



OPEN Context-dependent modulations in zebra finch distance calls revealed by a novel goal-directed vocalization paradigm

Zohreh Safarcharati^{1,2,7}✉, Amirreza Bahramani^{1,3,4,7}, Pouya Mokari Amjad^{1,7}, Mahsa Ravanbakhsh^{1,7}, Mohammad Reza Raoufy^{1,5,6}✉ & Mahdi Khademian¹✉

Songbirds are renowned for their complex vocal communication abilities; among them, zebra finches (*Taeniopygia guttata*) are a key species for studying vocal learning and communication. Zebra finches use various calls with different meanings, including the distance call, which is used for long-distance contact. Whether these calls are static with fixed meanings or flexible remains an open question. In this study we aimed to answer this question by designing a novel behavioral paradigm, in which we trained food-restricted zebra finches to use distance calls for food request. Nine out of ten birds learned this association and used their distance calls to obtain food when they were hungry. We then introduced a visually-separated audience and compared the distance calls used for food requests with those used for communication between birds. Our analyses revealed significant acoustic differences in power, pitch, and other spectral characteristics between the distance calls uttered in these two contexts, with calls directed at conspecifics exhibiting higher amplitude. Our findings suggest that zebra finches can use their distance call for different goals and also acoustically modulate it based on the context. Therefore, it demonstrates a level of vocal control thought to be exclusive to songs. This study enhances our understanding of vocal flexibility and its role in vocal communication.

Keywords Vocal communication, Zebra finch, Distance call, Operant conditioning, Vocalization paradigm, Acoustic flexibility

Vocal communication plays a fundamental role in social interactions, being crucial for survival and other aspects of life in many animal species¹. Songbirds, in particular, are well known for their ability to learn and produce complex songs, a trait that has made them key models for investigating the behavioral and neural mechanisms of vocal learning and communication². Among vast species of songbirds, zebra finch (*Taeniopygia guttata*) is a well-known model system whose neural circuitry and vocal repertoire has been extensively characterised³. Although much research has focused on their singing behavior, the significance of their calls—shorter and less complex vocalizations—has been comparatively overlooked⁴. Calls serve various functions in the social and environmental contexts of songbirds, including mating interactions, alerting conspecifics to danger, coordinating activities such as nest building³.

Among the various types of calls, distance calls are of particular interest. These calls are mainly produced when birds are visually separated and trying to maintain contact, but they are also used in other conditions such as during take-off, greetings, and between bouts of singing³. The distance call in zebra finches is distinct from other calls as they are louder, longer, and optimized for long-range communication⁵. Moreover, they are characterized by sexual dimorphism; females produce harmonic calls with a stable plateau, while males exhibit a higher-pitched tonal part followed by a frequency-modulated downsweep⁶. Another important aspect is that

¹School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran 1956836484, Iran. ²Department of Cognitive Neuroscience, Faculty of Interdisciplinary Sciences and Technologies, Tarbiat Modares University, Tehran, Iran. ³Department of Electrical Engineering, Sharif University of Technology, Tehran 1458889694, Iran. ⁴Sharif Brain Center, Sharif University of Technology, Tehran 1458889694, Iran. ⁵Department of Physiology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. ⁶Institute for Brain and Cognition, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. ⁷These authors contributed equally: Zohreh Safarcharati and Amirreza Bahramani. ✉email: zohreh.safarcharati@ipm.ir; raoufy@modares.ac.ir; khademian@ipm.ir

the male distance call is the only learned call among zebra finch call repertoire⁷. This characteristic leads to structural variations in the distance calls across different zebra finches^{8–10}.

Recent studies have begun to unveil the complexity and plasticity of songbird calls, challenging the earlier notion that these vocalizations are simple, invariant and involuntary responses. For instance some internal states such as stress¹¹ and access to water¹² can influence the acoustic features and dynamics of calls. Furthermore zebra finches can use the acoustic features of calls to recognize their mates^{13,14}, their parents¹⁵ and some other conspecific individuals¹⁶. These findings suggest that calls, like songs^{17,18}, possess a degree of flexibility and may convey more information than previously assumed. This raises an important question: Are calls fixed in meaning, or can they be adapted for different functions depending on the context?

To investigate this, traditional behavioral paradigms that rely on operant conditioning, such as key-pressing assays^{19–21} or operant discrimination tasks^{22–26} are insufficient. While these methods are valuable for many behavioral studies, they do not require the birds to use their vocalizations, which is essential for studying the flexibility and contextual use of calls. Recently, novel paradigms have been developed to approach this problem in which the bird was instructed to initiate vocalization^{27,28}, but to the best of our knowledge none have been designed specifically for zebra finches. Thus, a new paradigm is needed to address these questions.

In this study, we present the design of a novel behavioral paradigm that trains food-restricted zebra finches to use their distance calls to obtain food. This approach exemplifies “vocal usage learning”, as it repurposes an existing vocalization (distance call in this case) in a goal-directed context²⁹. Moreover, since operant conditioning can be indicative of voluntary motor control, this paradigm may contribute to the debate on whether non-human animals have volitional control over their vocalizations^{27,30}.

For our paradigm to be effective, we developed a robust and stable closed-loop vocalization detection and classification system. This system employs a vocal activity detection (VAD) algorithm to reliably identify vocalizations, which are then classified in real-time. Interestingly, we observed that the birds had the ability to learn using their distance calls for food request. After that, we investigated whether these calls differed from those used for long-distance contact. Our findings revealed significant differences in the acoustic features of distance calls associated with food acquisition compared to those used for responding to conspecific distance calls. Specifically, distance calls used for conspecific contact exhibited significantly higher power, which may suggest that birds modulate amplitude to effectively reach distant listeners. This amplitude modulation was previously reported in singing behavior, where increased amplitude was used where females were farther away¹⁸. Moreover, our analysis revealed a strong correlation between pitch and power in distance calls, suggesting that biomechanical constraints may inherently link these acoustic features.

These findings suggest that zebra finches can modulate their calls in a goal-directed manner, providing further evidence of call plasticity and contextual flexibility. This challenges the traditional view that calls are fixed in meaning and suggests that they can be adapted for different goals.

Results

Closed-loop VAD system for goal-directed vocalization paradigm

The goal of this paradigm was to train subjects to use specific vocalizations to obtain food. This required real-time analysis of the birds' vocalizations, identifying the appropriate ones, and using a closed-loop VAD system to trigger the feeder when the correct vocalization was detected (Fig. 1).

