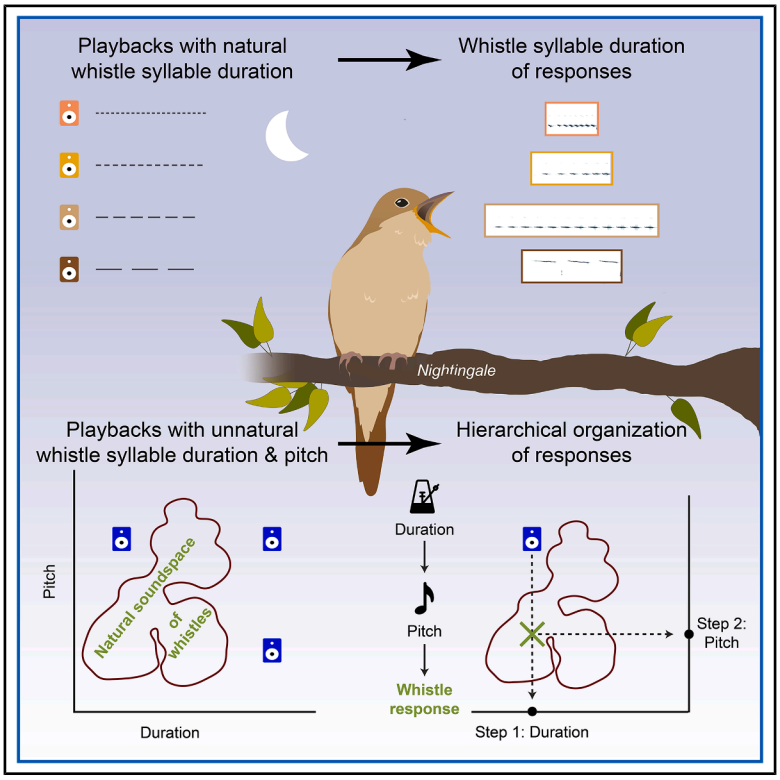
Report

Current Biology

Interplay between syllable duration and pitch during whistle matching in wild nightingales

Graphical abstract



Highlights

- Nightingales approximate whistle syllable duration when imitating whistle songs
- Nightingales match the pitch or duration of unnatural whistles, not both together
- Modeling reveals hierarchical organization of acoustic feature-matching

Authors

Juan Sebastián Calderón-García, Giacomo Costalunga, Tim P. Vogels, Daniela Vallentin

Correspondence

daniela.vallentin@bi.mpg.de

In brief

Calderón-García, Costalunga, et al. find that nightingales respond to whistle syllable playbacks of unnatural combinations of pitch and duration by matching either the spectral or temporal feature. Computational modeling reveals a hierarchical organization of acoustic features during whistle matching, with syllable duration imposing the initial constraint on the whistle response.

Calderón-García et al., 2026, Current Biology *36*, 791–798
February 2, 2026
https://doi.org/10.1016/j.cub.2025.12.025

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

Interplay between syllable duration and pitch during whistle matching in wild nightingales

Juan Sebastián Calderón-García,[1,3] Giacomo Costalunga,[2,3] Tim P. Vogels,[1] and Daniela Vallentin[2,4,*]
[1]Institute of Science and Technology Austria, Klosterneuburg 3400, Austria
[2]Neural Circuits for Vocal Communication Research Group, Max Planck Institute for Biological Intelligence, Eberhard-Gwinner-Str., Seewiesen 82319, Germany
[3]These authors contributed equally
[4]Lead contact
*Correspondence: daniela.vallentin@bi.mpg.de
https://doi.org/10.1016/j.cub.2025.12.025

SUMMARY

During complex vocal interactions, different features of acoustic stimuli are integrated to produce appropriate vocal responses,[1] such as copying sounds during vocal matching behavior in some animals.[2–12] However, little is known about the interplay and possible trade-offs between the different temporal and spectral acoustic features during these vocal exchanges.[2,13,14] Nightingales can flexibly match the pitch of their tonal "whistle songs" in real time during counter-singing duels.[15,16] Here, we show that the syllable duration of whistle playbacks could alter the song responses of wild nightingales, causing their whistle duration distribution to shift toward the presented stimulus duration. When exposed to whistle playbacks featuring unnatural combinations of pitch and duration, nightingales demonstrate a flexible trade-off between pitch matching and temporal imitation, yet they are constrained by their vocal repertoire. They selectively adapted their vocal responses to approximate these novel stimuli, aligning them with their natural whistle repertoire. We developed a computational model of nightingale whistle-matching behavior that revealed a hierarchical organization of acoustic feature production. During whistle matching, the feature integration process is constrained by the duration of syllables, and pitch matching follows within this temporal framework, forcing a trade-off between the two features. Our findings reveal a complex interplay between the spectral and temporal domains that shapes song-matching behavior.

RESULTS AND DISCUSSION

Few animal species, including some songbirds, are capable of vocal matching, adjusting their vocalizations to copy conspecifics.[2,4–8,10–12,15] During vocal imitation, spectral and temporal acoustic features of the matched vocalization are processed and produced during the matching response.[2,3,14] How the different features of the target vocalization are influencing vocal matching is largely unknown. We explored this phenomenon in wild nightingales, a species known for natural song-matching, during which they listen to and immediately replicate the songs of other nightingales.[3,17,18] We focused on their "whistle songs," sequences of unmodulated tonal syllables (Figure 1A), that nightingales are known to pitch-match during counter-singing duels.[15,16] Previous work presenting nightingales with playbacks demonstrated that the pitch of the stimuli is precisely integrated in the spectral features of the whistle responses.[15] However, nightingales' whistles' syllables can differ not only in pitch but also in duration, suggesting that the temporal component might also play a role in the matching response.

Nightingales approximate whistle syllable duration during whistle matching

Wild nightingales sing whistle syllables that fall into three duration clusters: short (<140 ms), medium (140–310 ms), and long syllables (>310 ms) (Figures 1A, 1B, and S1A–S1E; Audio S1; STAR Methods). We reanalyzed data from previous playback experiments that consisted of artificial whistles of different pitches but of the same duration (440 ms, within the normal distribution of long whistle syllables).[15] We observed a shift in the whistle syllable distribution toward the duration of the playback syllables, indicating an influence of the stimulus' temporal features in the pitch-matching responses (Figures S1F and S1G). Next, we systematically tested whether the duration of whistle syllables in the playbacks, independent of their pitch, influenced the birds' whistle responses. Therefore, we exposed wild nightingales to artificial whistle playbacks with a fixed pitch (2,390 Hz)[15] but varying whistle syllable durations (Figures 1B and 1C). To test whether nightingales would shift their natural whistle syllable-duration distribution, we presented playbacks containing syllables from the four less-frequent temporal regimes: very short, medium short, medium long, and very long (Figures 1B, 1C, and S1E; Audio S1). Although the overall range of whistle syllable durations in response was broad, we observed a clear shift in the distribution of these responses toward the duration of the very-short, medium-short, and very-long playbacks (Figures 1D and S2A–S2H). We developed an "occurrence index" (STAR Methods) to assess if whistles of specific durations occurred more or less frequently during playback conditions than expected by chance. An observed occurrence index greater than

