the Cold Spring Harbor Laboratory Animal Resource for their dedicated animal care. We thank the creators and maintainers of the open-source software used in this project. We also thank Taylor Sterry for visual illustrations of the mice and the PAIRId apparatus.

AUTHOR CONTRIBUTIONS

Conceptualization, A.B., X.M.Z., and C.E.H.; methodology, X.M.Z., C.E.H., and A.B.; investigation, X.M.Z., C.E.H., and M.B.D.; data curation, C.E.H. and X.M.Z.; software, X.M.Z.; formal analysis, X.M.Z. and C.E.H.; visualization, X.M.Z., C.E.H., and A.B.; writing—original draft, A.B.; writing—review & editing, X.M.Z., C.E.H., and A.B.; supervision, A.B.; project administration, A.B.; funding acquisition, A.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used OpenAI ChatGPT and Anthropic Claude in order to improve the readability of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animal statement
  - Animals
- METHOD DETAILS
  - Experimental Procedures
  - Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2025.10.036.

Received: May 14, 2025
Revised: September 12, 2025
Accepted: October 13, 2025
Published: November 17, 2025

REFERENCES

1. Ferreira-Pinto, M.J., Kanodia, H., Falasconi, A., Sigrist, M., Esposito, M.S., and Arber, S. (2021). Functional diversity for body actions in the mesencephalic locomotor region. Cell *184*, 4564–4578.e18. https://doi.org/10.1016/j.cell.2021.07.002.
2. Ruder, L., Schina, R., Kanodia, H., Valencia-Garcia, S., Pivetta, C., and Arber, S. (2021). A functional map for diverse forelimb actions within brainstem circuitry. Nature *590*, 445–450. https://doi.org/10.1038/s41586-020-03080-z.
3. Wang, L., Chen, I.Z., and Lin, D. (2015). Collateral pathways from the ventromedial hypothalamus mediate defensive behaviors. Neuron *85*, 1344–1358. https://doi.org/10.1016/j.neuron.2014.12.025.
4. Barkan, C.L., Kelley, D.B., and Zornik, E. (2018). Premotor neuron divergence reflects vocal evolution. J. Neurosci. *38*, 5325–5337. https://doi.org/10.1523/JNEUROSCI.0089-18.2018.
5. Briggman, K.L., and Kristan, W.B. (2008). Multifunctional pattern-generating circuits. Annu. Rev. Neurosci. *31*, 271–294. https://doi.org/10.1146/annurev.neuro.31.060407.125552.
6. Marder, E., O'Leary, T., and Shruti, S. (2014). Neuromodulation of circuits with variable parameters: single neurons and small circuits reveal principles of state-dependent and robust neuromodulation. Annu. Rev. Neurosci. *37*, 329–346. https://doi.org/10.1146/annurev-neuro-071013-013958.
7. Roberts, R.J.V., Pop, S., and Prieto-Godino, L.L. (2022). Evolution of central neural circuits: state of the art and perspectives. Nat. Rev. Neurosci. *23*, 725–743. https://doi.org/10.1038/s41583-022-00644-y.
8. Castellucci, G.A., Guenther, F.H., and Long, M.A. (2022). A theoretical framework for human and nonhuman vocal interaction. Annu. Rev. Neurosci. *45*, 295–316. https://doi.org/10.1146/annurev-neuro-111020-094807.
9. Hage, S.R., and Nieder, A. (2016). Dual neural network model for the evolution of speech and language. Trends Neurosci. *39*, 813–829. https://doi.org/10.1016/j.tins.2016.10.006.
10. Kelley, D.B., Ballagh, I.H., Barkan, C.L., Bendesky, A., Elliott, T.M., Evans, B.J., Hall, I.C., Kwon, Y.M., Kwong-Brown, U., Leininger, E.C., et al. (2020). Generation, coordination, and evolution of neural circuits for vocal communication. J. Neurosci. *40*, 22–36. https://doi.org/10.1523/JNEUROSCI.0736-19.2019.
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 *375*, 20190054. https://doi.org/10.1098/rstb.2019.0054.
12. Zhang, Y.S., and Ghazanfar, A.A. (2020). A hierarchy of autonomous systems for vocal production. Trends Neurosci. *43*, 115–126. https://doi.org/10.1016/j.tins.2019.12.006.
13. Banerjee, A., Phelps, S.M., and Long, M.A. (2019). Singing mice. Curr. Biol. *29*, R190–R191. https://doi.org/10.1016/j.cub.2018.11.048.
14. Hooper, E.T., and Carleton, M.D. (1976). Reproduction, growth and development in two contiguously allopatric rodent species, genus Scotinomys. Miscellaneous Publications*151* (Museum of Zoology, University of Michigan), pp. 1–52. http://deepblue.lib.umich.edu/handle/2027.42/56395.
15. Miller, J.R., and Engstrom, M.D. (2007). Vocal stereotypy and singing behavior in baiomyine mice. J. Mammal. *88*, 1447–1465. https://doi.org/10.1644/06-MAMM-A-386R.1.
16. Pasch, B., Bolker, B.M., and Phelps, S.M. (2013). Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice. Am. Nat. *182*, E161–E173. https://doi.org/10.1086/673263.
17. Fujishima, Y., and Long, M.A. (2025). Advertisement vocalizations support home-range defense in the singing mouse. Curr. Biol. *35*, 2960–2966.e5. https://doi.org/10.1016/j.cub.2025.04.034.
18. Banerjee, A., Chen, F., Druckmann, S., and Long, M.A. (2024). Temporal scaling of motor cortical dynamics reveals hierarchical control of vocal production. Nat. Neurosci. *27*, 527–535. https://doi.org/10.1038/s41593-023-01556-5.
19. 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. https://doi.org/10.1126/science.aau9480.
20. Jürgens, U. (2009). The neural control of vocalization in mammals: a review. J. Voice *23*, 1–10. https://doi.org/10.1016/j.jvoice.2007.07.005.
21. Jürgens, U. (1994). The role of the periaqueductal grey in vocal behaviour. Behav. Brain Res. *62*, 107–117. https://doi.org/10.1016/0166-4328(94)90017-5.
22. Fukushima, M., and Margoliash, D. (2015). The effects of delayed auditory feedback revealed by bone conduction microphone in adult zebra finches. Sci. Rep. *5*, 8800. https://doi.org/10.1038/srep08800.
23. Sterling, M.L., Teunisse, R., and Englitz, B. (2023). Rodent ultrasonic vocal interaction resolved with millimeter precision using hybrid beamforming. eLife *12*, e86126. https://doi.org/10.7554/eLife.86126.
24. Waidmann, E.N., Yang, V.H.Y., Luo, E., Doyle, W.C., and Jarvis, E.D. (2025). Mountable miniature microphones to identify and assign mouse ultrasonic vocalizations. Cell Rep. Methods *5*, 101081. https://doi.org/10.1016/j.crmeth.2025.101081.
25. Warren, M.R., Sangiamo, D.T., and Neunuebel, J.P. (2018). High channel count microphone array accurately and precisely localizes ultrasonic signals from freely-moving mice. J. Neurosci. Methods *297*, 44–60. https://doi.org/10.1016/j.jneumeth.2017.12.013.
26. Peterson, R.E., Tanelus, A., Ick, C., Mimica, B., Francis, N., Ivan, V.J., Choudhri, A., Falkner, A.L., Murthy, M., Schneider, D.M., et al. (2024). Vocal call locator benchmark (vcl) for localizing rodent vocalizations from multi-channel audio. In Advances in Neural Information Processing Systems, *37*, A. Globerson, L. Mackey, D. Belgrave, A.