To achieve this, we implemented a closed-loop VAD system. Although various methods have been used previously for this purpose^{31–34}, we aimed to develop a system that was more robust to environmental noise. This VAD system utilizes a unidirectional Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) with a many-to-many activation regime (Fig. 1). This configuration generates silence and vocalization activity class posteriors for each extracted feature vector in real-time. The trained system accurately detected 87.91% of zebra finch vocalizations, with an 11.54% false alarm rate, and only missed one vocalization (0.55%) in our test data, using a tuned confidence threshold (see methods for details). Upon detecting the correct vocalization (distance calls in this case), the system triggered the feeder. This closed-loop VAD system enabled us to effectively train the birds to use their distance calls to obtain food.

Zebra finches learned to use their distance calls for food request

We trained 10 zebra finches, placing each bird in the training chamber twice daily (see methods for details). Initially, during the pre-training phase, the birds emitted very few distance calls, as these are typically used for communication with conspecifics. To encourage the birds to use distance calls during training, we employed distance call playbacks, which have been shown to prompt vocal responses^{6,35–37}. This helped the birds understand that emitting a distance call would open the feeder.

After training, 9 out of 10 birds successfully learned to use their distance call to request food (watch Supplementary Video S1). The rate of distance calls significantly increased only under food-restricted condition after training (Fig. 2a). An overall Friedman test comparing distance call rates across four conditions—well-fed and food-restricted, both before and after training—revealed a that they don't have identical effects ($n=9$ birds; $Fr=251$, critical value=7.8). Post-hoc comparison of call rates in the restricted condition before and after training, showed a significant increase, supporting the effectiveness of training (post-hoc comparison: $p=0.0206$). Additionally, after training, birds emitted more distance calls when they were food-restricted compared to when they were well-fed (post-hoc comparison: $p=0.0429$). Full details on the distance call rates are available in Supplementary Table S1.

Before training, the call rate of the birds under food-restricted conditions was close to zero across five repetitions. After 16 days of training, the call rate for food requests increased and remained consistently high in five test repetitions (Fig. 2b). During the test phase, compared to the pre-training phase, the difference in distance call rates between food-restricted and well-fed conditions significantly increased, demonstrating that

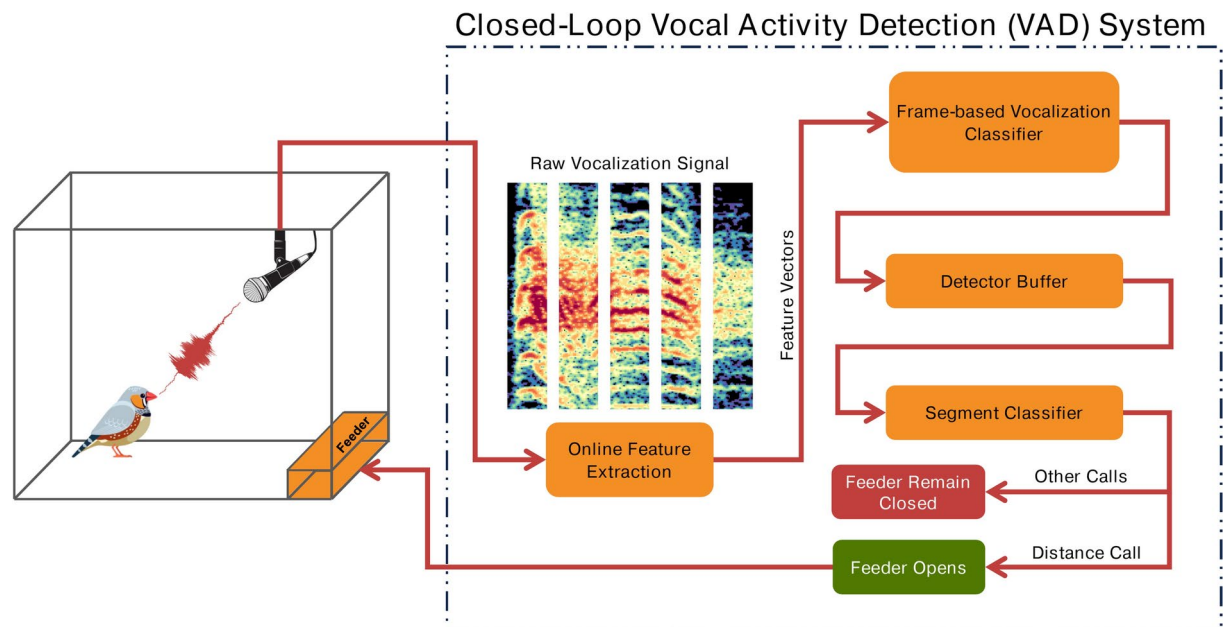


Fig. 1. Schematic of the experimental setup and closed-loop Vocal Activity Detection (VAD) system architecture. Initially, the microphone records the bird vocalizations and transmits the signal to the control computer. The VAD system then segments the vocalization into overlapping windows. Acoustic features are extracted from each part and the whole vocalization is classified as either a “distance call” or “other calls”. If the vocalization was identified as a distance call, the feeder automatically opens and remained open for five seconds, allowing the bird to access food.

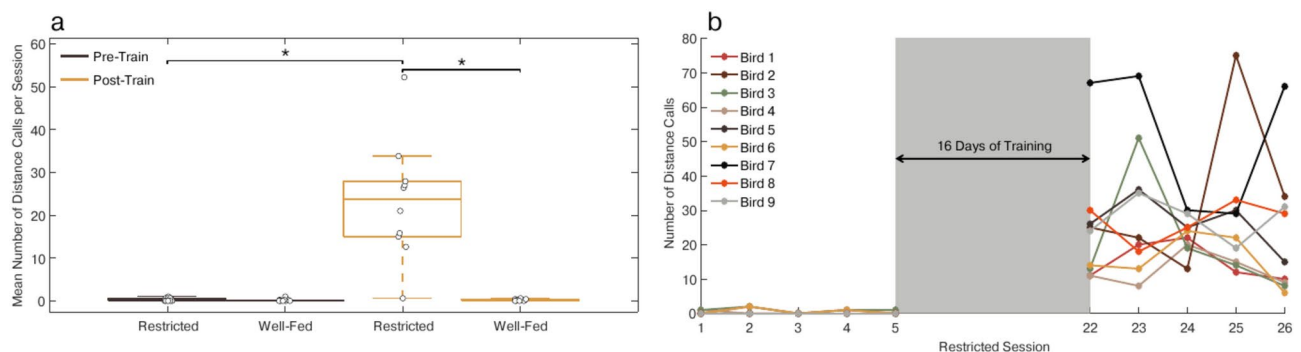


Fig. 2. Birds learned to use their distance calls for food request. (a) Box plots showing the mean number of distance calls produced by the birds averaged over five sessions across four conditions; from left to right: Pre-train Restricted, Pre-train Well-fed, Post-train Restricted, and Post-train Well-fed. (b) The number of distance calls over time during restricted sessions for all 9 learned birds. Pre-training and post-training phases are shown, while the training phase is omitted due to the significantly higher distance call rates caused by playback of distance calls (see “Methods”). Significant differences were assessed using the Friedman test, followed by post-hoc comparisons (* $p < 0.05$).

the zebra finches established a clear association between their distance calls and food requests. This data strongly suggests that this new novel paradigm was successfully learnt by the birds enabling them to use their distance calls to request food.

