Article

Convergent vocal representations in parrot and human forebrain motor networks

https://doi.org/10.1038/s41586-025-08695-8

Received: 16 July 2024

Accepted: 23 January 2025

Published online: 19 March 2025

Check for updates

Zetian Yang[1,2] & Michael A. Long[1,2]✉

Cortical networks for the production of spoken language in humans are organized by phonetic features[1,2], such as articulatory parameters[3,4] and vocal pitch[5,6]. Previous research has failed to find an equivalent forebrain representation in other species[7–11]. To investigate whether this functional organization is unique to humans, here we performed population recordings in the vocal production circuitry of the budgerigar (*Melopsittacus undulatus*), a small parrot that can generate flexible vocal output[12–15], including mimicked speech sounds[16]. Using high-density silicon probes[17], we measured the song-related activity of a forebrain region, the central nucleus of the anterior arcopallium (AAC), which directly projects to brainstem phonatory motor neurons[18–20]. We found that AAC neurons form a functional vocal motor map that reflects the spectral properties of ongoing vocalizations. We did not observe this organizing principle in the corresponding forebrain circuitry of the zebra finch, a songbird capable of more limited vocal learning[21]. We further demonstrated that the AAC represents the production of distinct vocal features (for example, harmonic structure and broadband energy). Furthermore, we discovered an orderly representation of vocal pitch at the population level, with single neurons systematically selective for different frequency values. Taken together, we have uncovered a functional representation in a vertebrate brain that displays unprecedented commonalities with speech-related motor cortices in humans. This work therefore establishes the parrot as an important animal model for investigating speech motor control and for developing therapeutic solutions for addressing a range of communication disorders[22,23].

Human speech requires exquisite control over vocal production, showing a high degree of flexibility while also maintaining the capacity to reliably produce words (Fig. 1a,b). Speech production engages several cortical regions[2,23–25], the activity of which often reflects the structure of spoken utterances[2,6,26]. Although humans are the only species capable of language, a wide range of vocal abilities have been observed in other animals[21]. For instance, the zebra finch, a species capable of vocal imitation, can learn to produce a short (approximately 0.5–1.0 s), stereotyped song composed of 'syllables' as well as a small set of calls[27] (Fig. 1c,d) whereas parrots, such as the budgerigar, can mimic conspecific and heterospecific sounds throughout life[15,16,28]. Budgerigar vocalizations comprise sequences of variable vocal elements called warble songs[12,29] and span a richer acoustic space (Fig. 1e,f and Extended Data Fig. 1). Because of the high degree of flexible control inherent in both the budgerigar song and human speech, we hypothesize that shared neural representations may underlie these vocalizations.

To investigate the mechanisms of vocal production in the budgerigar, we recorded the activity of a key forebrain structure, the AAC. Like the human speech motor cortex and a well-characterized vocal motor nucleus in the zebra finch, the robust nucleus of arcopallium (RA)[7,30,31] (Extended Data Fig. 2a), the AAC densely innervates vocal motor neurons[18–20] (Fig. 1g). Lesions[32] or electrical stimulation[33] in the AAC can influence vocal production, indicating a primary role in motor control. To test this idea, we recorded AAC activity (Fig. 1h and Extended Data Fig. 2b–e) with a chronically implanted high-density silicon probe[17] while birds produced flexible warble elements (n = 1,403, 522, 1,381 and 1,645 for each bird) and stereotyped calls (n = 40, 61, 27 and 43 for each bird; Fig. 1e). During vocalization (including both calls and warbles), AAC neurons (n = 220 from 4 birds; n = 49, 81, 43 and 47 for each bird) strongly increased their firing rates (baseline: 47.5 ± 16.0 Hz; vocal: 95.5 ± 46.1 Hz; two-sided Wilcoxon signed-rank test, $P = 2.7 \times 10^{-35}$) (Fig. 1h–j and Extended Data Fig. 3a). Importantly, changes in AAC activity often preceded the onsets and offsets of vocal elements (Extended Data Fig. 3b,c) indicating a premotor function. Therefore, AAC activity corresponded to the production of vocalizations.

To confirm that the observed responses in the AAC are specific for vocal motor production rather than auditory processing, we performed playback experiments with two birds (Methods). We found that AAC firing rates were significantly higher during vocalization than playback (vocal: 108.3 ± 44.8 Hz; playback: 55.3 ± 16.4 Hz; Extended Data Fig. 3d,e; two-sided Wilcoxon signed-rank test, $P < 10^{-7}$ for both birds). Compared with activity during a 10 s non-singing period (baseline: 52.0 ± 15.5 Hz), playback elicited a small but statistically significant increase in firing rate (Δfiring rate: 3.2 ± 3.8 Hz; two-sided Wilcoxon signed-rank test,

[1]NYU Neuroscience Institute and Department of Otolaryngology, New York University Langone Medical Center, New York, NY, USA. [2]Center for Neural Science, New York University, New York, NY, USA. ✉e-mail: mlong@med.nyu.edu

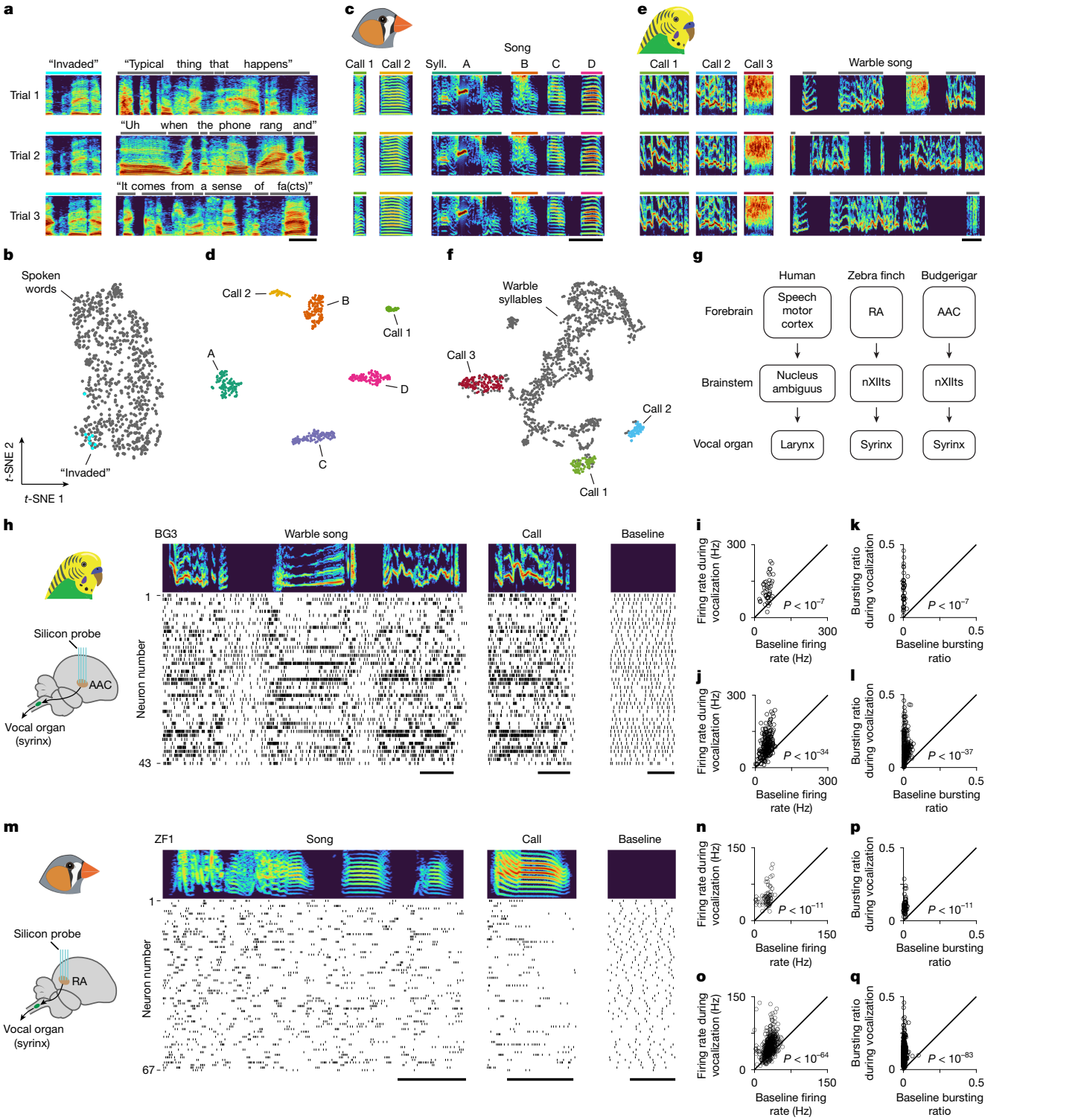


Fig. 1 | Comparative vocal behaviour and underlying neural activity. **a**, Example sonograms of a single human speaker excerpted from a 10-min session of the Switchboard corpus (Methods). Scale bar, 0.2 s. **b**, *t*-distributed stochastic neighbour embedding (*t*-SNE) representation of spectrograms of the spoken words from the session used in **a**, with all the instances of the word 'invaded' highlighted in cyan. **c**,**d**, Example sonograms (**c**) and *t*-SNE (**d**) of vocalizations by a zebra finch (ZF). Scale bar, 0.1 s. Syll., syllable. **e**,**f**, Example sonograms (**e**) and *t*-SNE (**f**) of vocalizations by a budgerigar (BG). Scale bar, 0.1 s. **g**, Vocal production pathway in the three species, highlighting direct projections from forebrain to brainstem vocal motor neurons. nXIIts, tracheosyringeal hypoglossal nucleus. **h**, Population recording in budgerigar AAC highlighting spiking activity during example vocalizations and non-vocal baseline periods. Scale bars, 100 ms. **i**,**j**, Mean firing rates of AAC neurons during vocalization versus during baseline for an example bird (**i**; BG3) and for data pooled from all budgerigars (**j**, $n = 4$ birds). Each circle represents a neuron. **k**,**l**, Bursting ratios of AAC neurons for an example bird (**k**; BG3) and for data pooled from all budgerigars (**l**, $n = 4$ birds). Each circle represents a neuron. **m**–**q**, Population recordings (**m**) from RA and associated quantification from a single zebra finch (**n**,**p**; ZF1) and from pooled data (**o**,**q**, $n = 7$ birds). Scale bars, 100 ms. Sonogram frequency range: 0.3–4 kHz in **a** and 0.3–7 kHz in other panels. In all panels, *P* values are from two-sided Wilcoxon signed-rank tests. See Extended Data Table 1 for exact *P* values, sample sizes and related information for statistical tests. Illustrations of the zebra finch in parts **c** and **m** are reproduced from ref. 60, Elsevier. Schematics of the brain in parts **h** and **m** are adapted from ref. 61, Springer Nature Limited.

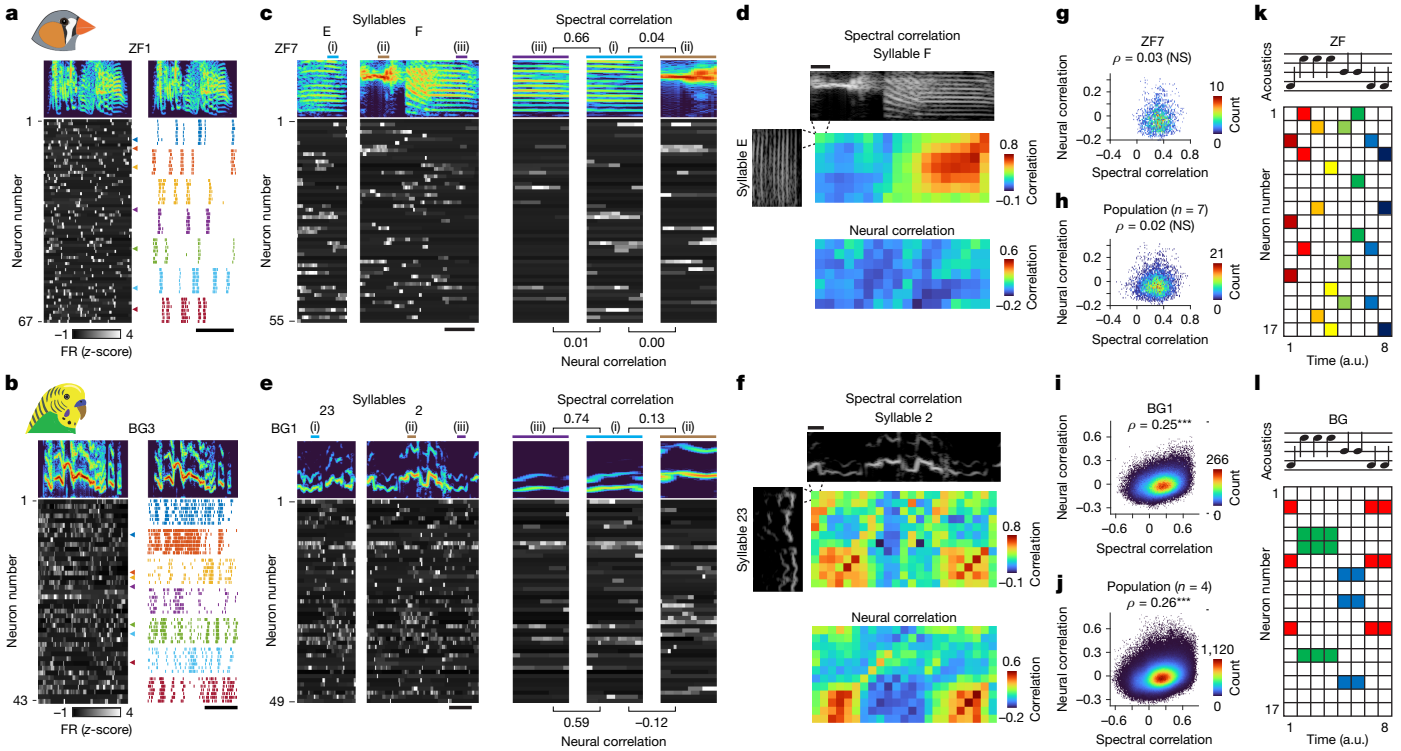


Fig. 2 | Different neural coding schemes in zebra finch RA and budgerigar AAC. a, Instantaneous firing rates (*z*-scored) of a population of RA neurons to an example song syllable (left) and spiking activity of a group of selected neurons across multiple renditions of that same syllable (right). FR, firing rate. Scale bar, 100 ms. **b**, AAC activity during a single call (left) and associated time-aligned activity of representative neurons across trials of that same call (right). Scale bar, 100 ms. Coloured triangles in **a** and **b** denote neurons shown on the right. **c**, Instantaneous firing rates of zebra finch RA neurons to two example syllables. Numbered coloured bars on top denote 20 ms segments expanded on the right along with accompanying Pearson correlations of the spectrograms (top) and neural activity (bottom) between each segment. Scale bar, 50 ms. **d**, Spectral (top) and neural (bottom) correlation matrices for syllables in **c**. Scale bar, 20 ms. **e**, Data from budgerigar AAC during production of two example syllables, with three indicated segments further analysed on the right. Scale bar, 50 ms. **f**, Spectral (top) and neural (bottom) correlation matrices for syllables in **e**. **g**–**j**, Correlation between spectral similarity and neural similarity matrices for an example zebra finch (**g**), an example budgerigar (**i**) and accompanying population data (**h**,**j**). In each panel, spectral and neural matrices were flattened and concatenated across syllable pairs before plotting. ρ denotes Spearman's correlation value. NS, not significant ($P > 0.05$), $^{***}P < 10^{-100}$ (two-sided permutation test). See Extended Data Table 1 for sample sizes and related information for statistical tests. **k**,**l**, Neural coding of vocal production by RA (**k**) or the AAC (**l**) population. Notes on top represent acoustic properties of vocalization; filled boxes indicate active neurons during vocal production. Neural activity was shifted forward by 12 ms to account for vocal motor delay (Methods). a.u., arbitrary unit. Illustration of the zebra finch in part **a** is reproduced from ref. 60, Elsevier.

$P < 10^{-4}$ for both birds; Extended Data Fig. 3d,f). During vocalization, many AAC neurons showed a strong increase in bursting activity as indicated by a 'bursting ratio' or the proportion of time a neuron spent bursting (baseline: 0.01 ± 0.01; vocal: 0.14 ± 0.09; two-sided Wilcoxon signed-rank test, $P = 7.7 \times 10^{-38}$; Fig. 1h,k,l and Extended Data Fig. 3g,h). We then compared our results with song-related activity in RA ($n = 502$ neurons from 7 birds; $n = 67, 80, 116, 104, 48, 32$ and 55 for each bird). Both the RA firing rate (baseline: 30.9 ± 8.8 Hz; vocal: 47.5 ± 19.6 Hz) and burst ratio (baseline: 0.005 ± 0.008; vocal: 0.11 ± 0.06; Fig. 1m–q and Extended Data Fig. 3i, j) increased during vocalization compared with silent baseline periods, albeit with a higher degree of sparseness than the AAC (RA: 0.21 ± 0.06; AAC: 0.32 ± 0.11; two-sided Wilcoxon rank-sum test, $P = 1.45 \times 10^{-37}$; Methods). These results support the notion that the AAC, similar to zebra finch RA[7,30,31], is a key vocal motor area in the budgerigar forebrain.

Budgerigar AAC reuses neural ensembles

Having established the AAC as a forebrain motor control region, we investigated the functional relationship between AAC population activity and vocal production. Previous research demonstrated that the burst timing of neurons in the zebra finch RA is precisely structured across song renditions[7,31,34,35] and our data are consistent with these findings (Fig. 2a). To assess whether AAC activity has a temporally precise relationship with behaviour, we aligned repeated vocalizations (for example, contact calls) and found that spiking was also highly stereotyped (Fig. 2b). This reliable motor representation enabled us to quantify the degree to which the neural activity could be related to the acoustic structure of vocalizations. Previous research has demonstrated that in zebra finch RA, acoustically similar portions of vocalizations were produced by distinct motor commands, establishing a 'degenerate' coding scheme[7]. Consistent with this previous result, we found low neural correlation of RA ensembles across syllables, even in cases with a high degree of spectral similarity (Fig. 2c,d and Methods).