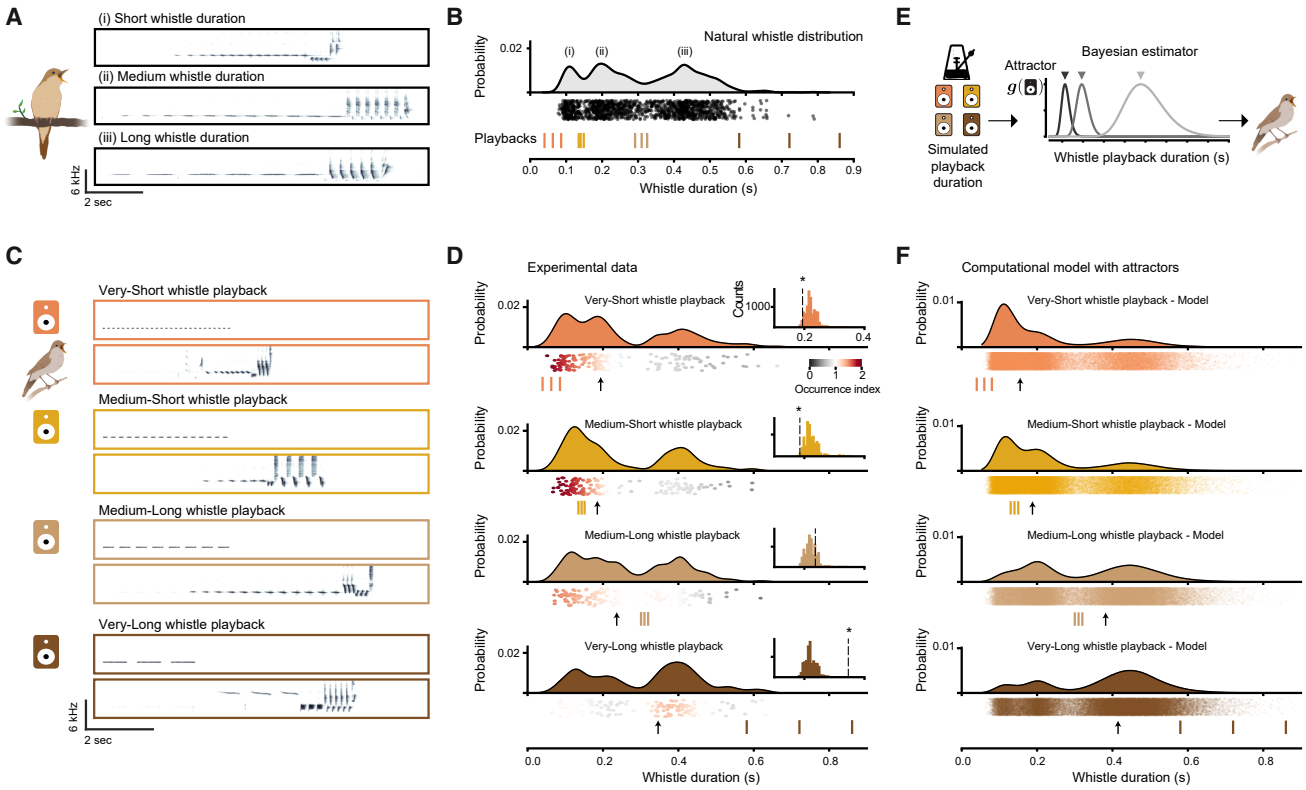


Figure 1. Nightingales' approximate whistle syllable duration during whistle matching

(A) Three examples of whistle song spectrograms with (i) short, (ii) medium, and (iii) long whistle syllable durations. See also Audio S1.

(B) Top: natural distribution of whistle syllable durations (0.294 ± 0.124 s; n = 1,237 whistles from N = 11 birds). The three modes of the probability distribution correspond to the short (i), medium (ii), and long (iii) syllable durations shown in (A). Middle: each black circle corresponds to the median whistle syllable duration of whistle syllables of individual whistle songs. Bottom: vertical color lines represent the durations of playback stimuli (orange, very-short; gold, medium-short; sand, medium-long; brown, very-long playbacks).

(C) Example spectrograms of playbacks and nightingale responses for the four different playback regimes. See also Audio S1.

(D) Probability distributions of nightingale whistle responses to each playback regime: very short (0.194 ± 0.096 s, n = 172 whistles from N = 7 birds), medium short (0.185 ± 0.075 s, n = 155, N = 7), medium long (0.236 ± 0.112 s, n = 153 whistles, N = 7 birds), and very long (0.346 ± 0.111 s, n = 152 whistles, N = 7 birds). Arrows represent the median. Individual whistle song responses were color-coded based on the relative frequency of occurrences within the song duration compared with the naturally observed distribution, calculated as the occurrence index (STAR Methods). Inserts show the distribution of medians, calculated from 10,000 permutations of shuffled data for each playback regime. Observed medians are indicated with black lines (very short, p = 0.0457; medium short, p = 0.02; medium long, p = 0.4718; and very long, p = 0.0001).

(E) Simulated playback durations act as attractor inputs that bias a trimodal Gaussian mixture of whistle durations. The Bayesian estimator modulates the Gaussian weights by means of an attractor function g that depends on the duration of the playback and its distance to the mean of each Gaussian. Triangles represent the mean of each Gaussian: 0.112, 0.222, and 0.444 s.

(F) Probability distributions of simulated whistle responses to each playback regime: very short (0.155 ± 0.053 s, n = 52,500 simulated whistles, N = 7 simulated birds), medium short (0.188 ± 0.0696 s, n = 52,500 simulated whistles, N = 7 simulated birds), medium long (0.355 ± 0.135 s, n = 52,500 simulated whistles, N = 7 simulated birds), and very long (0.414 ± 0.0833 s, n = 52,500 simulated whistles, N = 7 simulated birds). Arrows represent the median.

See also Figures S1–S3.

one in the vicinity of playback durations indicated a targeted shift of whistle syllable duration toward the stimuli duration (Figure 1D). Birds responded to playbacks of a given whistle duration with different whistle song types, indicating flexible temporal modulation independent of a specific song structure (Figures S2I and S2J). These results indicate that the temporal features of whistle stimuli are integrated in the production of whistle-matching responses, in addition to the previously described influence of the stimulus' pitch.[15]

To better understand these behavioral dynamics, we constructed a computational model in which the duration of simulated whistle responses depended on a Bayesian estimator using the distribution of natural whistle durations as a prior (Figures 1E, S3A, and S3B). Simulated playbacks acted as attractor inputs, biasing a trimodal Gaussian mixture model (GMM) that represented whistle durations. The weight of each Gaussian was updated by an attractor function that depended on the distance between the playback duration and each Gaussian's mean, increasing the likelihood of responses near the corresponding Gaussian while down-weighting more distant ones (STAR Methods). The strength of this bias was controlled by an attractor gain β and a parameter δ reflecting the drive of the bird to match the playback ("matching drive"; STAR Methods). Using this framework, the model successfully reproduced the

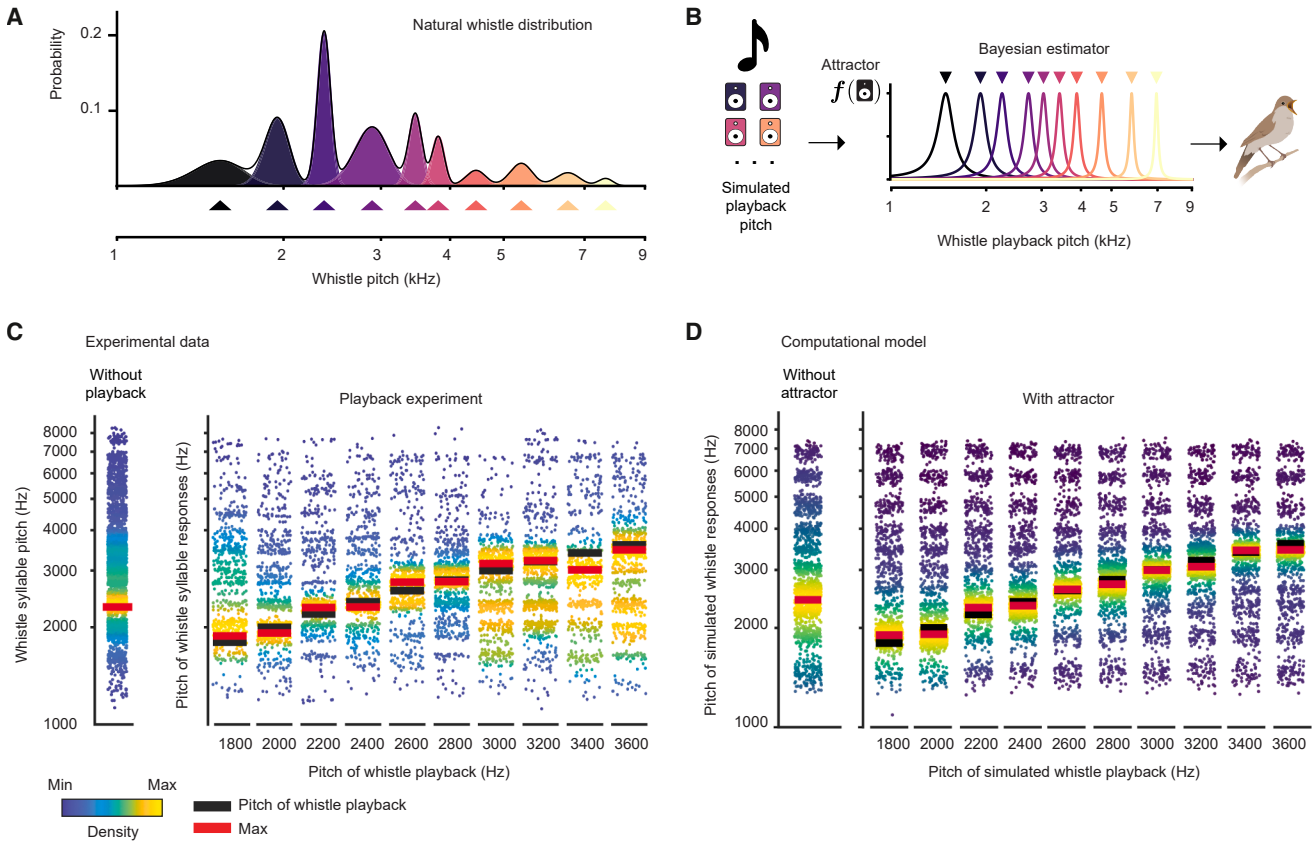


Figure 2. Computational model describing whistle pitch-matching behavior of nightingales

(A) Distribution of whistle pitches in natural whistle songs, modeled as a multimodal Gaussian mixture with ten Gaussians spanning low to high frequencies. Triangles indicate the mean of the ten Gaussians: 1.53, 1.94, 2.36, 2.89, 3.46, 3.80, 4.45, 5.37, 6.53, and 7.64 kHz.