Modulations in acoustic features of zebra finch distance calls based on the context

During the final testing sessions, we noticed a distinct difference between the distance calls used in response to other zebra finches (audience-directed distance call, or aDC) and those used to request food (goal-directed distance call, or gDC). This observation raised the question of whether the birds could modulate the acoustic features of their calls depending on the context. To explore this, we conducted additional tests with two previously trained restricted birds (Bird 1 and Bird 2), in which a conspecific audience, visually separated from the subject, attempted to make contact with him.

We manually labeled the distance calls; if a distance call from the audience occurred within 3 s before the subject's distance call, it was labeled as aDC; otherwise, it was labeled as gDC. Figure 3a shows the spectrogram of a 9-s example segment from the test session of Bird 1 (hear Supplementary Audio S1). The subject uttered a distance call without any preceding audience call (labeled as gDC) then the VAD correctly detected it and opened the feeder. Subsequently, the subject responded three times to two distance calls from the audience (labeled as aDCs). Figure 3b displays a zoomed-in view of the gDC, while Fig. 3c shows the third aDC. Notably, the aDC has higher power and is more harmonic, with smoother high-frequency components.

Our analysis exhibits significant differences in the acoustic features between aDCs and gDC in both birds. As shown in Fig. 3d, the distribution of vocalization powers differed significantly between the aDC and gDC (two-sample Kolmogorov–Smirnov test, Bird 1: $n_{\text{aDC}}=199$, $n_{\text{gDC}}=131$, $p=9.3313\text{e-}11$; Bird 2: $n_{\text{aDC}}=127$, $n_{\text{gDC}}=72$, $p=9.9341\text{e-}13$). Interestingly higher mean of aDC power suggests that the birds used louder vocalizations to reach distant audience. This is plausible given that the birds outside are farther away, and it has been shown that the auditory cortex of zebra finches can encode the distance of a propagated distance call³⁸. Additionally, the modulation of distance call amplitude aligns with prior findings showing that zebra finches can regulate the amplitude of their vocalizations depending on the social context³⁹. This observation is also consistent with previous research demonstrating that male zebra finches increase the amplitude of their songs as the distance to the targeted female increases¹⁸. To further validate this observation, we compared gDCs to the distance calls produced during test sessions in which no conspecific audience was present and the calls served solely to activate the feeder (gDC alone). As expected, there was no significant difference between their power (two-sample Kolmogorov–Smirnov test, Bird 1: $n_{\text{gDC alone}}=108$, $n_{\text{gDC}}=131$, $p=0.1830$; Bird 2: $n_{\text{gDC alone}}=49$, $n_{\text{gDC}}=72$, $p=0.1444$). These results suggest that the modulation in call amplitude is specifically employed to reach distant conspecifics.

Figure 3e shows that further analysis revealed that the distribution of vocalization pitch differed significantly between the aDC and gDC (two-sample Kolmogorov–Smirnov test, Bird 1: $n_{\text{aDC}}=199$, $n_{\text{gDC}}=131$, $p=8.0166\text{e-}14$; Bird 2: $n_{\text{aDC}}=127$, $n_{\text{gDC}}=72$, $p=3.3964\text{e-}10$). Similar to the power, no significant differences were observed between the pitch distributions of gDCs and gDCs alone (two-sample Kolmogorov–Smirnov test, Bird 1: $n_{\text{gDC alone}}=108$, $n_{\text{gDC}}=131$, $p=0.0877$; Bird 2: $n_{\text{gDC alone}}=49$, $n_{\text{gDC}}=72$, $p=0.1808$). Previous elegant studies have shown that zebra finches can alter the pitch of their song syllables^{31,34,40}, as well as their