We observed a categorically different code in the budgerigar AAC (Fig. 2e,f), where the vocal structure instead covaried with underlying neural response patterns (Spearman's $\rho = 0.26$, $P < 10^{-100}$), in contrast to the motor code observed in zebra finch RA (Spearman's $\rho = 0.02$, $P = 0.08$; Fig. 2g–j). Importantly, the relationship between neural and spectral correlation in budgerigars could not be explained by the recurrence of identical vocal units across different syllables (Methods and Extended Data Fig. 4). Our findings reveal a fundamental difference between the forebrain control of song in these two species. In the zebra finch, RA population activity changes throughout the course of the song, such that different configurations are active at different moments, representing an evolving population 'barcode'[36,37] (Fig. 2k).

By contrast, the AAC population activity in the budgerigar corresponds to the acoustic structure of song, with similar neural ensembles 'reused' to produce vocal elements that have similar spectral features (Fig. 2l).

Mapping of vocal categories in the AAC

In humans, the speech-related sensorimotor cortex is organized according to phonetic features[2,5,6]; speech sounds with shared articulatory parameters rely on reused motor commands[3,38]. Given the correspondence of vocal acoustic structure to neural activity in the budgerigar, we examined whether an equivalent feature-based representation is also present in populations of AAC neurons. Previous research has demonstrated that the budgerigar warble is composed of acoustically distinct elements[12]. For instance, 3 out of 4 budgerigars that we examined commonly produced distinct low-frequency syllables (n = 64, 49 and 40 for birds 1, 3 and 4, respectively) with a concentrated spectral energy below 700 Hz (Fig. 3a–c and Extended Data Fig. 5a). Moreover, warbles were composed of features ranging from 'consonant-like' noisy sounds with a broadband spectral distribution to 'vowel-like' harmonic elements that had a clear periodic acoustic structure[39] (Fig. 3a–c). To quantify these vocal types, we developed a 'harmonic index' in which these sounds represent two ends of a continuum (Fig. 3b,c and Methods). From these behavioural analyses, we have defined a set of distinct vocal components of the budgerigar repertoire.

We next investigated how these different acoustic categories are represented in the AAC. We found example neurons in which firing rates seemed to be linked to individual vocal types, including low-frequency (green), harmonic (red) and noisy (blue) sounds (Fig. 3d, also see triangles in Fig. 3e). We defined a selectivity index for each category (Methods and Extended Data Fig. 5b–d) and observed a diverse response profile across the AAC population (Extended Data Fig. 5e). To determine whether AAC neurons show distinct population response patterns associated with each acoustic category, we used principal component analysis (PCA) to map the AAC population responses over a window (−30 to −5 ms with respect to vocalization) into a three-dimensional neural state space (Methods and Extended Data Fig. 6a,b). We then labelled each neural state (Fig. 3f and Supplementary Video 1) on the basis of the vocalization produced (Fig. 3c). If the AAC population encodes acoustic categories, we would expect neural activity to form well-separated clusters in this neural state space. We found that responses underlying low-frequency vocalizations were distinct from all other neural states and responses to noisy and harmonic vocalizations were also clearly separated from each other (Fig. 3f and Supplementary Video 1), indicating a functional representation of vocal acoustic parameters in the budgerigar AAC.

In addition to noisy and harmonic sounds, birds produce vocalizations that fall between these two categories and have intermediate 'harmonic index' values (Fig. 3c, bottom). We returned to the state space representation to examine the neural ensemble activity during the production of these 'mixed' warble elements. One possibility is that such neural responses are heterogeneous and disorganized with representations scattered across the neural space. Alternatively, mixed neural responses could be localized in a specific region in this representation. We found that the latter was true: responses to mixed vocalizations were localized in between the neural representations for noisy and harmonic sounds (Fig. 3f and Supplementary Video 1). Moreover, when we colour coded each neural state on the basis of its harmonic index, we observed a strong relationship between that parameter and the position in the state space (Extended Data Fig. 6c). Supporting this notion, a linear regression model using neural states as predictors explained a substantial proportion of the variance in the harmonic index (R^2 = 0.45, 0.62, 0.51 and 0.52 for each budgerigar; Methods). These results reveal that AAC neural responses are organized according to the acoustic features of vocalizations, forming an orderly representation of vocal elements (that is, ranging from noisy to harmonic).

We further examined whether the organizing principles that we uncovered were similar across individuals. When we investigated the neural state clusters in our population of four budgerigars that had vocal repertoires with both shared and distinct components (Extended Data Fig. 7), we uncovered a consistent relationship between AAC neural space and the acoustics of produced vocal elements. For each bird, the distance of neural states in each category (for example, noisy–noisy) is significantly smaller than the distance between neural responses across categories (two-sided Wilcoxon rank-sum test, $P < 10^{-100}$; Fig. 3g and Extended Data Fig. 8a). This result is not due to the spike binning process, as neurons with shuffled spike times have significantly reduced separation of neural states across categories (Methods and Extended Data Fig. 8b). Our results also remain consistent after subsampling to prevent bin overlap (Methods and Extended Data Fig. 8c,d). A prediction of this neural mapping is that the AAC population travels across the state space during vocalization according to the current vocal acoustic category, as previously described in human speech-production circuits[2]. We found that neural states can shift rapidly in correspondence with changes in vocal categories (Extended Data Fig. 8e), consistent with the notion that the AAC ensemble dynamically controls vocal production.

Representation of vocal pitch in the AAC

The similarity in representation of vocal features across humans and budgerigars prompted us to investigate whether additional structure exists in the AAC population activity. Human studies have revealed that the speech motor cortex encodes vocal pitch[5,6] (that is, fundamental frequency). Budgerigars[40,41] and other parrot species (for example, cockatiels[42]) show excellent control of pitch, a parameter known to be related to syringeal tension[43]. To examine whether pitch is also explicitly represented in the AAC, we isolated vocalizations with a strong harmonic structure (Fig. 4a and Extended Data Fig. 9a) and estimated the fundamental frequency of these vocal elements (Methods and Extended Data Fig. 9a). Each bird exhibited an idiosyncratic distribution of pitches, broadly ranging from 1 to 5 kHz (Fig. 4b and Extended Data Fig. 9b). We next re-examined the AAC neural state space and found a highly structured representation related to the pitch of the produced vocalization (Fig. 4c). In all birds tested, responses were organized according to pitch, forming a monotonic gradient in neural space (Fig. 4c,d). We analysed this relationship by calculating a 'neural frequency axis' (Extended Data Fig. 9c) that traverses from low to high frequencies (Fig. 4c,d). After projecting neural responses onto this axis (Methods), we found that the vocal pitch in each bird seems to be linearly distributed along the neural axis, whereas no relationship is observed in the shuffled data (Fig. 4e and Extended Data Fig. 9d) or when a large temporal offset is added to the neural data (Extended Data Fig. 9e,f). Notably, the pitch of calls and warble syllables was represented in a similar manner (Extended Data Fig. 9g), suggesting a shared representation across both vocalization types. Our results indicate that neural responses in the AAC form an accurate representation of vocal pitch.

Given our results at the population level, we further examined whether single neurons in the AAC are tuned to vocal pitch. During the production of a frequency-modulated element, we observed that some AAC neurons changed their firing in an orderly manner, suggesting that different AAC neurons may be tuned to specific pitch values (Fig. 4f). To formalize this relationship, we calculated tuning curves for each individual neuron by plotting the mean neural responses as a function of pitch. We found neurons demonstrating strong tuning (Fig. 4g) as well as untuned neurons (Fig. 4g, bottom). Using a 'tuning index' that measures modulation of neural response across frequencies (Methods and Extended Data Fig. 10a), we found 51% (112 out of 220) of AAC neurons showed strong tuning for pitch (tuning index >6; Fig. 4h). Different pitch-tuned neurons had maximal responses at different

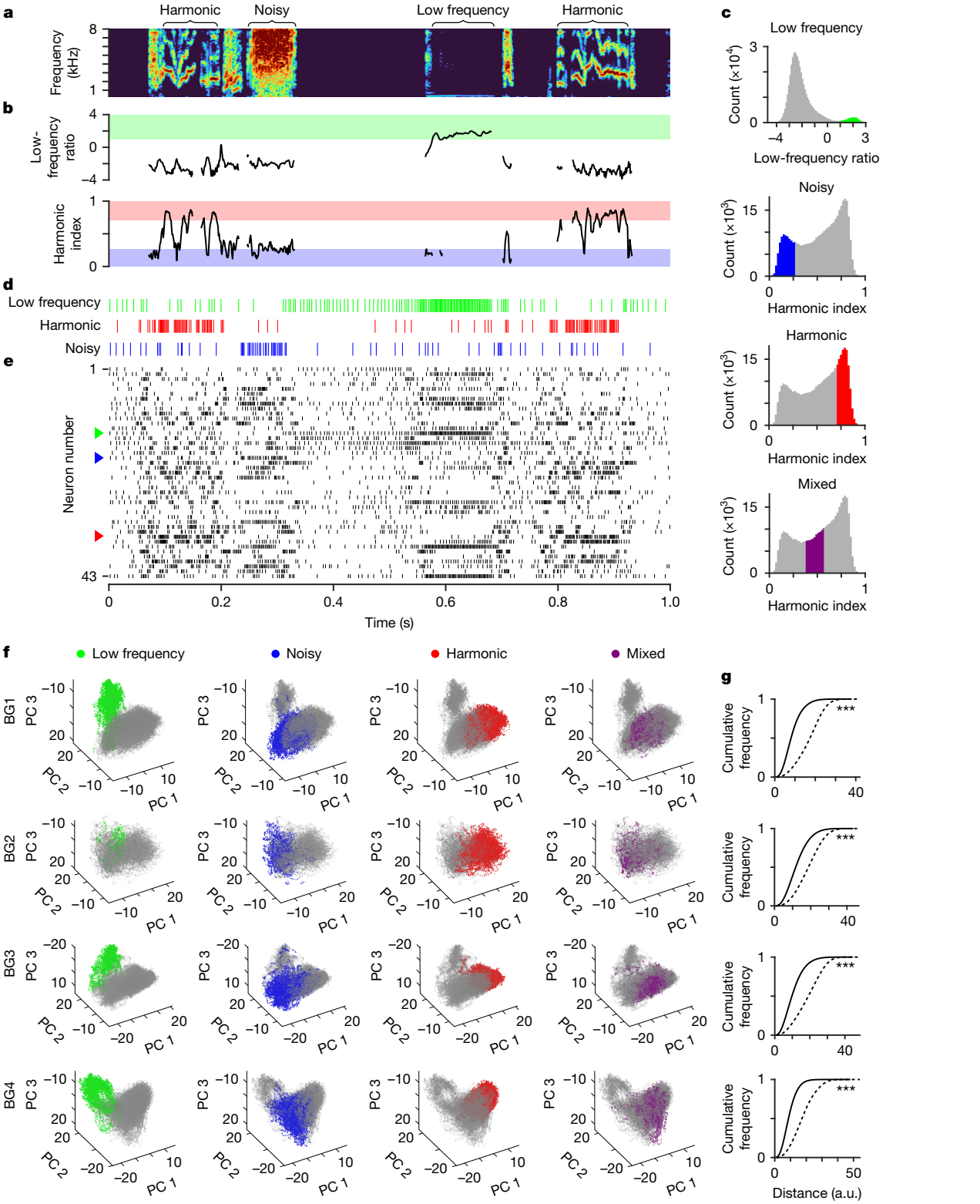


Fig. 3 | Distinct mapping of vocal acoustic types in the AAC state space. **a**, Example warble song from a budgerigar (BG3) with harmonic, low-frequency and noisy elements labelled at top. **b**, Low-frequency ratio (top) and harmonic index (bottom) of the vocal elements in **a**. Shading represents thresholds used to classify acoustic categories (Methods), with green, red and blue denoting low-frequency, harmonic and noisy types, respectively. **c**, Distribution of low-frequency ratio and harmonic index across all four budgerigars. In each panel, the coloured region represents a specific acoustic type. **d**,**e**, Spiking activity of AAC neurons, with a closer view of three example cells (**d**) taken from the population (**e**), as indicated by coloured triangles on the left. **f**, Population neural responses to vocalizations were mapped to a three-dimensional state space using PCA, with each point denoting a neural state associated with vocal production (Extended Data Fig. 6a). In each column, neural states corresponding to a specific acoustic category are coloured the same as in **c**. Each row represents data from a single budgerigar. PC, principal component. **g**, Cumulative distribution of Euclidean distances between neural states within (solid line) or between (dotted line) acoustic categories for each bird. $^{***}P < 10^{-100}$ (two-sided Wilcoxon rank-sum test). See Extended Data Table 1 for sample sizes and related information for statistical tests.

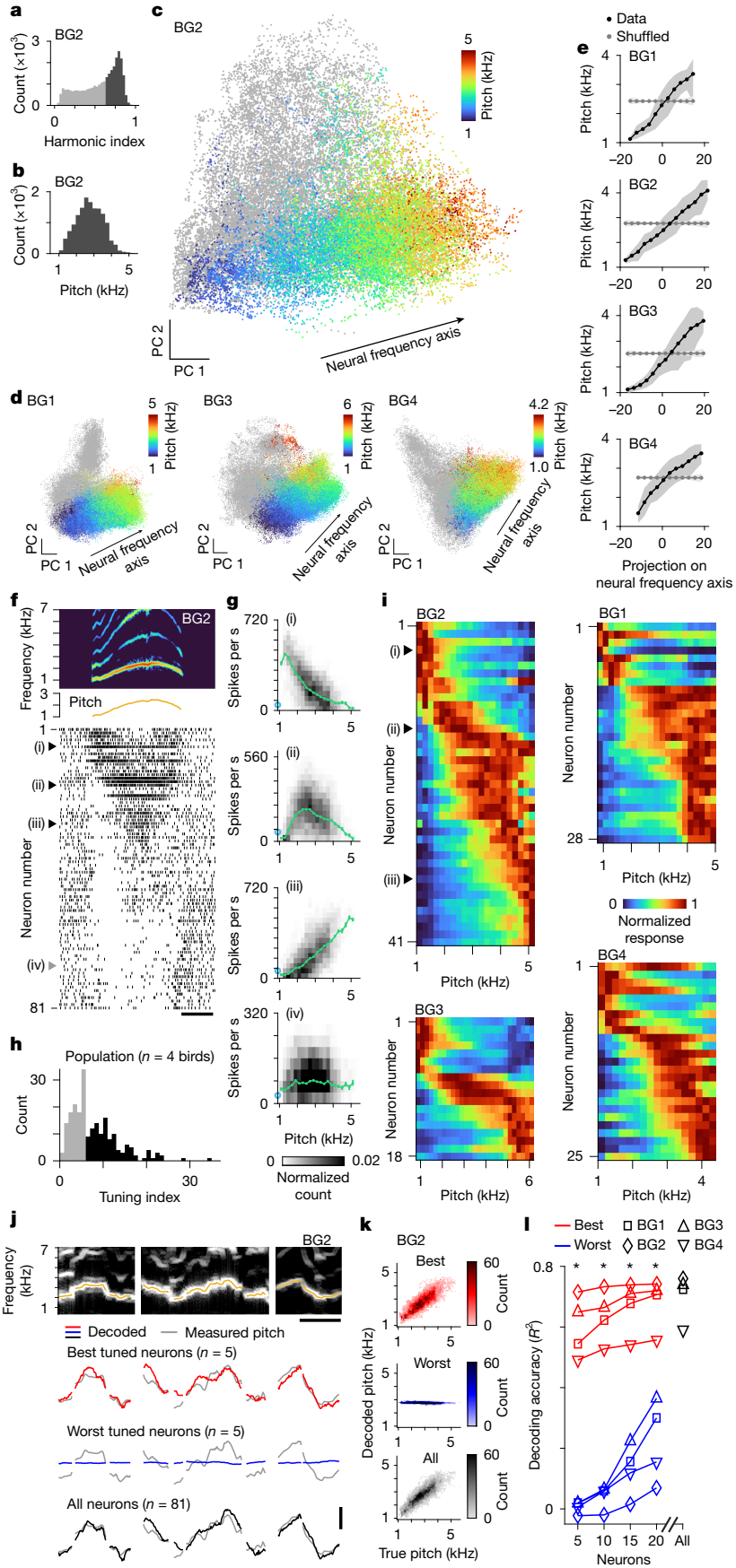


Fig. 4 | Neural representation of vocal pitch in the AAC. **a**, Vocalizations with clear fundamental frequencies are identified using the harmonic index. **b**, Pitch distribution for frequency-estimated vocal elements (dark grey in **a**) from budgerigar BG2. **c**, Population neural responses to vocalizations were mapped to a two-dimensional state space using PCA. Coloured dots represent neural states underlying frequency-estimated vocalizations (colour indicates pitch) and grey dots denote vocalizations with unmeasurable pitch. **d**, Neural state space for three other budgerigars. **e**, Relationship between pitch and the projection of neural states onto the neural frequency axis (black lines). Grey lines denote results from shuffled data. Lines represent the mean and shading indicates the 10th and 90th percentiles. **f**, AAC activity during the production of a frequency-modulated vocal element. Neurons sorted using Rastermap[62]. Scale bar, 100 ms. **g**, Pitch tuning in four example neurons from **f**, presented as binned scatter plots (grey) and averaged firing rates (mean ± s.e.m.) across pitches (green). Blue circles: baseline firing rates. **h**, Distribution of pitch-tuning index across four budgerigars. Pitch tuned (black): indices >6. **i**, Normalized response to pitches of all pitch-tuned neurons in each budgerigar. Triangles indicate neurons in **g**. **j**, Example sonograms (top) with measured and decoded pitch. Grey lines represent measured pitch (in yellow on sonogram) compared with cross-validated predictions of pitches using the five best pitch-tuned neurons (red), the five worst pitch-tuned neurons (blue) and all neurons (black). Image scale bar (top), 50 ms; vertical scale bar (bottom), 1 kHz. **k**, Relationship between decoded and measured pitches across all analysed vocalizations from BG2. Colours are the same as in **j**. **l**, Cross-validated decoding accuracy, quantified by the proportion of variance in measured pitch explained by the predicted pitch (R^2), for neural decoders incorporating different numbers of best tuned neurons (red), worst tuned neurons (blue) and all neurons (black). *$P < 0.01$ ($n = 4$ budgerigars, two-sided paired t-test; Extended Data Table 1).

frequency values. For instance, some neurons were most responsive when the bird produced low-frequency sounds (Fig. 4g, neuron (i)), whereas others were tuned to high frequency (Fig. 4g, neuron (iii)) or frequencies with intermediate values (Fig. 4g, neuron (ii)). Consistent with our previous result, we observed similar tuning profiles of single neurons across both calls and warbles (Extended Data Fig. 10b).