Fan, U. Paquet, J. Tomczak, and C. Zhang, eds. (Curran Associates, Inc.), pp. 106370–106382. https://proceedings.neurips.cc/paper_files/paper/2024/file/c00d37d6b04d73b870b963a4d70051c1-Paper-Datasets_and_Benchmarks_Track.pdf.

27. Sirotin, Y.B., Costa, M.E., and Laplagne, D.A. (2014). Rodent ultrasonic vocalizations are bound to active sniffing behavior. Front. Behav. Neurosci. *8*, 399. https://doi.org/10.3389/fnbeh.2014.00399.
28. Cox, S.S., Kearns, A.M., Woods, S.K., Brown, B.J., Brown, S.J., and Reichel, C.M. (2022). The role of the anterior insular during targeted helping behavior in male rats. Sci. Rep. *12*, 3315. https://doi.org/10.1038/s41598-022-07365-3.
29. Gachomba, M.J.M., Esteve-Agraz, J., Caref, K., Maroto, A.S., Bortolozzo-Gleich, M.H., Laplagne, D.A., and Márquez, C. (2022). Multimodal cues displayed by submissive rats promote prosocial choices by dominants. Curr. Biol. *32*, 3288–3301.e8. https://doi.org/10.1016/j.cub.2022.06.026.
30. von Merten, S., Hoier, S., Pfeifle, C., and Tautz, D. (2014). A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (mus musculus domesticus). PLoS One *9*, e97244. https://doi.org/10.1371/journal.pone.0097244.
31. Sales, G.D.S. (1972). Ultrasound and mating behaviour in rodents with some observations on other behavioural situations. J. Zool. *168*, 149–164. https://doi.org/10.1111/j.1469-7998.1972.tb01345.x.
32. Holy, T.E., and Guo, Z. (2005). Ultrasonic songs of male mice. PLoS Biol. *3*, e386. https://doi.org/10.1371/journal.pbio.0030386.
33. Castellucci, G.A., Calbick, D., and McCormick, D. (2018). The temporal organization of mouse ultrasonic vocalizations. PLoS One *13*, e0199929. https://doi.org/10.1371/journal.pone.0199929.
34. Sales, G.D. (2010). Ultrasonic calls of wild and wild-type rodents. In Handbook of Behavioral Neuroscience (Elsevier), pp. 77–88. https://doi.org/10.1016/b978-0-12-374593-4.00009-7.
35. Roberts, L.H. (1975). The rodent ultrasound production mechanism. Ultrasonics *13*, 83–88. https://doi.org/10.1016/0041-624x(75)90052-9.
36. Håkansson, J., Jiang, W., Xue, Q., Zheng, X., Ding, M., Agarwal, A.A., and Elemans, C.P.H. (2022). Aerodynamics and motor control of ultrasonic vocalizations for social communication in mice and rats. BMC Biol. *20*, 3. https://doi.org/10.1186/s12915-021-01185-z.
37. Mahrt, E., Agarwal, A., Perkel, D., Portfors, C., and Elemans, C.P.H. (2016). Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. Curr. Biol. *26*, R880–R881. https://doi.org/10.1016/j.cub.2016.08.032.
38. Riede, T., Borgard, H.L., and Pasch, B. (2017). Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. R. Soc. Open Sci. *4*, 170976. https://doi.org/10.1098/rsos.170976.
39. Fernández-Vargas, M., Riede, T., and Pasch, B. (2022). Mechanisms and constraints underlying acoustic variation in rodents. Anim. Behav. *184*, 135–147. https://doi.org/10.1016/j.anbehav.2021.07.011.
40. Riede, T., and Pasch, B. (2020). Pygmy mouse songs reveal anatomical innovations underlying acoustic signal elaboration in rodents. J. Exp. Biol. *223*, jeb223925. https://doi.org/10.1242/jeb.223925.
41. Welker, W.I. (1964). Analysis of sniffing of the albino rat 1. Behaviour *22*, 223–244. https://doi.org/10.1163/156853964X00030.
42. Deschênes, M., Moore, J., and Kleinfeld, D. (2012). Sniffing and whisking in rodents. Curr. Opin. Neurobiol. *22*, 243–250. https://doi.org/10.1016/j.conb.2011.11.013.
43. Roberts, L.H. (1972). Correlation of respiration and ultrasound production in rodents and bats. J. Zool. *168*, 439–449. https://doi.org/10.1111/j.1469-7998.1972.tb01360.x.
44. MacDonald, A., Hebling, A., Wei, X.P., and Yackle, K. (2024). The breath shape controls intonation of mouse vocalizations. eLife *13*, RP93079. https://doi.org/10.7554/eLife.93079.
45. 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.e4. https://doi.org/10.1016/j.neuron.2019.05.025.
46. Veerakumar, A., Head, J.P., and Krasnow, M.A. (2023). A brainstem circuit for phonation and volume control in mice. Nat. Neurosci. *26*, 2122–2130. https://doi.org/10.1038/s41593-023-01478-2.
47. Yang, C.F., Kim, E.J., Callaway, E.M., and Feldman, J.L. (2020). Monosynaptic projections to excitatory and inhibitory preBötzinger complex neurons. Front. Neuroanat. *14*, 58. https://doi.org/10.3389/fnana.2020.00058.
48. Park, J., Choi, S., Takatoh, J., Zhao, S., Harrahill, A., Han, B.X., and Wang, F. (2024). Brainstem control of vocalization and its coordination with respiration. Science *383*, eadi8081. https://doi.org/10.1126/science.adi8081.
49. Hartmann, K., and Brecht, M. (2020). A functionally and anatomically bipartite vocal pattern generator in the rat brain stem. iScience *23*, 101804. https://doi.org/10.1016/j.isci.2020.101804.
50. Sweeney, S.T., Broadie, K., Keane, J., Niemann, H., and O'Kane, C.J. (1995). Targeted expression of tetanus toxin light chain in Drosophila specifically eliminates synaptic transmission and causes behavioral defects. Neuron *14*, 341–351. https://doi.org/10.1016/0896-6273(95)90290-2.
51. Abhirami, S., Agarwalla, S., Bhattacharya, A., and Bandyopadhyay, S. (2023). Contribution of the ventral pouch in the production of mouse ultrasonic vocalizations. Phys. Rev. E *107*, 024412. https://doi.org/10.1103/PhysRevE.107.024412.
52. Smith, S.K., Burkhard, T.T., and Phelps, S.M. (2021). A comparative characterization of laryngeal anatomy in the singing mouse. J. Anat. *238*, 308–320. https://doi.org/10.1111/joa.13315.
53. Smith, S.K., Håkansson, J., Frazel, P.W., Long, M.A., Elemans, C.P.H., and Phelps, S.M. (2025). Mechanisms and control of a novel vocalization: The singing mouse song is a whistle that depends on air sac inflation. Preprint at bioRxiv. https://doi.org/10.1101/2025.05.16.654575.
54. Tripp, J.A., and Phelps, S.M. (2024). Females counter-sing, but response to male song differs by sex in alston's singing mouse. Biol. Lett. *20*, 20230484. https://doi.org/10.1098/rsbl.2023.0484.
55. Musolf, K., and Penn, D.J. (2012). Ultrasonic vocalizations in house mice. In Evolution of the House Mouse, M. Macholán, S.J.E. Baird, and P. Munclinger, eds. (Cambridge University Press), pp. 253–277. https://doi.org/10.1017/CBO9781139044547.012.
56. Steppan, S.J., and Schenk, J.J. (2017). Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS One *12*, e0183070. https://doi.org/10.1371/journal.pone.0183070.
57. Ribble, D.O., and Rathbun, G.B. (2018). Preliminary observations on home ranges and natural history of *Scotinomys teguina* in costa rica. Mammalia *82*, 490–493. https://doi.org/10.1515/mammalia-2017-0065.
58. Burkhard, T.T., Westwick, R.R., and Phelps, S.M. (2018). Adiposity signals predict vocal effort in alston's singing mice. Proc. Biol. Sci. *285*, 20180090. https://doi.org/10.1098/rspb.2018.0090.
59. Giglio, E.M., and Phelps, S.M. (2020). Leptin regulates song effort in neotropical singing mice (scotinomys teguina). Anim. Behav. *167*, 209–219. https://doi.org/10.1016/j.anbehav.2020.06.022.
60. Pasch, B., George, A.S., Campbell, P., and Phelps, S.M. (2011). Androgen-dependent male vocal performance influences female preference in neotropical singing mice. Anim. Behav. *82*, 177–183. https://doi.org/10.1016/j.anbehav.2011.04.018.
61. Burkhard, T.T., Sachs, E.R., and Phelps, S.M. (2023). Female preferences for high vocal effort in singing mice. Behaviour *160*, 275–297. https://doi.org/10.1163/1568539X-bja10203.
62. Castañeda-Rico, S., Maldonado, J.E., Hawkins, M.T.R., and Edwards, C.W. (2025). Unveiling hidden diversity: Phylogenomics of neotomine rodents and taxonomic implications for the genus peromyscus. Mol. Phylogenet. Evol. *203*, 108233. https://doi.org/10.1016/j.ympev.2024.108233.

