Parametric Modulation of a Shared Midbrain Circuit Drives Distinct Vocal Modes in a Singing Mouse

Xiaoyue Mike Zheng[1,2,3], **Clifford E. Harpole**[1,3], **Martin B. Davis**[1], **and Arkarup Banerjee**[1,2] ✉

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[2]Cold Spring Harbor Laboratory School of Biological Sciences, Cold Spring Harbor, NY, USA
[3]These authors contributed equally

Neural circuits capable of generating multiple outputs are essential for behavioral flexibility, yet their organizational principles remain poorly understood. Using vocal communication in singing mice (*Scotinomys teguina*), we investigated whether distinct vocal behaviors are controlled by separate pathways or by shared circuits operating under different parametric regimes. We developed a novel behavioral assay (PAIRId—Partial Acoustic Isolation Reveals Identity) that enables precise attribution of vocalizations during social interactions in singing mice. This approach revealed two major vocal modes: loud, temporally patterned songs used for long-distance communication and soft, unstructured ultrasonic vocalizations (USVs) employed during close-range interactions. Despite their dramatic acoustic and contextual differences, both vocal modes share peripheral sound production mechanisms and central neural control by the caudolateral periaqueductal gray (clPAG). We derived a simple mathematical model describing song rhythm as a linear progression of note rates, which captures song motor patterning with just three parameters and accurately predicts song duration across animals and conditions. Using this model, we demonstrate that progressive silencing of clPAG neurons systematically alters specific song parameters before eliminating all vocalizations. Notably, one of these parameters—which controls song termination—also accounts for natural sexual dimorphism in song production. Our findings reveal how differential amplitude and frequency modulation of shared neural circuits produces categorically distinct behavioral outputs and provide a mechanistic basis for how behavioral innovations can emerge through evolutionary tinkering of ancestral neural pathways.

Correspondence: ***abanerjee@cshl.edu***

Introduction

A fundamental question in neuroscience is whether diverse behaviors emerge from separate dedicated neural pathways or from parametric modulations within a common circuit. This becomes especially critical in the case of social behaviors, where animals may produce categorically different behaviors depending upon the context. Specific behaviors can be produced by dedicated motor circuits akin to labeled lines in sensory systems (1–3). In contrast, different behaviors may be instantiated through shared circuits capable of operating in different functional regimes (4–7). Determining the mechanisms by which multifunctional circuits implement alternative behavioral modes has significant implications for understanding both neural circuit architecture, function and evolution. For instance, shared circuits operating in multiple parametric regimes might offer a more efficient solution towards behavioral innovation than the deployment of entirely new neural pathways.

Vocal communication, which requires complex respiratory, laryngeal, and orofacial motor systems across different behavioral contexts, provides an ideal framework for investigating this circuit-level question (8–12). Vocalizations are particularly well-suited for such studies because they represent discrete, quantifiable outputs with measurable acoustic parameters that can be precisely characterized across different behavioral contexts (11, 13). Singing mice (*Scotinomys teguina*) — a highly vocal Neotropical rodent — is an attractive model system to explore the function and evolution of neural circuits underlying vocal behaviors in mammals (**Figure 1A, Video 1**)(14–17). Known for its distinctive songs that are used for vocal turn-taking, we have previously shown that this behavior is dependent upon motor cortical function using a combination of electrical stimulation, focal cooling, pharmacological silencing, and chronic silicon-probe recordings (18, 19). While this work established cortical contributions to vocal interactions, 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 (20, 21). As we will show, the midbrain PAG of the singing mouse provides an ideal testbed for addressing whether and how functionally distinct vocal behaviors are generated by separate dedicated pathways or parametric modulation of a common circuit—a fundamental question about neural circuit organization with broader implications for behavioral flexibility.

In this study, we address several key 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? By combining a novel behavioral paradigm with acoustic analysis, mathematical modeling, peripheral sound production experiments, targeted gain-of-function and loss-of-function manipulations of the cau-

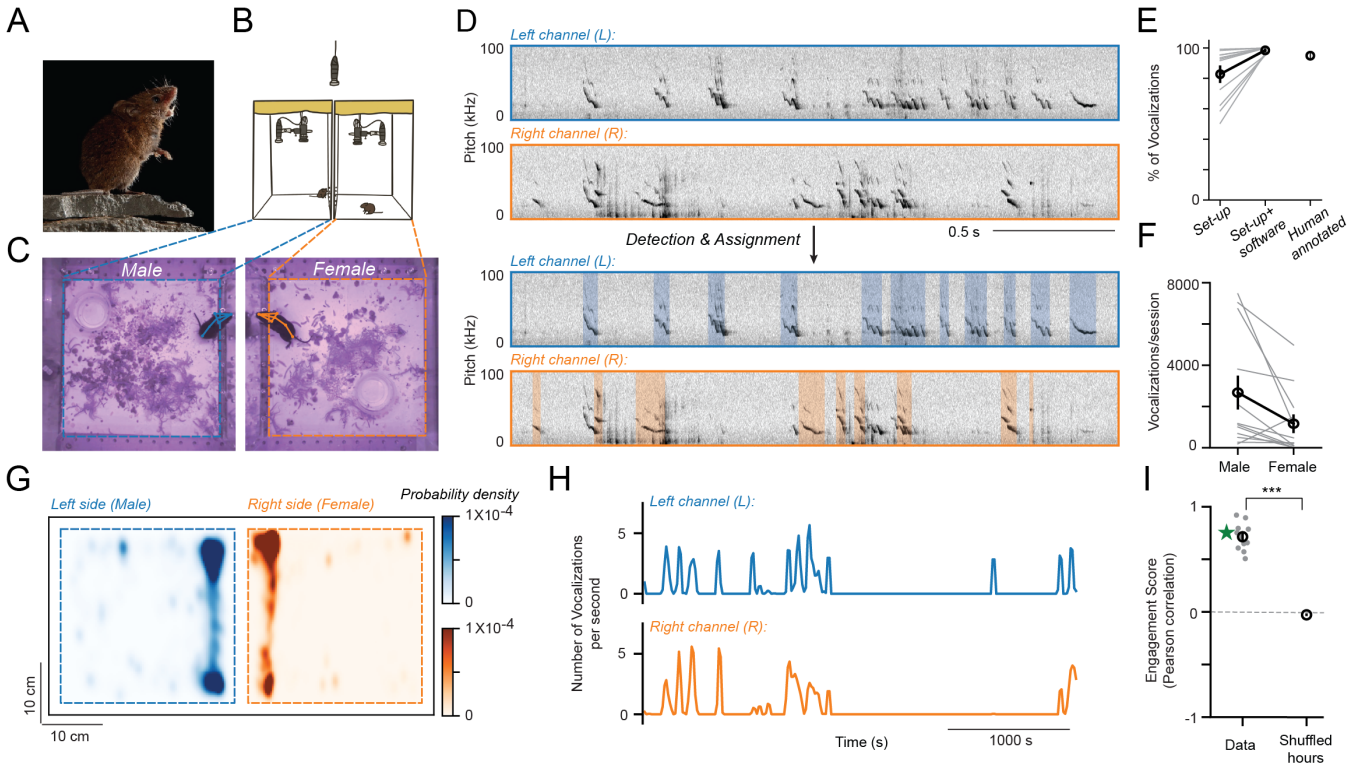

Fig. 1. PAIRId: a behavioral paradigm for identifying vocalizations of individual mice during social interactions. (A) An adult Alston's singing mouse (*S. teguina*). Photo: Christopher Auger-Dominguez. **(B)** Schematic of the PAIRId behavioral paradigm, where two mice housed in separate acoustically dampened enclosures—each equipped with its own microphone and camera—interact across a perforated plane. **(C)** Example video frames from both enclosures of the PAIRId paradigm. **(D)** Example audio data from the PAIRId paradigm before and after detection and assignment of vocalizations. *Top*: spectrograms of the two-channel audio inputs. *Bottom*: processed audio with vocalizations attributed to individual mice. **(E)** Proportion of vocalizations attributed to individual mice using hardware acoustic isolation (82.8 ± 5.0%) and the full analysis pipeline (98.4 ± 0.6%, n = 12 sessions of 5 hours each), compared with human annotations (94.9 ± 1.7%, n = 4 hours in separate sessions). **(F)** Number of vocalizations in PAIRId by the male (2673.3 ± 818.0, n = 12 sessions of 5 hours each, 6 animals) and the female (1162.3 ± 451.0, n = 12 sessions of 5 hours each, 4 animals). **(G)** Density of mouse locations (centroid) during vocalizations (n = 34,877 notes in 11 hours of high vocal engagement from 4 males and 3 females). **(H)** Rates of vocal production of two interacting mice (*top*: male; *bottom*: female) in an example hour with high vocal engagement. Epochs of vocal production were correlated between the two animals (Pearson correlation = 0.75). **(I)** Summary of Pearson correlations between the vocal rates of interacting mice during periods of high vocal engagement. The correlation coefficient for data (0.71 ± 0.04, n = 11 hours) is significantly higher than that of the shuffled control (P = 5.03e-8). Green star indicates the example hour in (H). Unless stated otherwise, values reported are Mean ± sem and hypothesis testing was performed using Mann-Whitney U test.

dolateral periaqueductal gray (clPAG), we demonstrate that singing mice use shared neural circuits for phonation operating under different amplitude modulation (AM) and frequency modulation (FM) regimes to generate categorically distinct vocal behaviors. Our findings provide a mechanistic basis for how behavioral innovations can emerge through evolutionary tinkering of ancestral neural pathways.

Results

PAIRId: A New Behavioral Assay for Individual Vocalization Attribution. Accurately attributing vocalizations to individual animals during social interactions is essential for understanding neural mechanisms of vocal communication. This challenge has prompted the development of various approaches, including wearable miniature microphones and microphone arrays with neural network analysis (22–24). We developed a complementary behavioral assay called PAIRId (**P**artial **A**coustic **I**solation **R**eveals **Id**entity), where two animals are placed in separate acoustically dampened enclosures equipped with private microphones and cameras while interacting across a perforated plane (**Figure 1B–C**, **Figure S1A** for video pipeline, **Video 2**).

The PAIRId paradigm leverages both hardware-level acoustic isolation and software-level signal processing to identify vocalizations of individual animals. Briefly, by detecting vocalizations in each channel and analyzing non-overlapping signals, we can directly assign the majority of vocalizations to their sources (82.8 ± 5.0%). For the remaining temporally overlapping vocalizations, we further compare spectrotemporal properties to disambiguate between coincident vocalizations and acoustic bleed-through (**Figure 1D–E**, **Methods**, **Figure S1B** for audio processing pipeline). This approach allows for the attribution of almost all vocalizations (98.4 ± 0.6%, 12 male-female pairs) with high accuracy (F1 score = 0.87 ± 0.02, 4 hours), comparable to human ground-truth annotations (**Figure 1E**, **Figure S1C**) and other available methods (22–24).

In our behavioral paradigm, male and female singing mice interact robustly, producing numerous vocalizations within a few body-lengths of each other (**Figure 1F–G**). These vocalizations form correlated bouts when both mice are engaged (**Figure 1H**), with each mouse showing stronger correlation with its actual partner (**Figure 1I**). PAIRId thus provides a robust platform for characterizing the rich vocal repertoires of individual mice during social encounters, establishing a