Across the population, pitch-tuned neurons had preferred frequencies that seemed to tile the entirety of the behaviourally relevant range (Fig. 4b,i and Extended Data Fig. 9b). To confirm that pitch is prominently encoded by AAC neurons, we built a generalized linear model (Methods) and found that pitch contributes significantly more than other acoustic factors (for example, entropy; Extended Data Fig. 10c).

As a last step, we trained a linear regression model to decode time-varying pitch from the spiking activity of AAC neurons. Using this approach, we attempted to predict the fundamental frequency of vocal elements that were held out during the training process (Methods). Using just five neurons with the largest tuning index, we could accurately predict pitch (Fig. 4j,k). By contrast, five untuned neurons (that is, lowest tuning index) did not enable any predictive power with respect to this parameter (Fig. 4j,k). As we incorporated more neurons into the model, decoding accuracy increased (Fig. 4j–l). Furthermore, decoders trained on warble syllables accurately predicted the pitch of calls (Extended Data Fig. 10d), supporting the notion that pitch is universally represented across vocal types in the AAC. This representation differs fundamentally from that observed in zebra finch RA, in which single neurons do not relate to a consistent pitch (Extended Data Fig. 11). Using decoders, we found no predictive value of the relationship between pitch and RA neuron spiking in the zebra finch ($n = 7$ birds, Pearson's $R = -0.02 \pm 0.20$; Extended Data Fig. 12a) compared with budgerigar AAC ($n = 4$ birds, Pearson's $R = 0.84 \pm 0.05$; Extended Data Fig. 12b). Taken together, our results confirm that AAC neurons systematically represent vocal pitch and imply that the AAC can exert precise control over this parameter.

Discussion

We investigated the neural mechanisms that underlie the production of flexible vocal elements in the budgerigar and found a highly structured representation of articulatory features at both single-cell and population levels in the forebrain area AAC. This mapping between neural population activity and vocal acoustics parallels the representation of phonetic features in the human speech cortex[2,4–6] and probably enables budgerigars to reuse and recombine existing neural solutions to produce new sounds, facilitating flexible vocalization and enabling rapid learning[44].

Several lines of evidence support the notion that AAC neurons encode parameters of the vocal musculature. First, we found contrasting neural population signatures for aperiodic (or noisy) and harmonic vocalizations. These two types of vocalization are linked to distinct oscillatory states of the avian vocal organ, the syrinx[45]. Although the anatomical structure of the budgerigar syrinx has been investigated[46], further ex vivo and modelling studies are needed to understand the syringeal mechanisms involved in vocal production[47]. An important consideration is that budgerigars and zebra finches exhibit key differences in the structure and mechanisms of the syrinx. In the zebra finch, the syrinx contains two sound sources[48], with each source independently controlled by the vocal motor nucleus and RA on the ipsilateral side[48,49]. By contrast, the budgerigar syrinx contains a single sound source[32,43], which could be controlled by their vocal motor nucleus and the AAC on both hemispheres[19,50]. Future research could investigate whether these differences in anatomical organization may relate to the distinct coding schemes observed between RA and the AAC. Second, the geometry and organization of the neural representation of vocalizations are highly consistent across individual birds (Figs. 3f and 4c,d). Because each budgerigar has a distinct vocal repertoire[51], our finding of a 'universal' motor representation in the budgerigar forebrain indicates that AAC neurons do not represent the individual vocalizations per se, but instead the underlying motor processes that generate those vocalizations. We thus propose that AAC neurons uniquely encode the acoustic outcome of motor commands using forebrain motor-control strategies resembling those observed in the human speech cortex. Future studies that simultaneously record AAC and syringeal muscle activity[52], together with pressure measurements of subsyringeal air sacs[53], can elucidate how AAC neurons coordinate breathing patterns with muscle contraction during vocal production.

The moment-to-moment control of articulatory parameters that we characterize in the budgerigar forebrain is fundamentally different from that observed in other non-human species studied to date. An important distinction can be made when comparing the budgerigar with the Bengalese finch (*Lonchura striata domestica*), in which changes to RA activity covary with song structure at specific moments[54], a process mediated by basal ganglia inputs critical for vocal learning[55]. As an ensemble, however, the activity observed in RA[7] as well as vocal forebrain regions in rodents[8] and non-human primates[9] does not seem to 'reuse' premotor commands for similar articulatory parameters, which suggests a distinct population code. Other vocal forebrain circuits featuring volitional signals[9,11] or information related to social context[10] also lack a representation of ongoing articulatory parameters. The vocal motor map in the budgerigar brain therefore allows a unique opportunity to investigate neural dynamics that functionally resemble those engaged during human speech production at both the circuit and cellular level. Differences still exist between these species, including the presence of only weak auditory responses in primary vocal production structures of the budgerigar (that is, the AAC), in contrast to the strong auditory responses observed in the human speech cortex[6,56]. Nevertheless, our results introduce the budgerigar as an exciting model for exploring mechanisms of vocal production and flexibility, and further experiments incorporating manipulation techniques have the potential to influence therapeutic approaches for a range of disorders that affect speech motor control[22,23].

In this study, we demonstrate that individual neurons in the budgerigar AAC represent motor commands with a distinct acoustic outcome. Similar to the encoding of colour in the retina[57], individual AAC neurons seem to feature broad tuning curves that span a wide frequency range (Fig. 4i), enabling an accurate frequency representation at the population level. How these AAC neurons are engaged by upstream structures remains poorly understood. The AAC receives inputs from several forebrain areas, including the central nucleus of the lateral nidopallium[19,20], a region lateral to the AAC, and two other regions in the anterior forebrain[18], the oval nucleus of the anterior mesopallium and the oval nucleus of the anterior nidopallium. Previous studies using targeted lesions[43] or inactivation[13] suggest that these two anterior areas are important for the production of frequency modulation in budgerigar calls. It is not yet known whether one of these regions primarily controls the AAC song production circuit or whether vocal production requires the coordinated activity of multiple upstream nuclei. Future investigations are needed to uncover how these regions may perform functions that resemble higher-level prefrontal areas involved in speech planning[58] and timing[59]. Taken together, we reveal a previously undescribed vocal motor interface in the budgerigar forebrain, raising the exciting possibility that features relevant for speech production may be mechanistically examined in an animal model.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-08695-8.

1. Eichert, N., Papp, D., Mars, R. B. & Watkins, K. E. Mapping human laryngeal motor cortex during vocalization. *Cereb. Cortex* **30**, 6254–6269 (2020).
2. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).

3. Chartier, J., Anumanchipalli, G. K., Johnson, K. & Chang, E. F. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron* **98**, 1042–1054 (2018).
4. Mugler, E. M. et al. Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri. *J. Neurosci.* **38**, 9803–9813 (2018).
5. Lu, J. et al. Neural control of lexical tone production in human laryngeal motor cortex. *Nat. Commun.* **14**, 6917 (2023).
6. Dichter, B. K., Breshears, J. D., Leonard, M. K. & Chang, E. F. The control of vocal pitch in human laryngeal motor cortex. *Cell* **174**, 21–31 (2018).
7. Leonardo, A. & Fee, M. S. Ensemble coding of vocal control in birdsong. *J. Neurosci.* **25**, 652–661 (2005).
8. Banerjee, A., Chen, F., Druckmann, S. & Long, M. A. Temporal scaling of motor cortical dynamics reveals hierarchical control of vocal production. *Nat. Neurosci.* **27**, 527–535 (2024).
9. Zhao, L. & Wang, X. Frontal cortex activity during the production of diverse social communication calls in marmoset monkeys. *Nat. Commun.* **14**, 6634 (2023).
10. Rose, M. C., Styr, B., Schmid, T. A., Elie, J. E. & Yartsev, M. M. Cortical representation of group social communication in bats. *Science* **374**, eaba9584 (2021).
11. Hage, S. R. & Nieder, A. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409 (2013).
12. Farabaugh, S. M., Brown, E. D. & Dooling, R. J. Analysis of warble song of the budgerigar *Melopsittacus undulatus*. *Bioacoustics* **4**, 111–130 (1992).
13. Zhao, Z. et al. Anterior forebrain pathway in parrots is necessary for producing learned vocalizations with individual signatures. *Curr. Biol.* **33**, 5415–5426 (2023).
14. Hile, A. G., Plummer, T. K. & Striedter, G. F. Male vocal imitation produces call convergence during pair bonding in budgerigars, *Melopsittacus undulatus*. *Anim. Behav.* **59**, 1209–1218 (2000).
15. Moussaoui, B., Overcashier, S. L., Kohn, G. M., Araya-Salas, M. & Wright, T. F. Evidence for maintenance of key components of vocal learning in ageing budgerigars despite diminished affiliative social interaction. *Proc. R. Soc. B* **290**, 20230365 (2023).
16. Lavenex, P. B. Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): the presence and implications of amplitude modulation. *J. Acoust. Soc. Am.* **106**, 491–505 (1999).
17. Egger, R. et al. Local axonal conduction shapes the spatiotemporal properties of neural sequences. *Cell* **183**, 537–548 (2020).
18. Durand, S. E., Heaton, J. T., Amateau, S. K. & Brauth, S. E. Vocal control pathways through the anterior forebrain of a parrot (*Melopsittacus undulatus*). *J. Comp. Neurol.* **377**, 179–206 (1997).
19. Paton, J. A., Manogue, K. R. & Nottebohm, F. Bilateral organization of the vocal control pathway in the budgerigar, *Melopsittacus undulatus*. *J. Neurosci.* **1**, 1279–1288 (1981).
20. Striedter, G. F. The vocal control pathways in budgerigars differ from those in songbirds. *J. Comp. Neurol.* **343**, 35–56 (1994).
21. Castellucci, G. A., Guenther, F. H. & Long, M. A. A theoretical framework for human and nonhuman vocal interaction. *Annu. Rev. Neurosci.* **45**, 295–316 (2022).
22. Utianski, R. L. et al. Prosodic and phonetic subtypes of primary progressive apraxia of speech. *Brain Lang.* **184**, 54–65 (2018).
23. Guenther, F. H. *Neural Control of Speech* (MIT Press, 2016).
24. Flinker, A. et al. Redefining the role of Broca's area in speech. *Proc. Natl Acad. Sci. USA* **112**, 2871–2875 (2015).
25. Dronkers, N. F. A new brain region for coordinating speech articulation. *Nature* **384**, 159–161 (1996).
26. Khanna, A. R. et al. Single-neuronal elements of speech production in humans. *Nature* **626**, 603–610 (2024).
27. Zann, R. A. *The Zebra Finch: a Synthesis of Field and Laboratory Studies* (Oxford Univ. Press, 1996).
28. Farabaugh, S. M., Linzenbold, A. & Dooling, R. J. Vocal plasticity in budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *J. Comp. Psychol.* **108**, 81–92 (1994).
29. Tu, H. W. & Dooling, R. J. Perception of warble song in budgerigars (*Melopsittacus undulatus*): evidence for special processing. *Anim. Cogn.* **15**, 1151–1159 (2012).
30. Simpson, H. B. & Vicario, D. S. Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.* **10**, 1541–1556 (1990).
31. Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
32. Heaton, J. T. & Brauth, S. E. Effects of lesions of the central nucleus of the anterior archistriatum on contact call and warble song production in the budgerigar (*Melopsittacus undulatus*). *Neurobiol. Learn. Mem.* **73**, 207–242 (2000).
33. Plummer, T. K. & Striedter, G. F. Auditory responses in the vocal motor system of budgerigars. *J. Neurobiol.* **42**, 79–94 (2000).
34. Elmaleh, M., Kranz, D., Asensio, A. C., Moll, F. W. & Long, M. A. Sleep replay reveals premotor circuit structure for a skilled behavior. *Neuron* **109**, 3851–3861 (2021).
35. Chi, Z. & Margoliash, D. Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* **32**, 899–910 (2001).
36. Chettih, S. N., Mackevicius, E. L., Hale, S. & Aronov, D. Barcoding of episodic memories in the hippocampus of a food-caching bird. *Cell* **187**, 1922–1935 (2024).
37. Schneidman, E., Bialek, W. & Berry, M. J. 2nd Synergy, redundancy, and independence in population codes. *J. Neurosci.* **23**, 11539–11553 (2003).
38. Mugler, E. M. et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. *J. Neural Eng.* **11**, 035015 (2014).
39. Mann, D. C., Fitch, W. T., Tu, H. W. & Hoeschele, M. Universal principles underlying segmental structures in parrot song and human speech. *Sci. Rep.* **11**, 776 (2021).
40. Manabe, K., Kawashima, T. & Staddon, J. E. Differential vocalization in budgerigars: towards an experimental analysis of naming. *J. Exp. Anal. Behav.* **63**, 111–126 (1995).
41. Manabe, K., Staddon, J. E. R. & Cleaveland, J. M. Control of vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*). *J. Comp. Psychol.* **111**, 50–62 (1997).
42. Seki, Y. Cockatiels sing human music in synchrony with a playback of the melody. *PLoS ONE* **16**, e0256613 (2021).
43. Brauth, S. E., Heaton, J. T., Shea, S. D., Durand, S. E. & Hall, W. S. Functional anatomy of forebrain vocal control pathways in the budgerigar (*Melopsittacus undulatus*). *Ann. NY Acad. Sci.* **807**, 368–385 (1997).
44. Moore, B. R. The evolution of learning. *Biol. Rev. Camb. Philos. Soc.* **79**, 301–335 (2004).
45. Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* **395**, 67–71 (1998).
46. Abdel-Maksoud, F. M., Hussein, M. M., Hamdy, A. & Ibrahim, I. A. Anatomical, histological, and electron microscopic structures of syrinx in male budgerigars (*Melopsittacus undulatus*). *Microsc. Microanal.* **26**, 1226–1235 (2020).
47. Elemans, C. P. et al. Universal mechanisms of sound production and control in birds and mammals. *Nat. Commun.* **6**, 8978 (2015).
48. Suthers, R. A., Goller, F. & Pytte, C. The neuromuscular control of birdsong. *Philos. Trans. R Soc. Lond. B* **354**, 927–939 (1999).
49. Wild, J. M. Neural pathways for the control of birdsong production. *J. Neurobiol.* **33**, 653–670 (1997).
50. Manogue, K. R. & Nottebohm, F. Relation of medullary motor nuclei to nerves supplying the vocal tract of the budgerigar (*Melopsittacus undulatus*). *J. Comp. Neurol.* **204**, 384–391 (1982).
51. Farabaugh, S. M. & Dooling, R. J. in *Ecology and Evolution of Acoustic Communication in Birds* (eds Kroodsma, D. E. & Miller, E. H.) Ch. 6 (Cornell Univ. Press, 1996).
52. Gaunt, A. S. & Gaunt, S. L. L. Electromyographic studies of the syrinx in parrots (Aves, Psittacidae). *Zoomorphology* **105**, 1–11 (1985).
53. Suthers, R. A., Goller, F. & Wild, J. M. Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *Proc. Natl Acad. Sci. USA* **99**, 5680–5685 (2002).
54. Sober, S. J., Wohlgemuth, M. J. & Brainard, M. S. Central contributions to acoustic variation in birdsong. *J. Neurosci.* **28**, 10370–10379 (2008).
55. Bottjer, S. W., Miesner, E. A. & Arnold, A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901–903 (1984).
56. Cheung, C., Hamiton, L. S., Johnson, K. & Chang, E. F. The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).
57. Solomon, S. G. & Lennie, P. The machinery of colour vision. *Nat. Rev. Neurosci.* **8**, 276–286 (2007).
58. Castellucci, G. A., Kovach, C. K., Howard, M. A. 3rd, Greenlee, J. D. W. & Long, M. A. A speech planning network for interactive language use. *Nature* **602**, 117–122 (2022).
59. Long, M. A. et al. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
60. Hozhabri et al. Differential behavioral engagement of inhibitory interneuron subtypes in the zebra finch brain. *Neuron* https://doi.org/10.1016/j.neuron.2024.11.003 (2024).
61. Moll, F. W. et al. Thalamus drives vocal onsets in the zebra finch courtship song. *Nature* **616**, 132–136 (2023).
62. Stringer, C. et al. Rastermap: a discovery method for neural population recordings. *Nat. Neurosci.* **28**, 201–212 (2025).

Methods

Animals

We used adult (>240 days after hatch) male budgerigars (*M. undulatus*) and adult (>90 days after hatch) male zebra finches (*T. guttata*) obtained from external breeders. The number of budgerigars was selected on the basis of previously published electrophysiological studies in birds[17,34,63]. All birds were maintained in a temperature- and humidity-controlled environment with a 12-h light/12-h dark schedule. All animal maintenance and experimental procedures conformed to the guidelines established by the Institutional Animal Care and Use Committee at the New York University Langone Medical Center.