(B) Simulated playback pitches act as attractor inputs that bias a multimodal Gaussian mixture of whistle pitches. The Bayesian estimator modulates the Gaussian weights by means of an attractor function f that depends on the pitch of the playback and its distance to the mean of each Gaussian. Triangles represent the mean of each Gaussian.

(C) Experimental data showing whistle syllable pitches sung without playback during the pre-playback control phase (left) and in response to whistle playbacks (right). The horizontal black and red bars represent the pitch of the whistle playback and the global maximum of the responses to the given playback, respectively. Adapted from Costalunga et al.[15]

(D) Model simulations without (left) and with (right) the attractor term, showing that the attractor reproduces the stimulus-dependent clustering of pitch responses observed in the experiment.

See also Figure S3.

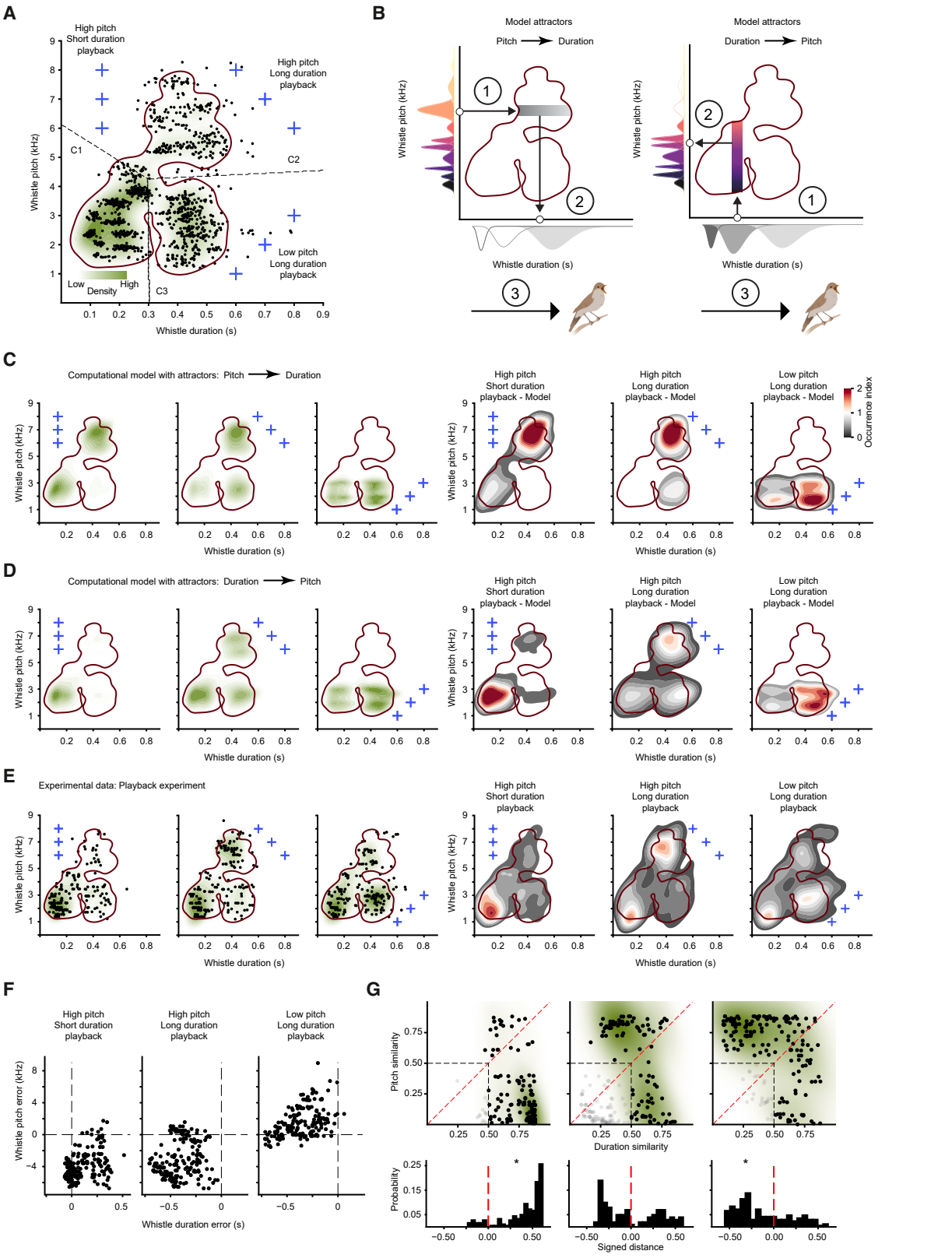
distribution of whistle durations across the four observed categories (very-short, medium-short, medium-long, and very-long playbacks) (Figures 1F, S3C, and S3D).

Just as whistle duration served as an attractor input in the temporal model, we next asked whether the same framework could generalize to also explain whistle pitch matching. We applied the model using the natural multimodal distribution of whistle syllable pitches as the prior, represented as a 10-component Gaussian mixture (Figures 2A and S3E–S3G). Playback pitches acted as attractor inputs that biased the Gaussian weights through a smooth distance-decay function, increasing the probability of responses near the corresponding pitch Gaussian while down-weighting more distant ones (Figure 2B). As in the temporal model, we included a matching drive parameter δ, with the attractor gain β controlling the strength of the bias. The model successfully reproduced the pitch-matching responses observed in previous work[15] (Figures 2C, 2D, and S3D–S3H), demonstrating that the attractor-based estimator can account for both temporal and spectral matching domains.

Hierarchical integration of whistle syllable duration and pitch during whistle matching

In the syllable-duration experiment, our playback stimuli consisted of the same pitch (2,390 Hz)[15] that naturally occurs in syllables of different durations. However, whistle songs naturally produced by singing nightingales comprise a broad and multimodal distribution of pitches and durations[15] (Figure 3A). To assess this multimodality, we applied k-means clustering to the distribution of whistle syllables in the soundspace (duration-pitch feature space). We found three main clusters (C1, low-pitched/short-duration syllables; C2, high-pitched syllables; and C3, low-pitched/long-duration syllables; Figures 3A, S4A, and S4B).

Based on the observed distribution of the data in the soundspace, we suspected a potential trade-off between pitch



(legend on next page)

matching and temporal imitation. Thus, we combined the duration and pitch models into a unified joint framework, in which each simulated bird is described simultaneously by a duration GMM and a pitch GMM (Figure 3B). Playback stimuli updated both temporal and spectral Gaussian weights through their respective attractor functions. To test how the temporal and spectral domains are hierarchically represented, we defined hierarchical support as the dimension that determines the direction of constraint in the joint response space. In the pitch→duration condition, playback pitch constrains the whistle syllable-duration responses to the restricted natural pitch-duration response region (Figures 3A and 3B, left). Conversely, in the duration→pitch condition, pitch responses were limited by the syllable-duration constraint imposed by the playback stimulus (Figures 3A and 3B, right). This setup allowed us to test two competing hypotheses of hierarchical organization: whether temporal structure constrains spectral responses or vice versa.

With this framework, we explored the interplay of spectral and temporal features when nightingales were whistle-matching playbacks consisting of a combination of pitches and durations that typically does not occur in their natural repertoire. Therefore, we simulated three representative playback types that combined common and uncommon temporal and spectral features: high pitch/short duration, high pitch/long duration, and low pitch/long duration (Figure 3A). Each playback type acted as attractor input under both hierarchical configurations (pitch→duration and duration→pitch). In the pitch→duration simulations, responses concentrated at pitch values close to the playback but spread across durations, leading to selective increases in the C2 cluster for high-pitch/short-duration and high-pitch/long-duration playbacks and in the C3 cluster for low-pitch/long-duration playbacks (Figure 3C). In the duration→pitch simulations, responses were anchored to the duration modes, with pitch spreading broadly: high-pitch/short-duration playbacks drove responses into the C1 cluster, high pitch/long duration into the C2 cluster, and low pitch/long duration into the C3 cluster (Figure 3D).