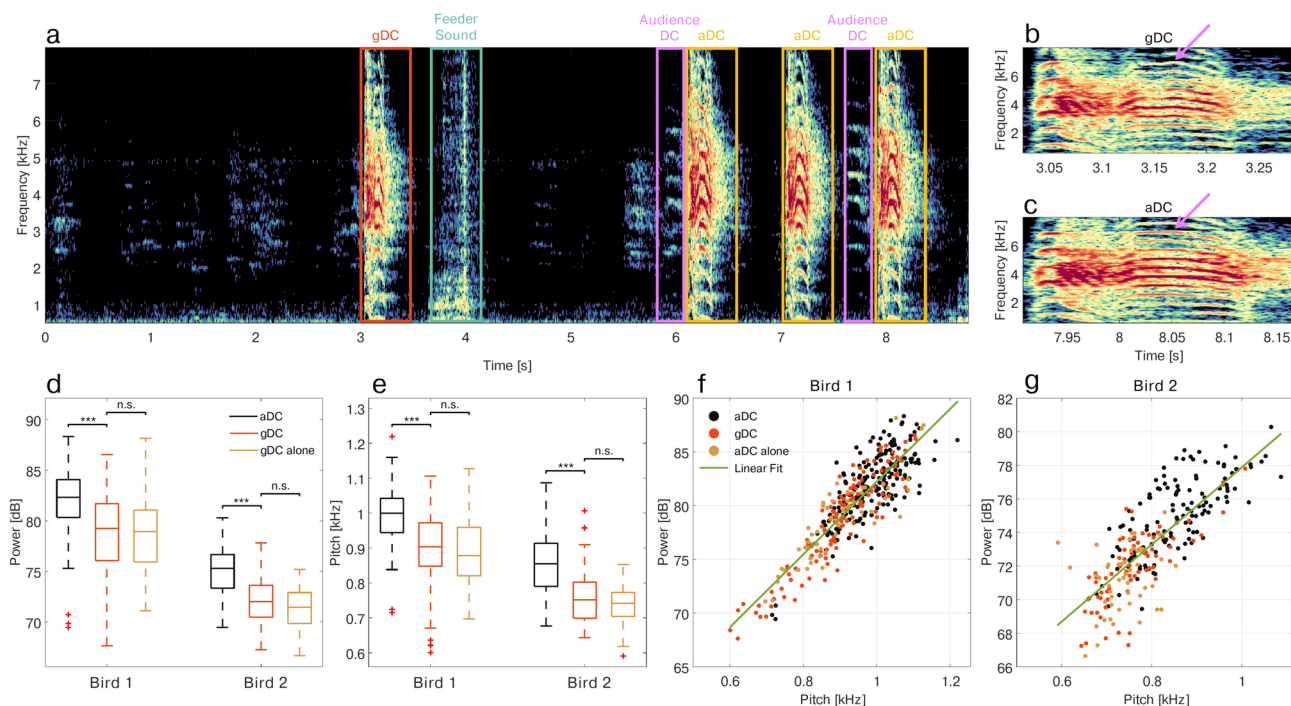


Fig. 3. Acoustic flexibility of distance calls. **(a)** Spectrogram of a 9-s segment from one of the recordings of the acoustic flexibility analysis experiment, with different vocalizations labeled. The audio for this segment is available in Supplementary Audio S1. Note that after the first distance call the feeder opened and remained open for five seconds. This is why there is no feeder sound after next distance calls. **(b)** Zoomed-in spectrogram of the goal-directed distance call (gDC) shown in panel (a). **(c)** Same as panel (b), but for the third audience-directed distance call (aDC). Note the distinct differences in power and harmonic structure (arrows) between the gDC and aDC. **(d–g)** Box plots comparing acoustic features of aDC, gDC, and gDC alone in two birds: **(d)** Power, and **(e)** Pitch. **(f)** Relationship between the power and the pitch of distance calls for Bird 1. **(g)** Same as panel (f), but for Bird 2. Significant differences were determined using the two-sample Kolmogorov–Smirnov test (** $p < 0.01$, *** $p < 0.001$). gDC goal-directed distance call, aDC audience-directed distance call, Audience DC distance call generated by the audience from outside.

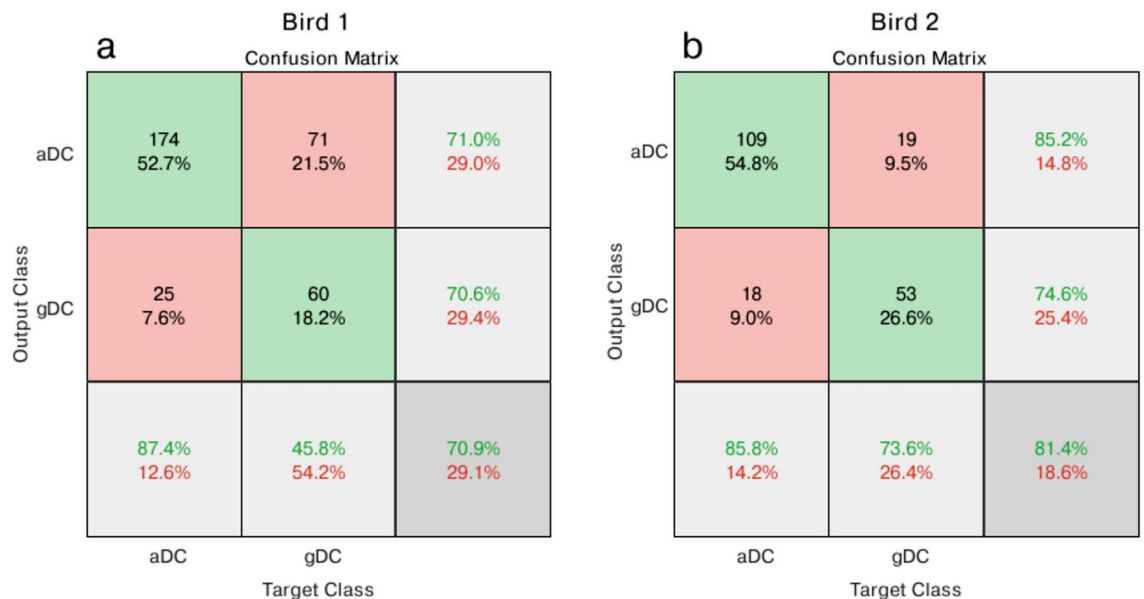


Fig. 4. Classification of two distance call types. **(a)** Confusion matrix illustrating the classification accuracy of the linear discriminant analysis (LDA) model for Bird 1. **(b)** Same as panel (c), but for Bird 2. *gDC* goal-directed distance call, *aDC* audience-directed distance call.

distance calls depending on the internal state¹¹. However, our analysis does not allow us to definitively determine whether the observed acoustic modulations specifically targeted call power, pitch, or both. Indeed, Fig. 3f and g demonstrate a strong correlation between pitch and power in distance calls (Bird 1: Pearson correlation coefficient = 0.8782; Bird 2: Pearson correlation coefficient = 0.7873), with linear fits yielding R^2 values of 0.7713 and 0.6198, respectively. This robust linear relationship, in conjunction with known biomechanical constraints, suggests that the birds may primarily modulate one feature, with consequent effects on the other one.

We also found that the distribution of goodness of pitch differed significantly between the aDC and gDC, though this was only significant for one bird (two-sample Kolmogorov–Smirnov test, Bird 1: n aDC = 199, n gDC = 131, $p = 2.8021 \times 10^{-7}$; Bird 2: n aDC = 127, n gDC = 72, $p = 0.1295$). Wiener entropy, another acoustic feature known to vary in corvids²⁷, did not show significant difference in our study (two-sample Kolmogorov–Smirnov test, Bird 1: n aDC = 199, n gDC = 131, $p = 0.3172$; Bird 2: n aDC = 127, n gDC = 72, $p = 0.3971$).

To further confirm acoustic difference between aDCs and gDCs, we used Linear Discriminant Analysis (LDA) to classify the vocalizations for each bird, using eight acoustic features (Power, Pitch, Goodness of Pitch, Wiener Entropy, Aperiodicity, Amplitude Modulation, Frequency Modulation, and Mean Frequency). The accuracies of the LDA classifiers were calculated using fivefold cross-validation. The classification accuracy for bird 1 was 73.03% and for bird 2 was 82.91%. Confusion matrices are presented in Fig. 4a and b. Together, these results indicate that zebra finches modulate their distance calls with respect to the context, suggesting the acoustic flexibility of distance calls.

Discussion

The transmission of information through vocal communication relies not only on the composition of vocal elements but also on the acoustic features of those elements. However, this form of information coding has not been extensively studied in animal models. To the best of our knowledge, we developed the first goal-directed vocalization paradigm in zebra finches, successfully training them to use distance calls to request food. More importantly, we indicated that zebra finches can modulate their vocalizations depending on context—distinguishing between goal-directed distance calls and those used for communication with conspecifics.

The paradigm we developed offers exciting potential for future research, as it allows zebra finches to be conditioned under more complex situations and enable researchers to further investigate “vocal usage learning”. This is in some ways similar to a prior work conducted with Bengalese finches, where birds were shown to modulate their song under task-based conditions²⁸. For instance, in a preliminary test, we extended the paradigm to zebra finch songs, finding that the birds also learned to use songs for food request (Supplementary Video S2). Interestingly, they terminated songs uncompleted at syllable boundaries just after the feeder opened, which could provide valuable insight into the neural mechanisms underlying song initiation and truncation⁴¹. Additionally, the paradigm could be employed to explore how zebra finches prioritize competing vocal behaviors, a key question in current research⁴².