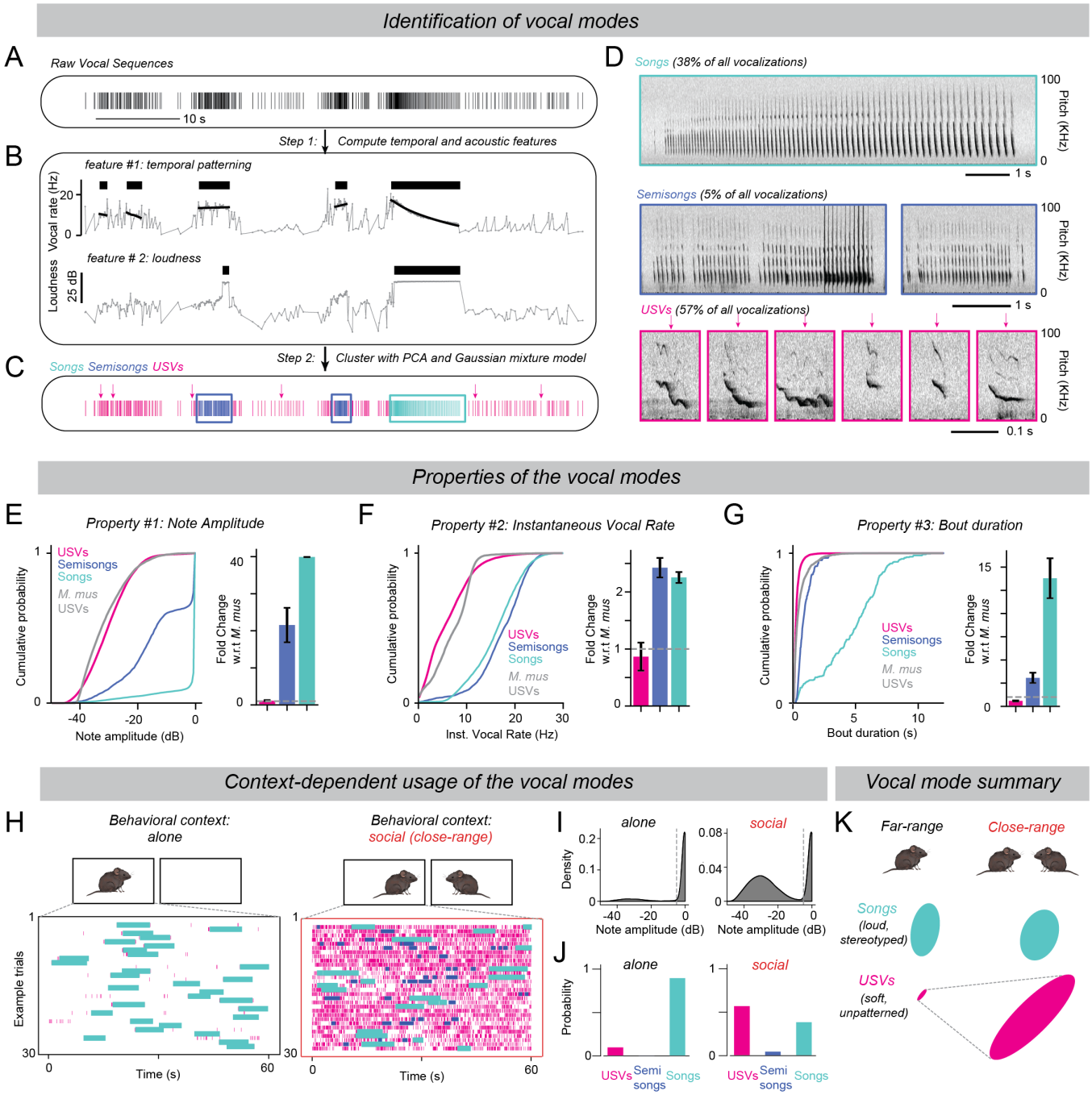

Fig. 2. Singing mice produce an extended repertoire of vocal modes during close-range social interactions (A) Raster plot illustrating vocalizations of a mouse in the PAIRId paradigm during an example 60-second period, with each line marking the onset of a vocalization. **(B)** Vocal sequences were characterized by substantial variation in temporal stereotypy (*top*) and loudness (*bottom*). *Top*: Temporal patterning features were extracted by analyzing the instantaneous vocal rate, with the black bar indicating long, stereotyped vocal sequences. *Bottom*: Amplitude features were computed for the vocalizations, with the black bar highlighting those with high loudness. **(C)** The same raster plot as in (A), with lines color-coded by the vocal modes (songs, cyan; semisongs, blue; USVs, magenta) identified by combining temporal stereotypy and loudness features. **(D)** Example spectrograms for each of the three identified vocal modes—songs, semisongs, and USVs—highlighted in (C). To account for the large loudness difference among vocalizations, the song spectrogram depicted here was computed from low-gain audio recording whereas the semisong and USV spectrograms were computed from high-gain audio recording. **(E)** 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 semisong: n = 2,100 notes from 9 mice, -13.06 ± 0.25 dB; singing mouse song: 17,681 notes from 10 mice, -2.27 ± 0.05 dB; lab mouse USV: -30.82 ± 0.09 dB). *Right*: Fold change of median note loudness per mouse relative to the median of lab mouse USVs (singing mouse USV: 1.1 ± 0.2, $P > 0.05$; singing mouse semisong: 21.6 ± 4.7, $P = 0.009$; singing mouse song: 39.9 ± 0.0, $P = 0.007$). **(F)** Same as (E), but for instantaneous note rate (singing mouse USV: 6.26 ± 0.03 Hz, fold change 0.9 ± 0.2, $P > 0.05$; singing mouse semisong: 17.24 ± 0.11 Hz, fold change 2.4 ± 0.2, $P = 0.009$; singing mouse song: 16.16 ± 0.04 Hz, fold change 2.3 ± 0.1, $P = 0.007$; lab mouse USV: 7.24 ± 0.05 Hz). **(G)** Same as (E), but for vocal bout duration (singing mouse USV: 0.31 ± 0.00 s, fold change 0.6 ± 0.1, $P = 0.007$; singing mouse semisong: 0.96 ± 0.05 s, fold change 3.1 ± 0.6, $P = 0.009$; singing mouse song: 4.64 ± 0.15 s, fold change 13.8 ± 2.2, $P = 0.007$; lab mouse USV: 0.56 ± 0.02 s). **(H)** Raster plots of 30 segments with the highest vocal activity when the mouse is alone (*left*) vs. during close-range social interactions in the PAIRId (*right*). Each row represents one minute, and each line indicates a vocalization, color-coded by its assigned mode. **(I)** Distribution of loudness of all vocalizations when mice are alone (*left*, n = 23,028 notes) versus when they interact socially in the PAIRId (*right*, n = 46,027 notes). **(J)** Distribution of vocal modes of all vocalizations across these two behavioral contexts (same as in (I)). **(K)** Cartoon depiction of the vocal repertoire in different social contexts. When alone, singing mice primarily produce songs, whereas during close-range social interactions they exhibit an expanded repertoire that includes both songs and USVs.

foundation for exploring underlying neural mechanisms.

Discovery of Distinct Vocal Modes. PAIRId revealed the richness of temporally sequenced streams of vocalizations that are produced by individual singing mice (**Figure 2A**). We quantified vocal sequence organization by hierarchically identifying patterns of temporal stereotypy—where mice produce notes at steady or smoothly varying rates (**Figure 2B** ***top***, **Methods**, **Figure S2A**). Stereotypy was measured by how well the instantaneous vocal rate (defined as the inverse of inter-note intervals) was explained by a linear model. As a first step, we identified locally contiguous vocalizations with high degree of stereotypy (**Figure S2C**). Next, such overlapping units were merged to form longer vocal sequences. Finally, unsupervised clustering of these resultant sequences using a gaussian mixture model revealed sub-clusters with distinct temporal patterns (**Figure S2B–E**). Complementary to the temporal dimension described above, we also observed substantial differences in loudness of individual notes (**Figure 2B** ***bottom***). Combining these loudness and temporal characteristics, we were able to define two major vocal modes—songs and ultrasonic vocalizations (USVs)—that together comprise 95% of the vocalizations (**Figure 2C, Methods**). In addition, we observed a rarer vocal mode (semisongs, 5% of vocalizations). Therefore, singing mouse vocal repertoire and sequences are much richer than previously anticipated.

Having established these vocal modes, we wondered how the singing mouse vocalizations compare with those of other rodents. We used laboratory mouse USVs as a comparative reference point due to their thorough characterization in the literature (25–27). Singing mouse USVs share acoustic properties with lab mouse USVs—including lower amplitude, lower temporal stereotypy, shorter bouts (**Figure 2E–G** ***magenta***), and variable note shapes (**Figure 1D** ***magenta***, **Figure S2E**). In contrast, songs are much louder, have faster tempos, and contain many more stereotyped notes over longer durations (**Figure 2E–G** ***cyan***). Acoustic properties of semisongs were found to be intermediate between songs and USVs (**Figure 2E–G** ***blue***). We conclude that songs possess acoustic characteristics in both amplitude and temporal domains that are categorically distinct from the USVs of both species.

This acoustic distinction is reflected in context-dependent usage. In solo contexts, mice produced songs almost exclusively (**Figure 2H–J**, **Figure S2B**). During close-range interactions, we found a considerable decrease in overall loudness of all vocalizations (**Figure 2I**). Moreover, the vocal repertoire became more complex, including many USVs, semisongs, and songs, with a significant shift toward USVs (**Figure 2H–J**). Interestingly, we found that the singing mice can switch between these vocal modes in rapid succession (**Figure 2H**, **Figure S3**). This constrains the underlying neural circuit mechanisms capable of producing smooth transitions between the multiple behaviorally relevant vocal modes. We conclude that softer USVs, which attenuate sharply over distance, are used for short-range communication, while songs predominate when mice are alone and are used for long-distance vocal turn-taking with conspecifics under visual occlusion, as we have previously shown (19). To summarize, singing mice thus use USVs during close-range social interactions similar to many other rodents, while employing a novel vocal mode (songs) for long-distance communication.

What might be the neural mechanisms driving these two distinct vocal modes—songs versus USVs? Given the categorical differences in temporal sequencing, loudness, bout durations, and context-dependent usage, it is conceivable that the neural circuits driving these two vocal modes are largely non-overlapping, parallel motor pathways. Alternatively, these two distinct vocal modes may share a common motor pathway operating in two different regimes. We tested these alternative models at three different levels of the vocal-motor hierarchy: peripheral mode of sound production, phonation-respiration coupling, and mechanisms of vocal gating by the caudolateral PAG in the midbrain.

Peripheral Mechanisms for Distinct Vocal Modes. We began by determining whether the biophysical mechanism for sound production differs between the two vocal modes. Rodents produce sounds using their larynx in two ways: by vibrating their vocal folds (similar to human speech) or by generating aerodynamic whistles within the larynx. While the precise laryngeal and aerodynamic mechanisms are still under investigation (28–30), these two broad classes can be distinguished by changing the air density in which the animals vocalize (31–33). To test the vibrational versus whistle models, we replaced the air in the behavioral enclosure with a mixture of helium and oxygen (heliox, 80% to 20% respectively). If sounds are produced by a whistle mechanism, the fundamental frequency (F0) should be sensitive to the medium's density; consequently, the pitch should increase in heliox compared to air. Fundamental frequency will not change if sounds are instead produced by a vibrational mechanism. Indeed, fundamental frequencies of both USVs and song notes increased significantly (by similar ratios) under heliox conditions, consistent with a whistle mechanism (**Figure 3A–D**). Crucially, we verified that singing mice are capable of producing sounds using a vibrational mechanism as well (i.e. squeaks) (**Figure 3E–F**). This demonstrates that both USVs and songs are produced by similar biophysical mechanisms at the vocal periphery.

Moving up the vocal motor hierarchy, we next tested the coupling between phonation and respiration during USVs versus songs (**Figure 3G–M**). We implanted a thermistor in the nasal cavity to continuously monitor temperature changes as a proxy of respiration while mice produce vocalizations during close-range social interactions (**Figure 3G–J**). We observed substantial differences in the distribution of respiration rates during USVs, songs, as well as silent (non-vocal) epochs (**Figure 3K**). During songs, the respiration rates were much higher than USVs and could easily exceed 20 cycles/second (**Figure 3K** ***cyan***). In singing mice, each song note was associated with a respiration cycle with phonations occurring exclusively during exhalations (**Figure 3I, L**). Similar to song notes, each USV, regardless of duration, is produced during exhalation with a consistent phase rela-

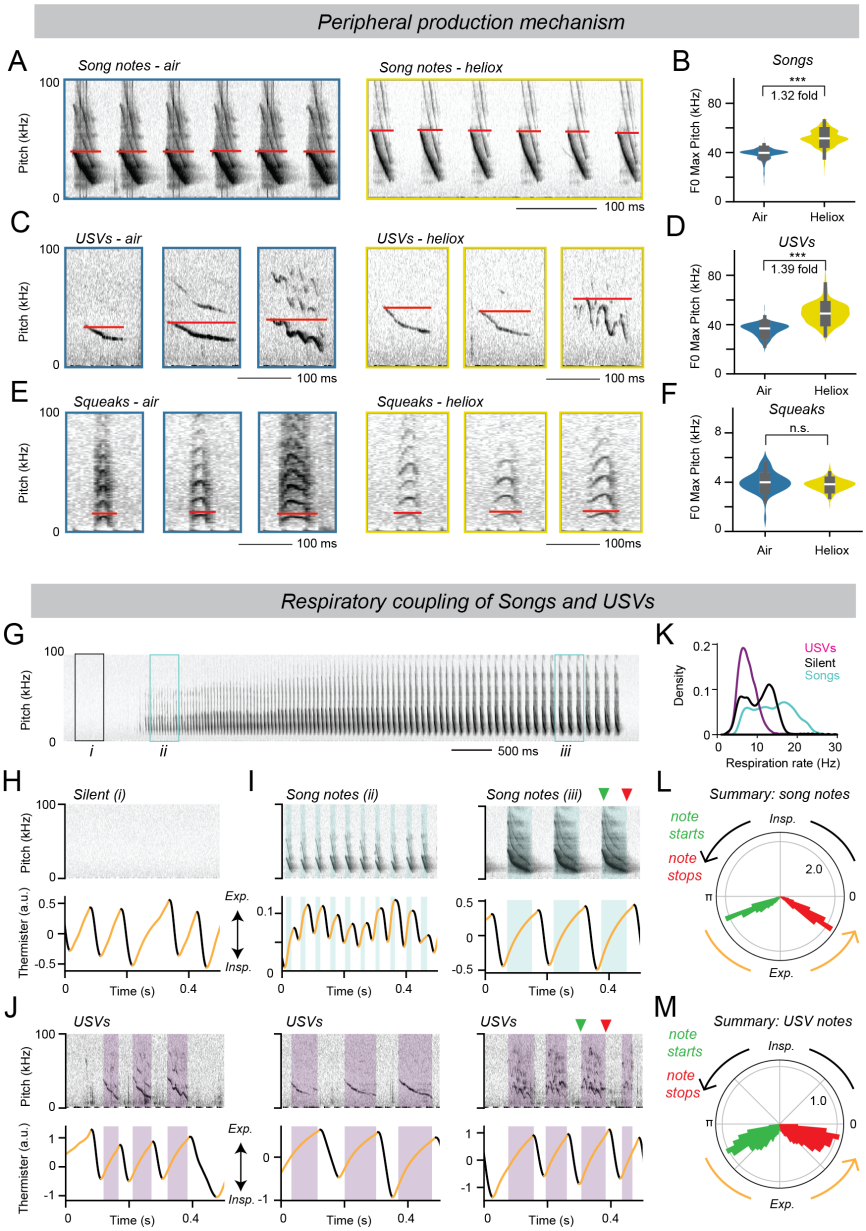

Fig. 3. Shared peripheral vocal production mechanisms between song notes and USVs (A) Example spectrograms of song notes produced in air and in a helium—oxygen mixture (heliox), illustrating an increase in the fundamental frequency (F0) in heliox. **(B)** The maximum F0 of song notes is significantly 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. **(C** and **D)** Same as (A) and (B), 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. **(E** and **F)** Same as (A) and (B), but for squeaks instead of song notes, showing 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. **(G)** Spectrogram of a song preceded by a silent period, recorded with simultaneous thermistor-based respiration monitoring. **(H)** Example respiration dynamics during silence (box i in (G)), with black indicating inhalation and orange indicating exhalation. **(I)** Same as (H), but during song production (boxes ii and iii in (G)). Song notes, shown in cyan, occur during exhalation. **(J)** Same as (H), but for when the mouse makes USVs. USVs, shown in magenta, also occur during the exhalation phase. **(K)** Distribution of respiration rate when the mouse is silent (black), making songs (cyan), or making USVs (magenta). **(L)** 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 = 2384 notes). **(M)** Same as (L), but for USVs (onset 3.93 ± 0.01 rad, offset 5.70 ± 0.01 rad, n = 1910 notes).

