The role of spectral features and song duration in zebra finch, Taeniopygia guttata, song recognition

Ning, Z.-Y.; Honing, H.; van Mil, H.; ten Cate, C.

Link to publication

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

The role of spectral features and song duration in zebra finch, *Taeniopygia guttata*, song recognition

Zhi-Yuan Ning [a] , Henkjan Honing [b] , Harald van Mil [a, c] , Carel ten Cate [a, d, *]

[a] *Institute of Biology Leiden, Leiden University, Leiden, the Netherlands*
[b] *Music Cognition Group, Institute for Logic Language and Computation, University of Amsterdam, Amsterdam, the Netherlands*
[c] *Mathematical Institute, Leiden University, Leiden, the Netherlands*
[d] *Leiden Institute for Brain and Cognition, Leiden University, Leiden, the Netherlands*

Zebra finch song perception is assumed to primarily involve a high sensitivity to fine spectral features of song elements while other features like element sequence and song duration do not seem to have a notable effect. However, the specific features that zebra finches focus on when identifying or discriminating sounds may not be as fixed as seems to be assumed and might depend on the characteristics of the stimuli. This apparent flexibility in auditory processing, along with the potential salience of differences in song duration for song perception, highlights the need for systematic research on the acoustic parameters that zebra finches can use to differentiate between songs. By employing a Go-Left/Go-Right operant task, we examined whether and how differences in song duration affect zebra finches' relative sensitivity for spectral features and duration in song recognition. Two groups of zebra finches were trained in a Go-Left/Go-Right operant task to discriminate between either two songs with similar durations ('Equal-duration group') or two songs with different durations ('Unequal-duration group'). We assessed to what extent the birds in the two experimental groups attend to the spectral characteristics and the absolute duration of the songs by measuring the responses to test stimuli consisting of spectral modifications or temporal changes. Our results showed that zebra finches use both spectral features and song duration to discriminate between two songs, but the importance of these acoustic parameters depended on whether the songs differed in duration or not. When duration can be used as an additional feature to distinguish two songs, spectral features have a less prominent role. This outcome shows that zebra finches have cognitive flexibility in their attention to different acoustic parameters.

Birdsongs convey important information that varies from individual identity to information about sex, age, individual quality or motivation. Meaningful communication requires that receivers be able to perceive and process the acoustic variation in songs. On the one hand, regardless of external conditions that may affect the transmission of song features, a receiver has to recognize a song as coming from the same singer. At the same time, the receiver must be able to discern meaningful variations within songs produced by the same singer, as well as being able to distinguish between songs from different individuals. This raises the question of the cognitive mechanisms through which songbirds recognize and classify songs and discriminate between different songs and song variants. Experimental studies have

addressed this topic in various ways, ranging from field studies using playback to psychophysical laboratory experiments using operant discrimination paradigms. Field studies examined, for example, the characteristics birds employ to recognize conspecific songs or to discriminate between conspecific and heterospecific ones (e.g. Dabelsteen & Pedersen, 1992; Naugler & Ratcliffe, 1992; Nelson, 1989). Psychophysical studies have been used to investigate the hearing ranges and the abilities of birds to detect specific details in the spectral or temporal structure of songs (e.g. Dooling & Prior, 2017; Kreutzer et al., 1990; MacDougall-Shackleton & Hülse, 1996; Neilans et al., 2010; Tu & Dooling, 2012). Such studies have provided important insights into the mechanisms underlying auditory perception and communication in birds. At the same time, studies on avian sound perception are relevant from a comparative perspective as they can reveal the presence of both similarities and differences in the acoustic features that are salient or noticeable by humans and those to which birds attend

---

* Corresponding author.
 E-mail address: c.j.ten.cate@biology.leidenuniv.nl (C. ten Cate).

(e.g. Dooling & Prior, 2017; Hoeschele, 2017; Hulse et al., 1984; Bregman et al., 2012).

Over the years, the zebra finch has emerged as a model species for examining the processing of complexly structured auditory stimuli at the level of behaviour as well as its underlying neurobiology. One area of research concerns the features that zebra finches can or do use to recognize or discriminate between songs. These features are often examined by using operant discrimination tasks. For instance, using a design in which zebra finches were first trained to respond to a single song type and not to respond to deviations, Dooling and collaborators (Dooling et al., 2002; Fishbein et al., 2021; Lawson et al., 2018; Lohr et al., 2006; Prior, Smith, Ball et al., 2018; Prior, Smith, Lawson et al., 2018; Vernaleo & Dooling, 2011; Vernaleo et al., 2010) examined the salience of various types of song changes on the identification of the target song. They showed that zebra finches are very sensitive to changes in the spectrotemporal structure of syllables but relatively insensitive to changes in syllable order in zebra finch song motifs. From these studies, they concluded that zebra finches primarily attend to spectral details such as the temporal fine structure (phase in the waveform over extremely short periods) within individual syllables. Also, several other studies (e.g. Geberzahn & Derégnaucourt, 2020; Mol et al., 2021; Uno et al., 1997; Vignal & Mathevon, 2011) indicated the prominent importance of spectral features for vocal discrimination in zebra finches, with low-frequency harmonics more important for song identification than high-frequency ones (Dent et al., 2016). A prominence of spectral features over syllable sequence for discriminating songs was also shown by Braaten et al. (2006) using a Go/Nogo paradigm. In another study, Nagel et al. (2010) trained adult female zebra finches to perform a classification task in a two-alternative forced choice paradigm to investigate the role of three acoustic parameters (pitch, tempo and amplitude) in discriminating between two male songs. Small changes in pitch ($\pm 2\%$) already affected song discrimination, while tempo alterations affected song discrimination only when these were substantial (>32%).

The above studies suggest that the main factors involved in sound perception in zebra finches are known and predictable: a high sensitivity for fine spectral features of acoustic stimuli with substantially less, if any, impact on other parameters, such as tempo (speed) or song duration. However, several other findings suggest that the features to which zebra finches attend when identifying or discriminating songs or other auditory stimuli are not as fixed as the experiments mentioned above suggest and may depend on the characteristics of the stimuli. For instance, in contrast to the study by Nagel et al. (2010), which suggested that zebra finches hardly attend to tempo changes of auditory stimuli equal to or less than 32%, other experiments demonstrated that zebra finches do respond to small tempo changes when two series of identical sound pulses could only be differentiated by attending to temporal features (van der Aa et al., 2015; ten Cate et al., 2016). Here a 25% change in tempo substantially reduced stimulus discrimination. The contrast of this finding with the limited impact of any tempo changes on song identification, as obtained by Nagel et al. (2010), may arise because Nagel et al. (2010) used songs with a similar song duration. If the duration of songs is similar, then duration might be an irrelevant parameter for song identification and therefore ignored. Spectral features are then the main distinguishing parameter, and zebra finches might focus their attention on such features to identify songs. This might explain a limited effect of tempo changes on song identification compared to the discrimination of auditory stimuli consisting of identical elements, differing in tempo only. Similarly, syllable sequence is not a prominent parameter when zebra finches are trained to discriminate two syllable strings consisting of different song syllables. However, when the two strings consist of the same syllables but in a different

sequence, zebra finches attend to the sequence in addition to the spectral structure of the syllables (Ning et al., 2023). This indicates that zebra finches are flexible in the auditory parameters they attend to and use those acoustic features that allow them to differentiate between the stimuli. This was also suggested by a study in which zebra finches were trained to discriminate between two sets of artificial vowel-like harmonic elements (Burgering et al., 2019). For one group of birds, the distinguishing feature of the spectra was the fundamental frequency (pitch), while for the other group this was the relative energy distribution over the harmonic spectrum across the elements, indicated as the 'spectral envelope'. Probe tests showed that the first group maintained the discrimination when the energy distribution over the spectra was changed, but the fundamental frequencies remained the same. The second group of birds ignored changes in the fundamental frequency of the spectra but maintained the discrimination when the harmonic sounds were replaced by a noise-vocoded sound. Such a manipulation divides the original sound into distinct frequency bands and replaces the spectral variation within each band by a noise signal with the same amplitude. The results of this experiment thus show that zebra finches can either ignore or use the fundamental frequency or the harmonic structure of the sound depending on which is relevant for acoustic discrimination. It also shows that zebra finches can attend to the shape of the spectral envelope, something that had not been tested before in zebra finches, but which had been demonstrated by Bregman et al. (2016) for starlings, *Sturnus vulgaris*, discriminating among more complex tone sequences. Bregman et al. (2016) suggested that the spectral envelope governs avian tone sequence recognition. The importance of this feature may have long gone unnoticed as many experiments on avian pitch perception used pure tones, for which the spectral band envelope corresponds directly to pitch. The findings of Burgering et al. (2019) and Bregman et al. (2016) may indicate that the sensitivity of zebra finches to pitch changes in songs need not necessarily indicate a sensitivity to pitch, but alternatively might result from being sensitive to the spectral envelope, something that has so far not been tested for song stimuli.

The apparent flexibility in the features used during auditory processing shown by zebra finches and the potential role of duration and spectral envelopes in song perception call for further research on the acoustic parameters that zebra finches can or do use to distinguish between songs. The present study aimed at exploring these parameters. Two groups of zebra finches were trained to discriminate two songs in a Go-left/Go-right task. For one group these songs were equal in duration (Equal-duration group), for the other group they were unequal in duration (Unequal-duration group). After being trained, zebra finches were tested with modified versions of training songs that were changed in one of the following ways: (1) increasing or decreasing the tempo, thus affecting the duration; (2) raising or lowering the pitch; (3) moving the entire song up in the frequency spectrum; or (4) replacing the harmonic spectrum by a noise-vocoded version. This design allowed us to examine two factors. The first is whether song duration is used as an additional factor to spectral features when zebra finches are trained to discriminate two songs that differ in duration. If the duration is used as an additional factor, we expect that learning will be easier and hence the training phase will be shorter when learning to discriminate between songs of different compared to similar duration. We also expect that the relative impact of modifying spectral features versus temporal ones on the ability to recognize and discriminate between songs will differ depending on whether training songs differ in duration. For birds of the Equal-duration group, song duration is not a distinguishing factor between the training songs, while it is for birds of the Unequal-duration group. Therefore, we expect that zebra finches

trained to discriminate between two songs of equal duration will be less sensitive to tempo changes of the songs than zebra finches trained with songs of different duration. In contrast, we expect that birds from the Equal-duration group will be more sensitive to changes in the spectral domain, because they can only use spectral features to discriminate between the training songs. Thus, the relative impact of tempo changes and spectral changes is expected to differ between the two experimental groups. The second factor we examined is the relevance of the spectral envelope versus pitch in song discrimination. If zebra finches, like starlings, attend to the spectral envelope rather than pitch for song recognition, we expect that vocoded songs may be easier to recognize than songs with pitch changes or songs moved up in frequency. For both the Equal-duration and the Unequal-duration groups we thus expect that if the birds attend more to the spectral envelope than to pitch, the vocoded version of the song will be considered more similar to the training songs than songs in which the pitch or frequency profile has been shifted.

METHODS

Subjects

We tested a total of 28 zebra finches (14 males, and 14 females; ages 215—720 days post hatching) originating from the in-house breeding colony at Leiden University. Before the experiment, the birds lived in single-sex groups of about 15—30 individuals in aviaries (2 × 2 m and 1.5 m high), in which food and water were available ad libitum. The birds were divided equally between two experimental groups, each consisting of seven males and seven females. Each group was trained with a different set of stimuli, and within each group half of the birds got one set of test stimuli ('series 1') and half another set of test stimuli ('series 2'), hence resulting in a total of four subgroups, each consisting of seven birds.

Operant Conditioning Cage

The birds were trained and tested individually in an operant conditioning cage (Skinner box; 70 × 30 cm and 45 cm high) containing three pecking keys (sensors) with a red LED light at the top/

Figure 1. Schematic front view of the operant conditioning apparatus (Skinner box) used for the experiment. A speaker (top of figure) is suspended from the ceiling above the cage. Within the cage, there are several perches (P) for the bird to sit on, a food hatch (F) located in the upper middle of the back panel and a lamp (L) is placed at the top of the cage. Two tubes with ad libitum water (W) are placed symmetrically on two sides of the cage, and three sensors (S1, S2, S3) with red LEDs are lined horizontally in the lower middle of the back panel.

bottom of each sensor (Fig. 1). Each operant cage was situated in a separate sound-attenuated chamber. The chamber was illuminated by a fluorescent lamp (Phillips Master TL-D 90 DeLuxe 18W/965, The Netherlands), which emitted a daylight spectrum following a 13.5:10.5 h light:dark schedule. Sound stimuli were played through a speaker (Vifa MG10SD09-08; frequency range 100—15 000 Hz) 1 m above the Skinner box. The volume of the speaker was adjusted to ensure that the sound amplitude in the Skinner box was approximately 65 dB (measured by an SPL meter—RION NL 15, RION), a level comparable to what the bird would be exposed to from a singing conspecific at the location of the bird. Sensors (S1, S2, S3), lamp, food hatch and speaker were connected to operant conditioning controller that also registered all sensor pecks.