63. Pigliucci, M. (2008). What, if anything, is an evolutionary novelty? Philos. Sci. *75*, 887–898. https://doi.org/10.1086/594532.

64. Ding, Y. (2025). Evolution of neural circuits in the origin of behavioral novelty. Curr. Opin. Behav. Sci. *63*, 101520. https://doi.org/10.1016/j.cobeha.2025.101520.

65. Jacob, F. (1977). Evolution and tinkering. Science *196*, 1161–1166. https://doi.org/10.1126/science.860134.

66. Sebastianelli, M., Lukhele, S.M., Secomandi, S., de Souza, S.G., Haase, B., Moysi, M., Nikiforou, C., Hutfluss, A., Mountcastle, J., Balacco, J., et al. (2024). A genomic basis of vocal rhythm in birds. Nat. Commun. *15*, 3095. https://doi.org/10.1038/s41467-024-47305-5.

67. Miles, M.C., Schuppe, E.R., and Fuxjager, M.J. (2020). Selection for rhythm as a trigger for recursive evolution in the elaborate display system of woodpeckers. Am. Nat. *195*, 772–787. https://doi.org/10.1086/707748.

68. Leiras, R., Cregg, J.M., and Kiehn, O. (2022). Brainstem circuits for locomotion. Annu. Rev. Neurosci. *45*, 63–85. https://doi.org/10.1146/annurev-neuro-082321-025137.

69. Baier, F., Reinhard, K., Nuttin, B., Sans-Dublanc, A., Liu, C., Tong, V., Murmann, J.S., Wierda, K., Farrow, K., and Hoekstra, H.E. (2025). The neural basis of species-specific defensive behaviour in peromyscus mice. Nature *645*, 439–447. https://doi.org/10.1038/s41586-025-09241-2.

70. Falkner, A.L., Wei, D., Song, A., Watsek, L.W., Chen, I., Chen, P., Feng, J.E., and Lin, D. (2020). Hierarchical representations of aggression in a hypothalamic-midbrain circuit. Neuron *106*, 637–648.e6. https://doi.org/10.1016/j.neuron.2020.02.014.

71. Lefler, Y., Campagner, D., and Branco, T. (2020). The role of the periaqueductal gray in escape behavior. Curr. Opin. Neurobiol. *60*, 115–121. https://doi.org/10.1016/j.conb.2019.11.014.

72. Silva, C., and McNaughton, N. (2019). Are periaqueductal gray and dorsal raphe the foundation of appetitive and aversive control? a comprehensive review. Prog. Neurobiol. *177*, 33–72. https://doi.org/10.1016/j.pneurobio.2019.02.001.

73. Tovote, P., Esposito, M.S., Botta, P., Chaudun, F., Fadok, J.P., Markovic, M., Wolff, S.B.E., Ramakrishnan, C., Fenno, L., Deisseroth, K., et al. (2016). Midbrain circuits for defensive behaviour. Nature *534*, 206–212. https://doi.org/10.1038/nature17996.

74. Yu, H., Xiang, X., Chen, Z., Wang, X., Dai, J., Wang, X., Huang, P., Zhao, Z.D., Shen, W.L., and Li, H. (2021). Periaqueductal gray neurons encode the sequential motor program in hunting behavior of mice. Nat. Commun. *12*, 6523. https://doi.org/10.1038/s41467-021-26852-1.

75. Kittelberger, J.M., Land, B.R., and Bass, A.H. (2006). Midbrain periaqueductal gray and vocal patterning in a teleost fish. J. Neurophysiol. *96*, 71–85. https://doi.org/10.1152/jn.00067.2006.

76. Brown, J.L. (1965). Vocalization evoked from the optic lobe of a songbird. Science *149*, 1002–1003. https://doi.org/10.1126/science.149.3687.1002.

77. Seller, T.J. (1980). Midbrain regions involved in call production in java sparrows. Behav. Brain Res. *1*, 257–265. https://doi.org/10.1016/0166-4328(80)90033-9.

78. Shimmura, T., Tamura, M., Ohashi, S., Sasaki, A., Yamanaka, T., Nakao, N., Ihara, K., Okamura, S., and Yoshimura, T. (2019). Cholecystokinin induces crowing in chickens. Sci. Rep. *9*, 3978. https://doi.org/10.1038/s41598-019-40746-9.

79. Schuller, G., and Radtke-Schuller, S. (1990). Neural control of vocalization in bats: mapping of brainstem areas with electrical microstimulation eliciting species-specific echolocation calls in the rufous horseshoe bat. Exp. Brain Res. *79*, 192–206. https://doi.org/10.1007/BF00228889.

80. Waldbillig, R.J. (1975). Attack, eating, drinking, and gnawing elicited by electrical stimulation of rat mesencephalon and pons. J. Comp. Physiol. Psychol. *89*, 200–212. https://doi.org/10.1037/h0076808.

81. Yajima, Y., Hayashi, Y., and Yoshii, N. (1980). The midbrain central gray substance as a highly sensitive neural structure for the production of ultrasonic vocalization in the rat. Brain Res. *198*, 446–452. https://doi.org/10.1016/0006-8993(80)90759-3.