To experimentally test the prediction from the model, we presented birds with the same set of playbacks as used in the model (high pitch/short duration, high pitch/very long duration, low pitch/very long duration) and assessed the birds' whistle-matching behavior (Figures 3A and 3E). The responses to the different playback types revealed a shift in the soundspace probability density distribution of whistles compared with the natural whistle distribution (Figures 3E and S4C–S4E; STAR Methods). For each playback type, we computed the proportion of responses attributable to each of the three clusters defined by the k-means algorithm and compared them to the expected proportions from shuffled datasets (Figure S4F; STAR Methods). When exposed to high-pitch/short-duration playbacks (Figure 3A), nightingales targeted the duration but not the pitch (Figure 3E, first), resulting in an increase in whistle responses in the C1 cluster and a reduction in the C2 cluster (Figure S4F). Conversely, for high-pitch/very-long-duration playbacks (Figure 3A), the birds matched the pitch (Figure 3E, second), increasing the production of whistles in the C2 cluster (Figure S4F). Finally, for low-pitch/very-long-duration playbacks (Figure 3A), nightingales matched the pitch and lengthened whistles, but not to the full syllable durations featured in the playback (Figure 3E, third), demonstrated by a reduction of whistles in the C1 and C2 clusters and an increase in cluster C3 (Figure S4F). These results closely align with the duration→pitch model predictions (Figure 3D) and support a hierarchical representation in which syllable-duration

Figure 3. Interplay between syllable duration and pitch during whistle matching

(A) Natural distribution of whistle syllable duration and pitch (n = 1,237 whistles, N = 11 birds). Each black circle corresponds to the median whistle syllable duration and pitch of whistle syllables in individual whistle songs. The blue crosses represent the duration and pitch of the three types of playback stimuli (high pitch/short duration, high pitch/very long duration, and low pitch/very long duration). The green shading represents the kernel density estimation (KDE) of the whistle responses in the soundspace (STAR Methods). The brown line encloses 90% of the control KDE density. The dashed lines represent the linear boundaries of the clusters (C1, C2, and C3) defined by a k-means algorithm (STAR Methods).

(B) In the combined model, both duration and pitch Gaussian weights are updated simultaneously by their attractor functions. Left: in the pitch→duration configuration, pitch provides the attractor input, and duration responses are constrained to the natural region of pitch-duration space observed in nightingale whistles. Right: in the duration→pitch configuration, duration provides the attractor input, and pitch responses are constrained to the same region. Numbers indicate hierarchical steps: pitch integration (1 or 2), duration integration (1 or 2), and song production (3).

(C) Left: distribution of simulated whistle durations and pitches in response to the three types of playback stimuli (high pitch/short duration, n = 1,050 simulated whistles, N = 7 simulated birds; high pitch/very long duration, n = 1,050 simulated whistles, N = 7 simulated birds; low pitch/very long duration, n = 1,050 simulated whistles, N = 7 simulated birds) represented as in (A). Right: density estimation of responses color-coded based on the relative density of occurrences compared with the naturally observed distribution, calculated as the occurrence index of observed to expected densities (STAR Methods). The brown line encloses 90% of the control KDE density.

(D) Left: distribution of simulated whistle duration and pitch in response to the three types of playback stimuli (high pitch/short duration, n = 1,050 simulated whistles, N = 7 simulated birds; high pitch/very long duration, n = 1,050 simulated whistles, N = 7 simulated birds; low pitch/very long duration, n = 1,050 simulated whistles, N = 7 simulated birds) represented as in (A). Right: density estimation of responses color-coded based on the relative density of occurrences compared with the naturally observed distribution, calculated as the occurrence index of observed to expected densities (STAR Methods). The brown line encloses 90% of the control KDE density.

(E) Left: distribution of whistle syllable durations and pitches in response to the three types of playback stimuli (high pitch/short duration, n = 193 whistles, N = 8 birds; high pitch/very long duration, n = 194 whistles, N = 8 birds; low pitch/very long duration, n = 190 whistles, N = 8 birds) represented as in (A). Right: density estimation of responses color-coded based on the relative density of occurrences compared with the naturally observed distribution, calculated as the occurrence index of observed to expected densities (STAR Methods). The brown line encloses 90% of the control KDE density.

(F) Distribution of differences in duration and pitch in response to the three types of playback stimuli. Dashed lines represent no difference in duration and pitch.

(G) Top: distribution of duration and pitch similarity in response to the three types of playback stimuli (STAR Methods). The green shading represents the KDE of the whistle responses in the similarity space. Bottom: distribution of signed distances between duration and pitch similarity scores (range 0–1) for responses with high similarity (i.e., similarity > 0.5 in either dimension; STAR Methods). The red dashed line marks zero distance (equal pitch and duration similarity) (Wilcoxon signed-rank tests: high pitch/short duration, $p < 0.001$; high pitch/very long duration, $p = 0.7129$; low pitch/very long duration, $p < 0.001$).

See also Figure S4.

constrains pitch responses within the natural pitch-duration space.

We further characterized the trade-off between syllable pitch and duration in response to our playback types by investigating the respective distance from the playbacks in pitch and duration (Figure 3F). For each response, we calculated a duration and pitch similarity index (Figures 3G and S4G; STAR Methods). When exposed to high-pitch/short-duration playbacks, the responses of the birds showed high duration similarity (Figure 3G, left). For high-pitch/very-long-duration playbacks, when nightingales exhibited high pitch similarity (Figure 3G, second), the whistle responses showed an increase also in duration similarity, without a preferred tendency toward one or the other feature (Figure 3G, middle). For low-pitch/very-long-duration playbacks, although the nightingales produced more long-duration syllables (Figure 3G, third), the birds responded with whistles that had higher pitch similarity than duration (Figure 3G, right). Our results suggest a flexible trade-off strategy for prioritizing acoustic features to emulate sounds outside of the natural whistle syllable repertoire, constrained by the vocal repertoire of the birds. Nightingales adjust their matching responses by projecting novel whistles into their bounded whistle soundspace through selective imitation of temporal and/or spectral features, providing evidence for performance constraints[13,19] during song-matching. We concluded that counter-singing nightingales exhibit vocal complexity not only through pitch matching[15] but also through integrating the duration of syllables in their whistle matches.

Modulating both spectral and temporal features of vocalizations is key to dynamic vocal coordination in humans as part of a broader phenomenon known as conversational entrainment or speech accommodation: speakers often unconsciously adapt features such as utterance duration, speech rate, and turn-taking timing to match their partner.[20–23] The ability to specifically imitate both the temporal and spectral features of whistle stimuli suggests a capacity for auditory discrimination, allowing nightingales to isolate individual vocal features from complex sounds. The hierarchical organization of pitch and duration observed in our modeling and experimental results may reflect interacting constraints at both the auditory discrimination and song production levels. When discriminating auditory inputs, nightingales may represent temporal patterns of whistle syllables categorically,[24,25] with only a few temporal attractors. Within each temporal mode, spectral processing might then operate at a finer resolution.[26,27] During song production, the coupling between respiratory and syringeal muscle control[28,29] may impose mechanical limits to the durations of whistle syllables. By contrast, pitch control may allow for more continuous and precise modulation once a temporal framework has been established. Investigating how auditory discrimination and song performance constraints of adult birds[13,19] interact with their song learning history[30,31] could reveal the proximate biomechanical, developmental, and neurophysiological mechanisms that influence whistle matching and vocal imitation. Our current understanding of the neural circuit underlying song production is largely based on songbird species with limited flexibility in singing behavior.[32,33] In zebra finches, for instance, persistent frequency or temporal changes in the song structure of adult birds can only be experimentally induced through long-term reinforcement learning.[34–36] In these birds, changes in the social context produce only minor adjustments in the singing performance.[37,38] This contrasts with the rapid song adjustments observed in nightingales, suggesting a species-specific adaptation in neural vocal control that enables real-time integration and production of songs[3] to imitate temporal and pitch variations in ambient whistle songs. Investigating such behavior and its neural correlates could shed light on broader mechanisms for song production, vocal learning, and real-time acoustic modifications during interactive communication.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Daniela Vallentin (daniela.vallentin@bi.mpg.de).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

- All data have been deposited at https://github.com/vallentinlab/NG-whistle-durations and are publicly available as of the date of publication.
- All original code has been deposited at https://github.com/vallentinlab/NG-whistle-durations and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We would like to thank J. Benichov and N. Hein for their help with fieldwork; M. Ramadas for helping with the segmentation analysis; T. Eliav, C. Chintaluri, G. Tkacik, and A. Navas for providing helpful comments to the project and manuscript; and A. Costalunga for the drawings of nightingales. Funding sources: The Joachim Herz Stiftung Add-on Fellowships for Interdisciplinary Life Science, awarded to G.C.; the ERC Consolidator Grant 819603 SYNAPSEEK, awarded to T.P.V.; and DFG Research Unit 5768–532521431, DFG Research Grant-547921981, DFG SFB 1315–327654276, and the ERC Starting Grant 757459 MIDNIGHT, awarded to D.V.