Stimuli

Training stimuli

A total of 24 natural song motifs were used. The song motifs were extracted from representative recordings of adult males from our breeding colony, but whose vocalizations had not been heard before by birds in this study. The training stimuli in this experiment were 14 stimulus pairs (seven pairs for each experimental group), each consisting of two different songs. Every stimulus pair was used twice, for two separate subgroups of birds (N = 7 birds/group). The two subgroups of birds per training stimulus pair were subjected to different series of test sounds: one subgroup to test series 1 and the other to test series 2 (see below). Of the 14 stimulus pairs, seven pairs consisted of songs of approximately equal duration, in which the shortest song always differed less than 5% from the duration of the longest song in a pair (mean duration of the shortest song was 98.21 ± 1.45% of the duration of the longest song). The group trained with these stimuli was the 'Equal-duration group'. For the other seven pairs the songs were of unequal duration, with the duration of song A being approximately 1.5 times longer than its paired song B (mean: 148.43 ± 6.50%). The group trained with these songs was the 'Unequal-duration group' (Fig. 2). Hence the experimental structure was that both the 'Equal' and the 'Unequal' group consisted of two subgroups of seven birds, each trained with the same stimulus set, but tested with a different set of test stimuli.

Within each song stimulus, the same motif was repeated three times with a silent gap between the motifs, thus simulating a natural song sequence. When played, the motifs were normalized such that the average intensity (RMS; calculated over the total duration of the stimulus) was the same for the two stimuli within a set but the amplitude variation of the original male zebra finch song was preserved. All training stimuli were bandpass filtered between 380 Hz and 22.5 kHz. The two stimuli from each training stimulus set were visually selected to differ in the spectral structure of the syllables (Fig. 2). All training stimuli were cut, synthesized, and filtered using Praat (version 6.0.54, http://www.praat.org). The amplitude of each stimulus was adjusted by using the 'Normalize' feature in Audacity (version 2.3.0, http://audacity.sourceforge.net).

Test stimuli

To test the impact of specific parameters that the birds may have used to discriminate the training stimuli, they were tested with modified versions of the training stimuli, which were grouped into two series of test stimuli (Table 1). The two series differed from each other in how strongly they modified specific parameters of the training stimuli. We expected that a stronger modification would have a stronger impact on song discrimination. Each subgroup of birds was tested with one series of sounds only. We used the Praat Vocal Toolkit (a Praat plugin with automated scripts for voice processing, www.praatvocaltoolkit.com) to edit each original training stimulus to produce a version with either spectral features
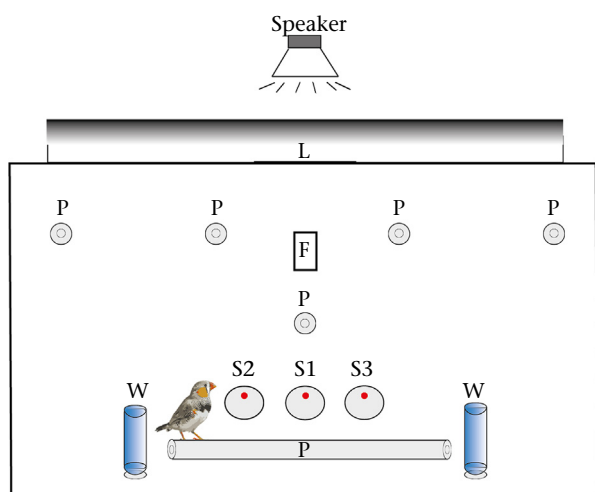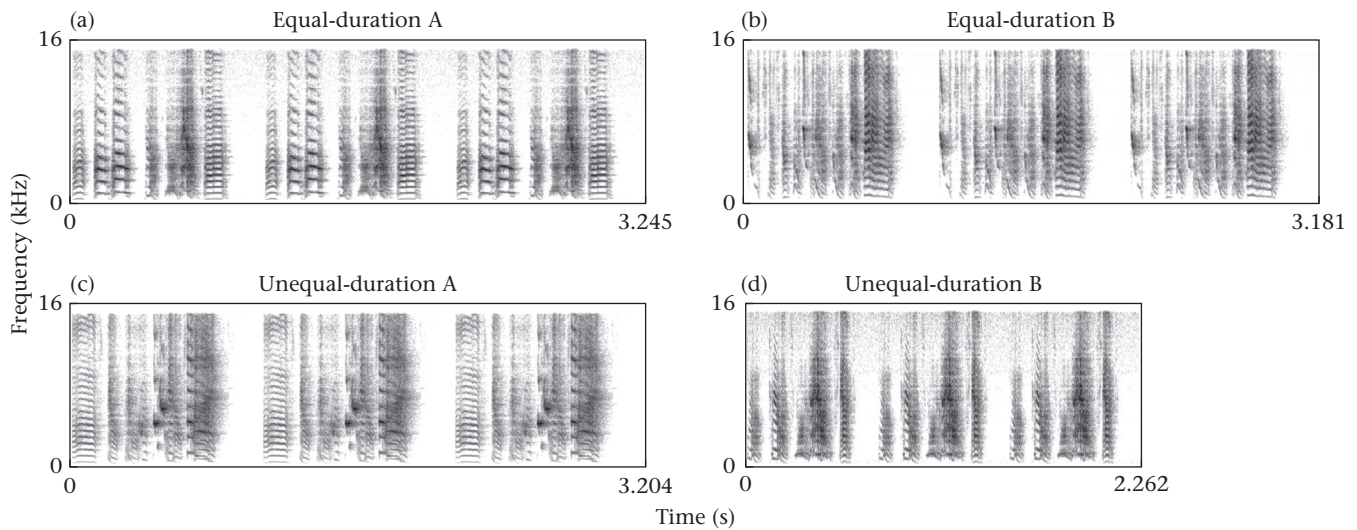
Figure 2. Spectrogram samples of training stimuli. Songs (a) Equal-duration A and (b) Equal-duration B form a pair of training stimuli used in the Equal-duration group, while songs (c) Unequal-duration A and (d) Unequal-duration B form a pair of training stimuli used in the Unequal-duration group.

Table 1
Overview of test stimuli used for the two experimental subgroups

	Frequency-shifted	Pitch-shifted upward	Pitch-shifted downward	Duration stretched	Duration compressed	Noise-vocoded
1st series	+1500 Hz	+20%	−20%	+50%	−50%	Spectral envelope maintained but spectral contour averaged
2nd series	+500 Hz	+8%	−8%	+20%	−20%	Spectral envelope & spectral contour maintained

Two test series were used for the subgroups of both the Equal-duration and Unequal-duration experimental groups of birds. The test stimuli differed in the degree to which pitch and duration were modified, and were more strongly changed in series 1 than in series 2 and in the scripts used for vocoding (series 1: vocoded version according to the script by Chris Darwin; series 2: vocoded version according to the script by Matt Winn. See text for details).

or the tempo was changed. For both the Equal-duration and the Unequal-duration training group, the test stimuli were always modified from the training stimuli in an identical way. We used the following set of test stimuli (Fig. 3, Table 1).

(1) Frequency-shifted. For this stimulus the whole frequency spectrum was shifted upwards linearly. By this manipulation, the harmonic relations between the frequencies are no longer preserved. This was obtained by using a Fresh plugin of Audacity (version 2.3.0, full buckets frequency shifter, www.fullbucket.de/music), adding a fixed value to the frequency of each component of the original sound signal. For the subgroup of birds tested with series 1 this value was 1500 Hz and for the subgroup of birds tested with series 2 this was 500 Hz.

(2) Pitch-shifted. The frequency spectrum was stretched or compressed on a log scale to produce a version in which the harmonic relationship between the frequencies in the song remained the same, but their absolute frequencies were changed. This version of the target sound was synthesized using the 'Change vocal trace' script of the Praat Vocal Toolkit by entering the specific formant shift ratio value in the options displayed in running this script. For the subgroup of birds tested with series 1, the frequency spectrum was stretched or compressed by 20%, and for the subgroup of birds tested with series 2 it was 8%. The choice of the values of 8% and 20% was based on the study by Nagel et al. (2010), in which an 8% change resulted in a reduced discrimination between two songs, although they were still discriminated above chance, while a 50% change resulted in lack of discrimination. The 20% value thus was intermediate between these.

(3) Time-scaled. The duration of the whole song was stretched or compressed proportionally without any change in the frequency domain. The 'change duration' script of Praat Vocal Toolkit was applied to obtain stretched and compressed song versions. For the subgroup of birds tested with series 1, the duration was stretched or compressed by 50%, and for the subgroup of birds tested with series 2 it was 20%. Here also the values of 20% and 50% were chosen based on the study by Nagel et al. (2010) in which a 20% change did not affect the degree of song discrimination, while a 50% change reduced (but not eliminated) the discrimination.

(4) Noise-vocoded. This modification maintains the spectral envelope (the overall shape of the frequency spectrum) of the elements within the motif, but averages the energy within specific frequency bands, thus removing any harmonic structure. To construct these stimuli, we used two different scripts to synthesize a vocoded morph of training stimuli: for the subgroup of birds tested with series 1, we used the modified Chris Darwin vocoded script (for the original version, see http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/Shannon) which also removed the within-syllable spectral contour (the shape of the sound's frequency components over time) of the song syllables (referred to as 'Contour-averaged Vocoded'), and for the subgroup of birds tested with series 2 we used the Matt Winn's Praat vocoded script (http://www.mattwinn.com/praat/vocode_all_selected_40.txt) which maintained the within-syllable spectral contour (referred to as 'Contour-maintained Vocoded'). Both scripts were set to divide cut-off frequency bandwidths equally for 15 bands contiguous with smooth transitions (1000 Hz bandwidth for one noise-vocoded band).
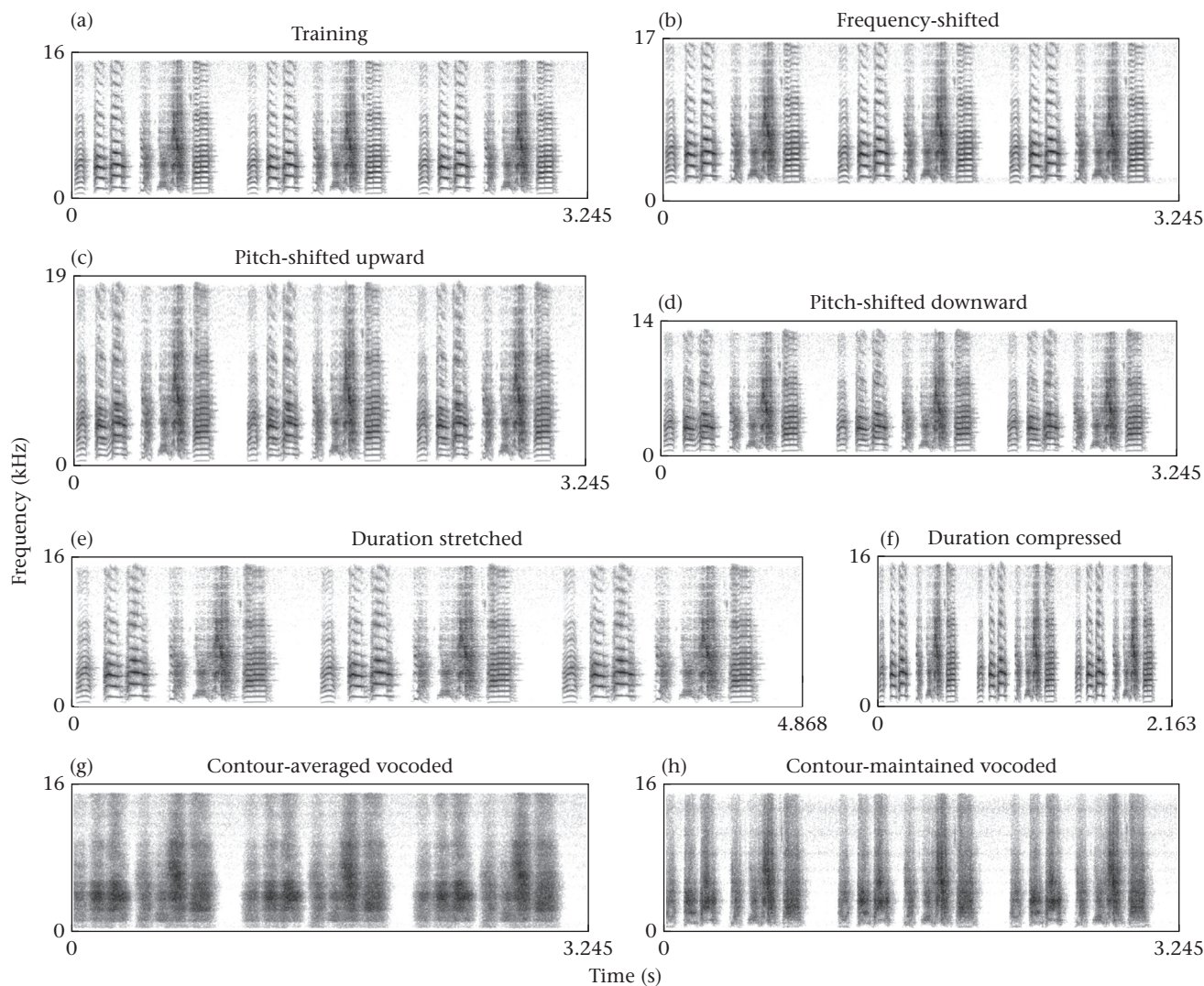
Figure 3. Examples of stimuli used in the test series, showing (a) the training stimulus, and its modified versions. The whole frequency spectrum of (b) the Frequency-shifted version was shifted upwards by 1500 Hz. The frequency spectrum of the Pitch-shifted stimulus was either (c) stretched (+20%) or (d) compressed proportionally (−20%). The duration of the Time-scaled stimulus was either (e) stretched (+50%) or (f) compressed (−50%). The Noise-vocoded versions were produced by using two scripts: (g) the modified Chris Darwin vocoded script (Contour-averaged Vocoded) and (h) the Matt Winn's Praat vocoded script (Contour-maintained Vocoded).