82. Brown, T.G. (1915). Note on the physiology of the basal ganglia and midbrain of the anthropoid ape, especially in reference to the act of laughter. J. Physiol. *49*, 195–207. https://doi.org/10.1113/jphysiol.1915.sp001703.

83. Jürgens, U., and Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. Exp. Brain Res. *10*, 532–554. https://doi.org/10.1007/BF00234269.

84. Magoun, H.W., Atlas, D., Ingersoll, E.H., and Ranson, S.W. (1937). Associated facial, vocal and respiratory components of emotional expression: An experimental study. J. Neurol. Psychopathol. *17*, 241–255. https://doi.org/10.1136/jnnp.s1-17.67.241.

85. Esposito, A., Demeurisse, G., Alberti, B., and Fabbro, F. (1999). Complete mutism after midbrain periaqueductal gray lesion. NeuroReport *10*, 681–685. https://doi.org/10.1097/00001756-199903170-00004.

86. Bazett, H.C., and Penfield, W.G. (1922). A study of the sherrington decerebrate animal in the chronic as well as the acute condition. Brain *45*, 185–265. https://doi.org/10.1093/brain/45.2.185.

87. Floody, O.R., and O'Donohue, T.L. (1980). Lesions of the mesencephalic central gray depress ultrasound production and lordosis by female hamsters. Physiol. Behav. *24*, 79–85. https://doi.org/10.1016/0031-9384(80)90017-7.

88. Jürgens, U., and Pratt, R. (1979). Role of the periaqueductal grey in vocal expression of emotion. Brain Res. *167*, 367–378. https://doi.org/10.1016/0006-8993(79)90830-8.

89. Kelly, A.H., Beaton, L.E., and Magoun, H.W. (1946). A midbrain mechanism for facio-vocal activity. J. Neurophysiol. *9*, 181–189. https://doi.org/10.1152/jn.1946.9.3.181.

90. Skultety, F.M. (1962). Experimental mutism in dogs. Arch. Neurol. *6*, 235–241. https://doi.org/10.1001/archneur.1962.00450210063007.

91. Vicario, D.S., and Simpson, H.B. (1995). Electrical stimulation in forebrain nuclei elicits learned vocal patterns in songbirds. J. Neurophysiol. *73*, 2602–2607. https://doi.org/10.1152/jn.1995.73.6.2602.

92. Ziobro, P., Woo, Y., He, Z., and Tschida, K. (2024). Midbrain neurons important for the production of mouse ultrasonic vocalizations are not required for distress calls. Curr. Biol. *34*, 1107–1113.e3. https://doi.org/10.1016/j.cub.2024.01.016.

93. 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*, e63493. https://doi.org/10.7554/eLife.63493.

94. 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. https://doi.org/10.1038/s41586-021-03403-8.

95. Wei, D., Talwar, V., and Lin, D. (2021). Neural circuits of social behaviors: Innate yet flexible. Neuron *109*, 1600–1620. https://doi.org/10.1016/j.neuron.2021.02.012.

96. Schwark, R.W., Fuxjager, M.J., and Schmidt, M.F. (2022). Proposing a neural framework for the evolution of elaborate courtship displays. eLife *11*, e74860. https://doi.org/10.7554/eLife.74860.

97. Isko, E.C., Harpole, C.E., Zheng, X.M., Zhan, H., Davis, M.B., Zador, A.M., and Banerjee, A. (2024). Selective expansion of motor cortical projections in the evolution of vocal novelty. Preprint at bioRxiv. https://doi.org/10.1101/2024.09.13.612752.

98. Merel, J., Botvinick, M., and Wayne, G. (2019). Hierarchical motor control in mammals and machines. Nat. Commun. *10*, 5489. https://doi.org/10.1038/s41467-019-13239-6.

99. Chakraborty, M., and Jarvis, E.D. (2015). Brain evolution by brain pathway duplication. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20150056. https://doi.org/10.1098/rstb.2015.0056.

100. Coleman, R.T., Morantte, I., Koreman, G.T., Cheng, M.L., Ding, Y., and Ruta, V. (2024). A modular circuit coordinates the diversification of

courtship strategies. Nature *635*, 142–150. https://doi.org/10.1038/s41586-024-08028-1.

101. Jourjine, N., and Hoekstra, H.E. (2021). Expanding evolutionary neuroscience: insights from comparing variation in behavior. Neuron *109*, 1084–1099. https://doi.org/10.1016/j.neuron.2021.02.002.
102. Katz, P.S. (2016). Evolution of central pattern generators and rhythmic behaviours. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150057. https://doi.org/10.1098/rstb.2015.0057.
103. Roemschied, F.A., Pacheco, D.A., Aragon, M.J., Ireland, E.C., Li, X., Thieringer, K., Pang, R., and Murthy, M. (2023). Flexible circuit mechanisms for context-dependent song sequencing. Nature *622*, 794–801. https://doi.org/10.1038/s41586-023-06632-1.
104. Seeholzer, L.F., Seppo, M., Stern, D.L., and Ruta, V. (2018). Evolution of a central neural circuit underlies Drosophila mate preferences. Nature *559*, 564–569. https://doi.org/10.1038/s41586-018-0322-9.
105. Bryja, J., and Konecny, A. (2003). Fast sex identification in wild mammals using PCR amplification of the Sry gene. Folia Zool.-Praha *52*, 269–274. https://www.scirp.org/reference/referencespapers?referenceid=923209.
106. Tachibana, R.O., Kanno, K., Okabe, S., Kobayasi, K.I., and Okanoya, K. (2020). USVSEG: A robust method for segmentation of ultrasonic vocalizations in rodents. PLoS One *15*, e0228907. https://doi.org/10.1371/journal.pone.0228907.
107. Pereira, T.D., Tabris, N., Matsliah, A., Turner, D.M., Li, J., Ravindranath, S., Papadoyannis, E.S., Normand, E., Deutsch, D.S., Wang, Z.Y., et al. (2022). SLEAP: A deep learning system for multi-animal pose tracking. Nat. Methods *19*, 486–495. https://doi.org/10.1038/s41592-022-01426-1.
108. Zhao, X., Ziobro, P., Pranic, N.M., Chu, S., Rabinovich, S., Chan, W., Zhao, J., Kornbrek, C., He, Z., and Tschida, K.A. (2021). Sex- and context-dependent effects of acute isolation on vocal and non-vocal social behaviors in mice. PLoS One *16*, e0255640. https://doi.org/10.1371/journal.pone.0255640.
109. Neunuebel, J.P., Taylor, A.L., Arthur, B.J., and Egnor, S.E.R. (2015). Female mice ultrasonically interact with males during courtship displays. eLife *4*, e06203. https://doi.org/10.7554/eLife.06203.
110. Malone, C.A., Ziobro, P., Khinno, J., and Tschida, K.A. (2025). Rates of female mouse ultrasonic vocalizations are low and are not modulated by estrous state during interactions with muted males. Sci. Rep. *15*, 6841. https://doi.org/10.1038/s41598-025-91479-x.
111. McAfee, S.S., Ogg, M.C., Ross, J.M., Liu, Y., Fletcher, M.L., and Heck, D.H. (2016). Minimally invasive highly precise monitoring of respiratory rhythm in the mouse using an epithelial temperature probe. J. Neurosci. Methods *263*, 89–94. https://doi.org/10.1016/j.jneumeth.2016.02.007.
112. Coffey, K.R., Marx, R.E., and Neumaier, J.F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology *44*, 859–868. https://doi.org/10.1038/s41386-018-0303-6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
NeuroTrace 435/455	Invitrogen/Thermo Fischer Scientific	Cat#N21479
Bacterial and virus strains		
AAV2/DJ-hSyn-flex-TeLC-eYFP	Fan Wang; Addgene plasmid; WZ Biosciences viral preparation	Cat#135391; RRID: Addgene_135391
AAV2/9-pENN.AAV.CamKII 0.4.Cre.SV40	James A. Willson; Addgene	Cat#105558-AAV9; RRID: Addgene_105558
AAV2/9-EF1α double-floxed-ChR2-mCherry	Karl Deisseroth; Addgene	Cat#20297-AAV9; RRID: Addgene_20297
Deposited data		
behavioral data	this study	DOI: 10.5061/dryad.ht76hdrw5
Experimental models: Organisms/strains		
Scotinomys teguina	laboratory colony	wild-derived, outbred from La Carpintera and San Gerardo de Dota, Costa Rica
Mus musculus C57Bl/6J	Jackson Laboratory	RRID: IMSR_JAX:000664
Oligonucleotides		
Sry primers	Bryja and Konečný[105]; Integrated DNA Technologies	N/A
Zfy-Zfx primers	Bryja and Konečný[105]; Integrated DNA Technologies	N/A
Software and algorithms		
Python	Python Software Foundation	https://www.python.org/
MATLAB	Mathworks	https://www.mathworks.com/products/matlab.html
Sound recording software	Avisoft Bioacoustics	Avisoft-RECORDER; https://avisoft.com/downloads/
USV detection software	Tachibana et al.[106]	USVSEG; https://github.com/rtachi-lab/usvseg
Markerless pose tracking software	Pereira et al.[107]	SLEAP; https://sleap.ai/
custom code	this study	https://github.com/singingmicelab/zheng-harpole-2025-vocalmodes-pag
Other		
Ultrasound microphone	Avisoft Bioacoustics	CM16/CMPA
Analog/digital converter	Avisoft Bioacoustics	UltraSoundGate 416H
Analog/digital converter	Avisoft Bioacoustics	UltraSoundGate 116H
Custom acrylic mouse enclosures	shopPOPdisplays	https://www.shoppopdisplays.com/
polyimide acoustic foam	Soundcoat	Soundfoam HTC
triggerable USB camera	Telodyne Flir	FLIR Blackfly BFS-U3-20S4C-C
camera lens	Edmund Optics	4.5 mm fixed-focal-length lens C Series #86-900