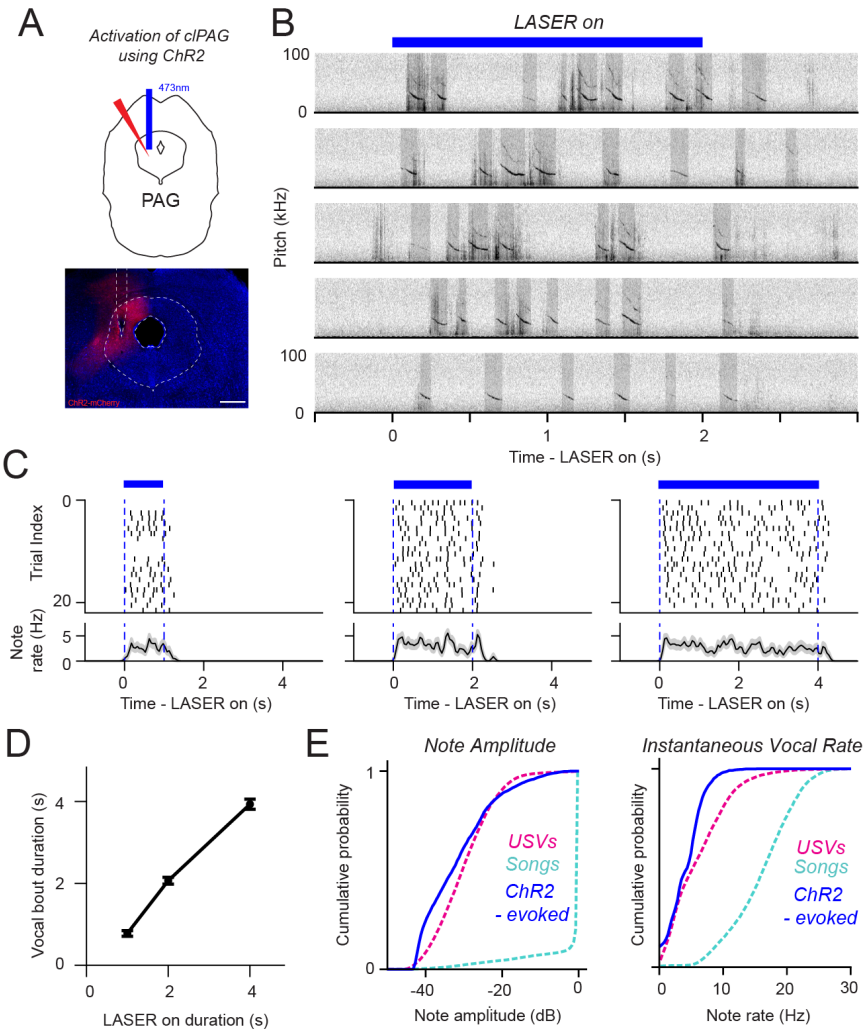

Fig. 4. Midbrain clPAG gates the production of USVs in singing mice. (A) The clPAG optogenetic activation experiment. *Top*: Schematic showing the unilateral virus injection into the clPAG to express ChR2-mCherry in CaMKII neurons, with the fiber implanted above. *Bottom*: Example image of the clPAG displaying virus expression and the fiber implant. **(B)** Spectrograms from five example trials of 2-second tonic optogenetic activation of the clPAG in an example singing mouse, showing evoked vocalizations during the stimulation period. **(C)** Vocalization raster and rate (mean ± sem) for optogenetic activations of 1 second (*left*), 2 seconds (*middle*), and 4 seconds (*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 (1s stimulation: 0.78 ± 0.07 s, n = 86 trials; 2s stimulation: 2.07 ± 0.08 s, n = 85 trials; 4s stimulation: 3.93 ± 0.12 s, n = 85 trials; n = 4 mice), illustrating that vocalizations persist for the full duration of the stimulation. **(E)** Cumulative distributions of note amplitude (*left*) and instantaneous note rate (*right*) for 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).

tionship (**Figure 3J, M**). This is in line with previous literature demonstrating that USVs are tightly coupled to underlying respiratory cycles and are produced exclusively during exhalations in lab mice (25, 34). Taken together, our results suggest that the two distinct vocal modes—USVs and songs—are produced by similar phonatory mechanisms, possibly sharing the well-characterized premotor and motor circuits in the hindbrain.

Vocal gating of USVs by the caudolateral PAG. Given the similarities between USVs in laboratory mice and singing mice in acoustic properties, behavioral context, and peripheral sound production mechanisms, we investigated whether vocal gating mechanisms in central brain circuits are also similar or not. Multiple studies have established the role of the midbrain periaqueductal gray (PAG) in innate behaviors, including vocalizations across vertebrates (21). Neurons in the caudolateral portion of the 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 have been demonstrated to be both necessary and sufficient for USV production in laboratory mice (35–40).

We relied on gross morphological similarities to identify the homologous PAG sub-region in the singing mouse. Optogenetic activation of ChR2-expressing clPAG neurons was sufficient to reliably elicit vocalizations (**Figure 4A–B**, **Video S3**). Vocalizations began immediately after light onset and scaled with the duration of photostimulation (**Figure 4B–D**, n = 4 mice). Acoustic analyses revealed that these vocalizations resembled USVs in both amplitude and instantaneous vocal rates (**Figure 4E**). Therefore, clPAG stimulation is sufficient to evoke USVs in singing mice. Conversely, synaptic silencing of clPAG neurons with Tetanus toxin light-chain

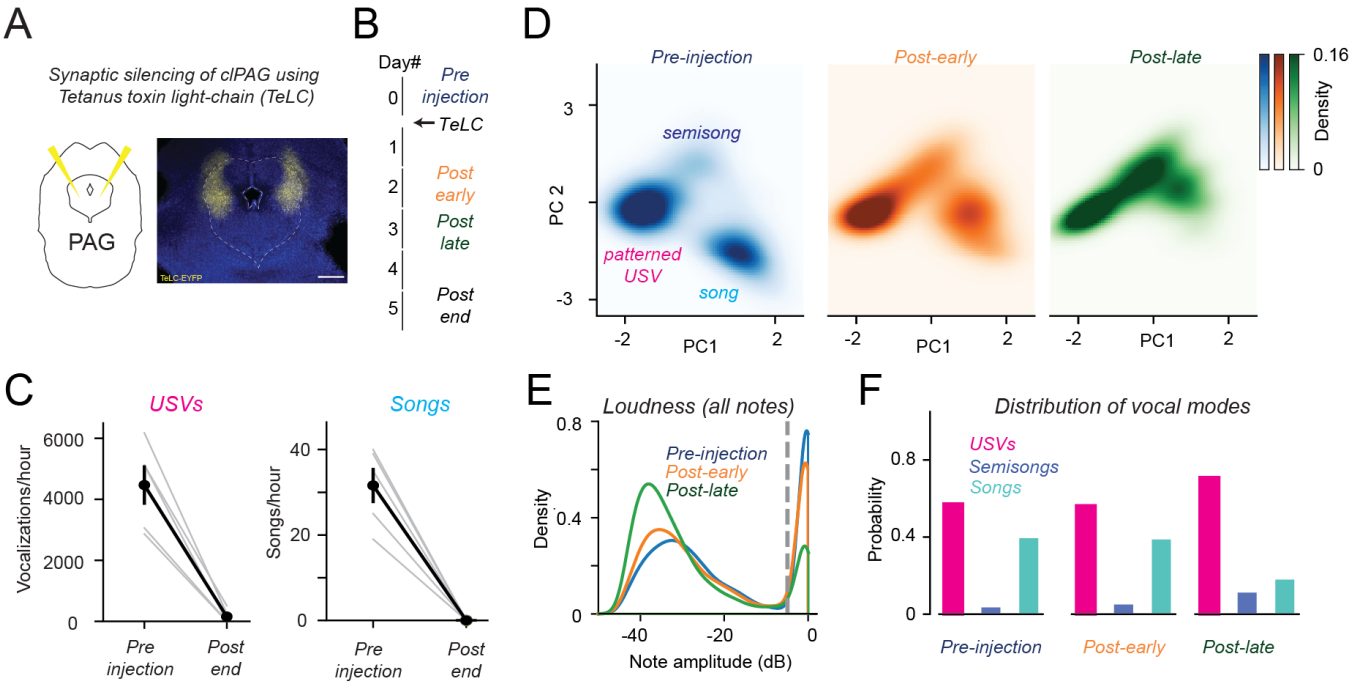

Fig. 5. Silencing the clPAG progressively degrades the vocal repertoire of singing mice (A) The clPAG silencing experiment with TeLC. Left: Schematic showing the bilateral virus injection into the clPAG to express TeLC-EYFP in CaMKII neurons. Right: Example image of the clPAG displaying virus expression. **(B)** Experimental timeline of the clPAG silencing experiment with tetanus toxin light chain (TeLC). The post-injection period was divided into two halves—post-early and post-late—based on systematic curation, with post-early covering the hours containing the first half of curated vocalizations and post-late the second half. **(C)** Number of USVs (left; pre-injection, 4469 ± 645 notes/hour, post-injection: 157 ± 102 notes/hour, 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 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 **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 post-late period (right; n = 252 segments from 21,084 notes), showing a shift in density from songs toward semisongs. **(E)** Distribution of loudness for all vocalizations before virus injection (n = 39,596 notes), during the post-early period (n = 34,826 notes), and during post-late period(n = 21,084 notes), showing an overall reduction in loudness. **(F)** Distribution of vocal modes of all vocalizations across these experimental timepoints (same as in (E)).

(TeLC) caused a severe reduction in USVs across all animals tested (**Figure 5A–C**, n = 5 mice). We conclude that clPAG neurons are both necessary and sufficient to generate species-typical USVs in the singing mouse.

Based on similarities in acoustic properties, behavioral context, peripheral production mechanisms, and central neural control by the clPAG, our findings strongly suggest 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 (41, 42). In contrast, songs represent a drastically divergent vocal behavior unique to the singing mice lineage, characterized by their significantly higher amplitude, complex temporal patterning, and specialized use in long-distance communication.

Parametric Control of Song Progression by clPAG. How does the neural circuitry governing the elaborate songs differ from USV control? We found that synaptic silencing of clPAG also eliminated songs along with USVs; in effect, the animals were rendered mute (**Figure 5B–C**). This observation rules out a parallel, independent mechanism for song production. Instead, it points to a shared vocal motor control circuit for both USVs and songs. Given the drastic differences in loudness and tempo during songs (compared to USVs), we hypothesized that song production involves driving the same neural circuit in a different operating regime. Consistent with this hypothesis, we observed a progressive shift from the song cluster to the semisong cluster during the post-early and post-late phases after TeLC injection (**Figure 5D**, **Figure S4A**). Concomitantly, we also observed progressive lowering of note loudness, consistent with the increasing proportion of USVs and semisongs in lieu of songs (**Figure 5E–F**). Tracking this progressive loss of song production offered precious clues into how the clPAG controls song production.

To understand the precise contribution of the clPAG to song production, we first needed to parameterize the behavior. We derived a mathematical equation to describe the temporal progression of individual notes during a song. During songs, singing mice string together a series of many notes over several seconds (**Figure 6A**). We serendipitously observed that during song progression, instantaneous note rate $r(i)$ varies linearly with note index i (**Figure 6B**, **Figure S4B**). 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 6B**). A linear model with just three parameters (start rate $r_{\max}$, slope m, and stop rate $r_{\min}$) adequately describes the instantaneous vocal rate throughout the song (**Figure 6B–C**, **Methods**):