Procedure

We used a Go-left/Go-right paradigm for training and testing. The procedure consisted of five phases: acclimation, pretraining, discrimination training, transition and probe testing. The birds stayed in the Skinner boxes during all phases of the experiment.

Acclimation phase

In the acclimation phase the birds were moved to the Skinner boxes (see Fig. 1). The food hatch remained open, so food was freely accessible in a container behind the hatch. The LED lights on the sensors were on. The goal of this phase was to acclimate the bird to the cage and show where to find food. The bird might also already learn to peck the sensors spontaneously. If in this stage the central sensor, S1, was stimulated by pecking, it would play song A or song B with a 50% chance on each. The side sensor S2 produced one of the two songs, and the other side sensor S3 produced the other song. The red LEDs of all three sensors were illuminated to attract the attention of the bird. After several hours to 1 day, with a median value of 26 (interquartile range, IQR 18—28) h, the next phase started by closing the food hatch.

Pretraining phase

The goal here was to familiarize and teach the bird training procedures. In this phase, the food hatch was closed, and the bird had to learn to peck all three sensors. Pecking the sensors in this phase led to the following effects: S1 (middle sensor) = sound A or B (no food), S2 (side sensor) = sound A + food hatch open (duration 12 s), S3 (side sensor) = sound B + food hatch open (12 s). This continued until the bird had learned to peck at each sensor, and that pecking the sensors resulted in access to the food. The bird might also already learn at this stage which song was related to S2 or S3. This process took several days, with a median value of 95 (IQR 68—122) h. If the bird did not peck the sensor spontaneously, the experimenter could turn on/off the LED to make it flash to stimulate the bird to pay attention to the sensor. Once the bird started pecking all the sensors regularly (i.e. pecking each of the three keys over 50 times in 1 day) for a day, the discrimination training phase began.

Discrimination training

In this phase, the bird had to learn to peck the sensor in the middle to elicit the playback sound, followed by pecking the sensor

on the left or right, depending on the playback sound. If the bird pecked the sensor linked to the particular stimulus being played, a response was rewarded with 12 s access to food. If the wrong sensor was pecked, the light was off for 3 s. Before any sensor was pecked, only the S1 LED was on. For example, when song A was played, pecking sensor S2 caused the food hatch to open while pecking sensor S3 resulted in the preset dark time, and vice versa. If the bird did not respond within 25 s, a test trial would automatically end without food reward or light-off penalty. Once the accuracy rate of pecking each sensor was greater than 0.60 per day, the duration of the light-off period went from 3 to 1 s, the food acquisition time from the initial stage of 12 s to the later stage of 10 s. The duration of this phase varied from bird to bird, with a median value of 456 (IQR 278–655) h. The proportion of correct responses out of all sounds that each bird responded to was calculated daily as the individual's discrimination rate for the sound stimuli. When a bird learned to associate the two training sounds with the corresponding correct sensors and had reached a discrimination score for the training stimuli greater than 0.75 for 3 consecutive days (general discrimination score >0.75, the accuracy rate of each sensor pecking >0.60 for 3 consecutive days), it was assumed that the bird was able to discriminate the trained song motifs and the training was switched to a transition phase.

Transition phase

During the transition training phase, training stimuli were identical to that in the discrimination training phase, but reinforcement by food reward or light-off period was reduced to occur randomly in 80% (instead of 100%) of trials. In the remaining 20% of trials (with stimuli identical to the training sounds), the subjects were not reinforced with either food or a light-off period. If the bird kept the same level of discrimination as in the training phase for 2 days, the test phase began. The duration of the transition phase had a median value of 47 (IQR 46–50) h.

Probe testing phase

In this phase, 16 test stimuli were introduced for 20% of pecks on S1. Twelve of these were novel stimuli (belonging to either series 1 or series 2). The remaining four test stimuli were nonrewarded training sounds used as control and were presented twice as often as the other test stimuli. These test stimuli (nonrewarded training sounds and novel stimuli) were never reinforced and were randomly interspersed between training stimuli. The remaining 80% were training stimuli with reinforcement. Each test sound was presented until it was given 40 trials. This process took 2–3 weeks, with a median value of 394 (IQR 339–549) h. After reaching this, the bird was transferred back to its aviary. The order of stimulus presentation was randomized across subjects.

Analysis

To examine whether the two training groups differed in the speed of discrimination learning, we used the total number of trials up to and including the day on which the learning criterion had been reached. A Mann–Whitney test (R Core Team, 2016) was used to detect differences between the two training groups on learning speed (required training trials) since the number of trials did not follow the normal distribution.

The reactions to the different test stimuli can be separated into three categories: a 'correct response' (i.e. the bird identifies the modified version of training stimulus A as A and the modified version of training stimulus B as B), an 'incorrect response' (responding with pecking the sensor for B if the stimulus was a modification of sound A and vice versa), and a 'nonresponse' (not pecking a key). For the statistical analyses, we examined the

proportion of correct responses as: Proportion Correct ('PC') = Count_Correct/(Count_Correct + Count_Incorrect + Count_Nonresponse). We found that there was a strong decline in responding to the test stimuli during the test phase: most birds reduced responding to each novel stimulus after 10 presentations (Fig. A1), indicating that the birds apparently learned to recognize the test stimuli as being different from the training ones and providing no reward. For this reason, we restricted our analyses of the responses to the different test stimuli to the first 10 test trials for each stimulus, as during this phase, the responses to the test stimuli were highest and therefore provided the best insights into whether there was variation in the proportion of correct responses between the experimental groups. To examine whether the birds still discriminated the test stimuli above chance, we examined whether the ratio of 'Count Correct/Count Incorrect' differed from 1. We did so by applying the log (Count_Correct/Count_Incorrect) (indicated as 'Log(Cor/Inco)' from now on) as the response variable against a log (odds ratio) = 0 in the model analysis. The nested structure of the data was also incorporated into the analysis since, for each experimental group of birds, one half was tested with test stimuli from series 1 and the other with test stimuli from series 2. In addition, one female individual in the Equal-duration training group exhibited responses that significantly deviated from those of the other individuals in the same group. During the probe testing phase, this bird's proportions of correct responses to two novel versions of stimuli ('Pitch-shifted +8%' and 'Contour-maintained Vocoded') exceeded 1.5 times the IQR above the upper quartile (Q3). Consequently, we identified this individual as an outlier and excluded its data from the model analyses (but, for completeness, it is shown in Figs 4 and 5).

For the spectrally changed treatments, the counts of the responses to modified sounds A and B were combined. For the Time-scaled treatments, the 'PC' and 'Log(Cor/Inco)' were calculated based on the response counts to the stimuli derived from training sound A and those derived from sound B separately. We analysed the data in this way because, for the Unequal-duration group, the 'Duration stretched 50%' sound B had a similar duration as training sound A (sound A was always the longer training sound), and the 'Duration compressed 50%' sound A had a similar duration as training sound B. Therefore, we expected that if stimulus duration was a parameter to which the birds were sensitive, that time scaling would differentially affect the responses to changes in the duration of training stimulus A and stimulus B. We thus did a separate analysis for the four Time-scaled treatments ('Duration stretched 20%', 'Duration compressed 20%', 'Duration stretched 50%' and 'Duration compressed 50%') and their corresponding training stimuli for two training groups, comparing the responses to training sounds A and B with those to the Time-scaled versions of sounds A and B.

To investigate the birds' ability to discriminate between various test sounds, generalized linear mixed-effects models (GLMMs) were utilized. These models incorporated 'Training_Group', 'Test_Treatment' and 'TrainingTrails_scaled' as fixed effects, with 'Bird_ID' as the random effect factor. Additionally, a fixed factor, 'Training_Sound', was included for the Time-scaled test treatments, encompassing four categories: 'Sound A—Equal-duration group', 'Sound B—Equal-duration group', 'Sound A—Unequal-duration group' and 'Sound B—Unequal-duration group'. As 'sex' had a negligible impact at the training group level it was not included in the model analysis. The analysis of these binomial models was carried out in R (R Core Team, 2016), utilizing the 'glmer' function from the lme4 package (Bates et al., 2015). Model selection was carried out using the Wald chi-square test. Finally, a post hoc analysis was conducted on the chosen model, incorporating false discovery rate (FDR) correction using the emmeans package (Lenth, 2023).

The experiment and procedures adhered to the European and Dutch legislation on animal experimentation and were approved by the Dutch Committee for Animal Experimentation (CCD—AVD number 1060020197507) and performed according to the guidelines of the Leiden University Committee for Animal Experimentation.

None of the birds had any experience with this experimental set-up or the stimuli preceding the experiment. Each experimental bird underwent a physical examination before being transferred to the Skinner boxes. When the birds were in the boxes, their condition was monitored daily by visual observation throughout all phases of the experiment. The standard checks included: freshwater intake, amount of food obtained in response to pecking sensors, activity and measuring weight when deemed necessary. The functioning of the operant equipment and stimulus playback were also checked on a daily base. The daily welfare checks were done by the experimenter (N.Z.) as well as the qualified animal caretaker (in possession of a so-called 'art.13f2' qualification: the qualification required by Dutch law), who also advised on the most suitable protocol. Food and water were refreshed three times per week, and the litter floor (containing hard paper and dry sand) of the Skinner box was cleaned once per week. The food used as reward in the operant chamber consisted of a standard seed mixture for small seed-eating birds (a commercial tropical seed mixture: Deli Nature 56-Foreign finches super, Schoten, Belgium) enriched with mineral and vitamin powder (GistoCal, Raalte, the Netherlands). Cuttlefish bone was also available. This was the same diet as in their home aviary.

If a bird did not operate the sensors for food for more than 18 h (a very rare event), the hatch would open automatically, allowing a bird to gain sufficient food (approximately equal to the amount of food it should have obtained otherwise), before switching back to the experimental protocol again. The food consumption was checked by recording the amount of food disappearing from the food container. The 18 h included the 10.5 h of darkness and meant that the birds would never have been without food for a full day. In addition, obtaining food from the food hatch always gave rise to seeds falling on the floor and this was thus available continuously.

The decision to keep the birds in the Skinner boxes for the entire duration of the experiment rather than taking the birds in and out of their experimental cage for daily sessions was discussed with and approved by the Leiden University animal welfare body. The considerations were that daily sessions would require catching and moving the birds, events considered stressful to the birds. Also, in our set-up the birds could get access to food whenever they wanted whereas otherwise some food restriction period would be necessary to keep the birds motivated. The training stimuli were normal zebra finch songs, which are known to be attractive to both male and female zebra finches. After finishing all phases of the experiment, the birds were returned to their home aviaries. Previous similar experiments showed that birds reintroduced to the aviaries after having been in the Skinner boxes for several weeks experienced no particular difficulties.