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animal statement

All animal care and experiments were conducted according to protocols approved by the Cold Spring Harbor Laboratory Institutional Animal Care and Use Committee, and comply with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

Animals

Adult, laboratory-reared, outbred male and female Alston's singing mice (*Scotinomys teguina*), aged 3–20 months, were selected from the colony maintained at Cold Spring Harbor Laboratory. This colony originated from the New York University Langone Medical Center colony,[19] which was descended from wild-captured *S. teguina* from La Carpintera and San Gerardo de Dota, Costa Rica. The sex of the singing mice was confirmed at weaning by genotyping the Y-chromosome's *Sry* gene, using the *Zfy-Zfx* genes as a positive

control.[105] Singing mice were kept at 20–22°C under a 12:12-hour light-dark cycle. They were housed in Thoren Systems #8 enclosures (30.80 × 40.60 × 22.23 cm; Worcester, MA) with corn cob bedding and enrichment items, including sphagnum moss (Galapagos Pet, Santa Barbara, CA), a running wheel (InnoDome + InnoWheel; Bio-Serv, Flemington, NJ), a paper hut (Bio-Hut; Bio-Serv), and a red transparent polycarbonate tube (10 cm × 5 cm, Mouse Tunnel; Bio-Serv). The singing mice were provided with food (a 1:1 mixture of Purina Cat Chow and Mazuri Exotic Animal Nutrition Insectivore pellets) and water *ad libitum*, supplemented with dried mealworms. In a different room from singing mice, C57Bl/6J laboratory mice were kept at 20–22°C in Thoren Systems #9 enclosures (19.56 × 30.91 × 13.34 cm) with corn cob bedding under a 12:12-hour light-dark cycle. Water and mouse chow were available *ad libitum*.

METHOD DETAILS

Experimental Procedures

PAIRId social dyad assay

To capture the vocal repertoire of rodents when they are close to one another and reliably identify which individual in a dyad produced each vocalization, we designed a ''two-enclosure'' behavioral assay which leverages partial acoustic isolation and two microphones (Figure 1). We refer to this assay as ''PAIRId'' (Partial Acoustic Isolation Reveals Identity). The enclosures were custom-built transparent acrylic rectangular boxes (outer dimensions: 12 × 12 × 18 inches, wall thickness: 0.25 inches; shopPopDisplays, Woodland Park, NJ). One face of each box was drilled with nine 0.25-inch holes using cutting drill bits: five holes positioned 2 inches above the inner floor, and four additional holes offset from the top row, 1.5 inches above the inner floor. A removable floor, covered with AlphaPad bedding (Shepherd Specialty Papers, Watertown, TN), was inserted into each enclosure to facilitate cleaning between sessions. Inside of a controlled acoustic environment box, we placed an aluminum breadboard designed to fit two of these enclosures and maintain a.932 cm gap between the outer dimensions of each. Considering the thickness of the two inner walls, the effective division between two rodents is 2.2 cm. The perforated faces of the acrylic boxes were facing the other across this gap, allowing the mice to acoustically interact with one another. To ensure proper ventilation, we fed external air into each box via tubing. To the breadboard we also attached two custom assembled cranes such that a microphone and a video camera could be lowered into each enclosure to the consistent height of 12 in between trials (crane materials from 80/20 Inc., Columbia City, IN). A third microphone was placed in the chamber and set to low gain to record the enclosure. Finally, when each enclosure contained their rodent, microphone, and camera, we inserted custom-cut 2-inch-thick polyimide acoustic foam into the opening (Soundfoam HTC, The Soundcoat Company, Deer Park, NY). The foam was cut to the outer perimeter of the acrylic box and lightly compressed to fit the inner perimeter, creating acoustic dampening between the enclosures. Each enclosure contained an Avisoft CM16/CMPA microphone, with a third microphone positioned outside the two enclosures. All three microphones were powered and recorded by an Avisoft UltraSoundGate 416H device, saving WAV files at a 250 kHz sampling rate using Avisoft-RECORDER software. Each enclosure also housed a FLIR Blackfly USB camera (BFS-U3-20S4C-C, Teledyne FLIR, LLC) equipped with a 4.5 mm fixed-focal-length lens (C Series #86-900, Edmund Optics). Frames were captured at 50 frames per second, triggered by a custom-programmed Arduino Mega 2560 R3. The same Arduino controlled the start and stop triggers for audio recording, ensuring synchronized audio and video.