Surgical procedures

All surgical procedures were performed under anaesthesia (1–3% isoflurane in oxygen) following established guidelines. We used similar approaches for implanting high-density silicon probes as those previously used by our group[17,34,63]. In brief, DiI (V22885, Thermo Fisher Scientific) was applied to probe shanks using a soft brush. The probe was then implanted into the target region by using stereotaxic coordinates relative to the bifurcation of the sagittal sinus. For recordings in budgerigars, we positioned the beak bar 17° down from horizontal and inserted the probe vertically. At this angle, the AAC coordinates were 5–6 mm lateral and 5–6 mm anterior. All AAC recordings were performed in the left hemisphere and in the dorsal subdivision of the AAC[18,20] (Extended Data Fig. 2b). A piezoelectric microphone (BU-33356-000, Knowles) was cemented to the skull overlying the right hemisphere using dental acrylic[64]. A ground wire was inserted under the skull approximately 3 mm posterior to the craniotomy site for the AAC. Throughout the implantation process, we continuously monitored electrical activity across all channels using the Intan recording system (RHD 1024ch Recording Controller and RHX data acquisition software, Intan Technologies). After successful identification of the AAC (depth: 3–3.6 mm), silicon elastomer (Kwik-Cast, WPI) was applied to the craniotomy. Our electrodes covered a spatial range of 450 μm anterior–posterior and 300 μm dorsal–ventral, enabling us to record from a substantial portion of the AAC (diameter: approximately 900 μm). After the recordings were completed, electrode positions were confirmed through histology (Extended Data Fig. 2a,b).

Recordings from zebra finch RA were collected as part of previously published studies by our group[34,63]. RA was located at 2.35 mm lateral and 0.1 mm posterior from the bifurcation of the superior sagittal sinus and 2.5–3.0 mm below the brain surface. RAs of both hemispheres were recorded in four zebra finches (ZF1–ZF4) and data were combined across hemispheres in these cases[63]. In the remaining three zebra finches (ZF5–ZF7), only RA of the right hemisphere was recorded[34].

Silicon probe and behavioural recordings

All AAC data were recorded in chronically implanted, freely moving budgerigars using integrated 128-channel high-density silicon probes (128-5, Diagnostic Biochips). Before implantation, a coated stainless-steel ground wire (0.0254-mm thickness, A-M Systems) was soldered to the reference contact of the headstage of the probe. The headstage was then secured in a customized protective casing made in-house using a 3D printer (Formlabs). Finally, the base of the shanks of the probe was fixed to a custom-designed microdrive by super glue (Loctite).

Chronic silicon probe recordings of budgerigars were performed in an arena in a sound-isolation chamber. The arena consisted of two conjoined cylindrical cages (Extended Data Fig. 13a) separated by a clear acrylic sheet. After probe implantation, the budgerigar was allowed to recover from anaesthesia and then put into one half of the arena, with four other familiar budgerigars housed in the other half, an arrangement that stimulated vocal production in recorded birds. We additionally supplemented the social environment with low-amplitude audio playback of budgerigar vocalizations through a nearby speaker[29]. Budgerigar vocalizations were recorded using a piezoelectric microphone (BU-33356-000, Knowles), which conducted sounds through the skull and provided high-quality and highly selective vocal signals with little contamination from environmental sounds[64] (Extended Data Fig. 13b).

We continuously recorded neural activity and vocalizations from the bird immediately after being placed in the arena. Neural and piezoelectric signals were transmitted to the Intan recording system (RHD 512ch Recording Controller, Intan Technologies) through an assisted electrical commutator (Doric Lenses), which facilitates free movement of the bird. Audio signals from both the piezoelectric and ambient microphones (AT803, Audio-Technica) were amplified by an analog preamplifier (TPS II, ART ProAudio) before being sent to the Intan controller. All data were sampled at 30 kHz. Playback experiments through a nearby speaker were performed on two budgerigars (BG3 and BG4). Three calls and three warble segments (duration: approximately 1 s) from the repertoire of the bird were played back during non-vocal periods. Each audio stimulus was played ten times in a randomized order.

Details pertaining to recording procedures in zebra finches have been described elsewhere[34,63]. Data from one zebra finch (ZF1) were collected acutely from a bird trained to sing while head-fixed[63]. In this bird, neural activity from RA was collected using a 64-channel silicon probe (64-H, probe obtained from S. Masmanidis). For all other zebra finches, neural recordings were performed chronically using the same 128-channel probes as used in budgerigar recordings. Vocalizations were elicited by female finches placed in a neighbouring cage and recorded using an omnidirectional microphone (AT803, Audio-Technica). To directly compare vocalizations of zebra finches with those of budgerigars (Fig. 1c–f), we implanted a piezoelectric microphone (BU-33356-000, Knowles) in a single zebra finch. After the surgery, the bird was placed in a cage on one side of the arena while a cage on the other side housed two female zebra finches to elicit songs from the implanted bird.

Histology

After recording, birds were perfused transcardially with PBS followed by 4% paraformaldehyde. Probes were lifted from the brain using the microdrive, and brains were extracted and left overnight in paraformaldehyde. Subsequently, the brains were sliced at a thickness of 100 μm using a vibrating microtome (VT1200S, Leica). Slices were mounted using Vectashield mounting medium (Vector Laboratories) and imaged with a slide scanner (VS120, Olympus) using the OlyVIA software.

Data analysis

Detection of vocalizations. For each budgerigar, we analysed data from the first day that birds produced warble songs following electrode implantation. Two budgerigars sang on the same day as the surgery, whereas the other birds initiated singing after one or two days. For each bird, we examined the initial period of audio recording containing at least 5 min of cumulative vocalizations, including both calls and warble songs but excluding the silent gaps between them. One of the four birds (BG2) vocalized less than 5 min during the recording session, so we analysed the entire audio recording.

Budgerigar vocalizations were detected using spectral features. Audio signals were first high-pass filtered (250 Hz) and then converted into spectrograms using the short-time Fourier transform (window size: 8.5 ms, step size: 1 ms). For each moment in the sonogram, a 3-ms analysis window was evaluated; vocalizations were distinguished from silent periods by a minimum set of 50 continuous spectrotemporal samples exceeding a threshold. Consecutive time points were then grouped into intervals as putative vocal periods and sounds with durations shorter than 2 ms were discarded as non-vocal. We then manually removed any false-positive sounds (for example, originating from mechanical noise). After this process, each sound was defined as a single vocal element, and warble songs were defined as vocal sequences comprising a minimum of three vocal elements separated by silent periods of less

than 1 s (ref. 65). The remaining vocal elements were classified as calls. For zebra finches, identities of vocalizations (that is, call 1, syllable A, and so on) were annotated manually[34,61].

Comparison of vocal repertoires across species. We compared vocal repertories between the zebra finch and budgerigar (Fig. 1c–f) using vocalizations recorded by piezoelectric microphones. For this analysis, additional instances of three budgerigar call types were included from an independent recording session to increase the total number of trials. For the human speech analysis, data were taken from the Switchboard-1 Telephone Speech Corpus (session ID: sw02012B)[66], which totalled 755 words spoken by an adult male English speaker using annotations provided by The Institute for Signal and Information Processing (https://isip.piconepress.com/projects/switchboard/).

For this initial analysis (Fig. 1b,d,f), dimensionality reduction techniques were applied to analyse the spectral structure of vocalizations from all species[61]. Background noise in each spectrogram was suppressed by applying a threshold set at 60% of the range of spectral energy. Spectrograms were then standardized to a uniform length of 300 ms through zero padding, after which they were projected into a 25-dimensional space using PCA. For visualization, *t*-SNE was used to embed each 25-dimensional vocal element using a perplexity parameter of 30.

We also quantified three acoustic features for each vocal element: duration, mean pitch and entropy variance (Extended Data Fig. 1). Mean pitch and entropy variance were calculated on the basis of pitch and entropy estimations using Sound Analysis Pro (SAP)[67].

Quantification of vocal-related neuronal activity. Spike sorting was performed using Kilosort[68] and sorted clusters were further manually curated using Phy[69] on the basis of criteria including interspike intervals, auto-correlogram, cross-correlogram and waveform shape, as we had previously done in the zebra finch[17,34,63]. Two metrics, mean firing rates and bursting ratios, were used to assess neuronal responses to vocalizations. The mean firing rate for each neuron was computed as the total number of spikes divided by the total duration of specific time intervals. The bursting ratio was defined as the percentage of time a given neuron exhibited bursting activity. To quantify bursting, spiking patterns were first converted into instantaneous firing rates at a 1-ms resolution by using the inverse of the interspike interval that immediately surrounds each time point[7]. Bursting was then defined as instantaneous firing rates exceeding a designated threshold (100 Hz for zebra finch RA[34,63] and 200 Hz for budgerigar AAC) on the basis of the distribution of interspike intervals (Extended Data Fig. 3h). These metrics were computed for both vocalization intervals (calls and song syllables) and for 10 s of baseline periods during which birds exhibited vocal quiescence.

To investigate the temporal relationship between AAC activity and vocal boundaries, we aligned the activity of each neuron to the onset of vocal elements that are preceded by a silent gap of at least 100 ms (Extended Data Fig. 3b). Furthermore, we aligned AAC activity with the end of vocal elements that are followed by a silent gap of at least 100 ms (Extended Data Fig. 3c).

For each neuron, sparseness[70] of its vocal activity was defined as

$$\text{Sparseness} = \left(\sum_{t=1}^{T} \frac{R_t}{T}\right)^2 \Bigg/ \left(\sum_{t=1}^{T} \frac{R_t^2}{T}\right) \quad (1)$$

where R_t was the instantaneous firing rate to the tth time point and T represented the summed duration in millisecond across vocal elements. Smaller sparseness is associated with more sparse neural activity.

Spectral and neural similarity analyses. We compared spectral and neural similarity between vocal elements. For zebra finches, all song syllables were included. For budgerigars, 50 syllables were randomly sampled from the pool of song syllables with durations longer than 100 ms. Any repeated syllables were then removed, leading to 38, 50, 43 and 45 unique warble syllables for each bird for subsequent analysis. This selection process was adopted only for this specific analysis (Fig. 2i,j). For each possible pair of syllables in each bird, spectrograms were segmented using a 20 ms sliding window with a step size of 10 ms. Pearson correlations were computed between spectral segments of two syllables to generate the spectral similarity matrix. A neural similarity matrix was calculated using segments of population neural activity (that is, *z*-scored instantaneous firing rates) during production of the two syllables. For each neuron, the instantaneous firing rates were converted to *z*-score using the mean and standard deviation of the firing rates across vocal elements. Neural activity was shifted forward by 12 ms before calculating similarity to account for vocal motor delay[7]. To quantify the relationship between spectral and neural structures, similarity matrices from all pairs of syllables were flattened and concatenated into a vector. Next, a Spearman correlation was performed between the concatenated spectral and neural similarity vectors.

To quantify similarity across renditions of the same vocal element, we analysed a dataset of zebra finch vocalizations recorded using the same piezoelectric microphone as for the budgerigars (BU-33356-000, Knowles; Extended Data Fig. 4a–c). We selected ten renditions of the four syllables produced by the zebra finch and segmented each syllable using identical parameters to those applied to the budgerigars (that is, 20-ms sliding window with a step size of 10 ms) and then calculated the correlation between all different renditions of the same segment, as well as the correlation between different segments across syllables.

Acoustic categorization of vocalizations. Acoustic properties of vocal elements were characterized by a low-frequency ratio and harmonic index. For every vocal time point (resolution: 1 ms), the low-frequency ratio was calculated as the logarithm of the ratio between mean spectral power below 700 Hz and that between 700 and 7,000 Hz. These two bands were chosen according to the fact that low-frequency syllables exhibit a fundamental frequency below 700 Hz (Extended Data Fig. 5a), whereas other syllables have fundamental frequencies exceeding this threshold and below 7,000 Hz (Extended Data Fig. 9b). The harmonic index was defined as HR × FI, where HR quantified the ratio of spectral energy that was harmonic and was calculated using the 'harmonicRatio' function in MATLAB (https://www.mathworks.com/help/audio/ref/harmonicratio.html; MATLAB Audio Toolbox, R2023a, Mathworks). FI, which represents the prominence of a fundamental frequency, was derived by first estimating the fundamental frequency of all vocalizations using SAP[67]. Mean spectral power was then computed for a fundamental band (400 Hz centred around the estimated fundamental frequency) and a nonfundamental band (400 Hz centred around 1.5 or 0.67 times the fundamental frequency for fundamental frequencies below or above 4.5 kHz, respectively). FI was then defined as the logarithm of the ratio between the mean spectral power of the two bands. FI was normalized to [0, 1] before calculating the harmonic index.

Vocal time points with low-frequency ratios above 1 were classified as low-frequency sounds. The remaining vocalizations were considered to have normal frequencies and displayed a bimodal distribution in the harmonic index (Fig. 3c). Vocal time points that fell in the bottom 20% and top 30% of harmonic indices pooled across birds were labelled as 'noisy' and 'harmonic' sounds, respectively, while those between the 30th and 50th percentiles were designated as 'mixed' sounds. To avoid the inclusion of short silent gaps in vocal elements, time intervals in any acoustic category with durations ≤4 ms were excluded.

Neural mapping of acoustic categories. To characterize single-neuron responses to the three acoustic categories, we defined a selectivity index for each category (Extended Data Fig. 5) as: $(R_{cat} - R_{base})/(R_{cat} + R_{base})$,

where R_{cat} was the mean response of each neuron to the category studied and R_{base} is the baseline firing rate of each neuron.

For population-level analyses, an N-dimensional (N: number of neurons) vector was computed by calculating the firing rate of each neuron during a window from −30 to −5 ms to each 1 ms time point during vocalization. These N-dimensional vectors were then organized into an $M \times N$ neural response matrix, where M represented the total number of vocal time points. PCA was applied to the matrix, through which each N-dimensional neural vector was mapped into a three-dimensional neural space (Extended Data Fig. 6a) and then colour coded according to the corresponding acoustic categories. Eigenvalues and explained variances of the principal components (PCs) were examined to determine the contribution of each PC (Extended Data Fig. 6b).

To quantify the relationship between neural state and the harmonic index, we build linear regression models associating neural PC scores with the harmonic index:

$$h_t = \beta_0 + \sum_{i=1}^{N} \beta_i s_{it} \quad (2)$$

where h_t was the harmonic index at each time point t (resolution: 1 ms), s_{it} was the score of the population response on the ith PC, and β_0, ..., β_N were regression coefficients to be estimated, with N being the total number of PCs. For each budgerigar, the model was fitted on half of the data and tested on the other half. Model performance was evaluated by R^2:

$$R^2 = 1 - \frac{\sum_t (H_t - h_t)^2}{\sum_t (H_t - \overline{H})^2} \quad (3)$$

where H_t was the observed harmonic index, h_t was the predicted index, and $\overline{H}$ was the mean of the observed index.

For computational efficiency, distance between neural states within and across categories were calculated on the basis of the first three PCs. Two control analyses were performed to exclude the possibility that the smaller neural distance within categories was owing to the binning procedure. We performed a permutation analysis (Extended Data Fig. 8b). In each permutation, we shifted the spike train of each neuron by a random time between 100 and 500 ms. We then calculated the neural response distances within and between different acoustic categories on the basis of the same binning process mentioned above. We defined the statistic ΔD as the difference between the mean distance of neural responses between categories and within categories: ΔD = mean (between-category distance) − mean (within-category distance). A larger ΔD indicates greater similarity in neural responses within categories compared with between categories. To generate a null distribution, we repeated the permutation process 1,000 times. We then calculated ΔD from the original unshuffled data.

Moreover, we performed a subsampling of the vocal time points, selecting those that were at least 25 ms apart to ensure that no N-dimensional neural vectors contained overlapping spikes. We then recalculated the neural distances within and between categories (Extended Data Fig. 8c,d).

Comparison of vocalizations across individual budgerigars. To compare the vocal repertoires of the four budgerigars, we first standardized each vocal element (including both calls and warbles) into spectrograms with a fixed duration of 125 ms. Vocal elements exceeding this length were segmented into 125-ms components using a sliding window with a step size of 5 ms. We then used a machine-learning-based algorithm, the variational autoencoder provided by the Autoencoded Vocal Analysis toolbox[71], to represent each vocal element by latent features[13]. These latent representations were subsequently embedded into a low-dimensional space using uniform manifold approximation and projection (UMAP; Extended Data Fig. 7).

Neural mapping of vocal pitch. For each budgerigar, vocalizations for which the fundamental frequency could be estimated reliably were isolated using the median as a threshold for the harmonic index (Extended Data Fig. 9a). All of the subsequent analyses were done on the basis of this set of frequency-tagged time periods. Pitch was estimated using SAP[67]. To visualize the representation of pitch in the neural ensemble, responses to these vocal time points were colour coded by the associated pitch in the two-dimensional PC space.

We quantified the geometry of the neural representation of pitch in the full-dimensional PC space. To avoid double dipping, all neural responses were first evenly split into odd and even halves (Extended Data Fig. 9c). In the first half, a neural frequency axis was computed by calculating the vector between the mean neural responses to the top 50% and bottom 50% of pitches. Neural responses from the second half were then projected onto this neural frequency axis using the dot product of these two vectors. The relationship between pitch and neural projections was measured by computing the mean pitch in a series of neural projection bins (bin size: 3). For reliability, only bins containing more than 20 samples were included.

A control analysis was performed by randomly shuffling the pitch values (Fig. 4e) across vocalizations followed by the calculation of the neural frequency axis described above. This shuffling process was repeated 5,000 times. For each shuffled sample, the slope of the relationship between mean pitch and neural projection was determined by fitting a linear model between the two variables. The distribution of slopes obtained from the shuffled data was then compared with the slope derived from the unshuffled data (Extended Data Fig. 9d).

We also investigated the neural representation of pitch using two windows (−125 to −100 ms and 100 to 125 ms), which were significantly shifted from our original window (−30 to −5 ms; Extended Data Fig. 9e,f). For these three representations, we evaluated the extent to which pitch could be predicted on the basis of PC1 and PC2. Specifically, for each point on the two-dimensional neural map, we trained a linear regression model using all remaining data points to associate the coordinates of PC1 and PC2 with pitch. This model was then used to predict the pitch value of the left-out point. The squared error between the actual and predicted pitch values was used as a measure of model performance.