While we demonstrated the ability to modulate distance calls based on context, an important question remains: Can other birds perceive these subtle changes in vocalizations? Numerous studies have consistently shown that zebra finches are highly sensitive to fine acoustic variations^{43–46}. Zebra finches are capable of individual vocal recognition even if the bird has to distinguish males that all produce an imitation of the same song²². They can

even discriminate between different renditions of the same song, even down to individual syllables⁴⁷. Moreover, they can use these fine variabilities to recognize different individuals in their group^{5,13–16,37,48}. Together all of these findings implicate that modulation of fine acoustic features can be perceived by zebra finches and can be used as a dimension in information communication.

Context-dependent vocal modulation is not a novel discovery. Many previous works demonstrated that the song in the contexts of courtship and territorial or aggressive encounters (“directed song”) is much more spectrotemporally ordered than songs produced in isolation or for daily training (“undirected song”) ^{17,49,50}. Some previous works have also showed context-dependent modulations in calls; a famous study shows that chickadees alarm calls contain information about predator size⁵¹. In zebra finches, vocal amplitude can be regulated based on social context³⁹ and the distance to the conspecifics¹⁸, which we also observed in this study. Moreover, previous work have shown that some internal states like food availability can be reflected in acoustic features of the song⁵². The ability to modulate the spectral features of distance calls has been previously reported¹¹. However, a key distinction in our findings is that while Perez et al. attributed this modulation to hormonal activity, our experiment indicates otherwise. Given that the aDC and gDC were vocalized in close temporal proximity, hormonal influence is unlikely. This shows a capacity for moment-by-moment control over the acoustic features of vocalizations, suggesting that this modulation may be cognitively driven.

A further question arises: Is this ability to modulate acoustic features restricted to learned vocalizations? To our knowledge, the capacity to modulate spectral features has only been observed in learned vocalizations, such as songs and male distance calls. This raises the question of whether modulation is tied to the vocal learning pathway. Utilizing this paradigm with other calls of male zebra finches or even female zebra finches, who do not learn their vocalizations, could help answer whether they can modulate simpler features like amplitude too, or if more complex modulations are exclusive to learned vocal behaviors.

Previous studies were successful in categorizing different calls into semantic categories with fixed meaning⁵³ that are produced in an almost reflexive manner in a specific behavioral context⁵⁴. However, in this study, by developing a novel goal-directed vocalization paradigm, we trained zebra finches to use their distance calls for food request, demonstrating that distance calls can be repurposed for a novel behavioral context, thereby acquiring a new functional meaning⁵³. This raises the intriguing possibility that similar reassignments might be feasible with other call types, potentially elucidating the distinctions between innate and volitional vocal production. For instance, if such operant conditioning is successful only with distance calls—which naturally convey a social need—it may reflect inherent differences in the neural substrates governing innate versus learned vocalizations.

It is also important to note that in our experiment, birds were free to initiate distance calls for food at their own volition, in the absence of external cues—a marked contrast to prior studies of volitional vocalization^{27,30}. While hunger likely played a role, the initiation and modulation of distance calls suggest the possibility of volitional control. The idea that distance calls could be cognitively controlled is supported by studies showing that Bengalese finches can adjust the sequencing of vocal elements in response to learned contextual cues^{28,55}. Furthermore, our results revealed that birds employed higher amplitude calls when addressing distant conspecifics. Notably, the rapid, moment-to-moment modulation of amplitude—as evidenced by the change in call amplitude within seconds in Fig. 3a—strongly suggests that these birds can exert cognitive control over this acoustic feature in response to contextual demands. Future research combining our task with existing paradigms used to explore volitional control over vocalization²⁷ could help address whether the observed modulations are indeed cognitively driven.

The neural circuitry involved in controlling vocal variability presents another exciting line of research. It is reasonable to hypothesize that circuits responsible for plasticity in song production are also involved in modulating distance calls. The anterior forebrain pathway, particularly the lateral magnocellular nucleus of the anterior nidopallium (LMAN), has been shown to regulate variability in song⁴⁰. Variability is transmitted to the vocal motor pathway through the robust nucleus of the arcopallium (RA) and HVC (proper name), both critical for song production and vocal learning^{56,57}. Interestingly, HVC and RA activity has been linked to food aversion, even in the absence of singing⁵⁸. It is also possible that different aspects of modulation, such as amplitude control versus spectral features, are governed by distinct circuits. This makes sense since the pathways for learned and unlearned vocalizations are different⁵⁹. Because this type of context-dependent modulations need some form of cognitive control, it appears that caudal nidopallium (NC) which have been shown to be the equivalent of mammalian prefrontal cortex⁶⁰ is also involved. Recent work has shown that executive control of songbird vocalizations is done by NCL (the lateral region of NC)⁶¹. In addition, it seems that subcortical regions have also an important role in call initiation, especially the motor thalamic nucleus Uvaeformis (Uva) which has been recently shown to have an important role in vocal initiation⁴¹. Altogether, the initiation and acoustic modulation of vocalizations may happen in a vast network including HVC, RA, LMAN, NC and Uva.

In summary, our study reveals compelling evidence that bird vocalizations are not just static and reflexive as previously thought. Instead, they can be dynamically modulated based on the context, showcasing an interesting flexibility. This presumed cognitive control over vocalization opens up new lines of research for studying animal communication. We hope that our goal-directed vocalization paradigm will offer valuable tools for further investigations. Also as proposed before⁶², we believe that distance calls—given their learned properties and use in social contexts—offer unique opportunities to explore vocal communication and learning.

Methods

Subjects

Ten adult male zebra finches (*Taeniopygia guttata*), averaging 240 days old at the start of the experiment, were sourced from the lab's breeding colony. These males were housed in a mixed-sex aviary before and after the experiment. During the experiment, they were kept individually in cages and separated from the others. The

birds were naïve to the experimental setup and training, and their social or genetic parentage, as well as genetic relatedness, was unknown. Throughout the experiment, the birds were maintained on a controlled feeding schedule and earned food during daily test sessions, with supplemental food provided as needed post-test. Water was provided ad libitum in both the aviary and test environments. The subjects were kept on a 12-h light–dark cycle (lights on from 07:00 to 19:00) with illumination provided by LED lamp.

Apparatus

Training sessions were conducted in a custom-built wire cage (60 cm × 40 cm × 45 cm) housed inside a sound-isolation chamber lined with sound-attenuating foam mats. The training chamber was illuminated by an overhead LED lamp. The cage featured a single perch, and a camera (Logitech HD Webcam C270) with a microphone (Boya BY-M1) was positioned on top to capture audio-visual recordings of the sessions. An online vocalization detection system was employed to monitor and classify the subjects' vocalizations in real time. Upon detecting a distance call, the system activated an automatic feeder that delivered food pellets as a reward. Additionally, a speaker (Logitech S-220) positioned above the cage was used to play auditory stimuli during training.