PAIRId performance control experiment

To evaluate the vocal attribution performance of PAIRId, we conducted experiments with one male mouse paired with an experimentally muted female (4 unique males, 1 stimulus female across 4 sessions). The female was muted by bilateral expression of tetanus-toxin light chain (TeLC) in the clPAG as in the viral delivery & cannula implantation section, which abolished all vocal output. Muting was confirmed by continuous recording and spectrogram inspection. To account for potential side biases in the microphones or apparatus, we alternated the intact males' position between the left and right sides of the PAIRId setup across sessions. Under these conditions, all vocalizations were expected to originate from the male, providing a benchmark against which to test PAIRId's attribution performance.

"Alone" assay

To capture the repertoire of a singing mouse in isolation while accounting for the novelty of the PAIRId enclosure, we used clear acrylic boxes with the same footprint as the PAIRId enclosure but with a shorter height (10 in) to fit inside soundproof acoustic-foam-lined MedAssociates cabinets (Fairfax, VT). Each box had a wire lid and was equipped with both high- and low-gain microphones outside but near the box. Audio was recorded using an Avisoft UltraSoundGate 416H device, saving WAV files at a 250 kHz sampling rate via Avisoft-RECORDER software.

Alone-social comparison experimental timeline

In two cohorts of 3 males and 2 females each (total of 6 males and 4 females aged 4-11 months), each combination of opposite sex dyads was subjected to the following experimental timeline: On Day 0, each singing mouse was removed from its home enclosure and placed in a short acrylic ''alone'' box to acclimate overnight. On Day 1, the singing mice were recorded individually in the alone boxes for 5 hours. After recording, they were transferred to the PAIRId enclosures, where they acclimated to the experimental setup in acoustic isolation from one another overnight. On Day 2, at approximately the same time, the two PAIRId enclosures were placed next to each other in the PAIRId setup, allowing the singing mice to interact while being recorded for 5 hours as described above. Following this, the session was complete, and the singing mice were transferred back to their home cages. All 12 combinations of opposite sex dyads sessions were recorded over the course of nine days on a schedule that allowed at least 24 hours of time spent

in home cage between sessions for individual singing mice. This design resulted in a dataset with each male represented in two sessions, one with each female of its cohort, and each female in three sessions, one with each male.

Laboratory mouse dyad experiment

To compare acoustic parameters of lab mice with those measured in singing mice, we re-analyzed the *Mus musculus* male-female dyad dataset from Isko et al.[97] In this dataset, a cohort was selected of three male and three female C57Bl/6J mice aged ∼two months old (58-70 days) such that the sexes were not littermates. All mice were singly housed and isolated for at least nine days before social exposure to increase the probability of vocalization.[108] In advance of recording, the female mouse was placed into a clean cage (Thoren Systems #8, Worcester, MA; 30.80 × 40.60 × 22.23 cm) lined with clean Alpha-pad cotton paper. After a period of acclimatization (min: 30 mins, max: 12 hours), a male was introduced to the cage with the female. Audio of the pair was recorded for 1 hour using two Avisoft UltraSoundGate 116H devices and two Avisoft CM16/CMPA microphones (with high and low gains), synchronized via a custom external trigger, with WAV files written with a 250 kHz sampling rate using Avisoft-RECORDER software. Female laboratory mice rarely vocalize in male-female dyads during social interactions of this length.[23,109,110] Consistent with this, visual inspection of spectrograms revealed no apparent instances of overlapping USVs that would indicate that the male and female mouse were vocalizing at the same time.

Surgical Procedures

Mice subjected to surgery were placed into an induction chamber with 1-2% isoflurane. Hair was clipped from the surgical site. The mouse was then placed onto a heating pad on a stereotaxic instrument (Kopf model 940, Tujunga, CA). The mouse's front teeth were latched onto a bite bar, the head secured with non-rupturing ear bars, and the head levelled. Nonsteroidal anti-inflammatory drug meloxicam was administered subcutaneously at 5 mg/kg. Following these preparations, the mouse was subjected to either 1) implantation of a thermistor, 2) injection of viral vector followed by implantation of an optogenetic fiber, or 3) injection of viral vectors.

Thermistor implantation

Chronic implantation of an intranasal thermistor is a well-established method for estimating respiration in rodents.[111] As obligatory nose-breathers, a rodent's nasal cavity warms during exhalation and cools when room-temperature air is inhaled due to the difference of a rodent's internal temperature and experimental conditions. We implanted thermistors, adapted from McAfee et al.[111] Briefly, after preparation (see above), a midline incision was made over the skull, to the anterior edge of the nasal bone. A cavity for the thermistor was opened by drilling the nasal bone (A/P 3.1mm, M/L 0.5mm from nasal suture), and the thermistor implanted within. The thermistor was sealed using Kwik-Cast silicone sealant, and the implant was secured to the skull with layers of Vitrebond, Metabond, and dental acrylic.

Viral delivery & cannula implantation

Adeno-associated viruses (AAVs) originally developed for laboratory mice also infect neurons and express their packaged cargo in singing mice. In this study, we leverage this to 1) activate neurons optogenetically using channelrhodopsin and 2) silence neuronal synaptic transmission using tetanus toxin light-chain (TeLC). We first identified the caudolateral PAG region of singing mice as 4.2 mm posterior and 0.6 mm lateral relative to bregma and 2.3 mm ventral from the brain's surface. Craniotomies were made using a dental handpiece and an FG carbide burr (Dentsply Sirona Midwest Tradition TL, Dentsply Sirona, Charlotte, NC). Viruses were injected using a Nanoject III (Drummond Scientific) at 2 nL/cycle with a 10 second interval. For the TeLC silencing experiments, five male singing mice (5-12 months old) were injected bilaterally in the clPAG with 80 nL of a 1:1 mixture of AAV2/DJ-hSyn-flex-TeLC-eYFP (Addgene #135391, custom packaged by WZ Biosciences) and AAV2/9-pENN.AAV.CamKII 0.4.Cre.SV40 (Addgene #105558). Three female singing mice also received the viral injections for TeLC-mediated silencing of the clPAG and were used as stimulus mice in the PAIRId accuracy control experiment and the vocal-respiratory coordination experiment. For the optogenetic activation experiments, four male singing mice (6-11 months old) were injected unilaterally in the clPAG with 150 nL of a 1:1:2 mixture of AAV2/9-EF1α double-floxed-ChR2-mCherry (Addgene #20297), AAV2/9-pENN.AAV.CamKII 0.4.Cre.SV40 (Addgene #105558) and sterile saline before fiber implantation. In the same surgery following this injection, an optogenetic cannula with a tapered tip (Optogenix,.39/200, active length 0.5mm, implant length 3mm, cLCF) was implanted and secured to the skull using Metabond. Subsequently, a headbar was also implanted and secured using Metabond, and the entire implant was protected by dental acrylic.

Histology

Mice were transcardially perfused with PBS followed by 4% paraformaldehyde (PFA), after which brains were dissected and post-fixed in 4% PFA overnight before being stored in PBS. Brains were sectioned into 100 μm coronal slices using a vibratome. To visualize tissue structure, select slices were stained with NeuroTrace 435/455 (Thermo Fisher Scientific, N21479) at a 1:30 dilution following the manufacturer's protocol. Stained slices were mounted on glass slides using ProLong Gold Antifade mounting medium (Thermo Fisher Scientific, P36930) and imaged with an epifluorescence microscope. Injection extent and cannula placement were verified histologically, using reporter spread or tissue damage relative to anatomical landmarks (Figure S4).