AUTHOR CONTRIBUTIONS

J.S.C.-G. and G.C. contributed equally. J.S.C.-G., G.C., and D.V. conceived, designed, and conducted the experiments; analyzed the data; and crafted the figures. J.S.C.-G. and G.C. wrote the first draft of the manuscript. All authors participated in writing and editing the manuscript. G.C., T.P.V., and D.V. acquired funding. T.P.V. and D.V. supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Generation of artificial playbacks
  - Audio recordings
  - Audio data processing and analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of previous pitch experiment
  - Clustering of whistle syllables duration (Experiment 1)
  - Clustering of whistle syllables duration and frequency (Experiment 2)
  - Whistle syllable duration variation within whistle songs

- Generating shuffled data
- Occurrence index of durations (Experiment 1)
- Occurrence index of whistle durations and pitch (Experiment 2)
- Durations and pitch similarity index (Experiment 2)
- Statistical analysis for Experiment 1
- Statistical analysis for Experiment 2

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2025.12.025.

Received: April 30, 2025
Revised: October 10, 2025
Accepted: December 9, 2025
Published: January 12, 2026

REFERENCES

1. Castellucci, G.A., Guenther, F.H., and Long, M.A. (2022). A theoretical framework for human and nonhuman vocal interaction. Annu. Rev. Neurosci. *45*, 295–316. https://doi.org/10.1146/annurev-neuro-111020-094807.
2. King, S.L., and McGregor, P.K. (2016). Vocal matching: the what, the why and the how. Biol. Lett. *12*, 20160666. https://doi.org/10.1098/rsbl.2016.0666.
3. Costalunga, G., Vallentin, D., and Benichov, J.I. (2024). A neuroethological view of the multifaceted sensory influences on birdsong. Curr. Opin. Neurobiol. *86*, 102867. https://doi.org/10.1016/j.conb.2024.102867.
4. Janik, V.M. (2000). Whistle matching in wild bottlenose dolphins (Tursiops truncatus). Science *289*, 1355–1357. https://doi.org/10.1126/science.289.5483.1355.
5. Balsby, T.J.S., and Bradbury, J.W. (2009). Vocal matching by orange-fronted conures (*Aratinga canicularis*). Behav. Processes *82*, 133–139. https://doi.org/10.1016/j.beproc.2009.05.005.
6. Krebs, J.R., Ashcroft, R., and Orsdol, K.V. (1981). Song matching in the great tit *Parus major* L.. Anim. Behav. *29*, 918–923. https://doi.org/10.1016/S0003-3472(81)80029-2.
7. Hinde, R.A. (1958). Alternative motor patterns in chaffinch song. Anim. Behav. *6*, 211–218. https://doi.org/10.1016/0003-3472(58)90053-8.
8. King, S.L., Sayigh, L.S., Wells, R.S., Fellner, W., and Janik, V.M. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. Proc. Biol. Sci. *280*, 20130053. https://doi.org/10.1098/rspb.2013.0053.
9. Balsby, T.J.S., Momberg, J.V., and Dabelsteen, T. (2012). Vocal imitation in parrots allows addressing of specific individuals in a dynamic communication network. PLOS One *7*, e49747. https://doi.org/10.1371/journal.pone.0049747.
10. Miller, P.J.O., Shapiro, A.D., Tyack, P.L., and Solow, A.R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. Anim. Behav. *67*, 1099–1107. https://doi.org/10.1016/j.anbehav.2003.06.017.
11. Schulz, T.M., Whitehead, H., Gero, S., and Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. Anim. Behav. *76*, 1977–1988. https://doi.org/10.1016/j.anbehav.2008.07.032.
12. Beecher, M.D., Campbell, S.E., Burt, J.M., Hill, C.E., and Nordby, J.C. (2000). Song-type matching between neighbouring song sparrows. Anim. Behav. *59*, 21–27. https://doi.org/10.1006/anbe.1999.1276.
13. Podos, J., and Nowicki, S. (2004). Performance limits on birdsong. Chapter 11. In Nature's Music, P. Marler, and H. Slabbekoorn, eds. (Academic Press), pp. 318–342. https://doi.org/10.1016/B978-012473070-0/50014-1.
14. Mercado, E.I.I.I., Mantell, J.T., and Pfordresher, P.Q. (2014). Imitating sounds: a cognitive approach to understanding vocal imitation. Comp. Cogn. Behav. Rev. *9*, 1–57. https://doi.org/10.3819/ccbr.2014.90002.
15. Costalunga, G., Carpena, C.S., Seltmann, S., Benichov, J.I., and Vallentin, D. (2023). Wild nightingales flexibly match whistle pitch in real time. Curr. Biol. *33*, 3169–3178.e3. https://doi.org/10.1016/j.cub.2023.06.044.
16. Naguib, M., Mundry, R., Hultsch, H., and Todt, D. (2002). Responses to playback of whistle songs and normal songs in male nightingales: effects of song category, whistle pitch, and distance. Behav. Ecol. Sociobiol. *52*, 216–223. https://doi.org/10.1007/s00265-002-0511-1.
17. Todt, D. (1971). Äquivalente und konvalente gesangliche Reaktionen einer extrem regelmäßig singenden Nachtigall (Luscinia megarhynchos L.). Z. Vergl. Physiol. *71*, 262–285. https://doi.org/10.1007/BF00298139.
18. Bartsch, C., Wenchel, R., Kaiser, A., and Kipper, S. (2014). Singing onstage: female and male common nightingales eavesdrop on song type matching. Behav. Ecol. Sociobiol. *68*, 1163–1171. https://doi.org/10.1007/s00265-014-1727-6.
19. Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). Evolution *51*, 537–551. https://doi.org/10.1111/j.1558-5646.1997.tb02441.x.
20. Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., and Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 11–19.
21. Chartrand, T.L., and Bargh, J.A. (1999). The chameleon effect: The perception–behavior link and social interaction. J. Pers. Soc. Psychol. *76*, 893–910. https://doi.org/10.1037//0022-3514.76.6.893.
22. Pardo, J.S. (2006). On phonetic convergence during conversational interaction. J. Acoust. Soc. Am. *119*, 2382–2393. https://doi.org/10.1121/1.2178720.
23. Levitan, R., and Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In pp. 3081–3084. https://doi.org/10.21437/Interspeech.2011-771.
24. Prather, J.F., Nowicki, S., Anderson, R.C., Peters, S., and Mooney, R.A. (2009). Neural correlates of categorical perception in learned vocal communication. Nat. Neurosci. *12*, 221–228. https://doi.org/10.1038/nn.2246.
25. Roeske, T.C., Tchernichovski, O., Poeppel, D., and Jacoby, N. (2020). Categorical rhythms are shared between songbirds and humans. Curr. Biol. *30*, 3544–3555.e6. https://doi.org/10.1016/j.cub.2020.06.072.
26. MacDougall-Shackleton, S.A., Hulse, S.H., Gentner, T.Q., and White, W. (1998). Auditory scene analysis by European starlings (Sturnus vulgaris): Perceptual segregation of tone sequences. J. Acoust. Soc. Am. *103*, 3581–3587. https://doi.org/10.1121/1.423063.
27. Bregman, M.R., Patel, A.D., and Gentner, T.Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. Proc. Natl. Acad. Sci. USA *113*, 1666–1671. https://doi.org/10.1073/pnas.1515380113.
28. Vicario, D.S. (1991). Contributions of syringeal muscles to respiration and vocalization in the zebra finch. J. Neurobiol. *22*, 63–73. https://doi.org/10.1002/neu.480220107.
29. Suthers, R.A., and Zollinger, S.A. (2004). Producing song: the vocal apparatus. Ann. N. Y. Acad. Sci. *1016*, 109–129. https://doi.org/10.1196/annals.1298.041.
30. Tchernichovski, O., Mitra, P.P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. Science *291*, 2564–2569. https://doi.org/10.1126/science.1058522.
31. Hultsch, H., and Todt, D. (1989). Memorization and reproduction of songs in nightingales (Luscinia megarhynchos): evidence for package formation. J. Comp. Physiol. *165*, 197–203. https://doi.org/10.1007/BF00619194.
32. Long, M.A., Jin, D.Z., and Fee, M.S. (2010). Support for a synaptic chain model of neuronal sequence generation. Nature *468*, 394–399. https://doi.org/10.1038/nature09514.
33. Hahnloser, R.H.R., Kozhevnikov, A.A., and Fee, M.S. (2002). An ultra-sparse code underliesthe generation of neural sequences in a songbird. Nature *419*, 65–70. https://doi.org/10.1038/nature00974.