RESULTS

Speed of Discrimination Learning

The discrimination training lasted until a bird reached the learning criterion of over 75% correct responses to both sound A and sound B for a consistent 3 days. All 28 birds finished the training and learned the discrimination on an average of 4209
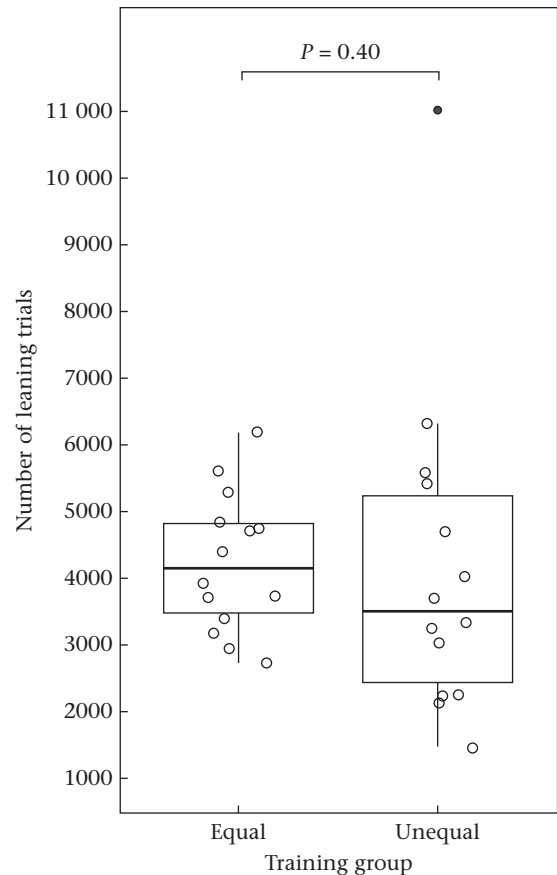
Figure 4. Number of learning trials needed to reach the learning criterion. Individual zebra finch results are shown with open dots. The black dot indicates the outlier. Box plots show median, first and third quartiles, and whiskers 1.5 × interquartile range.

(SD = 1840, N = 28) trials to reach the criterion. No significant difference (Z = 0.87, P = 0.40; Fig. 4) was found between the Equal-duration group (mean = 4243, SD = 1041) and the Unequal-duration group (mean = 4175, SD = 2439). Removal of one outlier (a female individual from the Unequal-duration training group requiring 11011 learning trials) did not change this outcome. It suggests that birds from the two training groups learnt approximately equally fast.

Responses to Test Stimuli

We examined the impact of the stimulus modifications in several ways. First, we examined whether spectral changes (frequency shifts as well as vocoding) had an impact on the proportion of correct identifications of the stimuli. Doing so, we addressed whether this impact was in the predicted direction of being larger in the Equal-duration than in the Unequal-duration group, based on the assumption that spectral changes might serve as primary cues in the Equal-duration group, given the (almost) identical song durations. Next, we examined whether there is a difference in impact among the various spectral modifications. Finally, while the proportion of correct responses may be affected by a modification, this need not imply that the birds can no longer discriminate between similar modifications of training songs A and B; they may still show more correct than incorrect responses. To address this, we examined whether the ratio of correct versus incorrect responses to a modified stimulus was still above chance. We used the same

analysis structure to examine the impact of the tempo changes on the birds' proportions of correct responses and discrimination rate.

The effect of spectral changes

Responses to spectrally changed stimuli differ between groups and between test stimuli. For the birds' responses to stimuli that are spectrally manipulated, the ANOVA Type III for models of both test series showed that the proportion of correct responses (PC) differed significantly between the Equal-duration and Unequal-duration training groups as well as between the different Test stimuli. Thus, the two factors 'Training_Group' and 'Test_Treatment', as well as their strong interaction effects for the response variable 'PC', were selected as fixed factors for models of both series (see (1) and (2) in Table 2), in addition, the factor 'TrainingTrails_scaled' and its interaction with 'Test_Treatment' were left in the model for the response variable 'PC' in series 1 since a significant effect was found in this interaction (see (1) in Table 2).

Spectral changes affect the equal- duration group most strongly. Fig. 5 shows that the Equal-duration group had a lower PC to all spectrally changed test stimuli compared to the Unequal-duration group. The pairwise comparisons between two training groups by the post hoc Tukey's HSD tests (Table A1) showed that this difference was significant for the stimuli 'Pitch-shifted -20%' ($P < 0.01$) and 'Contour-averaged Vocoded' ($P < 0.05$) in series 1 and for the stimuli 'Pitch-shifted -8%' ($P < 0.05$), 'Frequency-shifted 500 Hz' and 'Contour-maintained Vocoded' (both $P < 0.01$) in series 2. In addition, Pitch-shifted upward versions seem to have less impact on the between-groups difference than Pitch-shifted downward versions.

The observed differences between the groups are in line with our expectation that birds trained with Equal-duration stimuli are more sensitive to spectral changes than the birds trained with Unequal-duration stimuli. They also show that this effect is present in both test series.

Table 2
ANOVA (Type III Wald chi-square tests) table for selected GLMMs

Variable	χ^2	df	$Pr(>\chi^2)$
(1) MODEL PC for spectrally changed (sound A & sound B combined in series 1)			
Training_Group	4.0689	1	0.044 *
Test_Treatment	95.1957	4	<2.2e-16 ***
Training_Group:Test_Treatment	11.4205	4	0.022 *
TrainingTrails_scaled	0.0161	1	0.899
Test_Treatment:TrainingTrails_scaled	14.1089	4	0.007 **
(2) MODEL PC for spectrally changed (sound A & sound B combined in series 2)			
Training_Group	9.3598	1	0.002 **
Test_Treatment	101.5756	4	<2.2e-16 ***
Training_Group:Test_Treatment	9.7748	4	0.044 *
(3) MODEL PC for time-scaled (sound A & sound B separated in series 1)			
Training_Group	0.0053	1	0.942
Test_Treatment	48.2401	2	3.348e-11 ***
Training_Sound	0.0633	2	0.969
Training_Group:Test_Treatment	9.0091	2	0.011 *
Test_Treatment:Training_Sound	10.2016	4	0.037 *
(4) MODEL PC for time-scaled (sound A & sound B separated in series 2)			
Training_Group	1.2867	1	0.257
Test_Treatment	13.9044	2	0.001 ***
Training_Sound	0.5413	2	0.763
Training_Group:Test_Treatment	0.3930	2	0.822
Test_Treatment:Training_Sound	2.4069	4	0.661

All variables shown here were used as fixed factors for corresponding models, whether their P values were significant or not, because these were our variables of interest. 'Bird_ID' was used as the only random factor in all models.
*$P \le 0.05$; **$P \le 0.01$; ***$P \le 0.001$.

Differences in responses between test stimuli. In both series 1 and series 2, the birds responded with a higher PC to the training stimuli compared to all four spectrally changed stimuli in both training groups. For each training group, we examined whether there were differences in the PC of birds' responses between four spectrally changed stimuli for each test series (Table A1). In series 1 the birds of the Equal-duration group responded with a significantly higher PC to the 'Pitch-shifted $+20\%$' stimulus than to 'Contour-averaged Vocoded' ($P < 0.05$), and a clear trend to a difference between 'Frequency-shifted 1500 Hz' and 'Contour-averaged Vocoded' ($P = 0.06$). The birds of the Unequal-duration group responded with a significantly higher PC to the 'Pitch-shifted -20%' stimulus than to 'Contour-averaged Vocoded' and 'Frequency-shifted 1500 Hz' (both $P < 0.05$; Fig. 5a). In series 2, the birds of the Equal-duration group responded with a significantly higher PC to the 'Pitch-shifted $+8\%$' stimulus than to the other three spectrally changed stimuli ('Pitch-shifted -8%' ($P < 0.05$), 'Frequency-shifted 500 Hz' ($P < 0.01$) and 'Contour-maintained Vocoded' ($P < 0.001$). For the Unequal-duration group there is no significant difference in PC between the four spectrally changed stimuli (Fig. 5b). On the whole, these results show a weak tendency for pitch-shifted versions to have less impact on the PCs than vocoded versions. This implies that, if anything, the zebra finches were attending more strongly to precise spectral details of the song elements rather than to the spectral envelope. If they had attended more to the latter, vocoding would have had a lesser impact than the other manipulations.

Are spectrally changed stimuli still recognized?. If the birds are still capable of linking the modified stimuli to the respective training stimuli, the number of correct responses to the test stimuli should be higher than the number of incorrect responses. The birds of the Unequal-duration group responded above chance to all spectrally changed stimuli in both test series (Fig. 6a, b), while birds of the Equal-duration group responded above chance only to two of the spectrally changed stimuli ('Pitch-shifted $+8\%$' & 'Frequency-shifted 500 Hz') in series 2, and to none of the spectrally changed stimuli in series 1 (Table A2). This confirms the finding above that the birds from the Equal-duration group attended more strongly to spectral features than the birds from the Unequal-duration group. In addition, the Equal group showed a lower degree of recognition when the modifications were stronger (series 1) than when they were less strong (series 2).

The Effect of Duration Changes

Tempo changes affect the equal- and unequal-duration groups differently

In the ANOVA Type III model for responding to Time-scaled stimuli we also included the factors 'Training_Sound' (A or B) as fixed factor in addition to the factors 'Training_Group' and 'Test_Treatment', as well as the interactions of 'Training_Group' and 'Training_Sound' with 'Test_Treatment'. There were no significant differences in PC between the Equal-duration and the Unequal-duration training groups when the results for the Time-scaled versions of training stimuli A and B were combined. However, the results showed a significant interaction effect between 'Training_Group' and 'Training_Sound', as well as for 'Test_Treatment' and 'Training_Sound' for series 1 (see (3) in Table 2). Fig. 7 shows that this is due to the different responses of both groups to the various Time-scaled versions of training sounds A and B.

Differences in responses between test stimuli

The pairwise comparisons of the PC for the Time-scaled versions of sound A and sound B for the Equal- and Unequal-duration groups are shown in Table A3. In both series 1 and series 2, the PC shows no
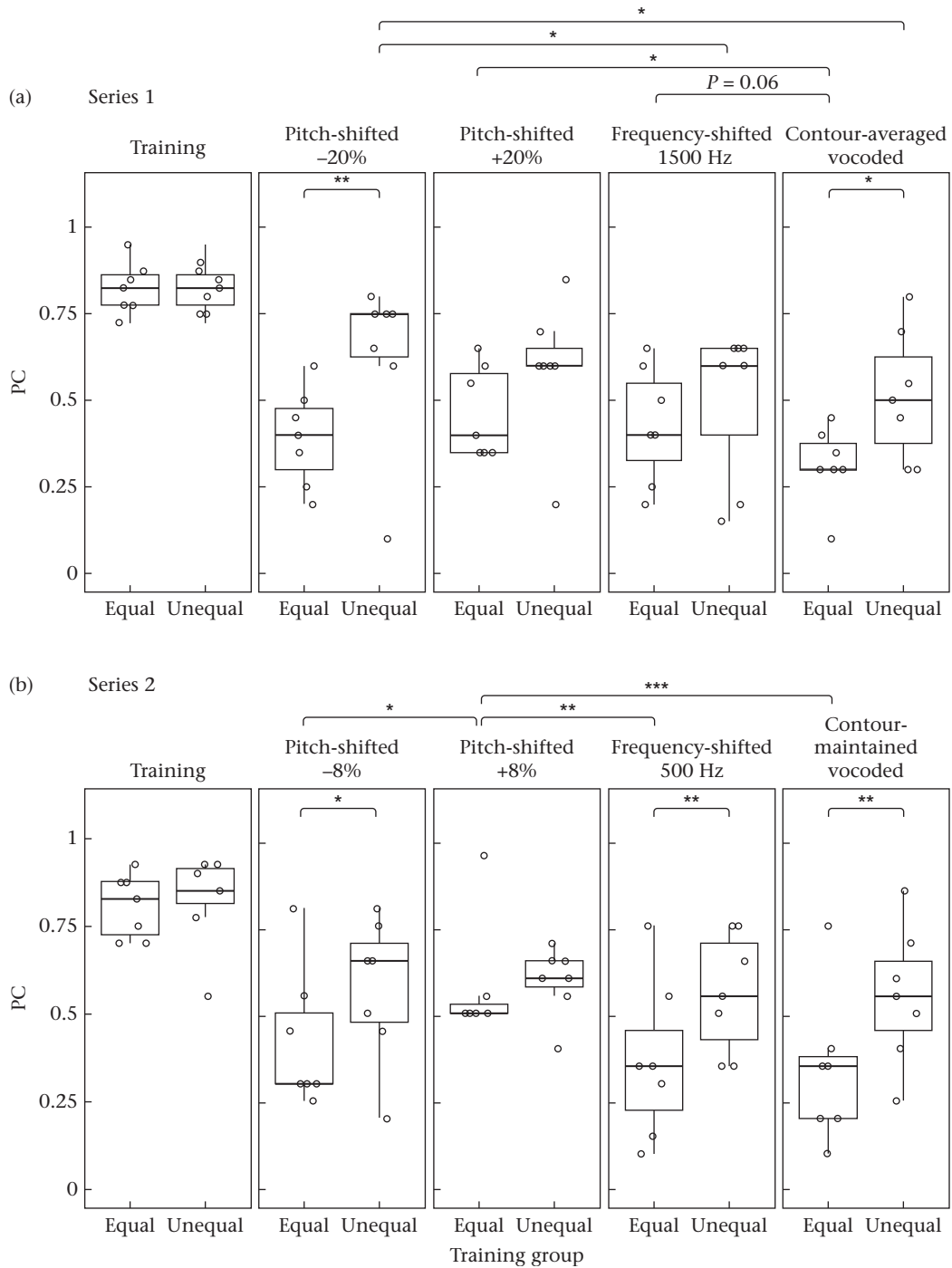
Figure 5. Proportion correct (PC) responses to spectrally changed stimuli of (a) series 1 and (b) series 2. The significant between-group and within-group differences are indicated, except for differences between the scores for the training stimuli and those for the other test stimuli. ***$P \leq 0.001$; **$P \leq 0.01$; *$P \leq 0.05$; for nonindicated comparisons: $P > 0.05$. Box plots show median, first and third quartiles, and whiskers $1.5 \times$ interquartile range.

significant differences between the A and B sound versions of the training and the two Time-scaled test sounds, i.e. PC for training sound A does not differ from that for training sound B; PC for 'Duration stretched 50%' A does not differ from that for 'Duration stretched 50%' B, etc.