Vocal activity detection system

Here we proposed a real-time vocalization detection and classification architecture for conditioning zebra finches to use their distance calls to obtain food. The proposed architecture consists of a one-stage detector/classifier recurrent neural network trained using the⁶³ dataset. The “AdultVocalizations” part of the dataset consists of 2969 call/song segments from 27 adult zebra finches; more specifically, we use six different zebra finch calls: Distance, Distress, Tet, Thuk, Tuck, Whine calls from this dataset.

The trained model is employed in the behavioral task of this study and can be further applied in other closed-loop experiments, potentially using other public or custom annotated datasets. The pipeline composition, data preparation, and model training are described below, with a quantitative evaluation of system performance provided in the results section.

The frame-based vocalization classifier is a left-to-right Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) that emits estimated vocalization class posteriors based on past and current acoustic features as follows:

$$p(x_t|y_{1:t})$$

where (x_t) is the vocalization class posterior, i.e., ($p(x_t = [\text{Sil}, \text{Distance}, \text{Distress}, \text{Tet}, \text{Thuk}, \text{Tuck}, \text{Whine}] | y_{1:t}) = [0.65, 0.06, 0.04, 0.13, 0.08, 0.04]$); the current frame, (y_t), is categorized as non-vocalization. The network is utilized and trained in a many-to-many operation mode, producing one posterior, (x_t), for each new input (acoustic feature), (y_t).

A total of 35 noise-augmented vocalization sequences were synthesized, each with a duration of approximately 720 s; 31 sequences are used for training and 4 of them for validation. For each sequence, selected target vocalizations from⁶³ were packed sequentially with randomized duration silences between two consecutive song/call vocalizations from the selected subset of 630 Distance, 51 Distress 613 Tet, 290 Thuk, 240 Tuck and 197 Whine calls. Each sequence contains vocalization/silence parts. Both vocalization and silence parts may contain background noise, randomly chosen with randomized signal-to-noise ratios. Short-term discrete-time Fourier analysis was performed on all synthesized sequences to extract spectrograms with a window length of 23 ms and a frame overlap of 11.5 ms. 60 linear filters were applied to each frame after applying a Hamming window, resulting in approximately 2 million feature vectors extracted as the training and validation set (about 88% for training and 12% for validation). Frame labels were selected from the silence and target vocalization set, relative to the frame's correspondence to the silence or vocalization part.

A five-layer deep LSTM was trained in a many-to-many operation mode to assign each frame to a class label, implicitly performing two tasks simultaneously: vocalization detection and classification. The input dimension of the model is the same as the feature vectors, i.e., 60, and the model output contains 7 softmax neurons with cross-entropy loss to classify silence and other vocalization types (Distance, Distress, Tet, Thuk, Tuck, and Whine calls). Approximately 88% of the data (31 out of 35 sequences) were used for model training, with Adam used as the training algorithm. The trained model correctly classified 96.33% of the validation data.

For our task we partitioned calls either as “distance call” or “other calls”. Different from the⁶³ dataset we recorded a test data for validating the VAD system. The system accurately detected 87.91% of zebra finch vocalizations, with an 11.54% false alarm rate, and only missed one vocalization (0.55%) in our test data, using a tuned confidence threshold. This closed-loop VAD system was developed using MATLAB R2022b.

Training procedure

Prior to the initiation of the experimental trials, the zebra finch subjects were acclimatized to the testing apparatus over a period of three days. This acclimatization phase was critical to mitigate any stress-induced variables that could potentially confound the behavioral outcomes. During this phase, the subjects were allowed to explore and become familiar with the confines of the device environment, thus reducing novelty-induced anxiety, which is known to affect performance in operant tasks. The birds did not receive any rewards during this phase. The experimental protocol consisted of three distinct phases: pre-training, training, and post-training.

1. *Pre-training* Before initiating the conditioning protocols, a pivotal assessment was conducted to ascertain the correlation between the zebra finches' vocalization patterns and their hunger levels. This phase was critical for establishing a control measure against which the efficacy of the operant conditioning could be evaluated. Additionally, this assessment ensured that the vocalizations were not a true reflection of the birds' hunger

state in innate behavior. Vocalization rates were measured within the apparatus under two distinct conditions: restricted and well-fed. This was performed five times for each condition. The order of restricted and well-fed experiments was randomized to prevent the response from becoming habitual due to repeating the same conditions. During the initial phase of food restriction, the zebra finches underwent a carefully regulated period of food deprivation lasting 5 h. This systematic method of limiting food intake was precisely formulated to minimize any stress that could arise from abrupt food shortages. By gradually reducing food access, our objective was to ensure that the birds retained a sense of reliability in the availability of food, thus averting any negative impacts on their welfare or conduct.

2. **Training** Subsequently, the subjects were placed in the apparatus twice daily, for 30-min sessions in the morning and afternoon. After each distance call, the feeder was opened. From the results obtained in the pre-training phase, the birds in the cage were usually silent; to enhance the learning process, playbacks of the distance call were employed as an auditory stimulus to provoke the subjects' vocal responses, which in turn opened the feeder. The playback sounds were pre-recorded distance calls from the same species. During this period, the birds had no access to food outside the apparatus, ensuring they had enough motivation to learn the task. If the bird received an insufficient amount of food in the apparatus, additional food was provided a few hours after each training session to prevent weight loss. Playback was administered randomly and without a fixed schedule to prevent conditioning effects, averaging 40 occurrences per session. Playback intervals were more frequent at the beginning of each session. This method aimed to determine if the bird would produce a distance call independently of the playback, indicating evidence of learning. Over time, the birds associated the distance call with obtaining food. After 16 days of training, the birds increased their calls without the playback trigger, at this time, the testing phase begins.
3. **Post-training** During this period, the birds did not hear any calls, and we investigated whether there was a correlation between the birds' vocalizations and their level of hunger. Similar to the pre-training phase, the test was conducted under two conditions—restricted and well-fed. The order of these conditions was randomized, and each condition was tested in 5 sessions.

To investigate our hypothesis about context-dependent acoustic flexibility in distance calls, we removed the sound attenuation chamber and introduced four visually separated conspecifics (both male and female) positioned two meters from the subject behind a non-acoustic barrier. The audience birds, housed in separate cages like in the previous experiment, attempted to contact the subject using distance calls. This experiment involved two trained, food-restricted birds.

Data analysis

Labeling the recorded signal was done online by the VAD system, and all subsequent analyses were performed offline using MATLAB R2022b. The analysis aimed to determine whether the four experimental conditions—pre-train restricted, pre-train well-fed, post-train restricted, and post-train well-fed—produced significant differences in the number of distance calls in zebra finches. First, the data of the unlearned bird was removed. Given the small sample size (9 subjects) and the repeated measures design we utilized the Friedman test, a non-parametric alternative to the repeated measures ANOVA, to assess differences across the four conditions. If the Friedman test indicated a significant difference between conditions, post-hoc pairwise comparisons were conducted by Wilcoxon signed rank test.