34. Ali, F., Otchy, T.M., Pehlevan, C., Fantana, A.L., Burak, Y., and Ölveczky, B.P. (2013). The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. Neuron *80*, 494–506. https://doi.org/10.1016/j.neuron.2013.07.049.

35. Tumer, E.C., and Brainard, M.S. (2007). Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature *450*, 1240–1244. https://doi.org/10.1038/nature06390.

36. Zai, A.T., Cavé-Lopez, S., Rolland, M., Giret, N., and Hahnloser, R.H.R. (2020). Sensory substitution reveals a manipulation bias. Nat. Commun. *11*, 5940. https://doi.org/10.1038/s41467-020-19686-w.

37. Hessler, N.A., and Doupe, A.J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. Nat. Neurosci. *2*, 209–211. https://doi.org/10.1038/6306.

38. Dunn, A.M., and Zann, R.A. (1996). Undirected song encourages the breeding female zebra finch to remain in the nest. Ethology *102*, 540–548. https://doi.org/10.1111/j.1439-0310.1996.tb01146.x.

39. Gu, N., Lee, K., Basha, M., Kumar Ram, S., You, G., and Hahnloser, R.H.R. (2024). Positive transfer of the whisper speech transformer to human and animal voice activity detection. In ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7505–7509. https://doi.org/10.1109/ICASSP48485.2024.10447620.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Code and datasets for analysis	Github	https://github.com/vallentinlab/NG-whistle-durations
Experimental models: Organisms/strains		
Nightingale (*Luscinia megarynchos*)	Wild populations	N/A
Software and algorithms		
Avisoft SASLab	R. Specht, Berlin, Germany	Pro 5.2
Audacity	https://www.audacityteam.org/	v.2.4.2
Python	https://www.python.org/	V 3.11.8
WhisperSeg	https://github.com/nianlonggu/WhisperSeg	N/A
Other		
battery-driven pre-amplifier	Roland, Japan	Duo-Capture EX
directional parabolic microphone	Telinga, Sweden	Stereo MK3
Speaker	JBL, Harman International Industries, USA	JBL

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We studied a wild population of Common nightingales (*Luscinia megarhynchos*) in the southwestern part of Brandenburg, Germany between April and May in 2020 and 2024. All recordings and playbacks were performed in accordance with the local authorities (Landesamt für Umwelt - Land Brandenburg LFU-N4 4730/14+5#181132/2021).

METHOD DETAILS

Generation of artificial playbacks

Artificial whistles generation

Artificial whistle syllables were synthesized using Python's NumPy and SciPy libraries. A white noise signal was generated at a 44.1 kHz sampling rate. This signal was then bandpass filtered with a 120 Hz bandwidth centered on the target pitch. A 0.01 s Hann window was applied for boundary fading. Finally, the resulting signal was amplitude-normalized. We establish the silent interval duration based on the relationship between whistle syllable duration and the duration of the subsequent silent interval in the control data (Figure S1E). Control data exhibited a positive correlation (R^2=0.43), indicating that shorter/longer whistle durations were associated with correspondingly shorter/longer silent intervals, following the linear relation:

$$Gap = 0.323 * Whistle\ duration + 0.06\ s$$

For both experimental paradigms, the inter-syllable silent interval within each playback was parametrically defined as a function of the respective whistle syllable duration. Subsequently, these silent intervals were appended to their corresponding whistle syllables, and the total number of whistle-interval units was given by the maximum number of entire units that fitted within 4 s.

Playback stimuli for experiment 1

Twelve playback stimuli were generated, each with distinct whistle durations and a constant frequency of 2390 Hz. Whistle durations were selected based on the trimodal distribution observed in the control dataset. According to cluster analysis of control whistle syllable durations (see: Clustering of whistle syllables durations (Experiment 1)), three duration categories were defined: short (<0.14 s), medium (0.14-0.31 s), and long (>0.31 s). Playback durations were then determined by sampling within each category, specifically including the minimum and maximum values, the 5th and 95th percentiles, and interpolated midpoints between selected values. This resulted in four distinct duration ranges for playback whistle syllables: Very-Short (0.04 s, 0.06 s, 0.08 s), Medium-Short (0.13 s, 0.14 s, 0.15 s), Medium-Long (0.30 s, 0.31 s, 0.32 s), and Very-Long (0.58 s, 0.72 s, 0.86 s).

Playback stimuli for experiment 2

Stimuli were designed with durations and/or pitches outside the natural soundspace of nightingale whistle songs. These regions were defined as areas where whistle syllable density fell below 10%. Specifically: for High-Pitch, Short-Duration three playbacks were created with a fixed whistle duration of 0.14 s and varying frequencies: 6 kHz, 7 kHz, and 8 kHz; for High-Pitch, Long-Duration three playbacks were generated with durations of 0.6 s, 0.7 s, and 0.8 s, and corresponding frequencies of 6 kHz, 7 kHz, and 8 kHz; for

Low-Pitch, Long-Duration three playbacks were produced with durations of 0.6 s, 0.7 s, and 0.8 s, and corresponding frequencies of 1 kHz, 2 kHz, and 3 kHz.

Audio recordings

Audio recordings (16-bit precision at 44.1 kHz sampling rate) from a total of 22 wild male nightingales were obtained during the mating season (April/May) of 2020 and 2024. Recording sessions were conducted between 11pm and 4am CET+1. Each nightingale was recorded with a directional parabolic microphone equipped with a windshield (Stereo MK3, Telinga, Sweden), connected to a battery-driven pre-amplifier (Roland Duo-Capture EX, Roland, Japan) and a laptop computer. Signals were acquired using Audacity v.2.4.2 (https://www.audacityteam.org/) and encoded as stereo.wav files (one channel for the playback and one for the bird).

Control recordings

Recordings from 11 naturally singing nightingales, acquired in 2020, were used as control data.

Song playback experiment

Playback experiments with 11 birds were performed as previously described.[15] In brief, a speaker (JBL, Harman International Industries, USA) was connected to the pre-amplifier and placed at ~10 meters from the singing bird. The playbacks were manually triggered depending on the singing behavior of the nightingales, to avoid overlaps with the birds' songs. Each bird received 20 repetitions of every playback stimulus. Experiment 1 consisted of 12 different stimuli that were randomly shuffled across the execution of the experiment for a total of 240 playback stimulations. Experiment 2 consisted of 9 different stimuli.

Audio data processing and analysis

Audio recordings were processed and analyzed using Audacity, Avisoft SASLab Pro 5.2 (R. Specht, Berlin, Germany) and Python.

As previously described,[15] stereo files were divided in mono files and high pass filtered (frequency=1000 Hz, roll-off=6 dB, High-Pass Filter build-in function of Audacity) and noise reduced to remove environmental noise (noise reduction=12dB, sensitivity=6.00, frequency smoothing=3, noise reduction build-in function of Audacity). Spectrograms were examined in Audacity for visual scoring of whistle responses. A song was considered a whistle response to playbacks when the following criteria were met (1) being a whistle song (containing frequency unmodulated whistle syllables) and (2) being an immediate response to the playback (i.e. the first or second song sung by the bird after the termination of the playback).

Whistle syllable segmentation

Onset and offset of these songs were visually inspected and manually annotated. Individual whistle syllables were manually segmented from whistle songs in Audacity. Avisoft was used for automatically measuring the average frequency and duration of individual syllables. The whistles were initially segmented manually by trained annotators. To assess and minimize potential bias, we validated these annotations using a semi-automated algorithm (WhisperSeg; https://github.com/nianlonggu/WhisperSeg)[39] applied to detect nightingale vocalizations across a range of background conditions. Manual and semi-automated annotations were compared for a subset of data (one bird as a proof of concept) by examining differences in onset, offset, and duration. Because our analyses focused on the median whistle duration per song, we quantified how much this statistic would change when using the semi-automated annotations instead of the manual ones and found negligible differences (mean=2 ms, standard deviation <16 ms). These results confirm that our manual segmentations were consistent and objective across conditions. While the semi-automated approach provided a useful benchmark, it still requires human oversight due to variable field recording conditions. Therefore, manual annotations were retained for the full dataset.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis and statistical computations were performed using Python. All values are reported as median ± median absolute deviation, if not noted otherwise.