The comparison of the PC of the training stimuli with that of the different test stimuli shows that, in series 1, the birds responded with a higher PC to the training stimuli compared to the Time-scaled versions of sound A and sound B (in both Equal-duration and Unequal-duration groups). This difference is significant in all

comparisons apart from the difference between training A and the 'Duration compressed 50%' A in the Equal-duration group, which showed a clear trend in the same direction ($P = 0.06$). The only difference in PC between the test stimuli is for birds of the Unequal-duration group, which responded with a significantly lower PC to the 'Duration compressed 50%' A than to the 'Duration stretched 50%' A ($P < 0.05$; Fig. 7a).

In series 2, the birds of the Equal-duration group responded with a significantly higher PC to training sound B than to 'Duration stretched 20%' B ($P < 0.01$), and there was a clear trend of difference

Figure 6. Visualization of Log (Correct/Incorrect) for birds responding to spectrally changed stimuli of (a) series 1 and (b) series 2. An asterisk indicates that the Log (Correct/Incorrect) of a test treatment is significantly different from 0; NS indicates that it overlaps 0. Box plots show median, first and third quartiles, and whiskers 1.5 × interquartile range. Horizontal dashed lines show the discrimination boundaries in which the proportion of correct responses is equal to the proportion of incorrect responses. The calculation of Log (Correct/Incorrect) was based on the counts of 'correct response' and 'incorrect response' from the same data set that was used for Fig. 5. Note that one bird's data point cannot be fully displayed in (b) because it made no incorrect responses to the 'Frequency-shifted 500 Hz' stimuli version, resulting in an infinitely large value after log-scaling.

between training sound B and 'Duration compressed 20%' B ($P = 0.06$). For the Unequal-duration group the PC did not differ between the training sounds and the 20% Duration changed stimuli.

To conclude, the '±50%' Time-scaled manipulation was noticed by birds from both Equal-duration and Unequal-duration groups, but this impact was weaker in the '±20%' Time-scaled manipulation. In addition, the Unequal-duration group responded differently to whether the 50% duration change concerned the long or the short song. This difference is meaningful and was expected if the birds in the Unequal-duration group attended to the song

duration for song recognition. Training sound A was always 50% longer than training sound B in the Unequal-duration training group. Therefore the 'Duration compressed 50%' of the sound A stimulus made this stimulus the same length as training stimulus B, while the 'Duration stretched 50%' of the sound B stimulus made this stimulus the same length as training stimulus A. This suggests that the similarities in duration between training songs and test songs resulted in reduced song recognition even when there were still differences in spectral features between the pair of sounds.
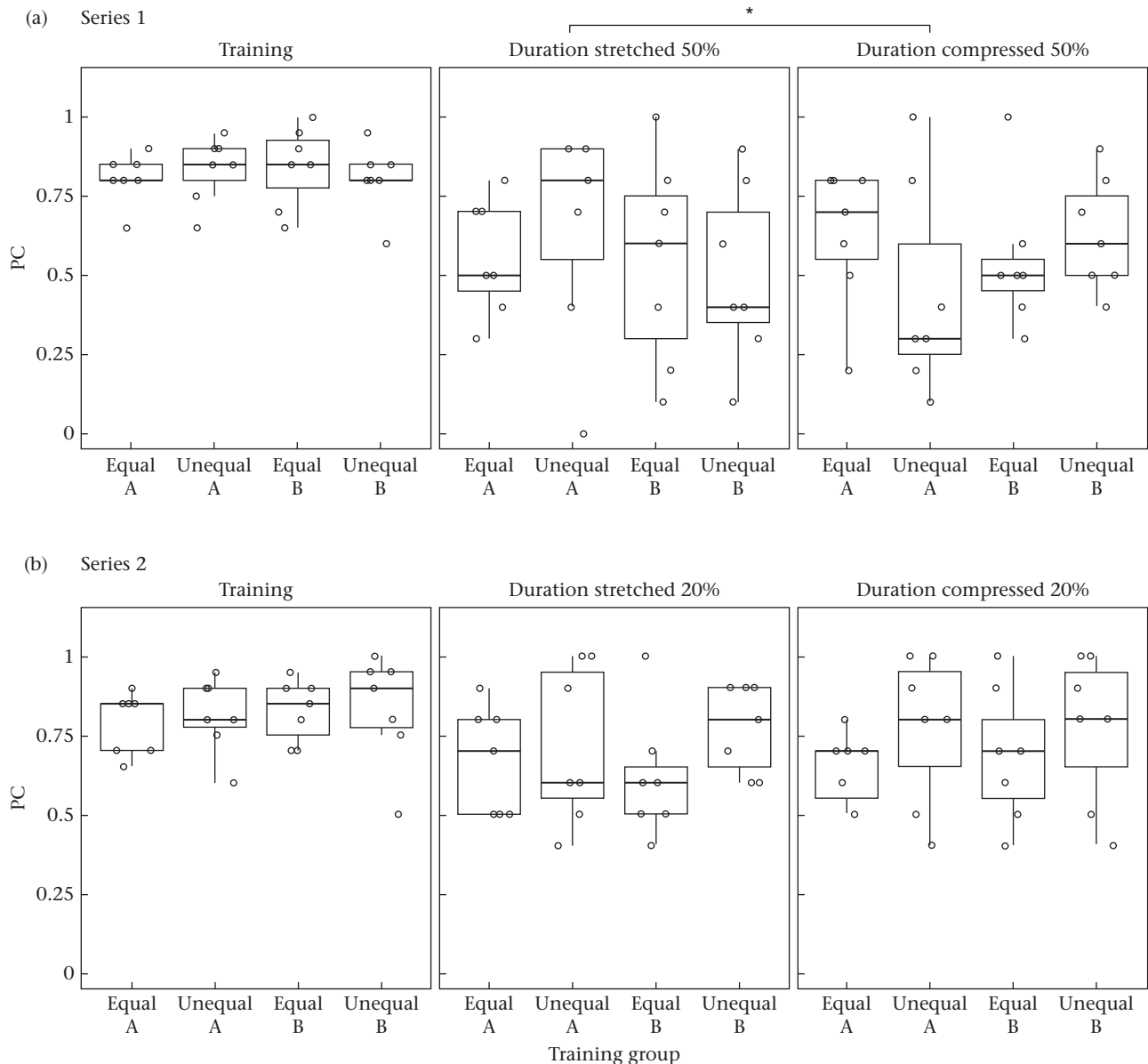
Figure 7. Proportion correct (PC) responses to Time-scaled stimuli of (a) series 1 and (b) series 2. For significant differences between training and Duration-changed stimuli see text. *$P \leq 0.05$. Box plots show median, first and third quartiles, and whiskers 1.5 × interquartile range.

Are time-scaled stimuli still recognized?

At the group level (if the birds' responses to sound A or sound B are not differentiated in the analysis), the responses of birds of both groups to all stimuli (Training stimuli and Time-scaled versions) are all different from 0, indicating they are recognized (Table A4). Similar to the analysis for the spectrally changed stimuli, we also examined for which of the Time-scaled stimuli the number of correct responses was higher than that of the incorrect responses, but now differentiating between the responses to the test stimuli derived from training sound A and those derived from B. The birds of the Equal-duration group responded correctly above chance on the 'Duration stretched 50%' and 'Duration compressed 50%' sound A, but not to the 'Duration stretched 50%' and 'Duration compressed 50%' sound B (Fig. 8a). However, as for this group training songs A and B were of equal duration and arbitrarily assigned to be either A or B, the difference between the responses can be ascribed to chance, also because there is no significant difference between the scores to the variants derived from training stimulus A and B

(Table A5). The birds of the Unequal-duration group responded significantly above chance to the 'Duration stretched 50%' sound A and 'Duration compressed 50%' sound B but responded to the 'Duration stretched 50%' sound B and 'Duration compressed 50%' sound A by chance (Fig. 8a). In line with the finding of a difference in impact on the proportion of correct responses, the difference between recognizing the stretched and the compressed versions of sounds A and B by birds from the Unequal-duration group confirms that these birds used song duration to distinguish training songs A and B. In series 2, the responses of birds of both groups to all Time-scaled sounds (no matter whether it was sound A or sound B) are statistically different from 0 in favour of a correct response (Fig. 8b).