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

Remarkably, this simple generative model works effectively

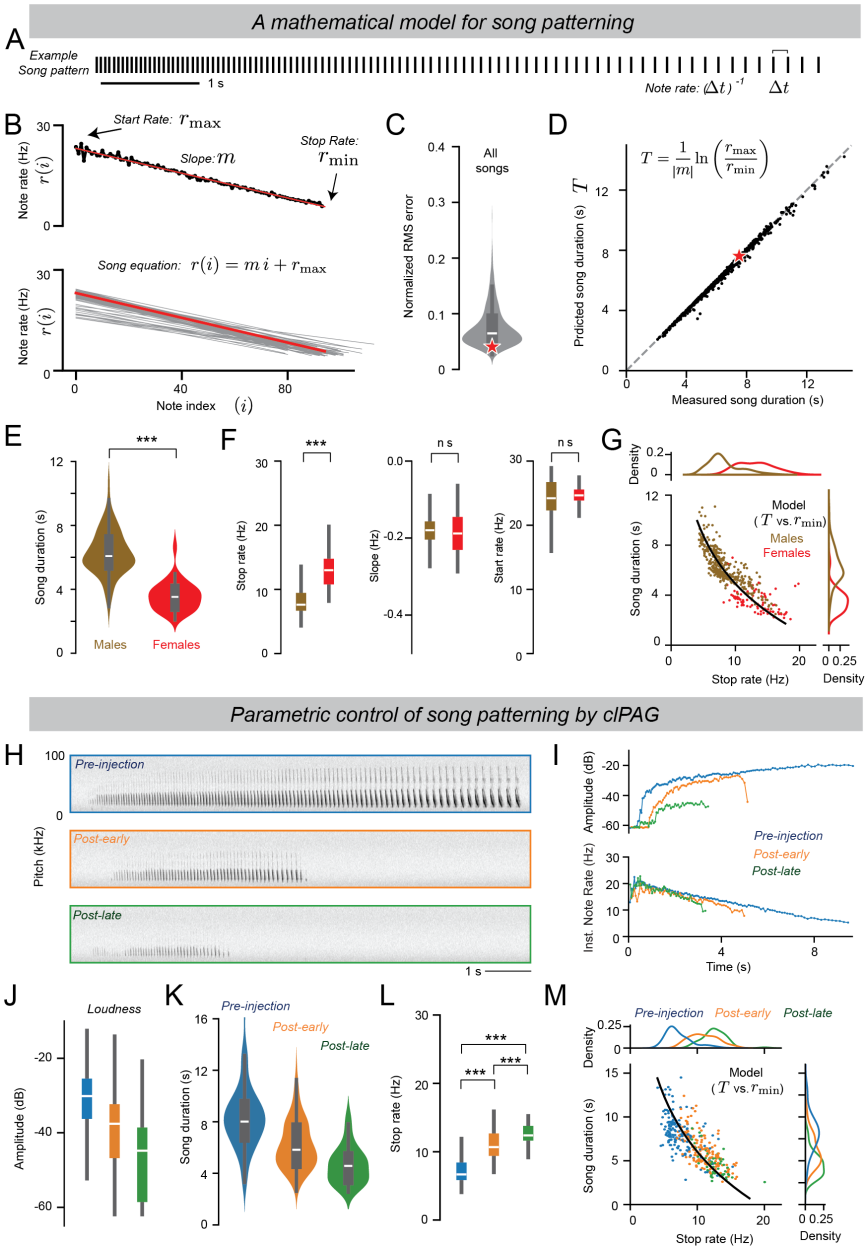

Fig. 6. Parametric modulation of song patterning by cIPAG (A) Raster plot of a song, with each line marking the onset of a song note. The annotation indicates that the instantaneous note rate is calculated as the inverse of the interval between note onsets. **(B)** Temporal patterning trajectories of example songs. Top: The trajectory of the 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 patterning trajectories for all songs produced by a single mouse (n = 60), with the red line indicating the song from the top panel. **(C)** Normalized RMSE of the linear model fit (n = 820 songs, 0.075 ± 0.001). **(D)** Model-predicted versus measured song duration (n = 820 songs from 12 mice across alone, social, and PAG TeLC pre- and post-injection conditions). The data, which largely align with the unity line, suggest that the three model parameters effectively describe songs of varying durations. **(E)** Song duration distribution for intact males and females when alone or socially interacting in the PAIRId (male: n = 370 songs, 6.28 ± 0.07 s; female: n = 86 songs, 3.58 ± 0.10 s; P = 6.65e-38). **(F)** Distribution of song patterning parameters of songs in (E), comparing males and females. Left: stop rate (male: 8.27 ± 0.13 Hz; female 13.10 ± 0.30 Hz; P = 1.21e-30); Middle: slope (male: -0.18 ± 0.00; female: -0.19 ± 0.01; P > 0.05); Right: start rate (male: 24.19 ± 0.15 Hz; female: 24.39 ± 0.19 Hz; P > 0.05). **(G)** Song duration versus song stop rate for male and female songs from (E). The black dashed line represents the model prediction, based on a mean slope of -0.18 and a mean start rate of 24.23 Hz. Overall, most songs cluster along this line, with males and females exhibiting distinct stop rates and song durations. **(H)** Spectrograms of example songs before the PAG TeLC virus injection (*top*), during the post-early period (*middle*), and during post-late period (*bottom*). **(I)** Trajectories of note amplitude (*top*) and instantaneous note rate (*bottom*) for the example songs shown in (H). **(J)** Cumulative distribution of note loudness in low-gain audio for all song notes across different PAG TeLC experimental timepoints (pre-injection: n = 15,583 notes, -32.53 ± 0.09 dB; post-early: n = 13,490 notes, -40.60 ± 0.10 dB; post-late: n = 3,774 notes, -46.43 ± 0.17 dB; pre-injection vs. post-early: P = 0.0; post-early vs. post-late: P = 5.90e-204; pre-injection vs. post-late: P = 0.0). **(K)** Same as (E), but for songs across different PAG TeLC experimental timepoints (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). **(L)** Same as (F) *left*, but for songs across different PAG TeLC experimental timepoints (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). **(M)** Same as (G), but for songs across different PAG TeLC experimental timepoints and with a model prediction based on a mean slope of -0.11 and a mean start rate of 19.15 Hz.

for all songs across all mice in our study (**Figure 6C, Figure S4C**). Additionally, by integrating the instantaneous rate in the time domain, we can accurately estimate song durations T across animals and experimental conditions (**Figure 6D, Figure S4C**):

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

The utility of this mathematical framework becomes apparent when examining the quantitative variation in song durations between male and female mice. In this species, male songs are longer by a few seconds (**Figure 6E**). Decomposing each song into its constituent parameters, we found that neither the start rate nor the slope significantly differed between sexes (**Figure 6F**). However, stop rates were significantly lower for males compared to females (**Figure 6F *left***). In fact, the relationship between song durations T and stop rates $r_{\min}$ for all songs precisely matched our model's prediction that T would vary inversely with the negative logarithm of the $r_{\min}$ (**Figure 6G**). Thus, male songs are longer specifically because of their lower stop rates. In other words, males are able to extend their song motor pattern and append long, loud notes at the end of songs that females do not produce (**Figure S4E–F**). This trait may correlate with male energy status, potentially serving as an honest signal of fitness under sexual selection (43–45).

Can this framework also explain the effect of synaptic silencing of clPAG on songs? Compared to pre-injection control songs, we observed a progressive reduction in note loudness after TeLC-mediated silencing of clPAG in example songs (**Figure 6H–I**) as well as across animals (**Figure 6J**). In the temporal domain, song duration also progressively decreased (**Figure 6H, K**). Visual inspection revealed that shortened songs followed the same initial trajectories (**Figure 6I**). Indeed, songs progressively became shorter as stop rates increased (**Figure 6K, L**), while the effect on the remaining two parameters were modest (**Figure S4D**). Strikingly, all of these perturbed songs followed the theoretical relationship predicted by our model (**Figure 6M**). The effect of TeLC-mediated silencing of clPAG on song patterning and duration phenocopied natural behavioral variation, initially transforming male songs to resemble those produced by females before eliminating vocalizations completely (**Figure 6G, M**, **Figure S4E–F**). Finally, songs progressively deteriorated over days, often breaking into multiple fragments before disappearing completely (**Figure S5**). Therefore, the parameter most affected by clPAG silencing—the stop rate—also accounts for natural sexual dimorphism in song production. 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 phonatory mechanism, but rather through amplitude modulation of individual notes and frequency modulation of motor patterning via a specific parameter (stop rate) of a shared vocal gating circuit (**Figure 7**).

Discussion

Summary. In this study, we developed a novel behavioral assay (PAIRId—Partial Acoustic Isolation Reveals Identity) that enables precise attribution of vocalizations to individual animals during social interactions (**Figure 1**). This approach revealed that singing mice employ two categorically distinct vocal modes: soft, unstructured USVs and loud, temporally patterned songs, often in quick succession during social encounters (**Figure 2**). Surprisingly, despite their dramatic differences in acoustic properties, temporal organization, and social context, both vocal modes share peripheral sound production mechanisms and central neural control by the caudolateral PAG (clPAG) (**Figure 3, 4, 5**). Using a mathematical model of song structure, we demonstrated how progressive silencing of clPAG neurons systematically alters specific aspects of song production before eliminating all vocalizations (**Figure 6**). These results reveal that differential amplitude modulation (AM) and frequency modulation (FM) of shared vocal motor circuits are used to produce categorically distinct behavioral outputs during social encounters (**Figure 7**). Collectively, our findings suggest how a single neural circuit might be functionally repurposed to support multiple behavioral outputs through modulation of specific parameters, pointing to a potential principle of multifunctional circuit reuse in the brain. This work also offers a window into how neural circuits could be modified during evolution to generate novel behaviors while preserving ancestral functions, potentially providing insights into the mechanistic basis of behavioral diversification across species.

Two Vocal Modes: Conserved USVs and Novel Songs. Our findings strongly suggest that USVs in singing mice represent an ancestral and conserved vocal behavior shared with other rodents. The acoustic properties of these vocalizations—their soft amplitude, reduced temporal stereotypy, and usage during close-range social interactions—closely parallel USVs observed in laboratory mice (25, 46), despite differences in fundamental frequency that can be attributed to species-specific variations in laryngeal morphology. The shared biophysical mechanisms of sound production through aerodynamic whistles and identical control by the caudolateral PAG further support this homology (33, 36). This conservation is particularly notable given the phylogenetic distance between singing mice (Cricetidae) and laboratory mice (Muridae) (47), 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 (27, 31, 48), providing strong evidence that USVs constitute an ancestral vocal mode that predates the divergence of these lineages approximately 20-25 million years ago.

In stark contrast, singing mouse songs represent a derived, novel vocal behavior unique to this lineage. While songs share the same peripheral phonation mechanism with USVs, they differ dramatically in their high amplitude, complex temporal patterning, and specialized usage in long-distance

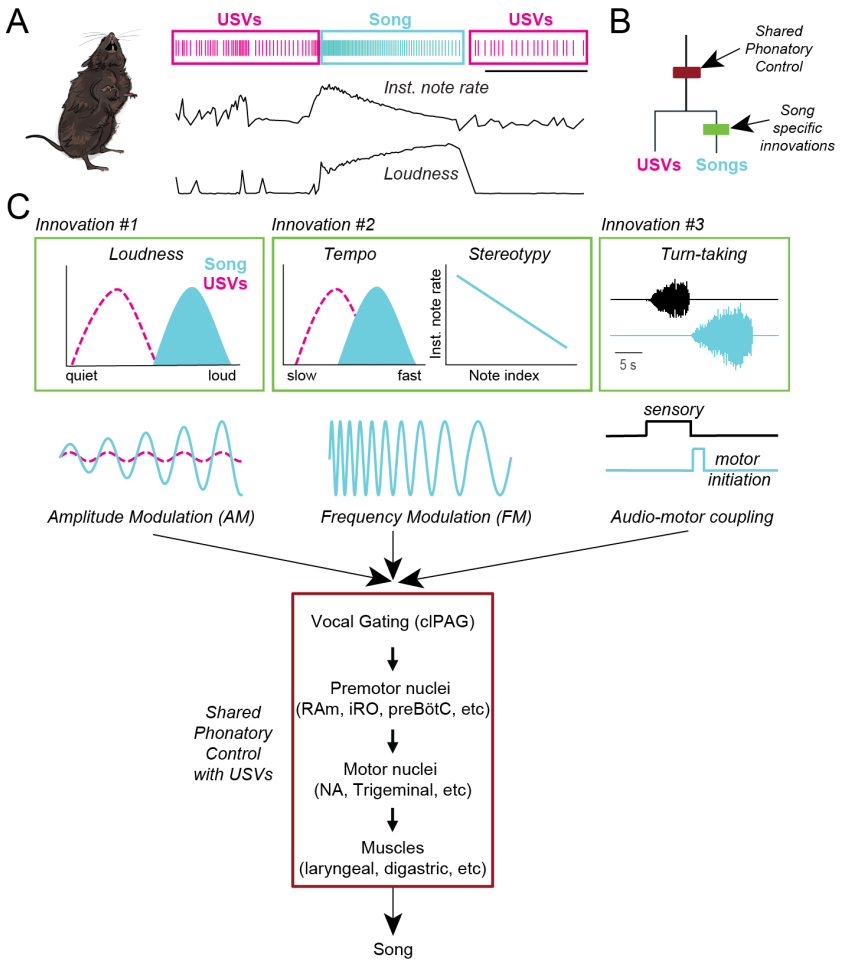

Fig. 7. Working model for vocal motor control in the singing mouse (A) Singing mice can produce two categorically distinct vocal modes in rapid succession–soft, unstructured USVs (magenta) and loud, patterned songs (cyan). **(B)** Songs and USVs share a peripheral sound production mechanism (maroon), while songs involve three separate innovations (green). **(C)** Songs exhibit three innovations (green): (1) the notes are much louder than USVs (cartoon data), (2) songs include notes produced at much higher rates than USVs (left cartoon) and arranged in highly stereotyped patterns (right cartoon), and (3) as demonstrated previously, songs can be used in tightly coupled antiphonal turn-taking behavior known as countersinging. Together, these behavioral innovations imply that songs arise from specific modulations to existing phonatory circuitry (maroon), including amplitude modulation for loudness, frequency modulation for patterning, and audio-motor coupling for turn-taking.

communication. This vocal innovation is restricted to species within the Baiomyini tribe, with *S. teguina* exhibiting the most elaborate song structure among its relatives (16, 32). The restricted phylogenetic distribution of songs suggests they evolved relatively recently (~6.5 million years ago (47, 49)) as a specialized adaptation, likely in response to selective pressures favoring long-distance communication in the montane cloud forest habitats these species occupy (15, 50).

Neural Substrates of Song Production. The periaqueductal gray (PAG) plays a conserved role in the control of instinctive behaviors across vertebrates (51–56), with particular importance for vocalization (20, 21). Stimulation of the PAG (also called the "central grey", and in birds, the "dorsomedial nucleus of the intercollicular complex") elicits species-typical vocalizations across vertebrates, including fish (57), birds (58–60), bats (61), rodents (62, 63), and primates (64–66). Further, bilateral PAG lesions cause mutism in both learned (speech) and innate vocalizations in humans (67), as well as mutism in many other species (57, 68–72).

Our findings in singing mice align with this body of evidence, confirming the critical role of clPAG in vocal control. However, our results extend beyond the established role of PAG in simple vocalizations by demonstrating its involvement in the control of a complex, temporally patterned vocal behavior—singing. Our mathematical model of song structure reveals how a simple parametric control mechanism—particularly the regulation of stop rates—can generate the remarkable variability in song duration observed both within and between sexes (16, 73). The dramatic amplitude differences that depend on clPAG (**Figure 5-6**) likely involve brainstem premotor and motor circuits operating on top of the core phonation mechanism (35–40). In addition, the distinctive temporal patterning of songs may result from the clPAG operating in a different functional regime (**Figure 7**), possibly through modified interactions with respiratory circuits and modulation by the orofacial motor cortex (OMC), which can functionally influence both song tempo as well as sensorimotor coupling during vocal interactions (18, 19).

In species with diverse vocal repertoires, vocal types appear to be differentially represented within the PAG: for example, partial PAG lesions in squirrel monkeys abolish some call

types but not others (70), and stimulation of distinct PAG subregions or cell-types can evokes distinct vocalizations (74–77). The organizational logic of how vocal gating of diverse modes can generate rapidly alternating vocal sequences during natural social interactions still remains unclear. For example, the intrinsic neural dynamics required for song control might be more complex than simple vocal gating for USVs. This is potentially akin to the ability to evoke calls and individual syllables but not full songs by experimental stimulation of motor pathways in songbirds (58, 59, 78, 79). While much remains to be done, the coexistence of both ancestral (USV, squeaks) and novel vocal behaviors (semisongs, songs) in the same animal provides a unique opportunity to investigate vocal motor control in the PAG at single-cell resolution with *in-vivo* electrophysiology. In addition, the framework we lay down here also provides a roadmap for future investigations into the circuit-level changes needed for song-specific evolutionary innovations.