Analysis of previous pitch experiment

To test whether whistle durations differed between control responses and pitch-matched responses, we performed a comparative analysis of duration distributions. For pitch-matched responses, we selected only whistle responses for which the median pitch was within 100 Hz of the playback pitch (formal definition of pitch match). The KDE and probability distribution were computed, and individual durations were plotted and color-coded by the occurrence index, providing a pointwise measure of how much more or less likely a duration was under pitch-matched conditions relative to control (Figure S1F). To statistically assess whether the two groups differed, we computed the observed difference in median duration between pitch-matched and control whistles. A permutation test was then conducted by randomly reassigning group labels 10,000 times and recalculating the median difference to generate a null distribution (Figure S1G). The two-sided p-value was defined as the proportion of permuted differences greater than or equal to the observed difference in absolute value. This approach allowed us to quantify both the magnitude and significance of duration changes in pitch-matching responses.

Clustering of whistle syllables duration (Experiment 1)

A Gaussian Mixture Model (GMM) was applied to analyze the distribution of whistle syllable durations in naturally singing nightingales, utilizing the Gaussian Mixture function from the scikit-learn (sklearn) Python library. GMMs were fitted with varying numbers of

components (1 to 9), and the Bayesian Information Criterion (BIC) was calculated for each model. The optimal number of components, k, was determined as the value that minimized the BIC score. This analysis yielded k=3, with component boundaries at 0.14 s and 0.31 s (Figure S1A). These boundaries facilitated the categorization of natural whistle durations into three distinct groups: short (<0.14 s), medium (0.14-0.31 s), and long (>0.31 s). Analysis of individual nightingales also revealed a trimodal distribution of whistle durations, consistent with the population data (Figure S1B)

Clustering of whistle syllables duration and frequency (Experiment 2)

To analyze the distribution of whistle syllables within the soundspace, k-means clustering was employed (Figure S1H). The optimal number of clusters (k) was determined using silhouette analysis, which assesses cluster consistency. The mean silhouette score was calculated for cluster counts ranging from 2 to 9, and the cluster count yielding the highest mean score was selected as optimal (k=3). Cluster boundaries (C1, C2, and C3) were defined by Voronoi regions, which encompass all points closer to a given centroid than to any other within the soundspace. Due to the Euclidean distance metric used by the k-means algorithm, these cluster boundaries are linear segments. Analysis of individual, naturally singing nightingales confirmed the preservation of the previously described whistle syllable distribution within the defined soundspace (Figure S1I).

Whistle syllable duration variation within whistle songs

Whistle syllable durations within a song exhibited some variability. When examining the duration difference between consecutive whistle syllables, a gradual transition was observed, with a mean difference of 0.008 s and a standard deviation of 0.044 s (Figure S1C). The duration difference between the first and last whistle syllables in each song revealed a mean duration increase of -0.047 s with a standard deviation of 0.1 s (Figure S1D). Consequently, the median syllable duration of each song was chosen for subsequent analysis. This metric preserved the previously identified short-medium-long clustered distribution (Figure 1B), provided a representative measure of individual songs, and ensured inter-song comparability.

Generating shuffled data

In the playback experiment, the median durations and pitches of observed whistle response syllables were randomly sampled with replacement from the responses dataset and assigned to the corresponding playback durations and pitches. The number of whistle-match occurrences was preserved for each playback pitch. To extract the respective parameters the shuffling was repeated 10000 times.

Occurrence index of durations (Experiment 1)

A comparative density analysis of median whistle durations between control and experimental conditions was conducted using Kernel Density Estimation (KDE) and subsequent KDE ratio calculations. KDE distributions were generated for whistle responses to playback stimuli, stratified by duration category. These experimental KDE distributions were then compared to a control KDE, derived from control data, by calculating a pointwise KDE ratio. The occurrence index is defined as the pointwise division of the experimental KDE (*KDE exp (duration)*) by the control KDE (*KDE control (duration)*):

$$Occurrence\ index\ (duration)\ =\ KDE\ exp\ (duration)\ /\ KDE\ control\ (duration) \quad \text{(Equation 1)}$$

This index captured relative density changes across playback conditions and informed how the experimental distributions for each playback category deviated from the control. An index below 1 indicates fewer whistles in the experimental group, an index equal to 1 indicates no difference, and above 1 indicates more whistles in the experimental group compared to the control.

This approach was employed for a detailed comparative analysis of distributional shifts, specifically highlighting regions of increased or decreased density relative to the control distribution. For visualization purposes, the occurrence index values were clipped to the range of 0 to 2.

Occurrence index of whistle durations and pitch (Experiment 2)

For each playback region (High-Pitch/Short-Duration, High-Pitch/Long-Duration, and Low-Pitch/Long-Duration) the joint probability density function of whistle duration and pitch was estimated using a Gaussian Kernel Density Estimation (KDE). This KDE was computed over a grid spanning the soundspace, with duration values ranging from 0 s to 0.9 s and pitch values ranging from 0 kHz to 9 kHz. This yielded the estimated experimental probability density, Z exp (duration, pitch). An analogous analysis was performed on the control data to obtain the control probability density, *Z control (duration, pitch)* at each grid point. Subsequently, an occurrence index was calculated as follows

$$Occurrence\ index\ (duration, pitch)\ =\ Z\ exp\ (duration, pitch)\ /\ [Z\ control\ (duration, pitch)\ +\ \beta] \quad \text{(Equation 2)}$$

The small constant β=0.0003 prevents division by zero. The occurrence index was clipped within 0 and 2. This index provides insights into how the distribution of experimental responses differs from the control. An index greater than 1 indicates areas where whistle responses are more concentrated for each of the different playback regions. Conversely, an index below 1 highlights region where responses are less frequent. When the index is close to 1, it suggests no significant difference between the experimental and control distributions.

Durations and pitch similarity index (Experiment 2)

To quantify how similar a bird's response was to the playback in terms of duration and pitch, we defined a continuous similarity index based on a sigmoid function. For each song, we computed the absolute error between the response and the playback in both duration ($|d_{median} - d_{playback}|$) and pitch ($|f_{median} - f_{playback}|$). These errors were then transformed into similarity values ranging from 0 to 1 using the function:

$$similarity = \frac{1}{1+exp\left(\frac{error - \mu}{\beta}\right)} \quad \text{(Equation 3)}$$

Where μ was set to the median error across the dataset, and β was set to 10% of the full error range for each feature (duration or pitch). In our case,$\mu_d = 0.33\ s, \mu_f = 1989\ Hz$ and$\beta_d = 0.1640\ s, \beta_f = 981.1\ Hz$. This transformation ensures that exact matches (error=0) yield similarity values close to 1 (max similarity equal to 0.88), while larger mismatches approach 0 in a gradual, tunable manner. This sigmoid-based approach provides a smooth, interpretable metric of acoustic similarity for our analyses.

Next, to quantify the relative similarity of responses in pitch and duration, we computed a signed distance metric based on the projection of each response onto the axis orthogonal to the identity line in similarity space. Specifically, for each response, we calculated its similarity to the playback in both duration and pitch using the above described sigmoid function. We then represented each response as a point in 2D space with duration similarity on the x-axis and pitch similarity on the y-axis. The signed distance was computed as the projection of this point onto the line perpendicular to the identity line (duration similarity=pitch similarity), given by

$$signed\ distance = \frac{1}{\sqrt{2}}(S_d - S_f) \quad \text{(Equation 4)}$$

where S_f and S_d are the frequency and duration similarity scores, respectively. Positive values indicate that the pitch similarity exceeds the duration similarity, whereas negative values indicate the reverse. This metric captures the directional asymmetry in how closely responses match playback stimuli along each acoustic dimension.

Statistical analysis for Experiment 1

The statistical significance of observed median whistle durations across temporal categories (Very-Short, Medium-Short, Medium-Long, and Very-Long) was assessed. A null distribution was generated through 10,000 bootstrap resamples of the data (see: Generating shuffled data), and the median was calculated for each category within each resampled dataset (example shuffled distribution in Figure S2B). P-values were then determined by comparing the observed median for each category to the null distribution of medians derived from the resampled data. For the Very-Short category, a left-tailed hypothesis test was employed, with the p-value representing the proportion of resampled medians less than or equal to the observed median. For the Very-Long category, a right-tailed test is used, where the p-value is the proportion of shuffled medians that are greater than or equal to the observed median. For the other two categories, Short-Med and Med-Long, a two-tailed test is applied, which accounts for deviations in both directions.

Statistical analysis for Experiment 2

The statistical significance of whistle responses to playback stimuli outside the nightingale's natural soundspace (High-Pitch/Short-Duration, High-Pitch/Long-Duration, and Low-Pitch/Long-Duration) was evaluated. For each playback region, whistle responses were classified into the three previously defined soundspace clusters (see: Clustering of whistle syllables duration and frequency (Experiment 2). The percentage of whistle responses within each cluster was calculated for each playback region. Subsequently, the data were subjected to 10,000 shuffle iterations (see: Generating shuffled data), and the percentage of whistle responses within each cluster was recalculated for each shuffled dataset (example shuffled distribution in Figure S2D). This process generated a null distribution of expected whistle response percentages for each cluster. The p-value in this analysis was computed using a two-sided test to determine whether the observed cluster proportions for each playback region significantly deviated from the null distribution (Figure S2E).