Vocal-respiratory coordination experiment

To estimate a singing mouse's respiration while vocalizing its full vocal repertoire, we implanted an intranasal thermistor. We used a muted female singing mouse as a stimulus to elicit the male's vocalizations, ensuring all recorded vocalizations originated from the implanted male. The stimulus female was muted, as in the viral delivery & cannula implantation section by bilaterally expressing TeLC in the clPAG via viral delivery. During an experiment, the thermistor of the implanted mouse was first connected to an overhead rotary joint (Adafruit, #736) and then routed into a custom-built amplifier circuit. The implanted male was recorded alone or with the presence of a singing mouse in a custom transparent acrylic cylindrical enclosure (12 inch diameter, 12 inch tall). The amplified thermistor signal was recorded via an Intan RHD 1024ch Recording Controller (Intan Technologies, Los Angeles, CA). Vocalizations were

captured using two Avisoft UltraSoundGate 116H devices and two Avisoft CM16/CMPA microphones, synchronized via a custom external trigger, with WAV files written with a 250 kHz sampling rate using Avisoft-RECORDER software. Synchrony with the thermistor signal was achieved via recording a copy of the trigger-on signal with the Intan RHD recorder.

Laryngeal phonation mechanism experiment

To determine the laryngeal phonation mechanism of the singing mice vocal repertoire, we recorded vocalizations from four male-female dyads (7-20 months, older singing mice freeze less and vocalize quickly after disturbance by an experimenter) in both air and heliox (80% He, 20% O 2). Each dyad was placed in a Thoren Systems #8 enclosure with a removable acrylic floor covered with clean AlphaPad bedding, beneath which a perforated clear PVC tube connected to the heliox tank was positioned. The enclosure was housed inside a MedAssociates cabinet lined with acoustic foam to facilitate heliox accumulation and acoustic isolation. The singing mice were allowed to vocalize for 45–60 minutes before heliox was introduced at a flow rate of 5 L/min, for an additional 45-60 minutes. Vocalizations were captured using two Avisoft UltraSoundGate 116H devices and two Avisoft CM16/CMPA microphones, synchronized via a custom external trigger, with WAV files written with a 250 kHz sampling rate using Avisoft-RECORDER software.

Optogenetic activation experiment

Two weeks following the surgery, clPAG neurons were optogenetically activated with blue light from a 473 nm laser with tonic light stimulation for 1, 2, or 4 seconds. The laser intensity was adjusted for each mouse to reliably elicit vocalizations—ranging from 0.4 to 1 mW at the fiber tip—before being fixed for experiments. Vocalizations were captured using two Avisoft UltraSoundGate 116H devices and two Avisoft CM16/CMPA microphones, synchronized via a custom external trigger, with WAV files written with a 250 kHz sampling rate using Avisoft-RECORDER software. Copies of the external synchronization trigger signal and the laser signal were recorded on an Intan RHD 1024ch Recording Controller (Intan Technologies, Los Angeles, CA) for synchronization.

Synaptic silencing experiment

In each experiment, a male-female singing mouse dyad was allowed to interact in the PAIRId setup for 24 hours as a ''pre-injection'' time point. After the baseline recording, the male of each dyad was subjected to injection of a virus mixture to silence neurons in the clPAG via tetanus toxin light chain. Following the injection and recovery from anesthetic on a heating pad (usually within a half hour), the perturbed singing mouse male was placed back into the PAIRId assay with its stimulus female and continuously recorded for four or five days.

Analysis

Detection of vocalizations

To analyze vocalizations, we first segmented biotic sounds from silence in audio files, a necessary step for vocal analysis in both PAIRId and other paradigms. We used a modified version of USVSEG software (usvseg09r2),[106] which implements a signal processing algorithm for the detection of typically quiet rodent sounds from background noise. Briefly, USVSEG extracts vocalization events by generating a stable spectrogram using the multitaper method, flattening it in the cepstral domain to remove noise, applying thresholding, and estimating onset/offset boundaries. This robust, species-agnostic software allowed us to adjust parameters to suit the acoustic profiles of singing mice specifically. We modified the open-source software slightly: to improve inter-file consistency, we adjusted the threshold calculation for detecting biotic sounds from a noise-based standard deviation per file to a fixed value optimized per setup. USVSEG performed well for detecting quiet vocalizations; however, the loud vocalizations emitted by singing mice were more reliably segmented using a custom Python-based method optimized to handle reverberations. This method segmented loud notes based on the signal-to-noise ratio in acoustic power, calculated from a spectrogram. A rolling estimate of background noise was used to dynamically adjust the noise threshold for segmentation. The detections herein were merged with those of USVSEG, with redundancies removed by giving priority to the detected loud notes. For data recorded in the PAIRId social assay, segmentation was performed for the left and right microphones separately before assignment. For the ''alone'', heliox, thermistor, optogenetic activation, and lab mouse dyad experiments that did not require individual assignment, these data would next be curated.

Assignment of vocalizations in PAIRId

To determine the source of each vocalization in the PAIRId social assay, we compared detections from both enclosures. If a detection occurred on one side without an overlapping detection on the other, the source was assigned to that side. For overlapping detections, we distinguished between simultaneous vocalizations (''coincidence'') and acoustic bleed-through. In cases of coincidence, the spectro-temporal shapes of the vocalizations differ between channels; in contrast, similar shapes indicate bleed-through. To extract the spectro-temporal shapes, we segmented the vocal fragments on each side by thresholding the spectrograms. If a fragment appeared in only one channel (consistent with coincidence), that channel was assigned as the source of that fragment. If the fragment appeared in both channels (consistent with bleed-through), the source was assigned to the louder channel or marked as unknown if there was little difference in acoustic power. After assigning the vocal fragments, we joined fragments on each side into notes respectively. Finally, we combined the non-overlapping notes and notes assigned from overlaps to produce the final output. See Figure S1A for a visual overview.

Quantification of the performance of PAIRId

We first quantified the performance of the PAIRId setup by evaluating its assignment step. The final output combines non-overlapping notes—which reflect the hardware's partial acoustic isolation performance—with notes assigned from overlapping detections using the algorithm. For each session, we calculated the percentage of notes assigned solely by the hardware, the percentage of total notes assigned by the full attribution workflow, and the percentage assigned by human annotation (Figure 1C). Next, we evaluated

the accuracy of the full analysis pipeline by benchmarking its performance against manual annotation in four hours from highly vocal sessions in the intact, social dataset (see Figures S1B–S1D). Additionally, we evaluated a highly vocal hour from each session of the control experiment where an intact male interacted with a muted female (via TeLC-mediated synaptic silencing of the clPAG) within PAIRId (see Figures S1E–S1G). During these sessions, we assessed the pipeline's performance by quantifying true positives, false positives, and false negatives. False positives for a given channel are defined as a PAIRId detection that lacks a match in human annotation "ground truth" for either that channel or unassigned (ambiguous channel assignment). False negatives for a given channel are defined as a human annotation "ground truth" that lacks a match in PAIRId detections for either that channel or unassigned (ambiguous channel assignment). Using these definitions, we computed precision, recall, and the F1 score for the annotated hours (Figures S1D and S1G).