The PAG as a Key Locus for Rapid Evolutionary Diversification. Evolutionary modification of ancestral mechanisms provides a pathway for rapid behavioral diversification. While co-option of existing elements is well-documented at molecular, genetic, and anatomical levels (80, 81), the neural mechanisms underlying behavioral innovations have remained more elusive. Our findings show that singing mice use the same phonation pathway for both close-range USVs and long-distance songs, indicating that novel vocal behaviors evolved by modifying ancestral circuit outputs—specifically amplitude, temporal patterns, and social context—rather than developing entirely new mechanisms. We identify the clPAG as a critical locus in process. We found that silencing the clPAG disrupted two novel features of the song mode: amplitude and duration. In particular, our mathematical model revealed that song duration is primarily controlled by the stop rate parameter, which was selectively affected by clPAG silencing. Interestingly, this same parameter accounts for the natural sexual dimorphism in song production, with males producing longer songs due to their lower stop rates compared to females. The convergence of naturally occurring variation and experimental manipulation on the same parameter suggests that stop rate is a key target for both evolution and neural control of this vocal behavior. This finding supports the idea that evolutionary modifications to specific parameters within shared neural pathways can yield dramatic phenotypic innovations—including sexually dimorphic displays—without requiring entirely new circuits. Thus, circuits within the PAG underpin the differential production of USVs and songs, in line with a recent proposal positing the PAG as a key driver in the evolution of innate motor (vocal) behaviors across species (82).

This pattern of neural circuit co-option may represent a general principle in the evolution of complex behaviors. Rather than evolving new neural circuits *de novo*—a process requiring coordinated changes across multiple organizational levels—evolution appears to favor modifying existing circuits to operate in new regimes (7, 51, 83–88). Our findings set up a case study of the mechanistic basis for this process, demonstrating that modest changes in specific parameters controlled by the same neural substrate can yield behaviors that appear categorically distinct. This principle may extend beyond vocal communication to other behavioral domains where innovations have emerged over short evolutionary timescales through the repurposing of ancestral neural circuits.

FUNDING

This work was supported by the National Institutes of Health BRAIN Initiative RF1-NS132046-01 (AB), Searle Scholars Program (AB), Pershing Square Foundation Innovator Fund (AB), , the Esther A. & Joseph Klingenstein Fund and the Simons Foundation Fellowship (AB), Cold Spring Harbor Laboratory (AB), the International Society for Neuroethology Konishi Research Award (CEH), and the George A. and Marjorie H. Anderson Fellowship (XMZ).

ACKNOWLEDGEMENTS

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 Benjamin Cowley 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 also thank Taylor Sterry for illustrating the PAIRId apparatus and mice, and Christopher Auger-Dominguez for the singing mouse photograph.

References

1. Manuel J Ferreira-Pinto, Harsh Kanodia, Antonio Falasconi, Markus Sigrist, Maria S Esposito, and Silvia Arber. Functional diversity for body actions in the mesencephalic locomotor region. *Cell*, 184(17):4564–4578.e18, August 2021. doi: 10.1016/j.cell.2021.07.002.
2. Ludwig Ruder, Riccardo Schina, Harsh Kanodia, Sara Valencia-Garcia, Chiara Pivetta, and Silvia Arber. A functional map for diverse forelimb actions within brainstem circuitry. *Nature*, 590(7846):445–450, February 2021. doi: 10.1038/s41586-020-03080-z.
3. Li Wang, Irene Z Chen, and Dayu Lin. Collateral pathways from the ventromedial hypothalamus mediate defensive behaviors. *Neuron*, 85(6):1344–1358, March 2015. doi: 10.1016/j.neuron.2014.12.025.
4. Charlotte L Barkan, Darcy B Kelley, and Erik Zornik. Premotor neuron divergence reflects vocal evolution. *J. Neurosci.*, 38(23):5325–5337, June 2018. doi: 10.1523/jneurosci.0089-18.2018.
5. K L Briggman and W B Kristan. Multifunctional pattern-generating circuits. *Annu. Rev. Neurosci.*, 31(1):271–294, 2008. doi: 10.1146/annurev.neuro.31.060407.125552.
6. Eve Marder, Timothy O'Leary, and Sonal Shruti. Neuromodulation of circuits with variable parameters: single neurons and small circuits reveal principles of state-dependent and robust neuromodulation. *Annu. Rev. Neurosci.*, 37(1):329–346, 2014. doi: 10.1146/annurev-neuro-071013-013958.
7. Ruairí J V Roberts, Sinziana Pop, and Lucia L Prieto-Godino. Evolution of central neural circuits: state of the art and perspectives. *Nat. Rev. Neurosci.*, 23(12):725–743, December 2022. doi: 10.1038/s41583-022-00644-y.
8. Gregg A Castellucci, Frank H Guenther, and Michael A Long. A theoretical framework for human and nonhuman vocal interaction. *Annu. Rev. Neurosci.*, 45:295–316, July 2022. doi: 10.1146/annurev-neuro-111020-094807.
9. Steffen R Hage and Andreas Nieder. Dual neural network model for the evolution of speech and language. *Trends Neurosci.*, 39(12):813–829, December 2016. doi: 10.1016/j.tins.2016.10.006.
10. Darcy B Kelley, Irene H Ballagh, Charlotte L Barkan, Andres Bendesky, Taffeta M Elliott, Ben J Evans, Ian C Hall, Young Mi Kwon, Ursula Kwong-Brown, Elizabeth C Leininger, Emilie C Perez, Heather J Rhodes, Avelyne Villain, Ayako Yamaguchi, and Erik Zornik. Generation, coordination, and evolution of neural circuits for vocal communication. *J. Neurosci.*, 40(1):22–36, January 2020. doi: 10.1523/jneurosci.0736-19.2019.
11. Andreas Nieder and Richard Mooney. The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 375(1789):20190054, January 2020. doi: 10/ggsbd8.
12. Yisi S Zhang and Asif A Ghazanfar. A hierarchy of autonomous systems for vocal production. *Trends Neurosci.*, 43(2):115–126, February 2020. doi: 10.1016/j.tins.2019.12.006.
13. Friedrich Ladich and Hans Winkler. Acoustic communication in terrestrial and aquatic vertebrates. *J. Exp. Biol.*, 220(13):2306–2317, July 2017. doi: 10.1242/jeb.132944.
14. Arkarup Banerjee, Steven M Phelps, and Michael A Long. Singing mice. *Curr. Biol.*, 29(6):R190–R191, March 2019. doi: 10.1016/j.cub.2018.11.048.
15. Emmet T Hooper and Michael D Carleton. Reproduction, growth and development in two contiguously allopatric rodent species, genus Scotinomys. *Miscellaneous Publications, Museum of Zoology, University of Michigan*, 151(15):1–52, 1976.
16. Jacqueline R Miller and Mark D Engstrom. Vocal stereotypy and singing behavior in baiomyine mice. *J. Mammal.*, 88(6):1447–1465, December 2007. doi: 10.1644/06-mamm-a-386r.1.
17. Bret Pasch, Benjamin M Bolker, and Steven M Phelps. Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice. *Am. Nat.*, 182(5):E161–73, November 2013. doi: 10.1086/673263.
18. Arkarup Banerjee, Feng Chen, Shaul Druckmann, and Michael A Long. Temporal scaling of motor cortical dynamics reveals hierarchical control of vocal production. *Nat. Neurosci.*, 27(3):527–535, March 2024. doi: 10.1038/s41593-023-01556-5.