Single-neuron tuning to vocal pitch. For a given neuron, activity (firing rates over a window from −30 to −5 ms) to different pitch bins (bin size: 200 Hz) were averaged within the bin to obtain a tuning curve (Fig. 4g). Bins containing fewer than 20 samples were not included. To quantify the strength of tuning, mean neural responses were first calculated for four pitch ranges (1–2 kHz, 2–3 kHz, 3–4 kHz and 4–5 kHz). A tuning index (Fig. 4h) was defined as the range of the four neural responses divided by the square root of the mean response to all pitches (Extended Data Fig. 10a).

To evaluate the relative contributions of pitch and other acoustic features to neuronal responses in the AAC, we performed a generalized linear modelling (GLM) analysis. For each neuron, we modelled responses to all vocal time points (1-ms resolution) using a Poisson GLM with an exponential link function. The four predictors included in the model are pitch, entropy, amplitude and frequency modulation. To test the significance of model fitting, we calculated the likelihood ratio test statistic:

$$\lambda = -2 \times (\ell_{null} - \ell_{model}) \quad (4)$$

where ℓ_{model} was the log-likelihood of the full model and ℓ_{null} was the log-likelihood of a null model in which no predictors were included. We then shuffled the neural responses 1,000 times and calculated a null distribution of λ. The model fitting is considered significant only when λ is larger than the 99th percentile of its null distribution.

For all neurons with a significant GLM model fitting, we calculated the relative contribution of each predictor (Extended Data Fig. 10c) using a previously described[72,73] equation: $\left(1-\frac{MR^2_{p,i}}{MR^2_f}\right)\Big/\sum_{j=1}^{4}\left(1-\frac{MR^2_{p,j}}{MR^2_f}\right)$, where MR^2_f was the McFadden's R^2 for the full model and $MR^2_{p,i}$ represented the McFadden's R^2 for a partial model with the ith predictor excluded.

Neural decoding of vocal pitch. Any vocal elements containing frequency-tagged periods were included in the decoding analysis. A leave-one-out approach[74] was used for training and predicting pitch. Assuming there were V qualified vocal elements, for each of them, population responses to the remaining V-1 vocal elements were used to build a linear regression model:

$$f_t = \beta_0 + \sum_{i=1}^{N} \beta_i \times r_{it} \quad (5)$$

where f_t was the fundamental frequency at a frequency-tagged time point t (resolution: 1 ms), r_{it} was the firing rate of the ith neuron over the time window of $t-30$ to $t-5$ ms and $\beta_0, \ldots, \beta_N$ were regression coefficients to be estimated, with N being the number of neurons. This model was then applied to the neural responses to the held-out vocal element to predict fundamental frequency.

Various populations of neurons were used to compare decoding performance, which was quantified by R^2:

$$R^2 = 1 - \frac{\sum_t (F_t - f_t)^2}{\sum_t (F_t - \overline{F})^2} \quad (6)$$

where F_t was the observed pitch, f_t was the predicted pitch, and $\overline{F}$ was the mean of the observed pitch.

To investigate whether warble syllables and calls are represented similarly in the AAC population, we performed an extra decoding analysis in which the linear model was trained on warble syllables only and then used to decode pitch of calls (Extended Data Fig. 10d).

Pitch representation in zebra finch RA. We examined pitch representation in RA at both the single-neuron and population level. For each RA neuron, all burst events occurring during song production were identified (Extended Data Fig. 11c). These burst intervals were then adjusted by a motor delay (12 ms) to determine the corresponding vocal intervals, from which mean pitch values were extracted. These burst-associated pitch values were then plotted across the whole population (Extended Data Fig. 11d). For each neuron, we calculated the mean and variance of its burst-associated pitch values. A random sample of pitch values, matching the number of burst-associated values, was drawn from the rest of the population to establish a null distribution for the mean and variance of burst-associated pitch. Moreover, we plotted the relationship between firing rates (measured in a window of −30 to −5 ms) and pitch values across syllables, calls and introductory notes in the vocal repertoire of a zebra finch.

At the population level, we performed pitch decoding similar to the approach used with the budgerigar. Specifically, for each song syllable, a linear decoder (as described by equation (5)) was trained to associate RA neuron firing rates with the pitch of the remaining syllables. This decoder was then used to predict the pitch of the withheld syllable (Extended Data Fig. 12a).

Statistics. Nonparametric statistical tests (for example, Wilcoxon rank-sum and permutation tests) were performed without assumptions regarding data distribution. In Fig. 4l, for which the sample size is small ($n = 4$), a parametric test was adopted. Exact P values, sample sizes and other information for each statistical test are detailed in Extended Data Table 1.

Reporting summary

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

Data availability

Data used in this study have been uploaded to Zenodo (https://doi.org/10.5281/zenodo.14057061)[75]. The Switchboard Telephone Speech Corpus is accessible through the Linguistic Data Consortium (https://catalog.ldc.upenn.edu/LDC97S62) and existing annotations of the corpus are provided by The Institute for Signal and Information Processing (https://isip.piconepress.com/projects/switchboard/). Source data are provided with this paper.

Code availability

The primary code used in this study is available at Zenodo (https://doi.org/10.5281/zenodo.14057061)[75].

63. Elmaleh, M., Yang, Z., Ackert-Smith, L. A. & Long, M. A. Uncoordinated sleep replay across hemispheres in the zebra finch. *Curr. Biol.* **33**, 4704–4712 (2023).
64. Fukushima, M. & Margoliash, D. The effects of delayed auditory feedback revealed by bone conduction microphone in adult zebra finches. *Sci. Rep.* **5**, 8800 (2015).
65. Tobin, C., Medina-Garcia, A., Kohn, G. M. & Wright, T. F. Does audience affect the structure of warble song in budgerigars (*Melopsittacus undulatus*)? *Behav. Processes* **163**, 81–90 (2019).
66. Godfrey, J. J. & Holliman, E. Switchboard-1 Release 2 LDC97S62. *Linguistic Data Consortium* https://doi.org/10.35111/sw3h-rw02 (1993).
67. Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167–1176 (2000).
68. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., Harris, K. D. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. Preprint at *bioRxiv* https://doi.org/10.1101/061481 (2016).
69. Rossant, C. et al. Spike sorting for large, dense electrode arrays. *Nat. Neurosci.* **19**, 634–641 (2016).
70. Rolls, E. T. & Tovee, M. J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **73**, 713–726 (1995).
71. Goffinet, J., Brudner, S., Mooney, R. & Pearson, J. Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife* **10**, e67855 (2021).
72. Engelhard, B. et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019).
73. Ito, B. S., Gao, Y., Kardon, B. & Goldberg, J. H. A collicular map for touch-guided tongue control. *Nature* **637**, 1143–1151 (2025).
74. Chang, L. & Tsao, D. Y. The code for facial identity in the primate brain. *Cell* **169**, 1013–1028 (2017).
75. Yang, Z. & Long, M. A. Data and code for 'convergent vocal representations in parrot and human forebrain motor networks'. *Zenodo* https://doi.org/10.5281/zenodo.14057061 (2025).

Acknowledgements We thank A. Bahle, C. Elemans, A. Fishbein, D. Jin, C. Kemere, M. Schmidt and current members of the Long laboratory for comments on earlier versions of this manuscript, B. Cooper, T. Movshon and D. Sanes for helpful conversations and A. Paulson for technical assistance. This research was supported by Simons Collaboration on the Global Brain (M.A.L.).

Author contributions Z.Y. and M.A.L conceived the study and designed the experiments; Z.Y. conducted the research; Z.Y. performed data analyses; Z.Y. and M.A.L. created the figures; Z.Y. wrote the initial draft of the manuscript; Z.Y. and M.A.L. edited and reviewed the final manuscript; M.A.L acquired funding and supervised the project.

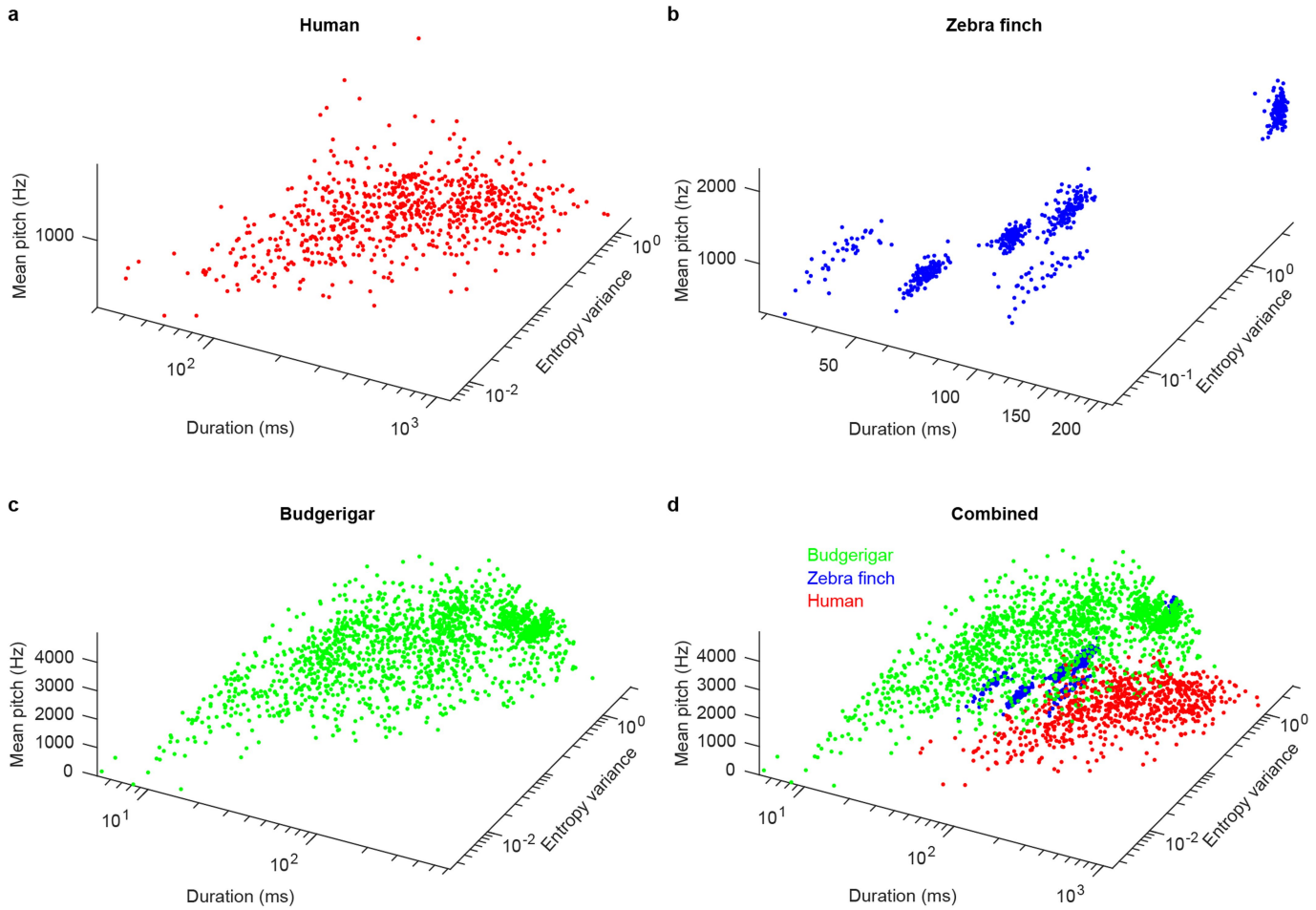
Competing interests The authors declare no competing interests.

Additional information

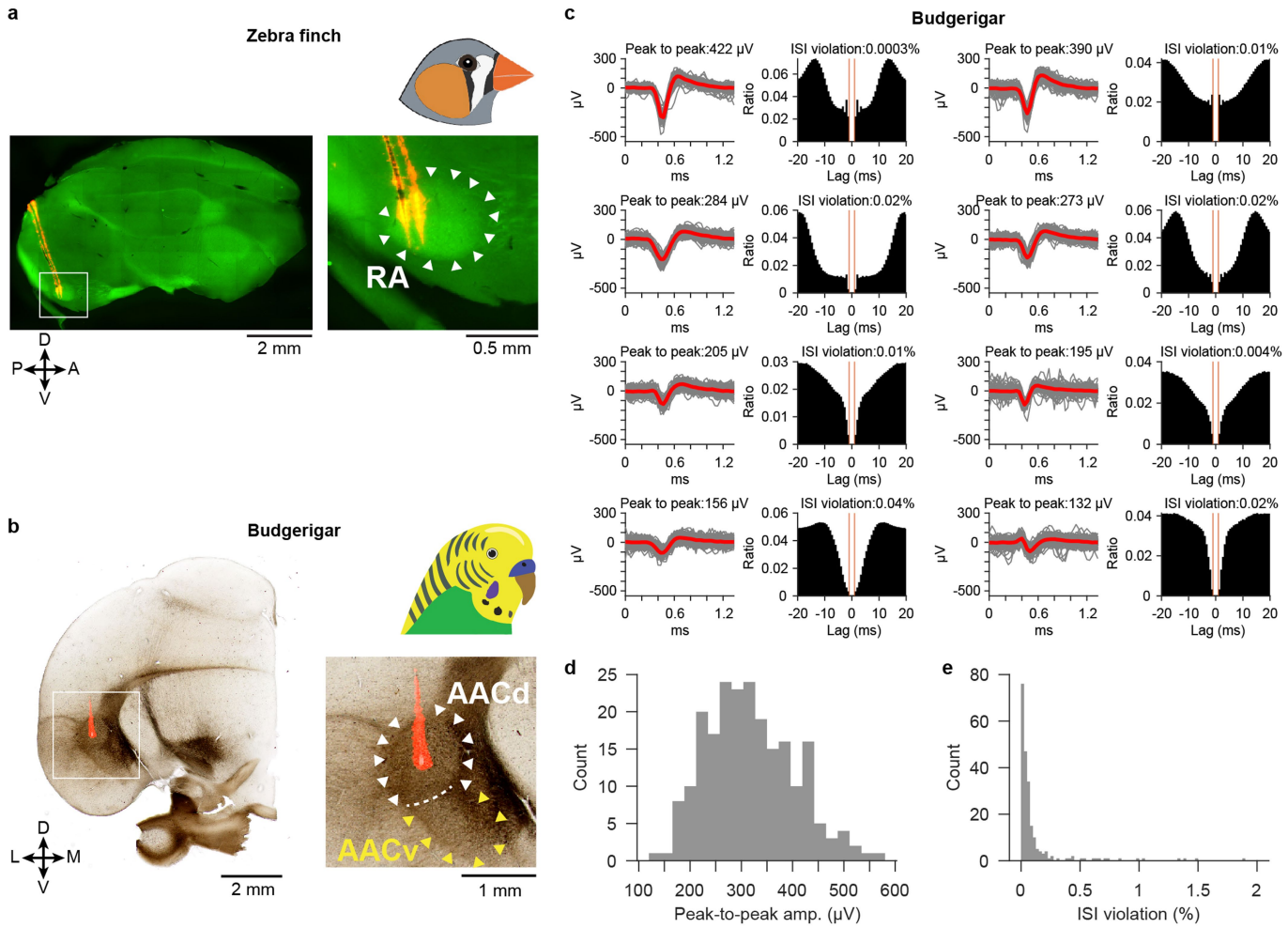
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-08695-8.