Model for temporal modulation

GMM for the median duration of natural singing nightingales. We set $y = \log x$ with x = median whistle duration of a whistle song (in seconds) and fit a k-component GMM (with k=3):

$$p(y) = \sum_{k=1}^{3} \omega_k N\left(y; \mu_{log,k}, \sigma^2_{log,k}\right) \quad \text{(Equation 5)}$$

Here $\mu_{log,k}$ and $\sigma_{log,k}$ are estimated in the log-space. The mixture weights ω_k represent the population fraction of observations assigned to component k, which in our case are short, medium and long duration components and are independent from the chosen units (Figure S1A). We modelled whistle durations of naturally singing nightingales on the log scale to avoid negative values for time.

Bird variability for modeling the production of whistle syllable duration. To capture across-bird variability, we first summarized each bird's component on the seconds' scale $(\mu_{sec,k}, \sigma_{sec,k})$ (obtained from a per-bird GMM). We then fit a 2D GMM to all birds' $(\mu_{sec,k}, \sigma_{sec,k})$ and find three clusters (Figure S3B) supporting a trimodal structure at the individual level (Figure S1B). For each

simulated bird (N=7), we sample one $(\mu_{sec,k}, \sigma_{sec,k})$ from each ordered cluster k (Fast, Medium, and Slow) and convert it to log space via:

$$\mu_{log} = \ln \mu_{sec}$$

$$\sigma_{log} = \sqrt{\ln\left(1+\left(\frac{\sigma_{sec}}{\mu_{sec}}\right)\right)^2} \quad \text{(Equation 6)}$$

The duration of a natural-singing bird b is thus described by a trimodal log-Gaussian mixture:

$$p(y;b) = \sum_{k=1}^{3} \omega_k N\left(y; \mu_{log,k}^{(b)}, \left(\sigma_{log,k}^{(b)}\right)^2\right) \quad \text{(Equation 7)}$$

With the global mixture weights ω_k preserved across birds.

Gaussian-selection probabilities during playback stimuli. We hypothesized that a playback duration d modulates the effective mixture weights through a quadratic attractor term:

$$r_k(d) = -\frac{1}{2}\left(\frac{\ln d - \mu_{log,k}^{(b)}}{\sigma_{log,k}^{(b)}}\right)^2 \quad \text{(Equation 8)}$$

and an attractor gain$\beta \geq 0$. The playback-dependent Gaussian probabilities were

$$\omega_k(d;\beta) = \frac{exp(\ln \omega_k + \beta r_k(d))}{\sum_{j=1}^{3} exp(\ln \omega_j + \beta r_j(d))} \quad \text{(Equation 9)}$$

Consequently, $\beta = 0$ represented the control/no-attractor case $\omega_k(d;0) = \omega_k$, and $\beta > 0$ yields engagement for the Gaussian whose $\mu_{log,k}$ was closest to $\ln d$.

Combining mode selection with within-mode variability gave the playback-conditioned response model:

$$p(y;b|d) = \sum_{k=1}^{3} \omega_k(d;\beta) N\left(y; \mu_{log,k}^{(b)}, \left(\sigma_{log,k}^{(b)}\right)^2\right) \quad \text{(Equation 10)}$$

With responses in seconds obtained by back-transforming $x = \exp(y)$.

Model for pitch modulation

GMM for the median pitch of natural singing nightingales. We fit a k-component GMM (with k=10)[15] with f = median whistle pitch of a whistle song (in Hz):

$$p(f) = \sum_{k=1}^{10} \nu_k N(f; \mu_k, \sigma_k^2) \quad \text{(Equation 11)}$$

The mixture weights ν_k represented the population fraction of observations assigned to component k, which in our case comprised the low to high pitch components of whistle songs.[15]

Bird variability for modeling the production of whistle syllable pitch. Instead of clustering in (μ, σ) space, we aggregate components across birds by their rank (i.e. left to right position in each bird's mixture). For each pitch mode (ranked left to right), we summarize across all birds the distribution of component centers (μ) and spreads (σ) using quartiles (Figure S3E). These quartiles define a diamond-shaped region in the (μ, σ)-plane: the bottom and top corners reflect the median center with lower and upper spreads (first quartile and third quartile of σ), while the left and right corners reflect the lower and upper quartiles of μ at the median spread. This "diamond" captures the typical range of parameter variability across birds for that mode. To simulate new birds, we then sample uniformly within these diamonds, ensuring that each synthetic bird's pitch components remain consistent with the natural variability observed in the population. The pitch of a natural-singing bird b is thus described by a multimodal Gaussian mixture:

$$p(f;b) = \sum_{k=1}^{10} \nu_k N\left(f; \mu_{,k}^{(b)}, \left(\sigma_k^{(b)}\right)^2\right) \quad \text{(Equation 12)}$$

With the global mixture weights ν_k preserved across birds.

Mode-selection probabilities during playback stimuli. We hypothesized that a playback pitch s modulates the effective mixture weights through a smooth distance-decay kernel in frequency:

$$K_k(s) = \frac{1}{1+\left(\frac{|s-\mu_k^{(b)}|}{\tau}\right)^2} \quad \text{(Equation 13)}$$

Where $\tau \geq 0$ controls the range of influence. The playback-dependent Gaussian probabilities were

$$\nu_k(s;\beta) = \frac{exp(\ln \nu_k + \beta \ln K_k(s))}{\sum_{j=1}^{10} exp(\ln \nu_j + \beta K_j(s))} \quad \text{(Equation 14)}$$

Consequently, $\beta = 0$ represented the control/no-attractor case $\nu_k(p;0) = \nu_k$, and $\beta > 0$ yielded attraction toward the Gaussian whose μ_k lied closest to the playback pitch s.

Combining mode selection with within-mode variability gave the playback-conditioned response model:

$$p(f;b|s) = \sum_{k=1}^{10} \nu_k(s;\beta) N\left(f;\mu_k^{(b)}, \left(\sigma_k^{(b)}\right)^2\right) \quad \text{(Equation 15)}$$

Model for temporal and pitch modulation

We combined the temporal and pitch models into a single joint framework, where each simulated bird was described by both a duration GMM and a pitch GMM as described above. In this setting, the Gaussian weights of both models were updated by their respective attractor functions (in temporal and spectral domains). What differs between the two simulation configurations was the support variable: in the Duration→Pitch condition, playback duration provided the attractor input, updating the duration weights, and pitch responses were then generated conditional on the simulated durations, by being restricted to fall within the whistle dictionary—the natural distribution of whistle responses in which songs consistently occupied the same region of pitch–duration space. In the Pitch→Duration condition, playback pitch provided the attractor input, updating the pitch weights, and duration responses were generated conditional on pitch under the same dictionary constraint. Thus, the two sets of mixture weights were modified simultaneously, but the causal direction of the attractor determines whether temporal structure constrained spectral responses or vice versa. This design allowed us to test the putative hierarchical representation of temporal versus spectral parameters in the nightingale's response system.

Matching drive parameter

In all model variants, we included the scenario in which birds engaged in a whistle-matching response based on the acoustic features of the input or not. To capture this, we introduced a "matching drive" parameter $\delta \in [0, 1]$ that linearly interpolates between the global baseline weights ((ω_k, ν_k) and the playback-modulated weights ($\omega_k(d;\beta)$ or $\nu_k(s;\beta)$):

$$\overline{\omega}_k = \delta \omega_k(d;\beta) + (1 - \delta)\omega_k \quad \text{(Equation 16)}$$

$$\overline{\nu}_k = \delta\, \nu_k(s;\beta) + (1 - \delta)\nu_k \quad \text{(Equation 17)}$$

Here, $\delta = 0$ corresponded to purely baseline probabilities (no influence of the attractor), while $\delta = 1$ corresponded to fully attractor-driven probabilities. Intermediate values interpolated smoothly between these extremes, allowing partial engagement of the attractor relative to the natural mixture weights. For both models, temporal and pitch modulation, we used $\delta = 0.5$.

In the joint model, we set $\delta = 0.5$, ensuring that the effective mixture probabilities for each dimension can reflect a combination of baseline tendencies and stimulus-driven attractor modulation. This formulation provides a flexible continuum from control behavior to strong playback-driven engagement, and allows us to evaluate how varying attractor strength and "matching drive" affect the hierarchy of temporal and spectral response organization.