19. Daniel E Okobi, Jr, Arkarup Banerjee, Andrew M M Matheson, Steven M Phelps, and Michael A Long. Motor cortical control of vocal interaction in neotropical singing mice. *Science*, 363(6430):983–988, March 2019. doi: 10.1126/science.aau9480.
20. U Jürgens. The neural control of vocalization in mammals: a review. *J. Voice*, 23(1):1–10, January 2009. doi: 10.1016/j.jvoice.2007.07.005.
21. U Jürgens. The role of the periaqueductal grey in vocal behaviour. *Behav. Brain Res.*, 62 (2):107–117, June 1994. doi: 10.1016/0166-4328(94)90017-5.
22. Max L Sterling, Ruben Teunisse, and Bernhard Englitz. Rodent ultrasonic vocal interaction resolved with millimeter precision using hybrid beamforming. *Elife*, 12, July 2023. doi: 10.7554/elife.86126.
23. Elena N Waidmann, Victor H Y Yang, William C Doyle, and Erich D Jarvis. Mountable miniature microphones to identify and assign mouse ultrasonic vocalizations. *bioRxiv*, February 2024. doi: 10.1101/2024.02.05.579003.
24. Megan R Warren, Daniel T Sangiamo, and Joshua P Neunuebel. High channel count microphone array accurately and precisely localizes ultrasonic signals from freely-moving mice. *J. Neurosci. Methods*, 297:44–60, March 2018. doi: 10.1016/j.jneumeth.2017.12.013.
25. Gregg A Castellucci, Daniel Calbick, and David McCormick. The temporal organization of mouse ultrasonic vocalizations. *PLoS One*, 13(10):e0199929, October 2018. doi: 10.1371/journal.pone.0199929.
26. Timothy E Holy and Zhongsheng Guo. Ultrasonic songs of male mice. *PLoS Biol.*, 3(12): e386, December 2005. doi: 10.1371/journal.pbio.0030386.
27. G D Sales. Ultrasound and mating behaviour in rodents with some observations on other behavioural situations. *J. Zool. (1987)*, 168(2):149–164, October 1972. doi: 10.1111/j.1469-7998.1972.tb01345.x.
28. Jonas Håkansson, Weili Jiang, Qian Xue, Xudong Zheng, Ming Ding, Anurag A Agarwal, and Coen P H Elemans. Aerodynamics and motor control of ultrasonic vocalizations for social communication in mice and rats. *BMC Biol.*, 20(1):3, January 2022. doi: 10.1186/s12915-021-01185-z.
29. Elena Mahrt, Anurag Agarwal, David Perkel, Christine Portfors, and Coen P H Elemans. Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. *Curr. Biol.*, 26(19):R880–R881, October 2016. doi: 10.1016/j.cub.2016.08.032.
30. Tobias Riede, Heather L Borgard, and Bret Pasch. Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. *R. Soc. Open Sci.*, 4(11):170976, November 2017. doi: 10.1098/rsos.170976.
31. Marcela Fernández-Vargas, Tobias Riede, and Bret Pasch. Mechanisms and constraints underlying acoustic variation in rodents. *Anim. Behav.*, 184:135–147, February 2022. doi: 10.1016/j.anbehav.2021.07.011.
32. Tobias Riede and Bret Pasch. Pygmy mouse songs reveal anatomical innovations underlying acoustic signal elaboration in rodents. *J. Exp. Biol.*, 223(Pt 12):jeb223925, June 2020. doi: 10.1242/jeb.223925.
33. L H Roberts. The rodent ultrasound production mechanism. *Ultrasonics*, 13(2):83–88, March 1975. doi: 10.1016/0041-624x(75)90052-9.
34. Laurence H Roberts. Correlation of respiration and ultrasound production in rodents and bats. *J. Zool. (1987)*, 168(4):439–449, December 1972. doi: 10.1111/j.1469-7998.1972.tb01360.x.
35. Alastair MacDonald, Alina Hebling, Xin Paul Wei, and Kevin Yackle. The breath shape controls intonation of mouse vocalizations. *Elife*, 13, July 2024. doi: 10.7554/elife.93079.2.
36. Katherine Tschida, Valerie Michael, Jun Takatoh, Bao-Xia Han, Shengli Zhao, Katsuyasu Sakurai, Richard Mooney, and Fan Wang. A specialized neural circuit gates social vocalizations in the mouse. *Neuron*, 103(3):459–472.e4, August 2019. doi: 10.1016/j.neuron.2019.05.025.
37. Avin Veerakumar, Joshua P Head, and Mark A Krasnow. A brainstem circuit for phonation and volume control in mice. *Nat. Neurosci.*, 26(12):2122–2130, December 2023. doi: 10.1038/s41593-023-01478-2.
38. Cindy F Yang, Euiseok J Kim, Edward M Callaway, and Jack L Feldman. Monosynaptic projections to excitatory and inhibitory preBötzinger complex neurons. *Front. Neuroanat.*, 14:58, September 2020. doi: 10.3389/fnana.2020.00058.
39. Jaehong Park, Seonmi Choi, Jun Takatoh, Shengli Zhao, Andrew Harrahill, Bao-Xia Han, and Fan Wang. Brainstem control of vocalization and its coordination with respiration. *Science*, 383(6687):eadi8081, March 2024. doi: 10.1126/science.adi8081.
40. Konstantin Hartmann and Michael Brecht. A functionally and anatomically bipartite vocal pattern generator in the rat brain stem. *iScience*, 23(12):101804, December 2020. doi: 10.1016/j.isci.2020.101804.
41. S Abhirami, Swapna Agarwalla, Anandaroop Bhattacharya, and Sharba Bandyopadhyay. Contribution of the ventral pouch in the production of mouse ultrasonic vocalizations. *Phys. Rev. E.*, 107(2-1):024412, February 2023. doi: 10.1103/physreve.107.024412.
42. Samantha K Smith, Tracy T Burkhard, and Steven M Phelps. A comparative characterization of laryngeal anatomy in the singing mouse. *J. Anat.*, 238(2):308–320, February 2021. doi: 10.1111/joa.13315.
43. Tracy Tran Burkhard, E Raney Sachs, and Steven M Phelps. Female preferences for high vocal effort in singing mice. *Behaviour*, 160(3-4):275–297, February 2023. doi: 10.1163/1568539x-bja10203.
44. Tracy T Burkhard, Rebecca R Westwick, and Steven M Phelps. Adiposity signals predict vocal effort in alston's singing mice. *Proc. Biol. Sci.*, 285(1877):20180090, April 2018. doi: 10.1098/rspb.2018.0090.
45. Bret Pasch, Andreas S George, Polly Campbell, and Steven M Phelps. Androgen-dependent male vocal performance influences female preference in neotropical singing mice. *Anim. Behav.*, 82(2):177–183, August 2011. doi: 10.1016/j.anbehav.2011.04.018.
46. Kerstin Musolf and Dustin J Penn. Ultrasonic vocalizations in house mice:. In Milos Macholan, Stuart J E Baird, and Pavel Munclinger, editors, *Evolution of the House Mouse*, pages 253–277. Cambridge University Press, Cambridge, July 2012. doi: 10.1017/cbo9781139044547.012.
47. Scott J Steppan and John J Schenk. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS One*, 12(8):e0183070, August 2017. doi: 10.1371/journal.pone.0183070.
48. Gillian D Sales. Ultrasonic calls of wild and wild-type rodents. In *Handbook of Behavioral Neuroscience*, Handbook of behavioral neuroscience, pages 77–88. Elsevier, 2010. doi: 10.1016/b978-0-12-374593-4.00009-7.
49. Susette Castañeda-Rico, Jesús E Maldonado, Melissa T R Hawkins, and Cody W Edwards. Unveiling hidden diversity: Phylogenomics of neotomine rodents and taxonomic implications for the genus peromyscus. *Mol. Phylogenet. Evol.*, 203(108233):108233, February 2025. doi: 10.1016/j.ympev.2024.108233.
50. David O Ribble and Galen B Rathbun. Preliminary observations on home ranges and natural history of *Scotinomys teguina* in costa rica. *Mammalia*, 82(5):490–493, September 2018. doi: 10/gd3w4d.
51. Felix Baier, Katja Reinhard, Victoria Tong, Julie Murmann, Karl Farrow, and Hopi E Hoekstra. The neural basis of defensive behaviour evolution in peromyscus mice. *bioRxivorg*, July 2023. doi: 10.1101/2023.07.04.547734.
52. Annegret L Falkner, Dongyu Wei, Anjeli Song, Li W Watsek, Irene Chen, Patricia Chen, James E Feng, and Dayu Lin. Hierarchical representations of aggression in a hypothalamic-midbrain circuit. *Neuron*, 106(4):637–648.e6, May 2020. doi: 10.1016/j.neuron.2020.02.014.
53. Yaara Lefler, Dario Campagner, and Tiago Branco. The role of the periaqueductal gray in escape behavior. *Curr. Opin. Neurobiol.*, 60:115–121, February 2020. doi: 10.1016/j.conb.2019.11.014.
54. Carlos Silva and Neil McNaughton. Are periaqueductal gray and dorsal raphe the foundation of appetitive and aversive control? a comprehensive review. *Prog. Neurobiol.*, 177:33–72, June 2019. doi: 10.1016/j.pneurobio.2019.02.001.
55. Philip Tovote, Maria Soledad Esposito, Paolo Botta, Fabrice Chaudun, Jonathan P Fadok, Milica Markovic, Steffen B E Wolff, Charu Ramakrishnan, Lief Fenno, Karl Deisseroth, Cyril Herry, Silvia Arber, and Andreas Lüthi. Midbrain circuits for defensive behaviour. *Nature*, 534(7606):206–212, June 2016. doi: 10.1038/nature17996.
56. Hong Yu, Xinkuan Xiang, Zongming Chen, Xu Wang, Jiaqi Dai, Xinxin Wang, Pengcheng Huang, Zheng-Dong Zhao, Wei L Shen, and Haohong Li. Periaqueductal gray neurons encode the sequential motor program in hunting behavior of mice. *Nat. Commun.*, 12(1): 6523, November 2021. doi: 10.1038/s41467-021-26852-1.
57. J Matthew Kittelberger, Bruce R Land, and Andrew H Bass. Midbrain periaqueductal gray and vocal patterning in a teleost fish. *J. Neurophysiol.*, 96(1):71–85, July 2006. doi: 10.1152/jn.00067.2006.
58. J L Brown. Vocalization evoked from the optic lobe of a songbird. *Science*, 149(3687): 1002–1003, August 1965. doi: 10.1126/science.149.3687.1002.
59. T J Seller. Midbrain regions involved in call production in java sparrows. *Behav. Brain Res.*, 1(3):257–265, June 1980. doi: 10.1016/0166-4328(80)90033-9.
60. Tsuyoshi Shimmura, Mai Tamura, Shosei Ohashi, Asuka Sasaki, Takamichi Yamanaka, Nobuhiro Nakao, Kunio Ihara, Shinsaku Okamura, and Takashi Yoshimura. Cholecystokinin induces crowing in chickens. *Sci. Rep.*, 9(1):3978, March 2019. doi: 10.1038/s41598-019-40746-9.
61. G Schuller and S Radtke-Schuller. 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(1):192–206, 1990. doi: 10.1007/bf00228889.
62. R J Waldbillig. Attack, eating, drinking, and gnawing elicited by electrical stimulation of rat mesencephalon and pons. *J. Comp. Physiol. Psychol.*, 89(3):200–212, May 1975. doi: 10.1037/h0076808.
63. Y Yajima, Y Hayashi, and N Yoshii. The midbrain central gray substance as a highly sensitive neural structure for the production of ultrasonic vocalization in the rat. *Brain Res.*, 198(2): 446–452, October 1980. doi: 10.1016/0006-8993(80)90759-3.
64. T G Brown. Note on the physiology of the basal ganglia and mid-brain of the anthropoid ape, especially in reference to the act of laughter. *J. Physiol.*, 49(4):195–207, May 1915. doi: 10.1113/jphysiol.1915.sp001703.
65. U Jürgens and D Ploog. Cerebral representation of vocalization in the squirrel monkey. *Exp. Brain Res.*, 10(5):532–554, June 1970. doi: 10.1007/bf00234269.
66. H W Magoun, D Atlas, E H Ingersoll, and S T Ranson. Associated facial, vocal and respiratory components of emotional expression: An experimental study. *J. Neurol. Psychopathol.*, 17(67):241–255, January 1937. doi: 10.1136/jnnp.s1-17.67.241.
67. A Esposito, G Demeurisse, B Alberti, and F Fabbro. Complete mutism after midbrain periaqueductal gray lesion. *Neuroreport*, 10(4):681–685, March 1999. doi: 10.1097/00001756-199903170-00004.
68. H C Bazett and W G Penfield. A study of the sherrington decerebrate animal in the chronic as well as the acute condition. *Brain*, 45(2):185–265, 1922. doi: 10.1093/brain/45.2.185.
69. Owen R Floody and Thomas L O'Donohue. Lesions of the mesencephalic central gray depress ultrasound production and lordosis by female hamsters. *Physiol. Behav.*, 24(1): 79–85, January 1980. doi: 10.1016/0031-9384(80)90017-7.
70. U Jürgens and R Pratt. Role of the periaqueductal grey in vocal expression of emotion. *Brain Res.*, 167(2):367–378, May 1979. doi: 10.1016/0006-8993(79)90830-8.
71. A H Kelly, L E Beaton, and H W Magoun. A midbrain mechanism for facio-vocal activity. *J. Neurophysiol.*, 9:181–189, May 1946. doi: 10.1152/jn.1946.9.3.181.
72. F M Skultety. Experimental mutism in dogs. *Arch. Neurol.*, 6(3):235–241, March 1962. doi: 10.1001/archneur.1962.00450210063007.
73. Joel A Tripp and Steven M Phelps. Females counter-sing, but response to male song differs by sex in alston's singing mouse. *Biol. Lett.*, 20(1):20230484, January 2024. doi: 10.1098/rsbl.2023.0484.
74. S Kyuhou and H Gemba. Two vocalization-related subregions in the midbrain periaqueductal gray of the guinea pig. *Neuroreport*, 9(7):1607–1610, May 1998. doi: 10.1097/00001756-199805110-00064.
75. Eric R Schuppe, Irene Ballagh, Najva Akbari, Wenxuan Fang, Jonathan T Perelmuter, Caleb H Radtke, Margaret A Marchaterre, and Andrew H Bass. Midbrain node for context-specific vocalisation in fish. *Nat. Commun.*, 15(1):189, January 2024. doi: 10.1038/s41467-023-43794-y.
76. Hari H Subramanian, Ron J Balnave, and Gert Holstege. The midbrain periaqueductal gray control of respiration. *J. Neurosci.*, 28(47):12274–12283, November 2008. doi: 10.1523/jneurosci.4168-08.2008.
77. Patryk Ziobro, Yena Woo, Zichen He, and Katherine Tschida. Midbrain neurons important for the production of mouse ultrasonic vocalizations are not required for distress calls. *Curr.*

Biol., 34(5):1107–1113.e3, March 2024. doi: 10.1016/j.cub.2024.01.016.

78. Jonathan I Benichov, Sam E Benezra, Daniela Vallentin, Eitan Globerson, Michael A Long, and Ofer Tchernichovski. The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr. Biol.*, 26(3):309–318, February 2016. doi: 10.1016/j.cub.2015.12.037.
79. Felix W Moll, Devorah Kranz, Ariadna Corredera Asensio, Margot Elmaleh, Lyn A Ackert-Smith, and Michael A Long. Thalamus drives vocal onsets in the zebra finch courtship song. *Nature*, 616(7955):132–136, April 2023. doi: 10.1038/s41586-023-05818-x.
80. James W Satterlee, David Alonso, Pietro Gramazio, Katharine M Jenike, Jia He, Andrea Arrones, Gloria Villanueva, Mariola Plazas, Srividya Ramakrishnan, Matthias Benoit, Iacopo Gentile, Anat Hendelman, Hagai Shohat, Blaine Fitzgerald, Gina M Robitaille, Yumi Green, Kerry Swartwood, Michael J Passalacqua, Edeline Gagnon, Rebecca Hilgenhof, Trevis D Huggins, Georgia C Eizenga, Amit Gur, Twan Rutten, Nils Stein, Shengrui Yao, Adrien Poncet, Clement Bellot, Amy Frary, Sandra Knapp, Mohammed Bendahmane, Tiina Särkinen, Jesse Gillis, Joyce Van Eck, Michael C Schatz, Yuval Eshed, Jaime Prohens, Santiago Vilanova, and Zachary B Lippman. Convergent evolution of plant prickles by repeated gene co-option over deep time. *Science*, 385(6708):eado1663, August 2024. doi: 10.1126/science.ado1663.
81. Neil Shubin, Cliff Tabin, and Sean Carroll. Deep homology and the origins of evolutionary novelty. *Nature*, 457(7231):818–823, February 2009. doi: 10.1038/nature07891.
82. Ryan W Schwark, Matthew J Fuxjager, and Marc F Schmidt. Proposing a neural framework for the evolution of elaborate courtship displays. *Elife*, 11, May 2022. doi: 10.7554/elife.74860.
83. Mukta Chakraborty and Erich D Jarvis. Brain evolution by brain pathway duplication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 370(1684):20150056, December 2015. doi: 10.1098/rstb.2015.0056.
84. Rory T Coleman, Ianessa Morantte, Gabriel T Koreman, Megan L Cheng, Yun Ding, and Vanessa Ruta. A modular circuit coordinates the diversification of courtship strategies. *Nature*, 635(8037):142–150, November 2024. doi: 10.1038/s41586-024-08028-1.
85. Nicholas Jourjine and Hopi E Hoekstra. Expanding evolutionary neuroscience: insights from comparing variation in behavior. *Neuron*, 109(7):1084–1099, April 2021. doi: 10.1016/j.neuron.2021.02.002.
86. Paul S Katz. Evolution of central pattern generators and rhythmic behaviours. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 371(1685):20150057, January 2016. doi: 10.1098/rstb.2015.0057.
87. Frederic A Roemschied, Diego A Pacheco, Max J Aragon, Elise C Ireland, Xinping Li, Kyle Thieringer, Rich Pang, and Mala Murthy. Flexible circuit mechanisms for context-dependent song sequencing. *Nature*, 622(7984):794–801, October 2023. doi: 10.1038/s41586-023-06632-1.
88. Laura F Seeholzer, Max Seppo, David L Stern, and Vanessa Ruta. Evolution of a central neural circuit underlies drosophila mate preferences. *Nature*, 559(7715):564–569, July 2018. doi: 10.1038/s41586-018-0322-9.

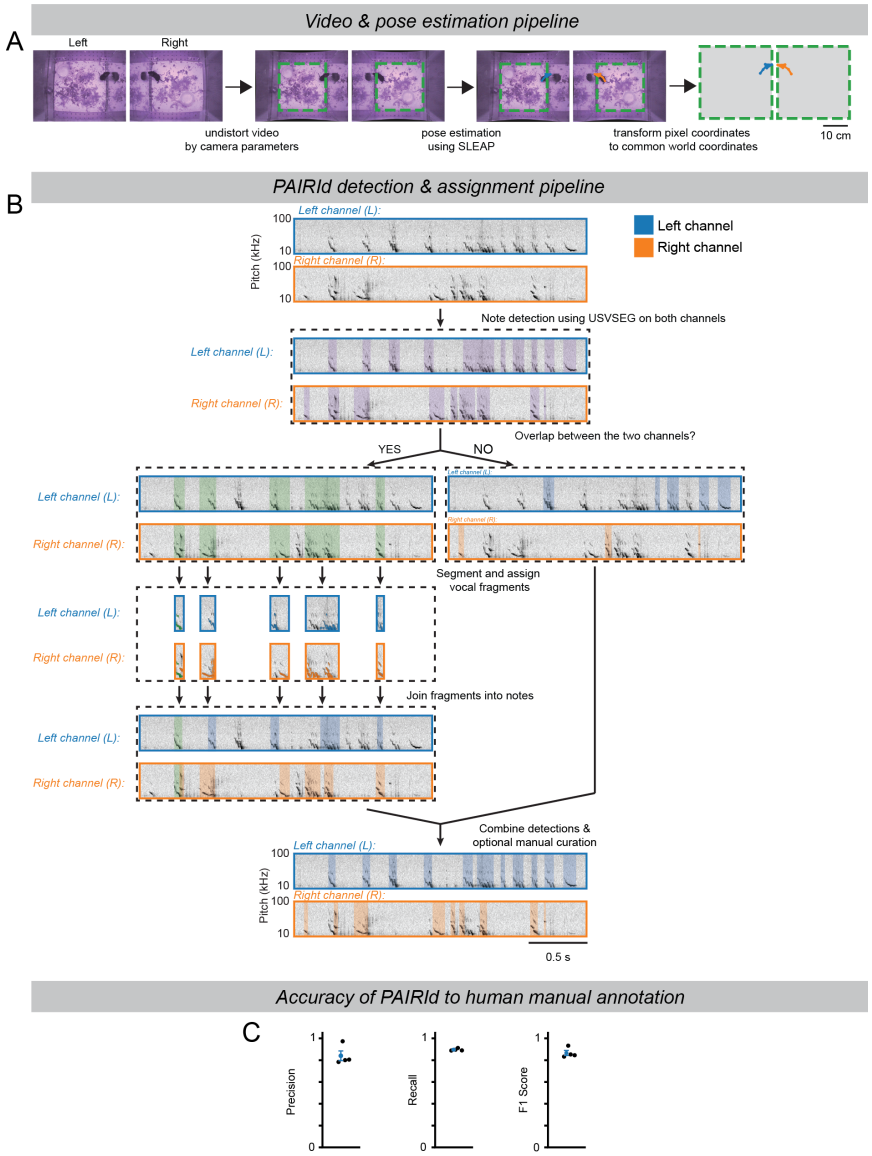

Fig. S1. The PAIRId processing pipelines, related to Figure 1. (A) The PAIRId video processing pipeline aligns two mice in common world coordinates. In steps, the distortion produced by the lens is removed relative to the enclosure floor, followed by pose estimation using SLEAP. The output positions of the 5-node skeleton are then transformed from pixel values to common world coordinates. **(B)** Visual overview of the PAIRId audio processing pipeline to attribute vocalizations to individual mice. **(C)** Accuracy performance metrics of PAIRId assignment compared with ground truth produced via human manual annotation. False positives for a given channel are defined as a PAIRId detection that lacks a match in ground truth detections for either the given channel or unassigned. False negatives for a given channel are defined as a ground truth detection that lacks a match in PAIRId detections for either the given channel or unassigned. *Left*: Precision, or the amount of true positives divided by the total of true positives and false positives, is 0.84 ± 0.04. *Center*: Recall, or the amount of true positives divided by the total of true positives and false negatives, is 0.89 ± 0.01. *Right*: F1 score, the harmonic mean of precision and recall as a combined performance metric, is 0.86 ± 0.02. For each, black circles are the values of the four ground truth hours compared with PAIRId algorithm, and blue circles are the mean ± sem for these four values.