For the acoustic flexibility analysis experiment, we labeled the distance calls manually. If a distance call from the audience occurred within three seconds before the subject's distance call, it was labeled as aDC; otherwise, it was labeled as gDC. It is worth noting that the VAD system recognized both aDC and gDC as distance calls and opened the feeder, so the labeling was not on the base of feeder openings. Then we used SAP2011⁶⁴ to extract acoustic features of distance calls. The extracted acoustic features were vectors with the same dimension as the distance call itself. The mean value of that vector was calculated so that every distance call had a single value for each of its acoustic features. To find significant differences between aDC and gDC, we used the two-sample Kolmogorov–Smirnov test because the assumptions of normality and homogeneity of variance were not met for all features. Furthermore, for calculating the correlation between the pitch and the power of the distance calls we used MATLAB's "corrcoef" function, and for the linear fit we used its "polyfit" function.

Finally, for the classification, a linear discriminant analysis (LDA) classifier was trained on the data using MATLAB's "fitcdiscr" function. We used 8 acoustic features (Power, Pitch, Goodness of Pitch, Wiener Entropy, Aperiodicity, Amplitude Modulation, Frequency Modulation, and Mean Frequency) as predictors. To assess the generalizability of the model, five-fold cross-validation was performed using MATLAB's "crossval" function. The confusion matrices were calculated by running the trained models on the entire datasets.

Data availability

The datasets for the current study are available from the corresponding authors upon reasonable request.

Received: 23 October 2024; Accepted: 4 March 2025

Published online: 12 March 2025

References

1. Elie, J. E. & Theunissen, F. E. The neuroethology of vocal communication in songbirds: production and perception of a call repertoire. In *The Neuroethology of Birdsong* Vol. 71 (eds Sakata, J. T. et al.) 175–209 (Springer, 2020).
2. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
3. Zann, R. A. *The Zebra Finch: A Synthesis of Field and Laboratory Studies* (Oxford University Press, 1996).

4. Marler, P. Bird calls: Their potential for behavioral neurobiology. *Ann. N. Y. Acad. Sci.* **1016**, 31–44 (2004).
5. Mouterde, S. C., Theunissen, F. E., Elie, J. E., Vignal, C. & Mathevon, N. Acoustic communication and sound degradation: How do the individual signatures of male and female zebra finch calls transmit over distance?. *PLoS ONE* **9**, e102842 (2014).
6. Vignal, C. & Mathevon, N. Effect of acoustic cue modifications on evoked vocal response to calls in zebra finches (*Taeniopygia guttata*). *J. Comp. Psychol.* **125**, 150–161 (2011).
7. Zann, R. Ontogeny of the zebra finch distance call: I. Effects of cross-fostering to bengalese finches. *Z. Für Tierpsychol.* **68**, 1–23 (1985).
8. Charles, B. F., Kovacevic, R., Tansinsin, S. L., Van Hoy, B. & Syud, F. A. The distance call of domesticated zebra finches (*Poephila guttata*). *Int. J. Comp. Psychol.* **8**, 867 (1995).
9. Runciman, D., Zann, R. A. & Murray, N. D. Geographic and temporal variation of the male zebra finch distance call. *Ethology* **111**, 367–379 (2005).
10. Zann, R. Structural variation in the zebra finch distance call. *Z. Für Tierpsychol.* **66**, 328–345 (1984).
11. Perez, E. C. et al. The acoustic expression of stress in a songbird: Does corticosterone drive isolation-induced modifications of zebra finch calls?. *Horm. Behav.* **61**, 573–581 (2012).
12. Ma, S., Maat, A. T. & Gahr, M. Power-law scaling of calling dynamics in zebra finches. *Sci. Rep.* **7**, 8397 (2017).
13. Vignal, C., Mathevon, N. & Mottin, S. Audience drives male songbird response to partner's voice. *Nature* **430**, 448–451 (2004).
14. Vignal, C., Mathevon, N. & Mottin, S. Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behav. Process.* **77**, 191–198 (2008).
15. Mulard, H., Vignal, C., Pelletier, L., Blanc, A. & Mathevon, N. From preferential response to parental calls to sex-specific response to conspecific calls in juvenile zebra finches. *Anim. Behav.* **80**, 189–195 (2010).
16. Elie, J. E. & Theunissen, F. E. Zebra finches identify individuals using vocal signatures unique to each call type. *Nat. Commun.* **9**, 4026 (2018).
17. Sossinka, R. & Böhner, J. Song types in the zebra finch *Poephila guttata castanotis*¹. *Z. Für Tierpsychol.* **53**, 123–132 (1980).
18. Brumm, H. & Slater, P. J. B. Animals can vary signal amplitude with receiver distance: Evidence from zebra finch song. *Anim. Behav.* **72**, 699–705 (2006).
19. Adret, P. Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Anim. Behav.* **46**, 149–159 (1993).
20. Rodríguez-Saltos, C. A. et al. Song preferences predict the quality of vocal learning in zebra finches. *Sci. Rep.* **13**, 605 (2023).
21. Terpstra, N. J., Bolhuis, J. J. & Den Boer-Visser, A. M. An analysis of the neural representation of birdsong memory. *J. Neurosci.* **24**, 4971–4977 (2004).
22. Geberzahn, N. & Derégnaucourt, S. Individual vocal recognition in zebra finches relies on the song syllable structure rather than on song syllable order. *J. Exp. Biol.* <https://doi.org/10.1242/jeb.220087> (2020).
23. Gess, A., Schneider, D. M., Vyas, A. & Woolley, S. M. N. Automated auditory recognition training and testing. *Anim. Behav.* **82**, 285–293 (2011).
24. Macedo-Lima, M. & Ramage-Healey, L. Auditory learning in an operant task with social reinforcement is dependent on neuroestrogen synthesis in the male songbird auditory cortex. *Horm. Behav.* **121**, 104713 (2020).
25. Park, T., Okanoya, K. & Dooling, R. Operant conditioning of small birds for acoustic discrimination. *J. Ethol.* **3**, 5–9 (1985).
26. Seki, Y., Osmanski, M. S. & Dooling, R. J. Failure of operant control of vocal learning in budgerigars. *Anim. Behav. Cogn.* **5**, 154–168 (2018).
27. Brecht, K. F., Hage, S. R., Gavrilov, N. & Nieder, A. Volitional control of vocalizations in corvid songbirds. *PLOS Biol.* **17**, e3000375 (2019).
28. Veit, L., Tian, L. Y., Monroy Hernandez, C. J. & Brainard, M. S. Songbirds can learn flexible contextual control over syllable sequencing. *ELife* **10**, e61610 (2021).
29. Vernes, S. C. et al. The multi-dimensional nature of vocal learning. *Philos. Trans. R. Soc. B* **376**, 0236 (2021).
30. Nieder, A. & Mooney, R. The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philos. Trans. R. Soc. B* **375**, 0054 (2020).
31. Ali, F. et al. The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. *Neuron* **80**, 494–506 (2013).
32. Gadagkar, V. et al. Dopamine neurons encode performance error in singing birds. *Science* **354**, 1278–1282 (2016).
33. Tumer, E. C. & Brainard, M. S. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* **450**, 1240–1244 (2007).
34. Xiao, L. et al. A basal ganglia circuit sufficient to guide birdsong learning. *Neuron* **98**, 208–221.e5 (2018).
35. Perez, E. C. et al. Physiological resonance between mates through calls as possible evidence of empathic processes in songbirds. *Horm. Behav.* **75**, 130–141 (2015).
36. So, N. L. T., Edwards, J. A. & Woolley, S. M. N. Auditory selectivity for spectral contrast in cortical neurons and behavior. *J. Neurosci.* **40**, 1015–1027 (2020).
37. Vicario, D. S., Naqvi, N. H. & Raksin, J. N. Sex differences in discrimination of vocal communication signals in a songbird. *Anim. Behav.* **61**, 805–817 (2001).
38. Mouterde, S. C., Elie, J. E., Mathevon, N. & Theunissen, F. E. Single neurons in the avian auditory cortex encode individual identity and propagation distance in naturally degraded communication calls. *J. Neurosci.* **37**, 3491–3510 (2017).
39. Cynx, J. & Gell, C. Social mediation of vocal amplitude in a songbird, *Taeniopygia guttata*. *Anim. Behav.* **67**, 451–455 (2004).
40. Andalman, A. S. & Fee, M. S. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl. Acad. Sci.* **106**, 12518–12523 (2009).
41. Moll, F. W. et al. Thalamus drives vocal onsets in the zebra finch courtship song. *Nature* **616**, 132–136 (2023).
42. Roeser, A. et al. Dopaminergic error signals retune to social feedback during courtship. *Nature* <https://doi.org/10.1038/s41586-023-06580-w> (2023).
43. Dooling, R. J., Best, C. T. & Brown, S. D. Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *J. Acoust. Soc. Am.* **97**, 1839–1846 (1995).
44. Ning, Z.-Y., Honing, H. & Ten Cate, C. Zebra finches (*Taeniopygia guttata*) demonstrate cognitive flexibility in using phonology and sequence of syllables in auditory discrimination. *Anim. Cogn.* **26**, 1161–1175 (2023).
45. Ohms, V. R., Escudero, P., Lammers, K. & Ten Cate, C. Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception. *Anim. Cogn.* **15**, 155–161 (2012).
46. Weisman, R. et al. Frequency-range discriminations: Special and general abilities in zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*). *J. Comp. Psychol.* **112**, 244–258 (1998).
47. Fishbein, A. R., Prior, N. H., Brown, J. A., Ball, G. F. & Dooling, R. J. Discrimination of natural acoustic variation in vocal signals. *Sci. Rep.* **11**, 916 (2021).
48. Mouterde, S. C., Elie, J. E., Theunissen, F. E. & Mathevon, N. Learning to cope with degraded sounds: Female zebra finches can improve their expertise at discriminating between male voices at long distance. *J. Exp. Biol.* <https://doi.org/10.1242/jeb.104463> (2014).
49. Sakata, J. T., Hampton, C. M. & Brainard, M. S. Social modulation of sequence and syllable variability in adult birdsong. *J. Neurophysiol.* **99**, 1700–1711 (2008).
50. Searcy, W. A. & Beecher, M. D. Song as an aggressive signal in songbirds. *Anim. Behav.* **78**, 1281–1292 (2009).
51. Templeton, C. N., Greene, E. & Davis, K. Allometry of alarm calls: Black-capped chickadees encode information about predator size. *Science* **308**, 1934–1937 (2005).