DISCUSSION

Our study demonstrates that zebra finches can use both spectral features and song duration when discriminating between two songs. However, the importance of these acoustic parameters
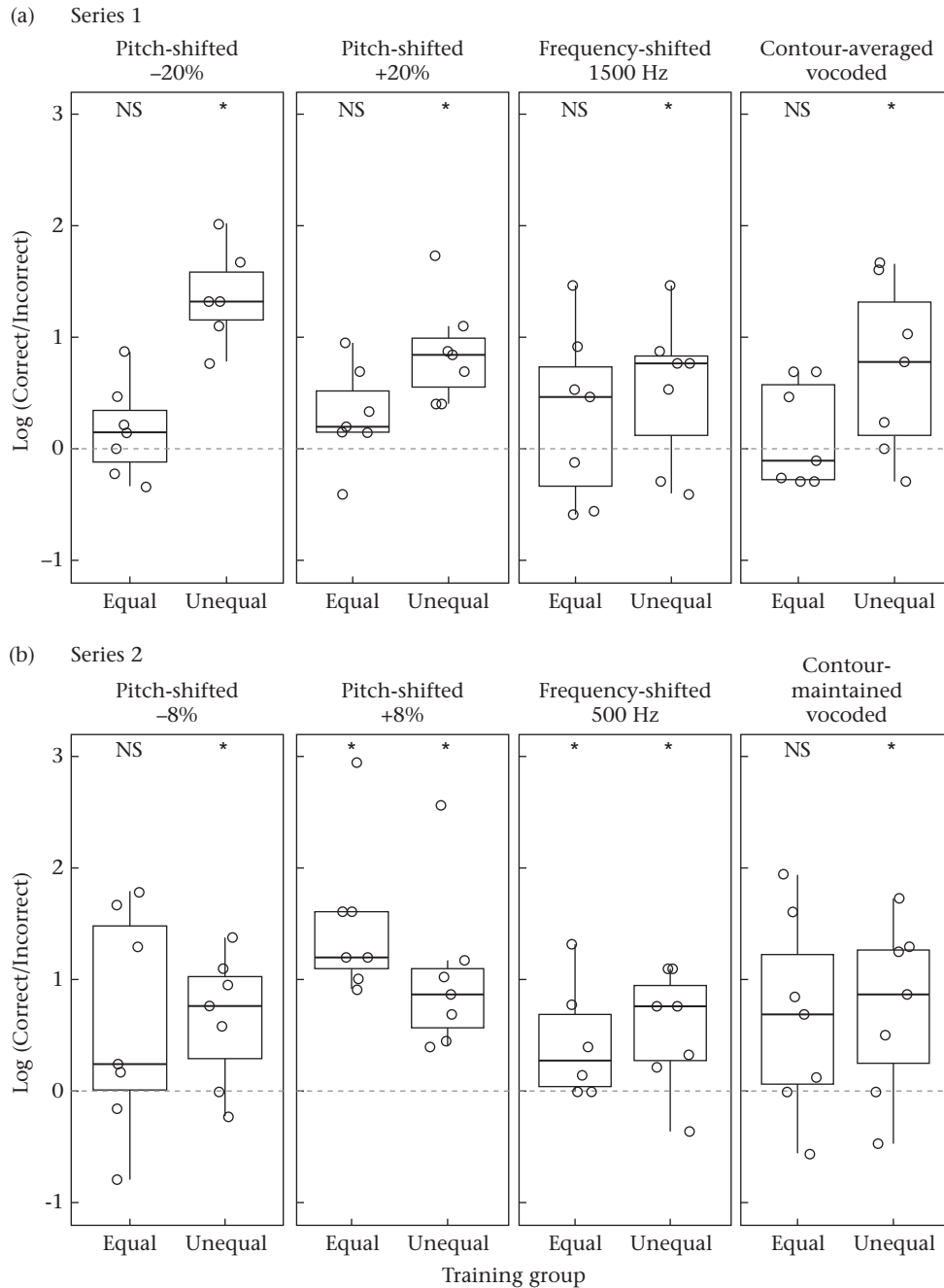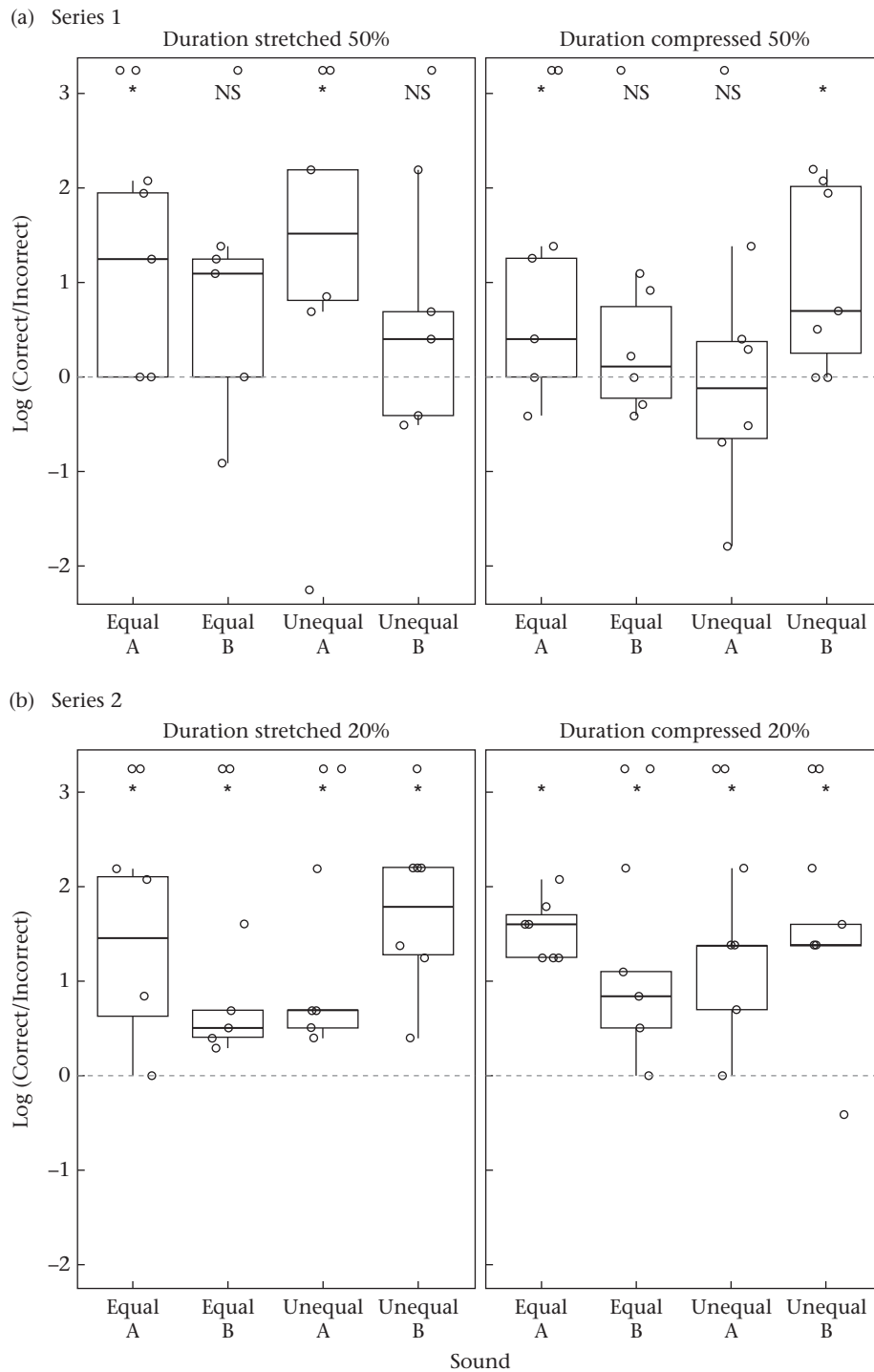
Figure 8. Visualization of Log (Correct/Incorrect) for birds responding to time-scaled stimuli of (a) series 1 and (b) series 2. An asterisk indicates that the Log (Correct/Incorrect) of a test treatment is significantly different from 0; NS indicates that it overlaps 0. Box plots show median, first and third quartiles, and whiskers 1.5 × interquartile range. Horizontal dashed lines show the discrimination boundaries in which the proportion of correct responses is equal to the proportion of incorrect responses. The calculation of Log (Correct/Incorrect) was based on the counts of 'correct response' and 'incorrect response' from the same data set that was used for Fig. 7.

depended on whether the songs differed in duration or not, with spectral features having a less prominent role when duration was available as an additional feature to distinguish two songs. Our results thus show that the acoustic parameters that zebra finches attend to are, at least partially, context driven, i.e. dependent on the degree to which these parameters differ between songs and as such support the hypothesis that zebra finches are cognitively flexible in

their attention to different acoustic parameters, related to the salience of the differences between songs.

Song Duration Does Not Affect Learning Speed

If zebra finches can use song duration as an additional cue for discrimination learning, then we may expect that this results in

faster song discrimination learning with songs of different compared to similar duration. However, in our current experiment, the learning speed of the birds trained on songs of unequal duration does not differ from that of the birds trained on songs of equal duration. Combined with our test results indicating that both experimental groups attended to spectral cues as well as song duration, albeit with a difference in weight, this suggests that both song features are considered right from the start of the learning process.

Spectral Cues or Song Duration?

Various studies (Dooling et al., 2002; Fishbein et al., 2021; Geberzahn & Derégnaucourt, 2020; Lawson et al., 2018; Lohr et al., 2006; Mol et al., 2021; Prior, Smith, Ball et al., 2018; Prior, Smith, Lawson et al., 2018; Vernaleo & Dooling, 2011; Vernaleo et al., 2010) concluded that when zebra finches discriminate between two songs they primarily attend to spectral details and temporal fine structure within individual syllables, and are far less sensitive to syllable sequence and temporal features of the whole song, such as song duration. In particular, a study by Nagel et al. (2010) showed that an 8% pitch shift already resulted in reduced discrimination between two songs, and that the songs were no longer discriminated between after a 32% pitch-shift. In contrast, stretching or compressing the songs by 32% in duration hardly affected discrimination, which was maintained even after a 64% change. In that study zebra finches had to discriminate between two songs of equal duration; hence the birds could not use song duration to recognize the songs. However, several recent studies indicated that the parameters that zebra finches can or do use in discriminating and recognizing sounds may depend, at least to some extent, on the difference between the sound stimuli (Burgering et al., 2018, 2019; Ning et al., 2023). Therefore, the main question underlying the current experiment was whether the importance of spectral parameters ('relative pitch' and 'spectral envelope'), and the temporal parameter 'duration', depended on whether the songs that had to be discriminated differed in overall duration. In both test series, the birds from both the Equal- and the Unequal-duration groups responded with a lower proportion of correct responses to all four spectrally changed stimuli than to the training stimuli, indicating that all birds were able to detect all the different types of spectral changes. However, the impact of the spectral changes was stronger in the Equal-duration than in the Unequal-duration group for both test series. The impact of the spectral changes was also stronger in series 1, in which the test sounds featured more substantial spectral modifications compared to the training sounds than in series 2. For the Equal-duration group, this even resulted in a loss of recognition of spectrally modified versions of training sounds for all spectral modifications in series 1 and half of them in series 2, while the Unequal-duration group maintained the recognition of all spectrally modified stimuli in both series. In response to changes in song duration, both groups also showed a lower proportion of correct responses and poorer discrimination when song durations were stretched or compressed by 50% (series 1), thus indicating that both groups attended to song duration. However, a 20% change in duration showed only a limited effect. These results are within the same ranges as observed by Nagel et al. (2010). Nevertheless, the importance of song duration for song discrimination was very noticeable in the Unequal-duration group. These birds no longer discriminated between the songs when the 50% compressed and stretched versions made the test song of the same length as the opposite training song, i.e. when the duration of the manipulated song A was similar to the duration of training song B and vice versa. For the Unequal-duration group, the 20% Time-scaled manipulation affected the discrimination substantially less than the 50% Time-

scaled manipulation. In this case, the temporal manipulation did not eliminate the differences in song duration between manipulated and training songs. To conclude, while our study confirms the important contribution of spectral features for song discrimination as obtained in earlier studies, it also shows that zebra finches use song duration as a prominent feature when songs are substantially different in duration, at the expense of attending to spectral features. Future studies may address whether the impact of song duration is related to the magnitude of the difference in duration between songs.

The finding that zebra finches are attending to the absolute duration of a stimulus also has relevance for studies examining rhythm perception in this species. The crucial test for being able to perceive a rhythmic pattern is whether humans or nonhuman animals can recognize a melody or tone sequence when this sequence is being speeded up or slowed down (e.g. Bouwer et al., 2021). Several studies have demonstrated that zebra finches could discriminate between a regular and an irregular pattern of song syllables or artificial tones (Lampen et al., 2014; van der Aa et al., 2015; ten Cate et al., 2016; Lampen et al., 2017; Rouse et al., 2021, 2023). However, this discrimination is reduced with a tempo change of the stimuli (van der Aa et al., 2015; ten Cate et al., 2016). This indicates that zebra finches attend more to the absolute duration of components of a stimulus, such as the duration of specific elements or intervals, rather than to the overall pattern of regularity (ten Cate et al., 2016), although it might be that with extensive training zebra finches might become more sensitive to the overall pattern (Rouse et al., 2021, 2023). The current finding that zebra finches show reduced discrimination when songs are compressed or stretched and attend to absolute song duration is thus in line with the results of the studies on zebra finch rhythm perception.

Impact of Various Spectral Changes

The second question we aimed to address in the current study concerns the relevance of spectral envelope and pitch in song discrimination. All our spectral manipulations maintained the absolute durations of syllables and songs but affected the spectral structure in different ways. The Frequency-shifted test stimuli moved the whole spectrum upward in a linear way. This maintained the frequency bandwidth but changed the harmonic relationships (with harmonic overtones being converted into inharmonic partials) among the frequencies within and between syllables. In the pitch-shifted stimuli the relative relationships among the frequencies within and between syllables are maintained, but the absolute pitch values have changed from those of the training stimuli. For the vocoded stimuli, the frequency ranges (spectral envelope) are identical to the training stimuli, but pitch information is removed and replaced by noise. Although the Unequal-duration group used the duration as a prominent cue and was less affected by the spectral changes, both groups showed decreased discrimination of all spectrally changed stimuli compared to discrimination of the training stimuli. Overall, the vocoded versions seem to reduce the discrimination more than the other stimuli, with at best a weak tendency for discrimination to be maintained best for the pitch-shifted stimuli. If we compare our data on the impact of pitch shifts on song discrimination with those obtained by Nagel et al. (2010), we found that birds in the Equal-duration group, which is most comparable to the experiment of Nagel et al. (2010), show a comparable outcome. In that study, an 8% pitch shift reduced but still maintained discrimination, but a 32% shift resulted in a lack of discrimination. These effects are in the same range as the reduced discrimination we obtained with an 8% pitch shift and lack of discrimination with a 20% change. The results of both our study and that of Nagel et al. (2010) also indicate that

zebra finches are more sensitive to pitch changes of songs than starlings are, which can still show discrimination of songs with pitch shifts up to ±40% (Bregman et al., 2012). Interestingly, starlings trained on piano melodies responded more strongly to pitch changes than those trained on songs, indicating that the nature of the stimuli may be a relevant factor in this songbird's sound discrimination (Bregman et al., 2012).