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

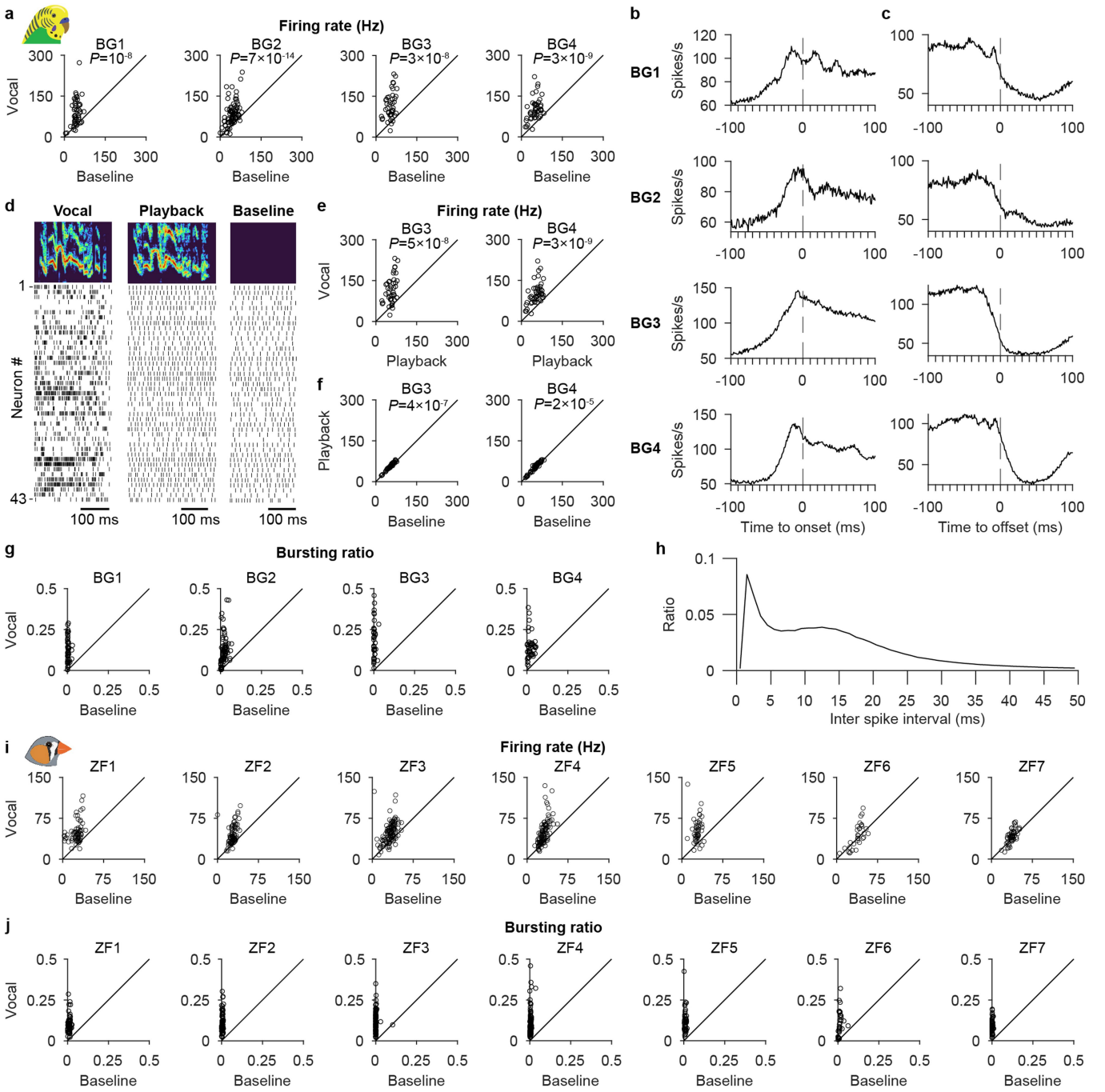
Peer review information *Nature* thanks Todd Troyer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



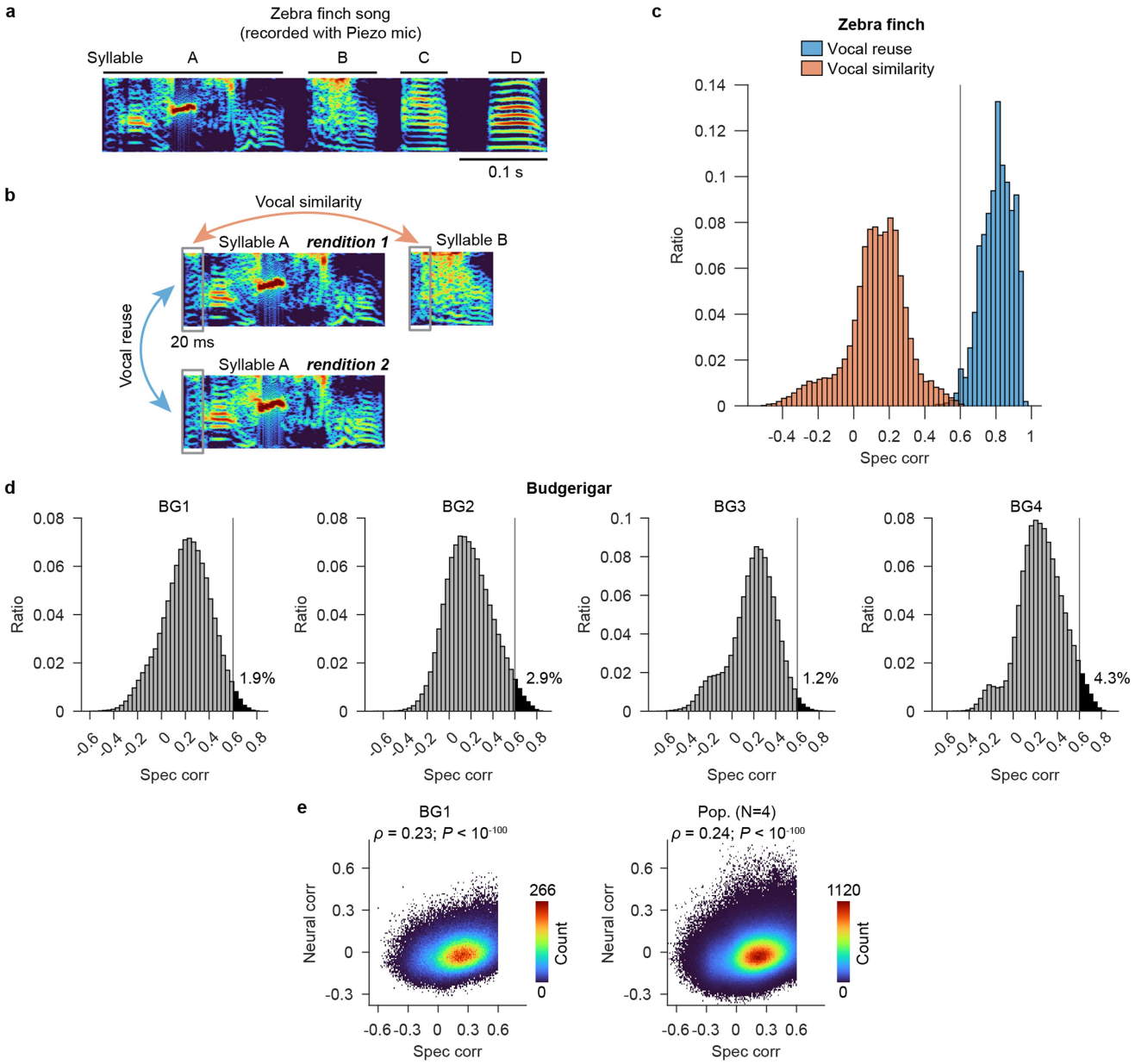
Extended Data Fig. 1 | Acoustic representation of vocal repertoires across species. a-c, Mean pitch, duration, and entropy variance from a human (**a**), a zebra finch (**b**), and a budgerigar (**c**). Data represent the same spoken words and vocal elements analyzed in Fig. 1a–f. **d**, Overlay of data from panels **a** through **c**.



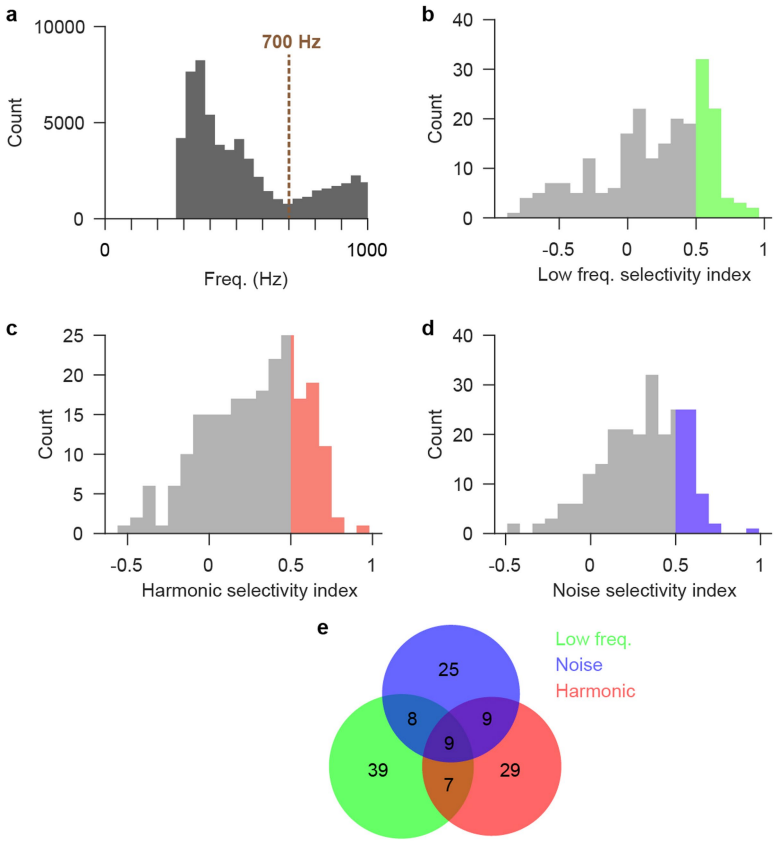
Extended Data Fig. 2 | Histology and single unit quality. a, Histological validation of silicon probe placement in zebra finch RA. *Left*, sagittal slice illustrating probe position, marked by DiI applied to the probe shank (red). *Right*, expanded view of the boxed region highlighting RA (white arrows) on the right. **b**, Histological validation of silicon probe placement in budgerigar AAC. *Left*, brightfield image of a coronal slice showing the position of the silicon probe, marked by DiI applied to the probe shank (red). *Right*, expanded view of the boxed region indicating dorsal AAC (AACd) and ventral AAC (AACv) by white and yellow arrows, respectively. **c**, Example single-unit waveforms and autocorrelograms from budgerigar AAC recordings. Mean waveforms (red) are overlaid on single-trial waveform traces, which are represented as gray lines. In the autocorrelogram, orange vertical lines indicate the refractory period (−1 to 1 ms). **d**, **e**, Distributions of peak-to-peak amplitude (**d**) and the percentage of refractory period violations (**e**) across all AAC single units (n = 220 from 4 birds). Illustrations of the zebra finch in **a** are reproduced from ref. 60, Elsevier.



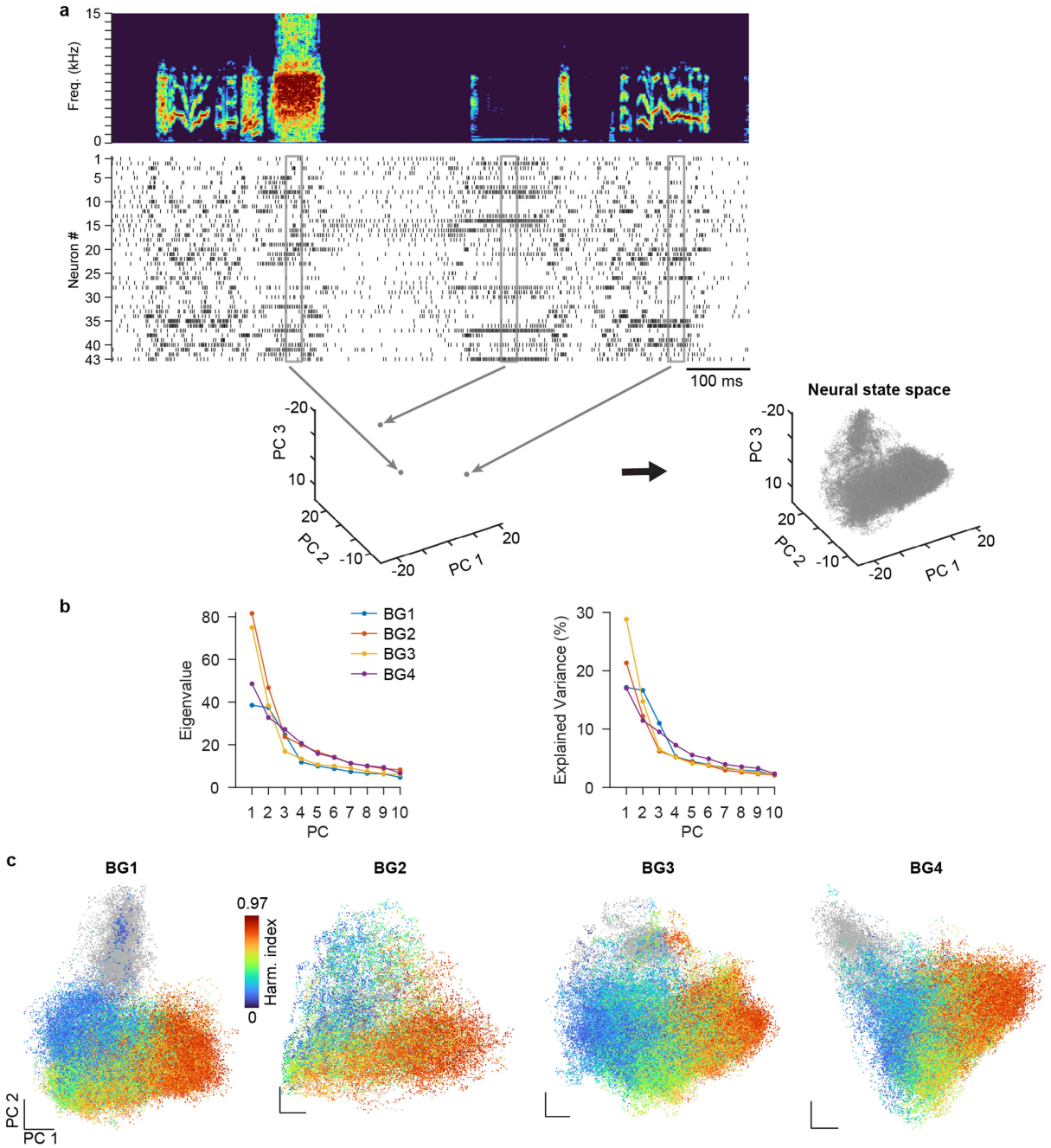
Extended Data Fig. 3 | Vocal motor responses in AAC and RA. a, Mean firing rates of AAC neurons during vocalization and baseline for each individual budgerigar. Exact p-values displayed within each plot are from two-sided Wilcoxon signed-rank tests (n = 49, 81, 43, 47 neurons for each plot). **b**, Population-averaged responses aligned to the onset of vocal elements that are preceded by a silent gap of at least 100 ms. **c**, Population-averaged responses aligned to the offset of vocal elements followed by a silent gap of at least 100 ms. **d**, Spiking activity from a population of AAC neurons (BG3) during example vocalization, playback, and quiet baseline periods. **e**,**f**, Mean firing rates of AAC neurons during vocalization compared with playback (**e**) and during playback compared with baseline (**f**). Exact p-values displayed within each plot are from two-sided Wilcoxon signed-rank tests (n = 43, 47 neurons for the BG3 and BG4 plots, respectively). **g**, Bursting ratios of AAC neurons during vocalization and baseline for each individual budgerigar. **h**, Distribution of interspike intervals of AAC neurons pooled across budgerigars. **i**,**j**, Mean firing rates (**i**) and bursting ratios (**j**) of RA neurons during vocalization and baseline for each zebra finch. Illustrations of the zebra finch in **i** are reproduced from ref. 60, Elsevier.



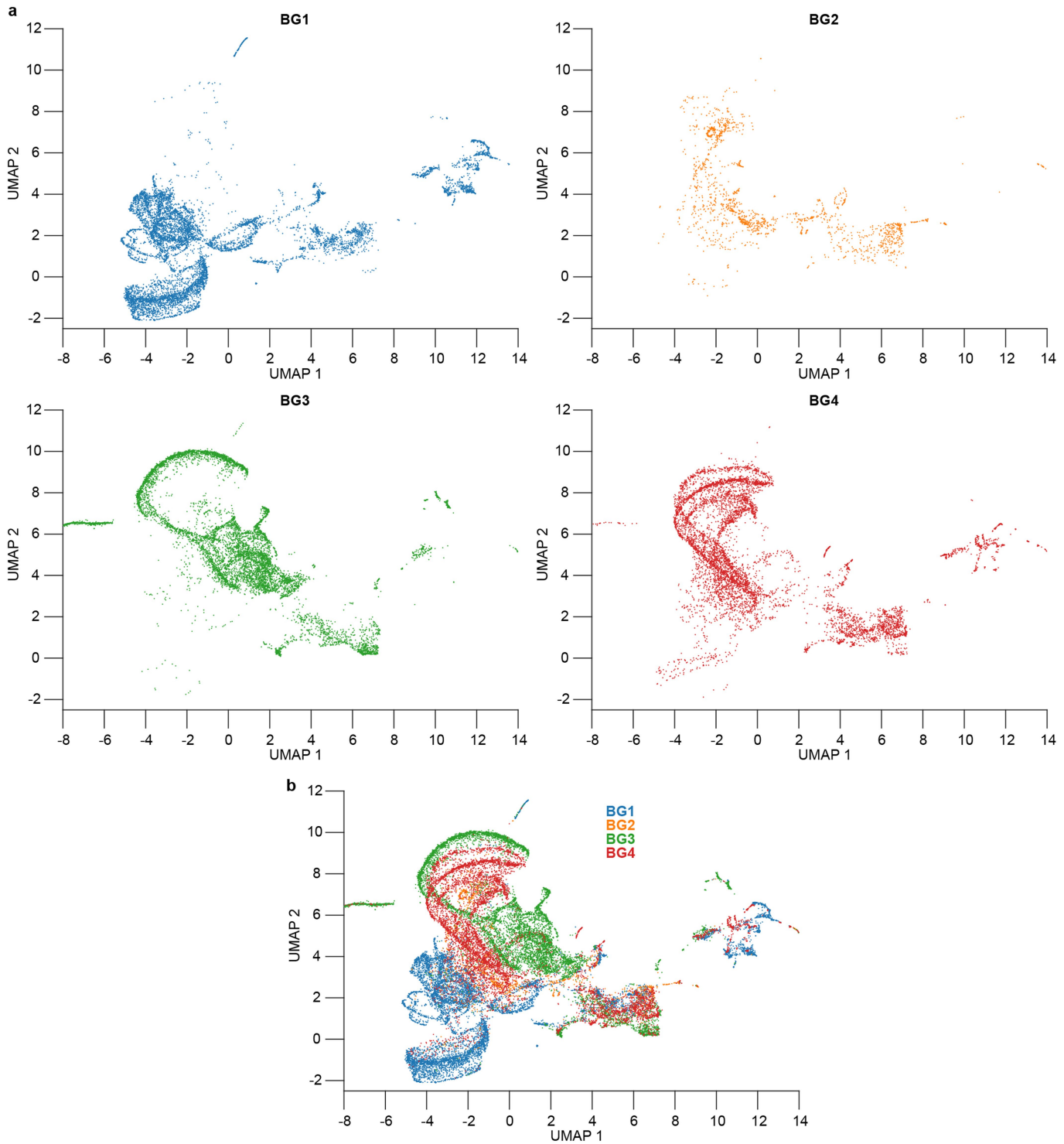
Extended Data Fig. 4 | Neural similarity in budgerigar AAC is not driven by vocal reuse of elemental components. a, Sonogram of an example zebra finch song recorded using the same piezoelectric microphone as used for budgerigar recordings. **b**, Schematic of the methodology for quantifying correlations between unique and repeated vocal segments. Each zebra finch syllable was decomposed into 20-ms segments using a sliding window with a 10-ms step size. Correlations for repeated segments (vocal reuse) were calculated between corresponding segments across 10 renditions (2 renditions shown for illustration) of each syllable. Correlations for unique segments (vocal similarity) were calculated between segments from different syllables. **c**, Distributions of correlation values between vocal similarity (orange) and reuse (blue), with 0.6 set as a threshold to distinguish between the two conditions. **d**, Distribution of correlation values across all vocal segments analyzed in Fig. 2 for each budgerigar, with percentage of correlations above 0.6 indicated on the right, representing potential vocal reuse. **e**, Correlation between spectral similarity and neural similarity matrices for an example budgerigar (left) and accompanying population data (right), excluding all cases with spectral correlation values exceeding 0.6. ρ denotes Spearman's correlation value. P-values displayed within each plot are from two-sided permutation tests (n = 216,848, 1,113,918 pairs of vocal segments for each plot, respectively).