52. Ritschard, M. & Brumm, H. Zebra finch song reflects current food availability. *Evol. Ecol.* **26**, 801–812 (2012).
53. Elie, J. E. & Theunissen, F. E. Meaning in the avian auditory cortex: Neural representation of communication calls. *Eur. J. Neurosci.* **41**, 546–567 (2015).
54. Seyfarth, R. M. & Cheney, D. L. Production, usage, and comprehension in animal vocalizations. *Brain Lang.* **115**, 92–100 (2010).
55. Tian, L. Y. & Brainard, M. S. Discrete circuits support generalized versus context-specific vocal learning in the songbird. *Neuron* **96**, 1168–1177.e5 (2017).
56. Hahnloser, R. H. R., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
57. Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Sci. New Ser.* **273**, 1871–1875 (1996).
58. Tokarev, K., Tiunova, A., Scharff, C. & Anokhin, K. Food for song: Expression of C-Fos and ZENK in the zebra finch song nuclei during food aversion learning. *PLoS ONE* **6**, e21157 (2011).
59. Simpson, H. & Vicario, D. Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.* **10**, 1541–1556 (1990).
60. Güntürkün, O. The avian ‘prefrontal cortex’ and cognition. *Curr. Opin. Neurobiol.* **15**, 686–693 (2005).
61. Brecht, K. F., Westendorff, S. & Nieder, A. Neural correlates of cognitively controlled vocalizations in a corvid songbird. *Cell Rep.* **42**, 112113 (2023).
62. Vicario, D. S. Using learned calls to study sensory-motor integration in songbirds. *Ann. N. Y. Acad. Sci.* **1016**, 246–262 (2004).
63. Elie, J. E. & Theunissen, F. E. The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.* **19**, 285–315 (2016).
64. 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).

Acknowledgements

We extend our deepest gratitude to Dr. Ali Ghazizadeh for his invaluable support in establishing the Birds Lab and for his insightful contributions throughout the experiments and manuscript preparation. We also wish to thank the IPM School of Cognitive Sciences for their essential support in the establishment of the Birds Lab.

Author contributions

Z.S. conceptualized the study; All authors designed the methodology; M.K. created the VAD software; Z.S., P.M. and M.R. collected the data; A.B. and Z.S. analyzed the data; A.B., Z.S. and M.K. wrote the manuscript; All authors reviewed the manuscript; M.K. and M.R.R. supervised the study.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

All procedures were approved by the Research Ethics Committee of the School of Cognitive Sciences (SCS) of the Institute for Research in Fundamental Sciences (IPM, protocol number 1402/40/1/2841). All methods in this study were performed in accordance with the ARRIVE guidelines.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-93105-2>.

Correspondence and requests for materials should be addressed to Z.S., M.R.R. or M.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025