Finally, we showed that both types of vocoded stimuli strongly reduced discrimination of the songs to a similar extent. It thus did not matter whether the spectral contour was maintained over the elements (Contour-maintained Vocoded) or not (Contour-averaged Vocoded). The impact of noise-vocoding on song recognition is surprising in light of earlier studies. For starlings, Bregman et al. (2016) showed that vocoded versions, but not pitch-shifted versions, of sequences of tones that varied in pitch and timbre maintained the discrimination between these sequences. This indicated that the sequences were discriminated between by their spectral envelope rather than pitch. Patel (2017) argued that this might also be a common characteristic across birds for the discrimination of natural vocalizations. However, so far, no study has examined how starlings respond to vocoded versions of conspecific songs and it hence remains to be explored whether such a stimulus would result in similar outcomes when compared to testing with artificial sounds. Nevertheless, the importance of the spectral envelope for auditory discrimination in birds seemed to be supported by a study in zebra finches, in which Burgering et al. (2019) trained zebra finches to distinguish between two sets of artificial harmonic tone stimuli, which could only be differentiated by attending to the spectral envelope. When these stimuli were noise-vocoded, maintaining the (absolute) spectral envelope but removing (absolute) pitch information, the discrimination was maintained, indicating that zebra finches indeed attended to the spectral envelope of the stimuli. Also, an extensive analysis of zebra finch vocalizations indicated that the shape of the frequency spectrum (spectral envelope) of the different vocalizations was an important potential information-bearing feature (Elie & Theunissen, 2016, 2020) for distinguishing various vocalizations. Hence, one would expect the spectral envelope to be important for discriminating between songs. Why this does not show up in the current study is not clear. One factor might be that the spectral envelope might be relevant to zebra finches when discriminating between calls or other shorter sounds, such as the single-element stimuli used by Burgering et al. (2019). In contrast, discrimination of songs might rely more on attending to other spectral features, including pitch and harmonic structure of the songs. Attending to such features has been demonstrated in a range of studies (e.g. Dooling & Lohr, 2006; Lohr et al., 2006; Okanoya & Dooling, 1990; Prior, Smith, Ball et al., 2018; Prior, Smith, Lawson et al., 2018; Uno et al., 1997; Vignal & Mathevon, 2011).

Conclusion and Outlook

To conclude, our study shows that the acoustic parameters that zebra finches use to distinguish between different songs depend on the dimensions in which these songs differ. As we demonstrated here, this could be spectral features, but also song duration. Similarly, in another study we showed that although zebra finches do not usually give much attention to the sequential order of the syllables when discriminating between songs, they can very readily use this sequence if needed (Ning et al., 2023), while Burgering et al. (2018, 2019) demonstrated that attending to either the fundamental frequency (absolute pitch) or the energy distribution of a harmonic spectrum (spectral envelope) also varied depending on the task. These results thus contribute to expanding evidence that zebra finches are cognitively flexible: when faced with the task

of discriminating between different acoustic stimuli; they appear to focus on the most salient features distinguishing these stimuli. That the importance of different parameters for sound discrimination may depend on the nature of the stimuli and on the task the birds are facing is also recognized by others (e.g. Patel, 2017). However, this does not imply that there is no bias in this ability, but it indicates that there may be a difference between which features an animal does use to discriminate stimuli in a particular context and which it can use. Our study also shows that both the features in which stimuli differ and the magnitude of those differences affect their importance in discrimination. Future studies might explore other potential cues for song discrimination. Such investigations will contribute to a more nuanced understanding of how birds perceive and utilize various song features as a discriminative cue. At the same time, comparing the results of our study with those obtained in starlings (Bregman et al., 2016) suggests important differences between avian species, differences that call for further exploration.

Author Contributions

Z.-Y.N. and C.t.C. conceived and designed the research. H.H. contributed to the design of the study. Z.-Y.N. carried out the experiments. H.v.M. designed the statistical model and the computational framework. H.v.M. and Z.-Y.N. analysed the data. C.t.C. and H.H. contributed to the interpretation of the results. Z.-Y.N. and C.t.C. wrote the initial draft of the manuscript, and H.H. and H.v.M. provided feedback. All authors contributed to the overall framework of the paper and edited the final manuscript.

Data Availability

Samples of the audio stimuli and the data sets analysed during the current study are provided as Supplementary material.

Declaration of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

Supplementary Material

Supplementary material associated with this article is available in the online version at https://doi.org/10.1016/j.anbehav.2024.03.011.

References

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bouwer, F. L., Nityananda, V., Rouse, A. A., & ten Cate, C. (2021). Rhythmic abilities in humans and non-human animals: A review and recommendations from a methodological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1835). https://doi.org/10.1098/rstb.2020.0335

Braaten, R. F., Petzoldt, M., & Colbath, A. C. (2006). Song perception during the sensitive period of song learning in zebra finches (*Taeniopygia guttata*). *Journal*

of Comparative Psychology, 120(2), 79−88. https://doi.org/10.1037/0735-7036.120.2.79

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2012). Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition, 122*(1), 51−60. https://doi.org/10.1016/j.cognition.2011.08.008

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences of the United States of America, 113*(6), 1666−1671. https://doi.org/10.1073/pnas.1515380113

Burgering, M. A., ten Cate, C., & Vroomen, J. (2018). Mechanisms underlying speech sound discrimination and categorization in humans and zebra finches. *Animal Cognition, 21*(2), 285−299. https://doi.org/10.1007/s10071-018-1165-3

Burgering, M. A., Vroomen, J., & ten Cate, C. (2019). Zebra finches (*Taeniopygia guttata*) can categorize vowel-like sounds on both the fundamental frequency ('pitch') and spectral envelope. *Journal of Comparative Psychology, 133*(1), 106−117. https://doi.org/10.1037/com0000143

Dabelsteen, T., & Pedersen, S. B. (1992). Song features essential for species discrimination and behaviour assessment by male blackbirds (*Turdus merula*). *Behaviour, 121*(3/4), 259−287. http://www.jstor.org/stable/4535030.

Dent, M. L., Martin, A. K., Flaherty, M., & Neilans, E. G. (2016). Cues for auditory stream segregation of birdsong in budgerigars and zebra finches: Effects of location, timing, amplitude, and frequency. *Journal of the Acoustical Society of America, 139*(2), 674−683. https://doi.org/10.1121/1.4941322

Dooling, R. J., Leek, M. R., Gleich, O., & Dent, M. L. (2002). Auditory temporal resolution in birds: Discrimination of harmonic complexes. *Journal of the Acoustical Society of America, 112*(2), 748−759. https://doi.org/10.1121/1.1494447

Dooling, R. J., & Lohr, B. (2006). Auditory temporal resolution in the zebra finch (*Taeniopygia guttata*): A model of enhanced temporal acuity. *Ornithological Science, 5*(1), 15−22. https://doi.org/10.2326/osj.5.15

Dooling, R. J., & Prior, N. H. (2017). Do we hear what birds hear in birdsong? *Animal Behaviour, 124*, 283−289. https://doi.org/10.1016/j.anbehav.2016.10.012

Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition, 19*(2), 285−315. https://doi.org/10.1007/s10071-015-0933-6

Elie, J. E., & Theunissen, F. E. (2020). Author correction: Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications, 11*(1). https://doi.org/10.1038/s41467-020-16398-z

Fishbein, A. R., Prior, N. H., Brown, J., Ball, G. F., & Dooling, R. J. (2021). Discrimination of natural acoustic variation in vocal signals. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-020-79641-z

Geberzahn, N., & Derégnaucourt, S. (2020). Individual vocal recognition in zebra finches relies on the song syllable structure rather than on song syllable order. *Journal of Experimental Biology.* https://doi.org/10.1242/jeb.220087

Hoeschele, M. (2017). Animal pitch perception: Melodies and harmonies. *Comparative Cognition & Behavior Reviews, 12*, 5−18. https://doi.org/10.3819/ccbr.2017.120002

Hulse, S. H., Cynx, J., & Humpal, J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *Journal of Experimental Psychology: General, 113*(1), 38−54. https://doi.org/10.1037/0096-3445.113.1.38

Kreutzer, M., Dooling, R. J., Brown, S. D., & Okanoya, K. (1990). A comparison of song syllable perception by species of birds. *International Journal of Comparative Psychology, 4*(2). https://doi.org/10.46867/c4h599

Lampen, J., Jones, K. M., McAuley, J. D., Chang, S., & Wade, J. (2014). Arrhythmic song exposure increases ZENK expression in auditory cortical areas and nucleus taeniae of the adult zebra finch. *PLoS One, 9*(9), Article e108841. https://doi.org/10.1371/journal.pone.0108841

Lampen, J., McAuley, J. D., Chang, S. E., & Wade, J. (2017). ZENK induction in the zebra finch brain by song: Relationship to hemisphere, rhythm, oestradiol and sex. *Journal of Neuroendocrinology, 29*(12). https://doi.org/10.1111/jne.12543

Lawson, S. L., Fishbein, A. R., Prior, N. H., Ball, G. F., & Dooling, R. J. (2018). Relative salience of syllable structure and syllable order in zebra finch song. *Animal Cognition, 21*(4), 467−480. https://doi.org/10.1007/s10071-018-1182-2

Lenth, R. V. (2023). *Estimated marginal means, aka least-squares means* [R package emmeans version 1.9.0] https://cran.r-project.org/package=emmeans.

Lohr, B., Dooling, R. J., & Bartone, S. (2006). The discrimination of temporal fine structure in call-like harmonic sounds by birds. *Journal of Comparative Psychology, 120*(3), 239−251. https://doi.org/10.1037/0735-7036.120.3.239

MacDougall-Shackleton, S. A., & Hülse, S. H. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology, 110*(2), 139−146. https://doi.org/10.1037/0735-7036.110.2.139

Mol, C., Bolhuis, J. J., & Moorman, S. (2021). Vocal learning in songbirds: The role of syllable order in song recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1836). https://doi.org/10.1098/rstb.2020.0248

Nagel, K. I., McLendon, H. M., & Doupe, A. J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of Neurophysiology, 104*(3), 1426−1437. https://doi.org/10.1152/jn.00028.2010

Naugler, C., & Ratcliffe, L. (1992). A field test of the sound environment hypothesis of conspecific song recognition in American tree sparrows (*Spizella arborea*). *Behaviour, 123*(3/4), 314−324. http://www.jstor.org/stable/4535075.

Neilans, E. G., Welch, T. E., Maddox, R. K., Shinn-Cunningham, B., & Dent, M. L. (2010). The importance of syllable position and combination in the perception of birdsong by budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America, 128*(Suppl. 4), 2319. https://doi.org/10.1121/1.3508173

Nelson, D. A. (1989). The importance of invariant and distinctive features in species recognition of bird song. *Condor, 91*(1), 120−130. https://doi.org/10.2307/1368155

Ning, Z., Honing, H., & Ten Cate, C. (2023). Zebra finches (*Taeniopygia guttata*) demonstrate cognitive flexibility in using phonology and sequence of syllables in auditory discrimination. *Animal Cognition, 26*(4), 1161−1175. https://doi.org/10.1007/s10071-023-01763-4

Okanoya, K., & Dooling, R. J. (1990). Temporal integration in zebra finches (*Poephila guttata*). *Journal of the Acoustical Society of America, 87*(6), 2782−2784. https://doi.org/10.1121/1.399069

Patel, A. D. (2017). Why doesn't a songbird (the European starling) use pitch to recognize tone sequences? The informational independence hypothesis. *Comparative Cognition & Behavior Reviews, 12*, 19−32. https://doi.org/10.3819/ccbr.2017.120003

Prior, N. H., Smith, E. W. L., Ball, G. F., & Dooling, R. (2018). Temporal fine structure and zebra finch vocalizations. *Journal of the Acoustical Society of America, 143*(Suppl. 3), 1897. https://doi.org/10.1121/1.5036170

Prior, N. H., Smith, E. W. L., Lawson, S. L., Ball, G. F., & Dooling, R. J. (2018). Acoustic fine structure may encode biologically relevant information for zebra finches. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-24307-0

R Core Team. (2016). *R: A language and environment for statistical computing.* Retrieved from https://www.R-project.org/.