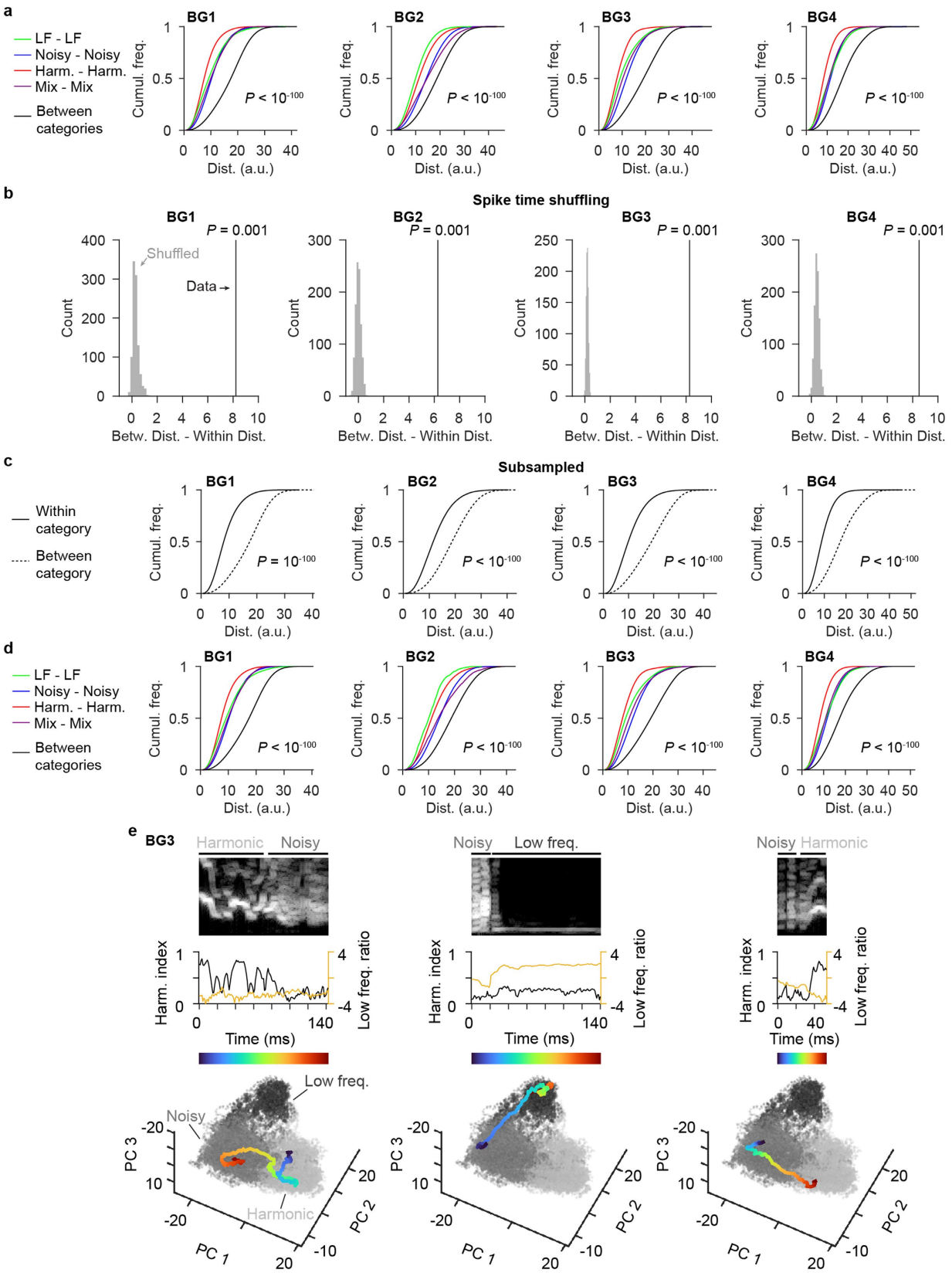
Extended Data Fig. 5 | Single-neuron representation of vocal acoustic categories. a, Distribution of fundamental frequencies in budgerigar vocalizations shows a clear minimum at 700 Hz. **b-d**, Distribution of the selectivity index for low frequency (**b**), harmonic (**c**), and noisy (**d**) vocalizations across the AAC population. Colored bins indicate neurons with a selectivity larger than 0.5 for each index. **e**, Venn diagram displaying the number of neurons with selectivity indices larger than 0.5 for each of the three categories.



Extended Data Fig. 6 | Constructing the neural state space. a, Schematic illustrating the construction of the neural state space. Population neural response within 25-ms time windows were mapped to a three-dimensional space using PCA (see Methods). **b**, Scree plots depicting the eigenvalues and explained variances for the first 10 neural PCs. Each line represents a single budgerigar. **c**, Harmonic index (indicated by color of each dot) plotted on state space diagram for all 4 budgerigars. Low frequency notes (that is, gray dots) not included in this analysis.

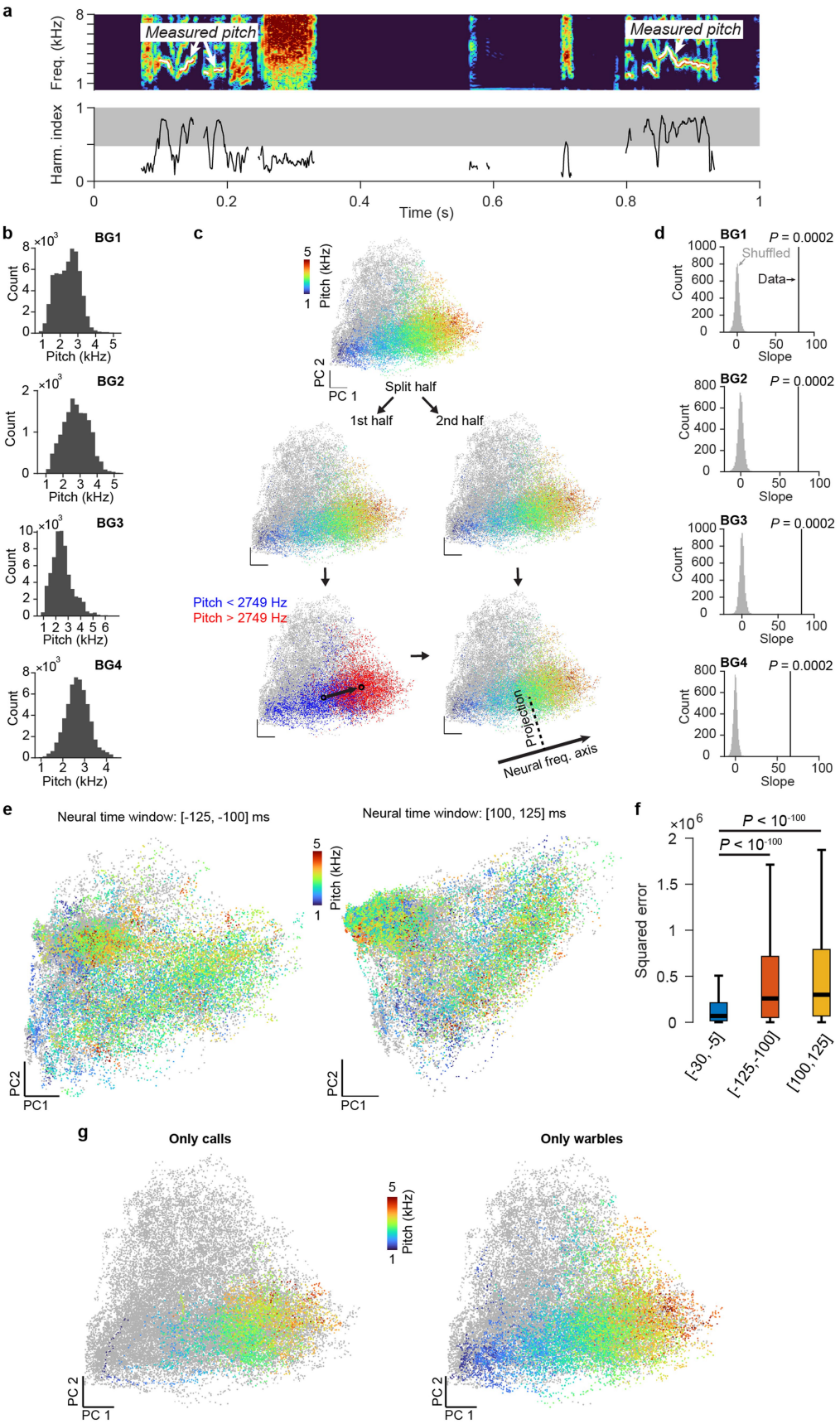


Extended Data Fig. 7 | Comparison of vocal repertoires across budgerigars. a, UMAP projections of latent representations of budgerigar vocalizations. Each plot represents an individual bird. **b**, Combined UMAP plot displaying data from all budgerigars.



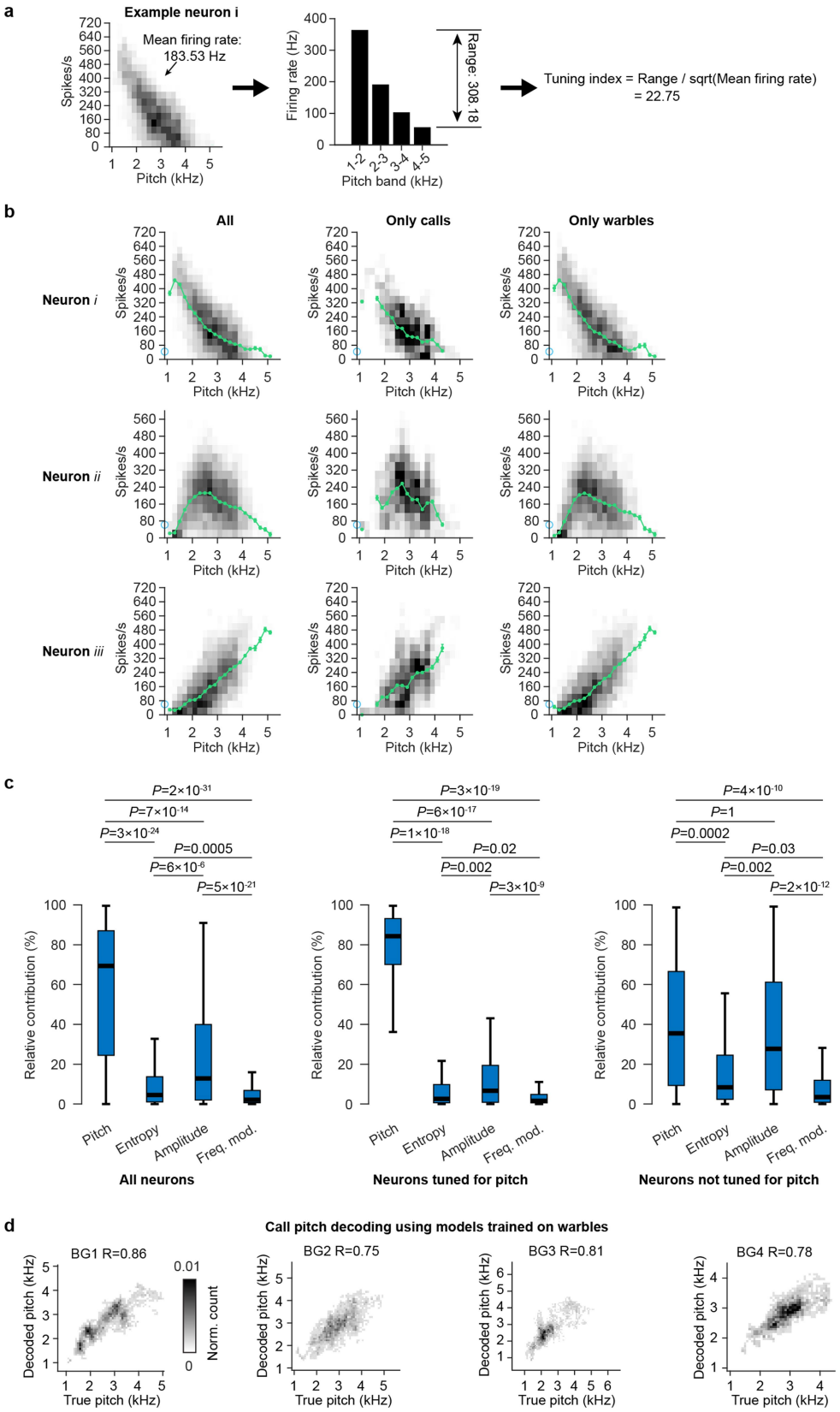
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Populational representation of acoustic categories and neural dynamics during vocal production in AAC. a, Cumulative distribution of Euclidean distances between neural states within (colored line) or across (black line) acoustic categories for each budgerigar. Distance within each category is significantly smaller than between categories ($P < 10^{-100}$ for all tests). P-values shown within each plot are from two-sided Wilcoxon rank-sum tests (n = 66,349,440, 144,270,591, 671,739,531, 44,391,753, 1,854,597,921 instances of distances [1st plot]; n = 613,278, 7,657,741, 57,829,635, 3,801,903, 101,861,288 instances of distances [2nd plot]; n = 13,512,201, 318,087,253, 270,688,278, 94,772,028, 1,578,197,393 instances of distances [3rd plot]; n = 20,196,190, 86,638,866, 710,136,141, 37,840,650, 1,317,021,524 instances of distances [4th plot]). **b**, Mean difference between within-category and between-category neural distances derived from the true data (black vertical lines) plotted against the null distribution generated by randomly shifting spike times (see Methods). Exact p-values displayed in each plot are from one-sided permutation tests (n = 1,000 permutations). **c**, Cumulative distribution of Euclidean distances between neural states within (solid line) or across (dotted line) acoustic categories, based on a subsample of non-overlapping neural states (see Methods). P-values shown within each plot are from two-sided Wilcoxon rank-sum tests (n = 1,435,416 within and 2,902,569 between category instances of distances [1st plot]; n = 119,859 within and 168,561 between category instances of distances [2nd plot]; n = 1,113,510 within and 2,508,576 between category instances of distances [3rd plot]; n = 1,326,700 within and 2,000,210 between category instances of distances [4th plot]). **d**, Similar to **c** but with each within-category distances plotted separately as colored lines. P-values shown within each plot are from two-sided Wilcoxon rank-sum tests (n = 106,491, 216,811, 1,038,961, 73,153, 2,902,569 instances of distances [1st plot]; n = 990, 11,628, 101,025, 6,216, 168,561 instances of distances [2nd plot]; n = 21,528, 503,506, 445,096, 143,380, 2,508,576 instances of distances [3rd plot]; n = 30,381, 125,250, 1,113,778, 57,291, 2,000,210 instances of distances [4th plot]). $P < 10^{-100}$ for all tests. **e**, Neural trajectories for three example vocal elements spanning two acoustic types. Color indicates time within the element. Grayscale dots are the neural states corresponding to the three acoustic types (see labels in panel at left).