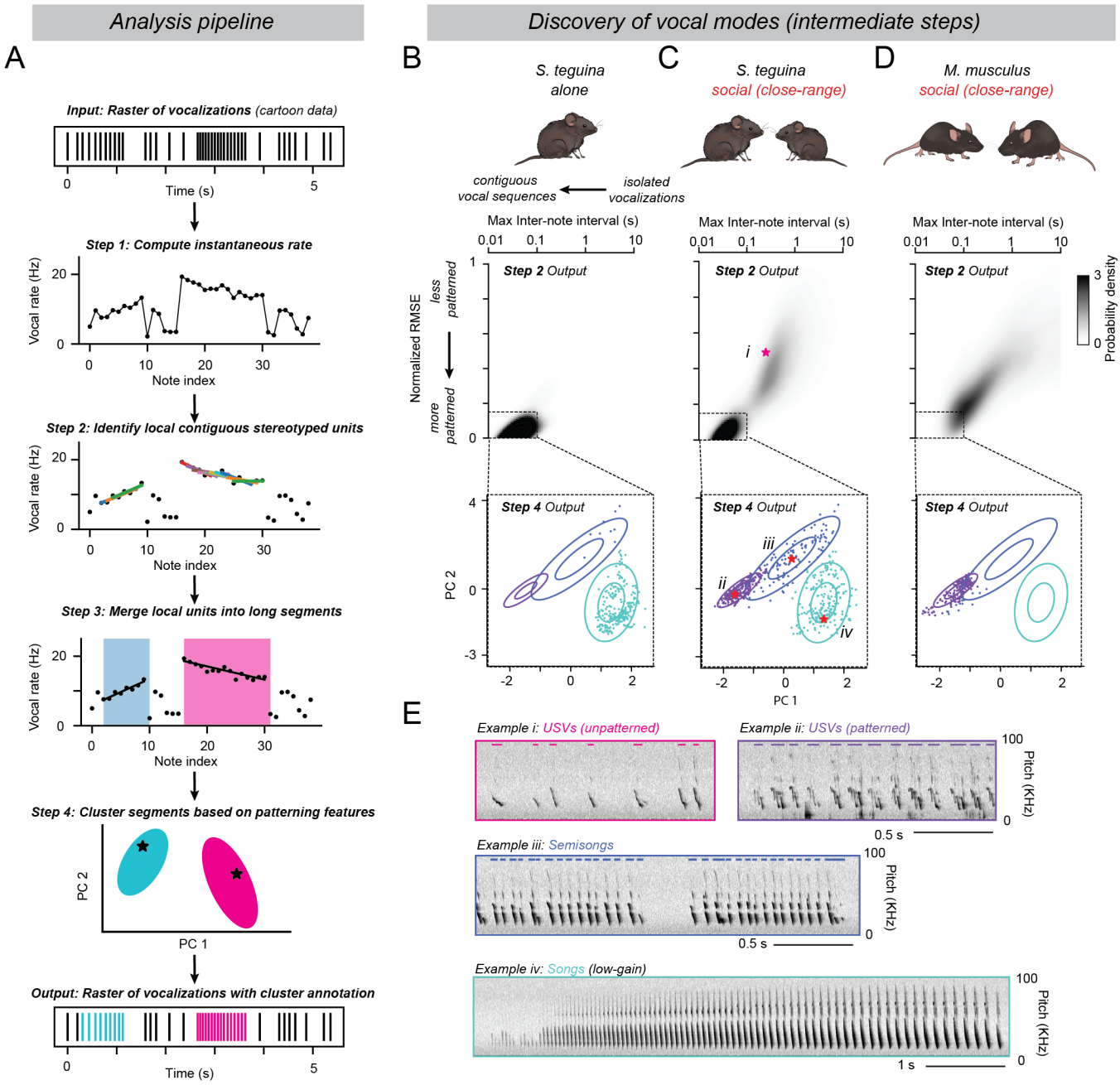

Fig. S2. Temporal patterning analysis hierarchically identifies distinct vocal modes, related to Figure 2. (A) Visual overview of the analysis pipeline for identifying stereotyped vocal production, using cartoon vocalizations as a demonstration. (1) First, instantaneous note rate of vocalizations—the inverse of the difference between two consecutive note-start times—was computed and plotted against note index. (2) Then, using a sliding window of notes, local contiguous units with stereotypy are identified (colored lines). A window is deemed temporally patterned if it exhibits both temporal contiguity (via a low max inter-note interval), and stereotyped note sequencing (via a low NRMSE). (3) Consecutive windows (overlapping and/or with 0.5 s maximum gap separation) meeting these criteria are merged into longer segments (blue and red highlights). Merged segments are characterized via RANSAC regression to robustly fit a linear model (blue and red lines). (4) Key patterning features of the linear model (maximum rate, minimum rate, and number of notes from RANSAC) are used for clustering the patterned segments with a Gaussian mixed model (GMM) on principal components. In this cartoon, the blue segment falls into the cyan cluster and the pink segment into the magenta cluster. (output) Notes from the identified segments are assigned a category based on the clusters from the temporal patterning analysis (cyan, magenta), while others are considered unpatterned (black). **(B-D)** Intermediate steps from the temporal patterning analysis of all vocalizations of the listed conditions: (B) singing mice (*Scotinomys teguina*) alone, (C) singing mice with nearby conspecifics in PAIRId, and (D) laboratory mice (*Mus musculus*) in a nearby social setting. The top plots are the values for all windows from step 2, and dotted lines the thresholds for patterning and contiguity. The bottom plots are the values of the principal components of the features of linear models of patterned, merged segments. Ellipses represent the first and second standard deviations of the clusters found via GMM. **(E)** Representative example spectrograms of vocalizations produced by singing mice of the 4 clusters—(i) unpatterned USVs, (ii) patterned USVs, (iii) semisongs, and (iv) songs. Overlaid colored lines indicate vocalizations of the focal mouse. The temporal features of examples i-iv are denoted with stars in C.

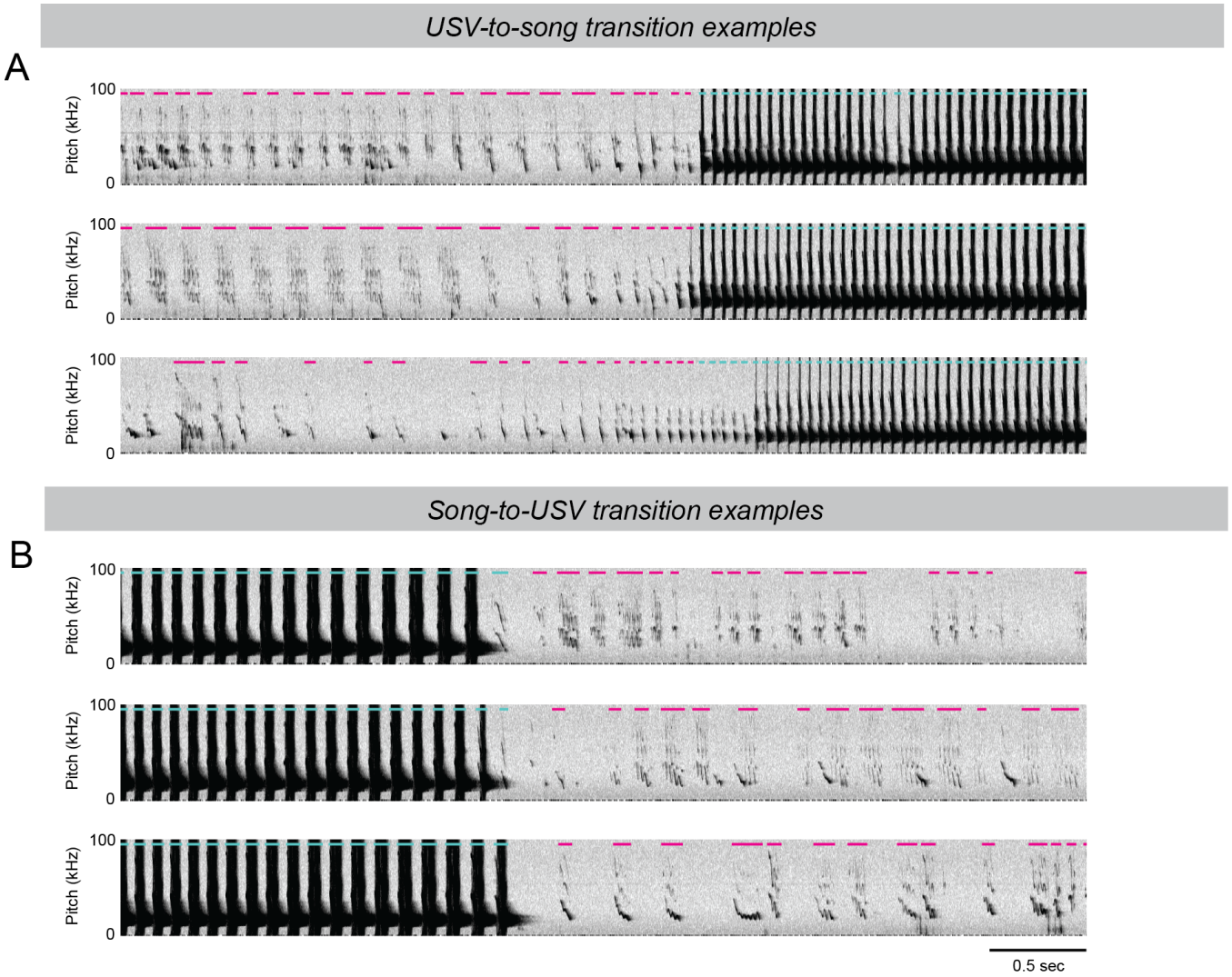

Fig. S3. Rapid, smooth transitions between USVs and songs, related to Figure 2. (A) Representative example spectrograms displaying a singing mouse transitioning from USVs to songs. **(B)** Representative example spectrograms displaying a singing mouse transitioning from the final notes of a song to USVs. Overlaid colored lines indicate vocalizations of the focal mouse, categorized as USV (magenta) or song (cyan).