Curation of vocalizations

Detections were manually curated using a customized spectrogram browser adapted from the open-source MATLAB graphical user interface DeepSqueak.[112] The browser was modified to display, edit, and export the associated detections for two aligned audio files. Curation involved correcting biotic sound boundary errors and removing abiotic false positives. Particular attention was given to correcting the boundaries of quiet but rapidly emitted vocalizations, which are especially challenging to segment automatically. PAIRId assignment detections were further reviewed to resolve "unassigned" vocalizations wherever possible. Four curated hours of PAIRId data with high vocal activity and different individual mice were designated as "ground truth" for evaluating assignment methods. Every hour from each dataset was curated, except for the long-term TeLC pre- and post-perturbation recordings, where only select hours were curated for the focal mouse.

Characterization of vocalizations

Vocalizations were characterized at both the individual note and temporal patterning levels. At the note level, several acoustic features were computed, including note duration, note amplitude, and pitch. Note duration was directly extracted from the detections. Note amplitude was determined as the peak amplitude within the 10 kHz to 120 kHz frequency range; when applicable, measurements from both high- and low-gain microphones were used. For the laryngeal phonation mechanism (heliox) experiments, the fundamental frequency was further quantified. For each note, a human annotator visually inspected the spectrogram and identified the lowest continuous trace and recorded the highest pitch of that trace. At the temporal patterning level, we first computed note-to-note temporal features. These features included the inter-event interval (the time between the end of the current event and the start of the next event), the log-transformed inter-event interval, the inter-start interval (the time between the start of the current event and the start of the next event), and the instantaneous rate (defined as the reciprocal of the inter-start interval). Next, to identify supra-note patterns of arbitrary length, we applied the following heuristics. First, a fixed-size 7-note sliding window was applied across all vocalizations. Within each window, two metrics were computed: the maximum log-transformed inter-event interval (representing the largest temporal gap) and the normalized root mean square error (NRMSE) from a linear regression of instantaneous rate versus note index. A window was deemed temporally patterned if it exhibited both a low maximum log inter-event interval ($\leq$100 ms), indicating temporal contiguity, and a low NRMSE ($\leq$0.15), reflective of stereotyped note sequencing. Consecutive windows meeting these criteria were merged into longer segments - either through direct overlap or, when non-overlapping, if the time gap between them did not exceed 0.5 seconds - to form supra-note patterns. Each resulting segment was then further characterized using RANSAC regression to robustly fit a linear model that accounted for outliers, thereby enabling the identification of long patterns with simple linear stereotypy. Once long patterns were delineated, we extracted key patterning features - specifically, the maximum rate (*r_max*), minimum rate (*r_min*), and the number of notes from the RANSAC fit. These features were subjected to principal component analysis (PCA) for dimensionality reduction to 2D, and a Gaussian mixture model (GMM) was then applied to cluster the segments into three distinct groups. Each note was subsequently assigned a category based on whether it belonged to a segment in one of the three patterning clusters or was unpatterned. Finally, by examining both note amplitude and spectrograms, we defined the four categories: songs, semisongs, and patterned and unpatterned USVs (combined into a single category: USVs). See Figure S2 for visuals accompanying these methods. To further characterize the temporal properties of songs and USVs, we also computed an alternative, complementary method for defining vocal bouts. Vocal bouts were defined as sequences of notes with inter-event intervals shorter than 100 ms, consistent with the definition of "group" in Castellucci et al.[33] Using this criterion, we identified song bouts and USV bouts and extracted their durations.

A mathematical model of the song rhythm

The song is composed of a series of progressively longer notes that evolve predictably over 6–10 seconds. We derived a simple mathematical model of the stereotyped temporal patterning of the song. We observed that the instantaneous note rate $r(i)$ decreases linearly with the note index i, and postulated the linear model:

$$r(i) = m(i-1) + r_{\max}, \text{for } i = 1, 2, \ldots, N-1.$$

where

- $r_{\max} = r(1)$ is the maximum instantaneous rate (i.e. the start rate),
- $m < 0$ is the slope (indicating a decrease in rate as i increases),
- N is the number of notes.

Then, we also defined the minimum rate by extrapolating this linear relationship to the last note:

$$r_{\min} = r(N) = m(N-1) + r_{\max}.$$

Thus, the three parameters *r_max*, *r_min*, and *m* fully characterize the temporal patterning of a song. From these three parameters, we next derived the total duration of the song *T*, a global property of the song. The total duration *T* (i.e., the time difference between the first and last note) is given by

$$T = \sum_{i=1}^{N-1} \frac{1}{r(i)}$$

To obtain an analytic expression, we approximated this sum by the integral

$$T = \int_1^N \frac{1}{r(i)} di = \frac{1}{m} \ln\left(\frac{r_{\min}}{r_{\max}}\right)$$

Since $m < 0$ and $r_{\min}/r_{\max} < 1$, the logarithm is negative, and division by the negative *m* yields a positive duration *T*. Alternatively, one may write:

$$T = \frac{1}{|m|} \ln\left(\frac{r_{\max}}{r_{\min}}\right)$$

Thus, this model enabled us to analyze song variability (e.g., duration) in terms of these three interpretable parameters.

Video analysis

Using the overhead camera synchronized with audio in the PAIRId social assay enclosures, we quantified inter-animal distance for each dyad. We preprocessed videos of each enclosure by correcting the lens distortion using calibrated camera parameters. To estimate pose, we generated a SLEAP model using 950 manually labeled training frames and the multi-animal top-down pipeline with a single instance.[107] The model estimated a skeleton with 6 points (nose, each ear, back of head, middle of spine, base of tail). The output nodes were then converted from pixel space in each video to common world coordinates. The middle-of-spine node was considered the centroid of each animal and used to calculate inter-animal distance. See Figure 2A for a visual overview.

Analysis of USV usage in PAIRId

To examine whether USVs were produced during social engagement, we identified the locations of USV production by mapping mouse positions in common world coordinates obtained from video analysis. To further assess temporal engagement, we analyzed the correlation of USV bouts over time. We restricted analyses to hours of robust vocal activity, defined as those with more than 1,000 USVs per hour based on the distribution across sessions, which yielded 8 high-activity hours. Within these hours, we computed USV vocalization rates for each side using a 30-second rolling window with a 15-second step size, and then calculated the Pearson correlation coefficients between the left and right rates. As a control, we generated shuffled datasets by pairing left-side rates from one hour with right-side rates from different hours (e.g., hour 1 left vs. hour 2 right). This preserves the temporal structure within each hour while disrupting the correspondence between the two partners, providing a null distribution to assess whether they were vocally coordinated.

Respiration analysis

To identify respiratory events, we analyzed the signal from the intranasal thermistor. The raw signal was first downsampled to 500Hz and then filtered using a 4th order Butterworth bandpass filter (0.5 - 50Hz). Inhalation onsets were detected as prominent peaks (high temperature), while offsets (marking the start of exhalation) were identified as corresponding troughs (low temperature). Putative respiratory cycles were defined by pairing each inhalation onset with the unique subsequent inhalation end occurring between consecutive onsets. Cycles with a duration shorter than 0.5 seconds were considered valid and retained for downstream analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical comparisons were non-parametric (Mann–Whitney U test, $\alpha = 0.05$). All analyses were conducted in Python. Key findings are highlighted in the results section; full statistical reporting is provided in the figure legends. Error bars represent standard error of the mean.