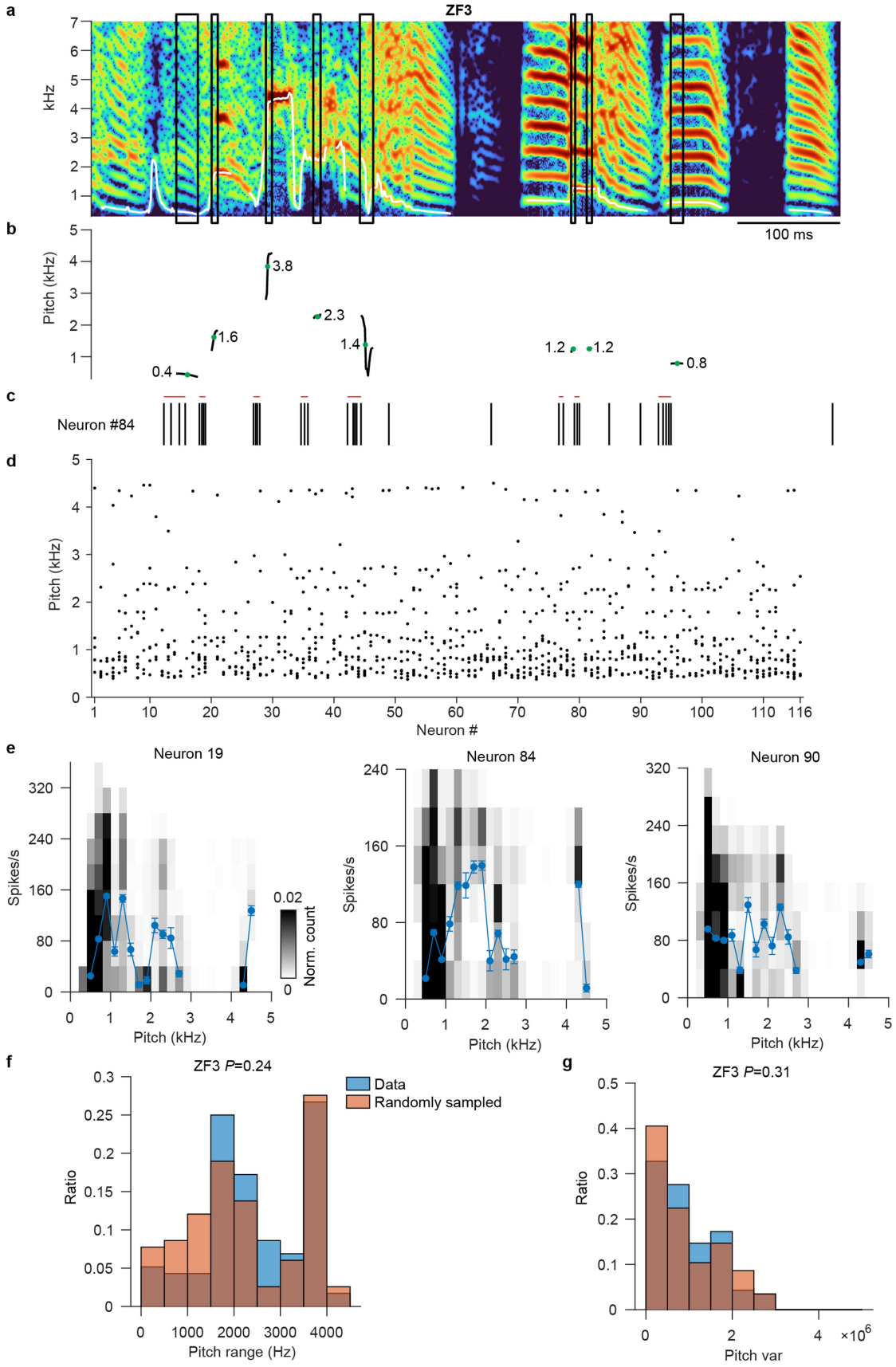
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | A neural frequency axis in AAC state space. a, Pitches (white lines on top) were estimated from vocalizations with harmonic indices that exceeded the median of the distribution (shaded region). **b**, Distribution of estimated pitches for each individual budgerigar. **c**, Schematic of the calculation of the neural frequency axis. The axis was computed as the vector between the mean neural responses to low and high pitches in half of the data. The other half of the data was then used to calculate the relationship between pitch and projection onto the axis (see Methods). **d**, Slope of the relationship between pitch and projections onto the neural frequency axis for each budgerigar. Black vertical lines denote slope from data; gray distributions indicate slopes from pitch-shuffled data. Exact p-values displayed in each plot are derived from one-sided permutation tests (n = 5,000 permutations). **e**, Population neural responses at different time windows relative to vocalizations were mapped to a two-dimensional state space using PCA. Colored dots represent neural states underlying vocalizations with frequency estimated (color indicates pitch), and gray dots are associated with less harmonic vocalizations whose pitch was difficult to estimate. **f**, Performance of a linear model in predicting pitch values using scores of the first two PCs for neural responses calculated within a motor time window (−30 to −5 ms), compared to two other time windows with large temporal shifts (−125 to −100 ms and 100 to 125 ms). P-values displayed within the plot are from two-sided Wilcoxon rank-sum tests (n = 14,566 neural states) with Bonferroni correction. Boxplot elements: center line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum values excluding outliers. **g**, Population-level pitch representation (Fig. 4c) displayed separately for calls and warble syllables. Colored dots denote neural states associated with pitches in calls (left) or warble syllables (right). Gray dots represent neural states for other vocalizations.



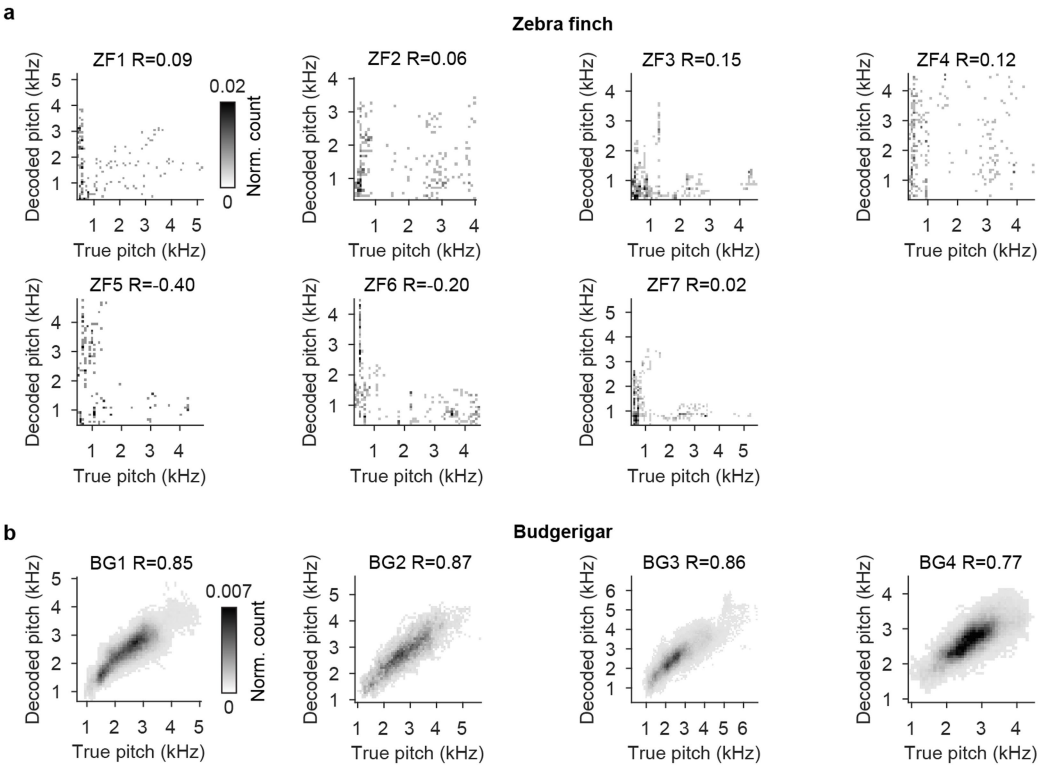
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Pitch encoding in budgerigar AAC. a, Schematic of pitch tuning index calculation. **b**, Pitch tuning in three example neurons from Fig. 4g, presented as binned scatter plots (gray) and averaged firing rates (mean ± SEM) across pitches (green) for all vocal elements (left), only calls (middle), and only warble syllables (right). Blue circles represent baseline firing rates. **c**, Relative contributions of four acoustic features in GLM models explaining single AAC neuron responses shown for all neurons (left), pitch-tuned neurons (pitch tuning index > 6; middle), and non-tuned neurons (pitch tuning index <6; right), respectively. Boxplot elements: center line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum values excluding outliers. Exact p-values shown within each plot are from Bonferroni-corrected, two-sided Wilcoxon signed-rank tests (n = 220, 112, 108 neurons for each plot, respectively). **d**, Relationship between decoded and measured pitches across all calls for each budgerigar, shown as binned scatter plots. Linear decoders trained on warble syllables were used to predict pitch in calls. R values represent Pearson correlation between decoded and measured pitches.

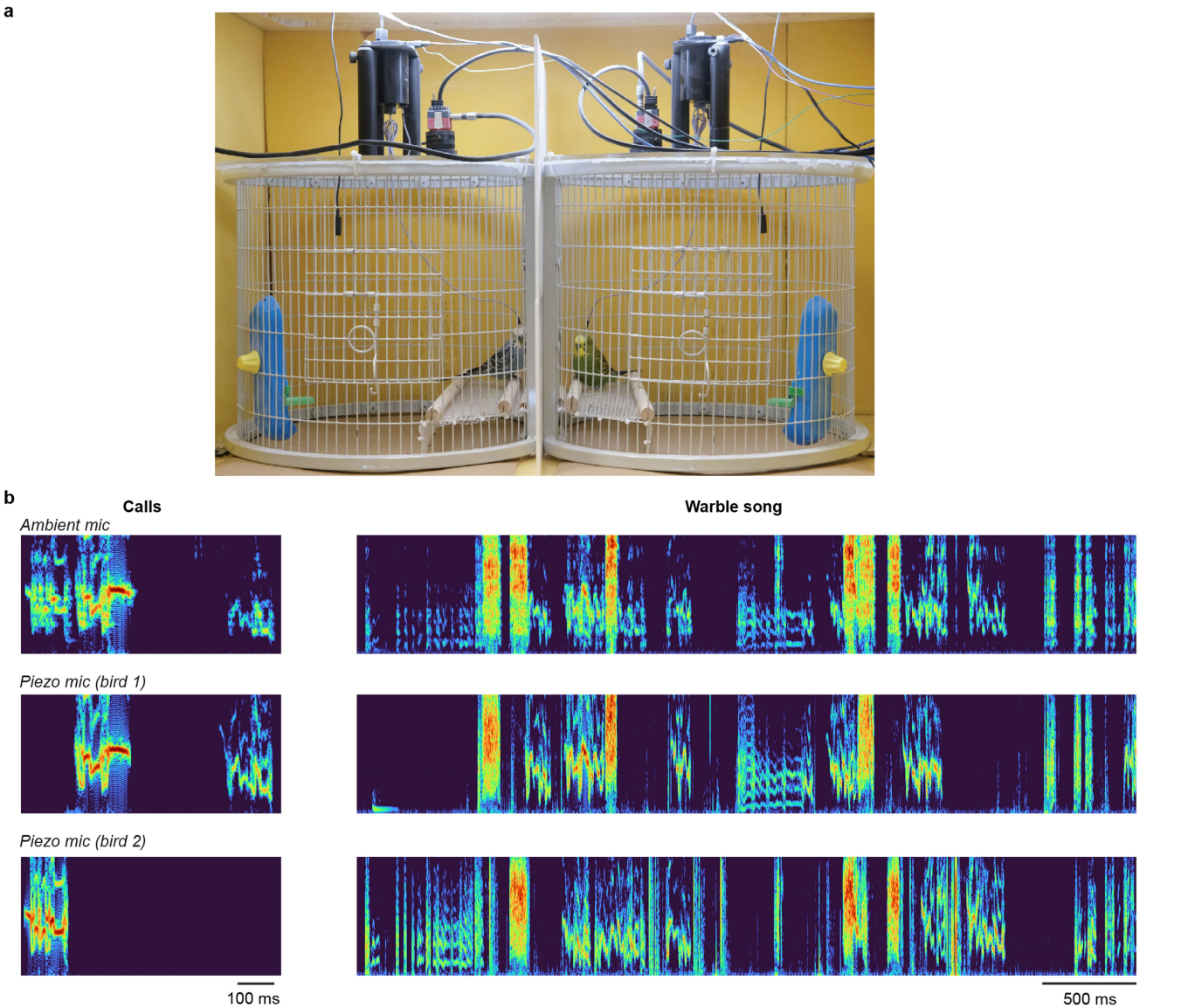


Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Neurons in zebra finch RA do not appear to encode pitch within the song. a-c, Burst-associated pitch values for an example RA neuron. Black boxes in **a** denote vocalization periods associated with bursting events (indicated by red bars in **c**), with vocal periods shifted by 12 ms relative to burst events. Pitch contour within each period is highlighted in **b**, with the mean pitch value marked by a green dot and a number next to the contour. **d**, Pitch values for all bursting events across the RA population. Each dot represents a burst event and its associated pitch. **e**, Relationship between firing rates and pitch for three example RA neurons, presented as binned scatter plots (gray) and averaged firing rates (mean ± SEM) across pitches (blue). **f**, Distributions of the range of burst-associated pitch values for observed data (blue) and for randomly sampled pitch values from **d**. **g**, Distributions of the variance of burst-associated pitch values for observed data (blue) and for randomly sampled pitch values from **d**. In **f** and **g**, Exact p-values shown on top are derived from two-sided Wilcoxon rank-sum tests ($n = 116$ samples).



Extended Data Fig. 12 | Pitch decoding using population activity. **a**, Relationship between decoded and measured pitches using all RA neurons for each zebra finch shown as binned scatter plots. **b**, Relationship between decoded and measured pitches using all AAC neurons for each budgerigar. In **a** and **b**, R values represent Pearson correlation between decoded and measured pitches.



Extended Data Fig. 13 | Behavioral recording methodology. a, Photograph of recording arena for budgerigars with one budgerigar tethered on each side. **b**, Example recordings from an omnidirectional ambient microphone (top) and two piezoelectric microphones implanted on two budgerigars (bottom). Piezoelectric microphones enable excellent isolation of individual vocalizations which are mixed in the ambient microphone.

Extended Data Table 1 | Information for statistical tests

Figure	Test name	Sample size	Independent or paired measurements	*P* value
Fig. 1i	Two-sided Wilcoxon signed-rank test	43 neurons	Paired	3.0×10^{-8}
Fig. 1j	Two-sided Wilcoxon signed-rank test	220 neurons	Paired	2.7×10^{-35}
Fig. 1k	Two-sided Wilcoxon signed-rank test	43 neurons	Paired	1.1×10^{-8}
Fig. 1l	Two-sided Wilcoxon signed-rank test	220 neurons	Paired	7.7×10^{-38}
Fig. 1n	Two-sided Wilcoxon signed-rank test	67 neurons	Paired	2.6×10^{-12}
Fig. 1o	Two-sided Wilcoxon signed-rank test	502 neurons	Paired	3.1×10^{-65}
Fig. 1p	Two-sided Wilcoxon signed-rank test	67 neurons	Paired	1.1×10^{-12}
Fig. 1q	Two-sided Wilcoxon signed-rank test	502 neurons	Paired	6.1×10^{-84}
Fig. 2g	Two-sided permutation test	2,343 pairs of vocal segments	Paired	0.12
Fig. 2h	Two-sided permutation test	6,323 pairs of vocal segments	Paired	0.08
Fig. 2i	Two-sided permutation test	221,147 pairs of vocal segments	Paired	$< 10^{-100}$
Fig. 2j	Two-sided permutation test	1,144,092 pairs of vocal segments	Paired	$< 10^{-100}$
Fig. 3g panel 1	Two-sided Wilcoxon rank-sum test	926,751,315 within & 1,854,597,921 between category instances of neural distance	Independent	$< 10^{-100}$
Fig. 3g panel 2	Two-sided Wilcoxon rank-sum test	69,902,557 within & 101,861,288 between category instances of neural distance	Independent	$< 10^{-100}$
Fig. 3g panel 3	Two-sided Wilcoxon rank-sum test	697,059,760 within & 1,578,197,393 between category instances of neural distance	Independent	$< 10^{-100}$
Fig. 3g panel 4	Two-sided Wilcoxon rank-sum test	854,811,847 within & 1,317,021,524 between category instances of neural distance	Independent	$< 10^{-100}$
Fig. 4l 1st column	Two-sided paired t-test	4 budgerigars	Paired	0.002
Fig. 4l 2nd column	Two-sided paired t-test	4 budgerigars	Paired	0.002
Fig. 4l 3rd column	Two-sided paired t-test	4 budgerigars	Paired	0.004
Fig. 4l 4th column	Two-sided paired t-test	4 budgerigars	Paired	0.008

nature portfolio

Corresponding author(s): Michael A. Long

Last updated by author(s): Dec 16, 2024

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	We used the RHX data acquisition software (ver. 3.1.0) from Intan Technologies to collect neural and audio data. The OlyVIA software (ver. 2.9.1) from Olympus was used to collect histological data.
Data analysis	Spikes were sorted by Kilosort (ver. 1.0 and 2.5; https://github.com/MouseLand/Kilosort) and curated with Phy (ver. 2.0b6; https://github.com/cortex-lab/phy). Sound Analysis Pro (ver. 2011; http://soundanalysispro.com/matlab-sat) was used to extract fundamental frequencies of vocalizations. Rastermap (ver. 0.0.4; https://github.com/MouseLand/rastermap) was used to sort neurons in Fig. 3f. Comparison of budgerigar vocal repertoire was performed using the variational autoencoder provided by Autoencoded Vocal Analysis (ver. 0.3.1; https://github.com/pearsonlab/autoencoded-vocal-analysis). Harmonic ratio was calculated by the 'harmonicRatio' function from the Matlab Audio Toolbox (ver. 2023a; https://www.mathworks.com/help/audio/ref/harmonicratio.html). Other analyses were performed using custom code (https://doi.org/10.5281/zenodo.14057061) written in Matlab 2023a.

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

Data

Policy information about availability of data

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

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

Data used in this study has been uploaded to Zenodo (https://doi.org/10.5281/zenodo.14057061). The Switchboard Telephone Speech Corpus is accessible through the Linguistic Data Consortium (https://catalog.ldc.upenn.edu/LDC97S62), and annotations of the corpus are provided by The Institute for Signal and Information Processing (https://isip.piconepress.com/projects/switchboard/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Numbers of birds (four budgerigars and seven zebra finches) were chosen according to previously published electrophysiological studies in birds (Egger et al., 2020; Elmaleh et al., 2021 and 2023). Numbers of vocalizations (see main text) are commensurate with a previous single neuronal study of human speech production (Khanna et al., 2024).
Data exclusions	All well-isolated AAC neurons were included for analysis. Detailed statements are made in the methods, if analyses focus on a subset of vocalizations (e.g., on vocal elements where pitch could be estimated).
Replication	Only one experiment was performed in a single budgerigar. In total, data were collected from four budgerigars. Results are highly consistent across the four budgerigars.
Randomization	There was only one experimental group in the study, to which all four budgerigars were assigned. In the playback experiment, the order of vocalizations was randomized. The budgerigars were originally obtained from an outside breeder (open colony).
Blinding	There is no group allocation in the experimental design. Cross validation techniques were adopted to avoid overfitting and double-dipping.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Animals and other research organisms

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

Laboratory animals	Four adult (> 240 days posthatch) male budgerigars (Melopsittacus undulatus) and seven adult (> 90 days posthatch) male zebra finches (Taeniopygia guttata) were used in this study.
Wild animals	This study does not involve any wild animals.
Reporting on sex	Data were only collected from male birds in this study.
Field-collected samples	We did not use any field-collected samples.
Ethics oversight	All animal maintenance and experimental procedures conformed to the guidelines established by the Institutional Animal Care and Use Committee at the New York University Langone Medical Center.

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*