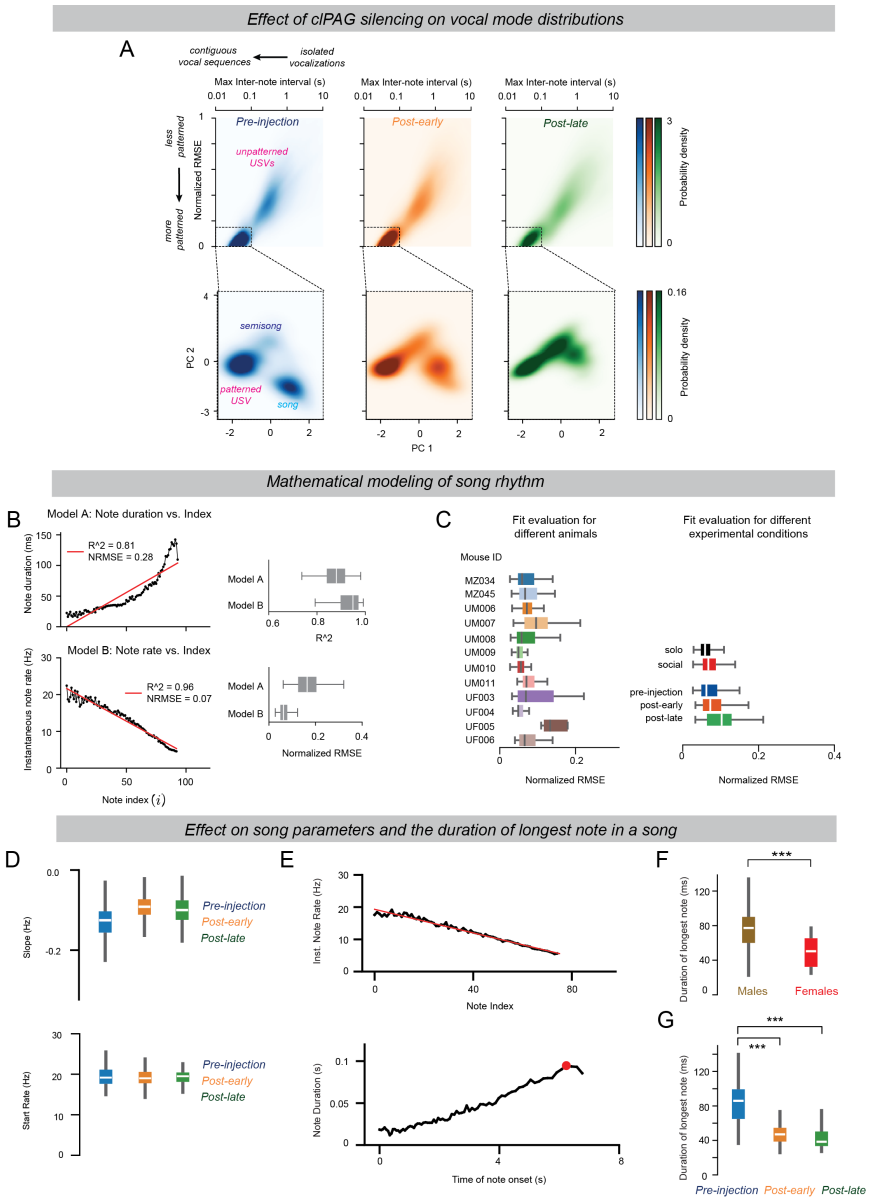

Fig. S4. Effects of TeLC-Mediated Silencing of the clPAG on the Vocal Repertoire, with a Focus on Songs, related to Figures 5-6. (A) Intermediate steps from the temporal patterning analysis of all vocalizations before the TeLC virus injection in the clPAG (left; n = 428 segments from 39,596 notes), during 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). The top panels show the stereotypy and contiguity of local units, and the bottom panels show the temporal features of the merged long segments. **(B)** Alternative parametrizations of song patterning; *top left*: note duration plotted against note index for an example song, with overlaid linear fit; *bottom left*: same as *top left*, but for instantaneous note rate plotted against note index; *top right*: the R^2 values of the two parametrizations for all songs when mice are alone or socially interacting (n = 456 songs; Model A - duration vs. index: 0.86 ± 0.01; Model B - inst. rate vs. index: 0.90 ± 0.01; P = 3.54e-36); *bottom right*: same as *top right*, but for normalized RMSE (Model A: 0.17 ± 0.00; Model B: 0.07 ± 0.00; P = 5.35e-120). **(C)** Normalized RMSE of all songs, grouped by mice (*left*) or by experimental condition (*right*). **(D)** Distribution of song patterning parameters of songs before TeLC virus injection in the clPAG (n = 157 songs), during the post-early period (n = 153 songs), and during the post-late period (n = 54 songs). Top: slope (pre-injection: -0.13 ± 0.00; post-early: -0.09 ± 0.00; post-late: -0.09 ± 0.01; pre-injection vs. post-early: P = 1.15e-14; post-early vs. post-late: P > 0.05; post-early vs. post-late: P = 1.43e-5). Bottom: start rate (pre-injection: 19.38 ± 0.18 Hz; post-early: 18.97 ± 0.18 Hz; post-late: 19.00 ± 0.29 Hz; pre-injection vs. post-early: P > 0.05; post-early vs. post-late: P > 0.05; post-early vs. post-late: P > 0.05). **(D)** Trajectories of an example song visualized using two parameterizations. *Top*: Instantaneous note rate is plotted against note index, same as in Figure 6B. *Bottom*: Note duration is plotted against note onset time. The red circle highlights the longest note in the song. **(E)** Distribution of the duration of the longest note for songs by males and females (male: n = 370 songs, 75.1 ± 1.1 ms; female: n = 86 songs, 49.2 ± 1.3 ms; P = 2.72e-24). **(F)** Same as (E) but for songs before TeLC virus injection in the clPAG, during the post-early period, and during the post late period (pre-injection: n = 157 songs, 84.1 ± 1.8 ms; post-early: n = 153 songs, 48.3 ± 1.0 ms; post-late: n = 54 songs, 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-21).

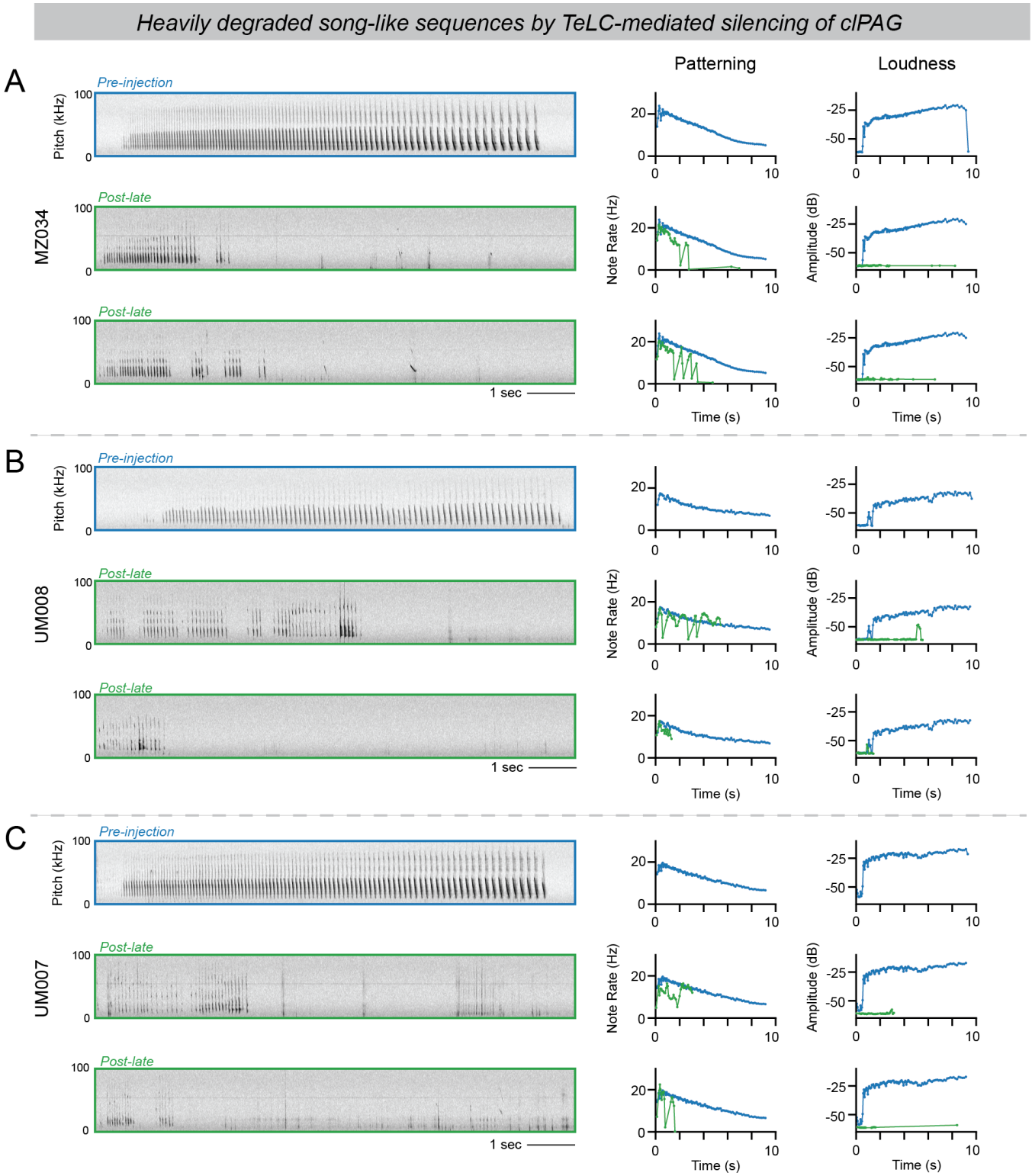

Fig. S5. Examples of final song-like sequences by TeLC-mediated silencing of the clPAG, related to Figures 5-6. (A-C) Examples of the final song-like sequences wherein multiple song properties are heavily degraded as a result of TeLC expression in the clPAG. Each triplet represents a different mouse: the top panel shows a representative pre-injection song with blue-outlined spectrograms and corresponding patterning and loudness traces; the subsequent panels display two final song-like sequences from the post-late period (in green), with the pre-injection traces provided as a reference.

Supplementary Video 1 **A song of an Alston's singing mouse (*S. teguina*), related to Figure 1 and 6.** Top: video recording of an adult male mouse singing. Bottom: spectrogram of the song, illustrating the temporal patterning of note progression.

Supplementary Video 2 **Two singing mice vocalizing in the PAIRId paradigm, related to Figure 1.** Top left: video recordings of each enclosure. Top right: the poses of both mice in common world coordinates. Bottom: spectrograms from each enclosure, overlaid with vocalizations detected and attributed to individual mice. The pitch of high-gain audio was reduced by a factor of eight for human hearing range.

Supplementary Video 3 **Optogenetic activation of the clPAG in a singing mouse elicited USVs, related to Figure 4.** Top: Video recording of the mouse, displaying the 2-second tonic 473 nm laser stimulation. Bottom: Spectrogram of the mouse's vocalizations, demonstrating USVs throughout the stimulation period. The pitch of high-gain audio was reduced by a factor of eight for human hearing range.

Materials & Methods

Experimental Model.

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 (1), 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 (2). 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.*

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" (**p**artial **a**coustic **i**solation **r**eveals **id**entity). 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.

"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 (3). 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 (4). 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 (5, 6). 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 (7). 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 2016. 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.

Stereotaxic viral injection & optogenetic 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 ¼ 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). 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.

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 (described below) as a stimulus to elicit the male's vocalizations, ensuring all recorded vocalizations originated from the implanted male. 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. 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.

Tetanus toxin light chain inactivation 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) (8), 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 the PAIRId setup. 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 S1B for a visual overview.

Quantification of the performance of PAIRId setup. 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 both the percentage of notes assigned solely by the hardware and the percentage of total notes assigned (Figure 1E). Next, we evaluated the full analysis pipeline by benchmarking it against four manually-annotated "ground truth" hours. During these sessions, we assessed the pipeline's performance by quantifying true positives, false positives, and false negatives. For instance, a false positive on the left was defined as an assignment on the left that did not correspond to any ground truth event in either the left or unknown category. Using these definitions, we computed precision, recall, and the F1 score for the annotated hours (Figure S1C).

Curation of vocalizations. Detections were manually curated using a customized spectrogram browser adapted from the open-source MATLAB graphical user interface DeepSqueak (9). 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, semi-songs, and patterned and unpatterned 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 (10). 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:

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

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 (11). 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 S1A for a visual overview.

Social interactions in the PAIRId setup. To determine if the two mice are socially engaged in the PAIRId setup, we first compared the number of vocalizations in the PAIRId setup to those in the "alone" condition. We defined "socially active" hours as those with note counts exceeding the threshold of the mean plus three times the standard deviation in alone hours. For these socially active hours, we further examined both the spatial and temporal organization of vocalizations. Spatially, we examined the locations of the mice in common world coordinates during vocalizations. Temporally, we computed the vocalization rate for each side using a rolling window (30-second window with a 15-second step size) and then calculated the Pearson correlation between the rates from the two sides. We also generated shuffled controls by mismatching the left and right rates from different hours for comparison.

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.

Methods References

1. Daniel E Okobi, Jr, Arkarup Banerjee, Andrew M M Matheson, Steven M Phelps, and Michael A Long. Motor cortical control of vocal interaction in neotropical singing mice. *Science*, 363(6430): 983–988, March 2019. doi: 10.1126/science.aau9480.
2. Josef Bryja and Adam Konecny. Fast sex identification in wild mammals using PCR amplification of the Sry gene. *FOLIA ZOOLOGICA-PRAHA-*, 52(3):269–274, 2003.
3. Emily C Isko, Clifford E Harpole, Xiaoyue Mike Zheng, Huiqing Zhan, Martin B Davis, Anthony M Zador, and Arkarup Banerjee. Selective expansion of motor cortical projections in the evolution of vocal novelty. *bioRxivorg*, October 2024. doi: 10.1101/2024.09.13.612752.
4. Xin Zhao, Patryk Ziobro, Nicole M Pranic, Samantha Chu, Samantha Rabinovich, William Chan, Jennifer Zhao, Caroline Kornbrek, Zichen He, and Katherine A Tschida. Sex- and context-dependent effects of acute isolation on vocal and non-vocal social behaviors in mice. *PLoS One*, 16(9):e0255640, September 2021. doi: 10.1371/journal.pone.0255640.
5. Joshua P Neunuebel, Adam L Taylor, Ben J Arthur, and S E Roian Egnor. Female mice ultrasonically interact with males during courtship displays. *Elife*, 4, May 2015. doi: 10.7554/elife.06203.
6. Max L Sterling, Ruben Teunisse, and Bernhard Englitz. Rodent ultrasonic vocal interaction resolved with millimeter precision using hybrid beamforming. *Elife*, 12, July 2023. doi: 10.7554/elife.86126.
7. Samuel Stuart McAfee, Mary Cameron Ogg, Jordan M Ross, Yu Liu, Max L Fletcher, and Detlef H Heck. Minimally invasive highly precise monitoring of respiratory rhythm in the mouse using an epithelial temperature probe. *J. Neurosci. Methods*, 263:89–94, April 2016. doi: 10.1016/j.jneumeth.2016.02.007.

8. Ryosuke O Tachibana, Kouta Kanno, Shota Okabe, Kohta I Kobayasi, and Kazuo Okanoya. USVSEG: A robust method for segmentation of ultrasonic vocalizations in rodents. *PLoS One*, 15(2): e0228907, February 2020. doi: 10.1371/journal.pone.0228907.
9. Kevin R Coffey, Ruby E Marx, and John F Neumaier. DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology*, 44(5):859–868, April 2019. doi: 10.1038/s41386-018-0303-6.
10. Gregg A Castellucci, Daniel Calbick, and David McCormick. The temporal organization of mouse ultrasonic vocalizations. *PLoS One*, 13(10):e0199929, October 2018. doi: 10.1371/journal.pone.0199929.
11. Talmo D Pereira, Nathaniel Tabris, Arie Matsliah, David M Turner, Junyu Li, Shruthi Ravindranath, Eleni S Papadoyannis, Edna Normand, David S Deutsch, Z Yan Wang, Grace C McKenzie-Smith, Catalin C Mitelut, Marielisa Diez Castro, John D'Uva, Mikhail Kislin, Dan H Sanes, Sarah D Kocher, Samuel S-H Wang, Annegret L Falkner, Joshua W Shaevitz, and Mala Murthy. SLEAP: A deep learning system for multi-animal pose tracking. *Nat. Methods*, 19(4):486–495, April 2022. doi: 10.1038/s41592-022-01426-1.