Rouse, A. A., Patel, A. D., & Kao, M. H. (2021). Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences of the United States of America, 118*(29). https://doi.org/10.1073/pnas.2026130118

Rouse, A. A., Patel, A. D., Wainapel, S., & Kao, M. H. (2023). Sex differences in vocal learning ability in songbirds are linked with differences in flexible rhythm pattern perception. *Animal Behaviour, 203*, 193−206. https://doi.org/10.1016/j.anbehav.2023.05.001

ten Cate, C., Spierings, M. J., Hubert, J., & Honing, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Frontiers in Psychology, 7.* https://doi.org/10.3389/fpsyg.2016.00730

Tu, H., & Dooling, R. J. (2012). Perception of warble song in budgerigars (*Melopsittacus undulatus*): Evidence for special processing. *Animal Cognition, 15*(6), 1151−1159. https://doi.org/10.1007/s10071-012-0539-1

Uno, H., Maekawa, M., & Kaneko, H. (1997). Strategies for harmonic structure discrimination by zebra finches. *Behavioural Brain Research, 89*(1−2), 225−228. https://doi.org/10.1016/s0166-4328(97)00064-8

Van der Aa, J., Honing, H., & ten Cate, C. (2015). The perception of regularity in an isochronous stimulus in zebra finches (*Taeniopygia guttata*) and humans. *Behavioural Processes, 115*, 37−45. https://doi.org/10.1016/j.beproc.2015.02.018

Vernaleo, B. A., & Dooling, R. J. (2011). Relative salience of envelope and fine structure cues in zebra finch song. *Journal of the Acoustical Society of America, 129*(5), 3373−3383. https://doi.org/10.1121/1.3560121

Vernaleo, B. A., Dooling, R. J., & Leek, M. R. (2010). Relative salience of acoustic features in zebra finch song. *Journal of the Acoustical Society of America, 127*(Suppl. 3), 1863. https://doi.org/10.1121/1.3384460

Vignal, C., & Mathevon, N. (2011). Effect of acoustic cue modifications on evoked vocal response to calls in zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology, 125*(2), 150−161. https://doi.org/10.1037/a0020865

Appendix

Table A1
Post hoc test results of the binomial GLMM for the interaction of spectrally changed Test_Treatment & Training_Group

Stimuli	Training_Group	Series	Estimate	SE	z	P
Proportion correct (sounds A+B in two training groups)						
Pairwise comparisons between training groups: series 1						
Training	Equal−Unequal	1	0.0676	0.407	0.166	0.868
Pitch-shifted −20%	**Equal − Unequal**	**1**	**−1.1345**	**0.359**	**−3.165**	**0.008**
Pitch-shifted +20%	Equal−Unequal	1	−0.6006	0.351	−1.711	0.145
Frequency-shifted 1500 Hz	Equal−Unequal	1	−0.2342	0.349	−0.671	0.628
Contour-averaged Vocoded	**Equal−Unequal**	**1**	**−0.8465**	**0.355**	**−2.387**	**0.043**
Pairwise comparisons between test stimuli: series 1						
Training−Pitch-shifted −20%	**Equal_Duration**	**1**	**2.1898**	**0.294**	**7.456**	**<0.0001**
Training−Pitch-shifted +20%	**Equal_Duration**	**1**	**1.8516**	**0.290**	**6.384**	**<0.0001**
Training−Frequency-shifted 1500 Hz	**Equal_Duration**	**1**	**1.9422**	**0.290**	**6.688**	**<0.0001**
Training − Contour-averaged Vocoded	**Equal_Duration**	**1**	**2.4698**	**0.298**	**8.301**	**<0.0001**
Pitch-shifted −20%−Pitch-shifted +20%	Equal_Duration	1	−0.3382	0.250	−1.352	0.252
Pitch-shifted −20%−Frequency-shifted 1500 Hz	Equal_Duration	1	−0.2476	0.250	−0.989	0.359
Pitch-shifted −20%−Contour-averaged Vocoded	Equal_Duration	1	0.2800	0.258	1.084	0.348
Pitch-shifted +20%−Frequency-shifted 1500 Hz	Equal_Duration	1	0.0906	0.246	0.368	0.713
Pitch-shifted +20%−Contour-averaged Vocoded	**Equal_Duration**	**1**	**0.6182**	**0.254**	**2.431**	**0.030**
Frequency-shifted 1500 Hz − Contour-averaged Vocoded	Equal_Duration	1	0.5276	0.254	2.074	0.064
Training−Pitch-shifted −20%	**Unequal_Duration**	**1**	**0.9877**	**0.296**	**3.341**	**0.002**
Training−Pitch-shifted +20%	**Unequal_Duration**	**1**	**1.1834**	**0.292**	**4.059**	**0.0002**
Training−Frequency-shifted 1500 Hz	**Unequal_Duration**	**1**	**1.6404**	**0.290**	**5.652**	**<0.0001**
Training−Contour-averaged Vocoded	**Unequal_Duration**	**1**	**1.5557**	**0.289**	**5.374**	**<0.0001**
Pitch-shifted −20%−Pitch-shifted +20%	Unequal_Duration	1	0.1957	0.259	0.755	0.500
Pitch-shifted −20%−Frequency-shifted 1500 Hz	**Unequal_Duration**	**1**	**0.6527**	**0.257**	**2.537**	**0.022**
Pitch-shifted −20%−Contour-averaged Vocoded	**Unequal_Duration**	**1**	**0.5680**	**0.256**	**2.215**	**0.045**
Pitch-shifted +20%−Frequency-shifted 1500 Hz	Unequal_Duration	1	0.4570	0.253	1.810	0.101
Pitch-shifted +20%−Contour-averaged Vocoded	Unequal_Duration	1	0.3724	0.252	1.480	0.174
Frequency-shifted 1500 Hz−Contour-averaged Vocoded	Unequal_Duration	1	−0.0846	0.249	−0.339	0.734
Pairwise comparisons between training groups: series 2						
Training	Equal−Unequal	2	−0.252	0.387	−0.652	0.515
Pitch-shifted −8%	**Equal−Unequal**	**2**	**−0.889**	**0.339**	**−2.623**	**0.015**
Pitch-shifted +8%	Equal−Unequal	2	−0.358	0.335	−1.069	0.357
Frequency-shifted 500 Hz	**Equal−Unequal**	**2**	**−1.095**	**0.344**	**−3.187**	**0.004**
Contour-maintained Vocoded	**Equal−Unequal**	**2**	**−1.230**	**0.348**	**−3.539**	**0.002**
Pairwise comparisons between test stimuli: series 2						
Training−Pitch-shifted −8%	**Equal_Duration**	**2**	**1.9396**	**0.296**	**6.545**	**<0.0001**
Training−Pitch-shifted +8%	**Equal_Duration**	**2**	**1.3160**	**0.291**	**4.518**	**<0.0001**
Training−Frequency-shifted 500 Hz	**Equal_Duration**	**2**	**2.2073**	**0.302**	**7.306**	**<0.0001**
Training−Contour-maintained Vocoded	**Equal_Duration**	**2**	**2.3733**	**0.307**	**7.731**	**<0.0001**
Pitch-shifted −8%−Pitch-shifted +8%	**Equal_Duration**	**2**	**−0.6236**	**0.265**	**−2.350**	**0.027**
Pitch-shifted −8%−Frequency-shifted 500 Hz	Equal_Duration	2	0.2677	0.277	0.966	0.371
Pitch-shifted −8%−Contour-maintained Vocoded	Equal_Duration	2	0.4337	0.282	1.537	0.156
Pitch-shifted +8%−Frequency-shifted 500 Hz	**Equal_Duration**	**2**	**0.8912**	**0.272**	**3.279**	**0.002**
Pitch-shifted +8%−Contour-maintained Vocoded	**Equal_Duration**	**2**	**1.0572**	**0.277**	**3.815**	**0.0003**
Frequency-shifted 500 Hz − Contour-maintained Vocoded	Equal_Duration	2	0.1660	0.288	0.576	0.565
Training−Pitch-shifted −8%	**Unequal_Duration**	**2**	**1.3031**	**0.286**	**4.548**	**<0.0001**
Training−Pitch-shifted +8%	**Unequal_Duration**	**2**	**1.2099**	**0.287**	**4.214**	**0.0001**
Training−Frequency-shifted 500 Hz	**Unequal_Duration**	**2**	**1.3646**	**0.286**	**4.768**	**<0.0001**
Training − Contour-maintained Vocoded	**Unequal_Duration**	**2**	**1.3952**	**0.286**	**4.877**	**<0.0001**
Pitch-shifted −8%−Pitch-shifted +8%	Unequal_Duration	2	−0.0932	0.249	−0.374	0.888
Pitch-shifted −8%−Frequency-shifted 500 Hz	Unequal_Duration	2	0.0615	0.248	0.248	0.893
Pitch-shifted −8%−Contour-maintained Vocoded	Unequal_Duration	2	0.0921	0.248	0.372	0.888
Pitch-shifted +8%−Frequency-shifted 500 Hz	Unequal_Duration	2	0.1547	0.249	0.622	0.888
Pitch-shifted +8%−Contour-maintained Vocoded	Unequal_Duration	2	0.1853	0.249	0.745	0.888
Frequency-shifted 500 Hz−Contour-maintained Vocoded	Unequal_Duration	2	0.0306	0.247	0.124	0.902

Response variables in GLMMs: the proportion of correct responses if birds respond to one of two sounds. Each of the corrected pairwise multiple comparison tests is separated into clusters within the table. Bold indicates significant differences.

Table A2
Estimates and 95% confidence intervals for the correct identification of spectrally changed stimuli

Training_Group	Stimuli	Series	Emmean	SE	CL (95%)	
					Lower	Upper
Log(Cor/Inco)~Training_Group+Stimuli+Training_Group:Stimuli+(1|BirdID), (sound A & sound B combined)						
Pairwise comparisons: series 1						
Equal_Duration	Pitch-shifted −20%	1	0.201	0.202	−0.1958	0.5976
Equal_Duration	Pitch-shifted +20%	1	0.283	0.191	−0.0908	0.6564
Equal_Duration	Frequency-shifted 1500 Hz	1	0.310	0.200	−0.0813	0.7021
Equal_Duration	Contour-averaged Vocoded	1	0.002	0.215	−0.4196	0.4237
Unequal_Duration	**Pitch-shifted −20%**	**1**	**1.075**	**0.213**	**0.6573**	**1.4921**
Unequal_Duration	**Pitch-shifted +20%**	**1**	**0.833**	**0.202**	**0.4373**	**1.2284**
Unequal_Duration	**Frequency-shifted 1500 Hz**	**1**	**0.635**	**0.206**	**0.2314**	**1.0380**
Unequal_Duration	**Contour-averaged Vocoded**	**1**	**0.693**	**0.205**	**0.2900**	**1.0954**
Pairwise comparisons: series 2						
Equal_Duration	Pitch-shifted −8%	2	0.252	0.244	−0.2258	0.7307
Equal_Duration	**Pitch-shifted +8%**	**2**	**1.234**	**0.280**	**0.6850**	**1.7834**
Equal_Duration	**Frequency-shifted 500 Hz**	**2**	**0.625**	**0.294**	**0.0492**	**1.2013**
Equal_Duration	Contour-maintained Vocoded	2	0.344	0.286	−0.2174	0.9056
Unequal_Duration	**Pitch-shifted −8%**	**2**	**0.682**	**0.207**	**0.2856**	**1.0885**
Unequal_Duration	**Pitch-shifted +8%**	**2**	**0.885**	**0.216**	**0.4613**	**1.3093**
Unequal_Duration	**Frequency-shifted 500 Hz**	**2**	**0.564**	**0.202**	**0.1678**	**0.9606**
Unequal_Duration	**Contour-maintained Vocoded**	**2**	**0.727**	**0.213**	**0.3097**	**1.1443**

Lower CL and upper CL represent the lower and upper 95% confidence limits of the confidence interval. If zero is part of that confidence interval, the treatment combination Training_Group and Stimuli is not significantly different from 0. If both confidence levels are positive, then there is a bias towards correct responses. If they are both negative, then they are biased towards incorrect responses. Each of the corrected pairwise multiple comparison test sets is separated into clusters within the table. Bold indicates significance.

Table A3
Post hoc test results of binomial GLMM for the interaction of Time-scaled Test_Treatment and Training_Sound

Stimuli	Training_Sound	Series	Estimate	SE	z	P
Proportion correct (sounds A & B separated)						
Pairwise comparisons between training sounds: series 1						
Duration stretched 50%	EqualA−EqualB	1	0.0594	0.345	0.172	0.863
Duration stretched 50%	UnequalA−UnequalB	1	0.6801	0.355	1.915	0.111
Duration compressed 50%	EqualA−EqualB	1	0.3645	0.350	1.043	0.396
Duration compressed 50%	UnequalA−UnequalB	1	−0.7906	0.353	−2.238	0.101
Pairwise comparisons between test stimuli: series 1						
Training−Duration stretched 50%	**EqualA**	**1**	**1.10e+00**	**0.383**	**2.867**	**0.012**
Training−Duration compressed 50%	EqualA	1	7.92e-01	0.387	2.047	0.061
Duration compressed 50%−Duration stretched 50%	EqualA	1	3.05e-01	0.350	0.871	0.384
Training−Duration stretched 50%	**EqualB**	**1**	**1.44e+00**	**0.402**	**3.573**	**0.001**
Training−Duration compressed 50%	**EqualB**	**1**	**1.44e+00**	**0.402**	**3.573**	**0.001**
Duration compressed 50%−Duration stretched 50%	EqualB	1	−4.05e+05	0.344	0.000	1.000
Training−Duration stretched 50%	**UnequalA**	**1**	**9.58e-01**	**0.412**	**2.325**	**0.020**
Training−Duration compressed 50%	**UnequalA**	**1**	**1.88e+00**	**0.407**	**4.620**	**<0.0001**
Duration compressed 50%−Duration stretched 50%	**UnequalA**	**1**	**−9.20e-01**	**0.357**	**−2.582**	**0.015**
Training−Duration stretched 50%	**UnequalB**	**1**	**1.44e+00**	**0.391**	**3.693**	**0.001**
Training−Duration compressed 50%	**UnequalB**	**1**	**8.93e-01**	**0.395**	**2.259**	**0.036**
Duration compressed 50%−Duration stretched 50%	UnequalB	1	5.50e-01	0.352	1.563	0.118
Pairwise comparisons between training sounds: series 2						
Duration stretched 20%	EqualA−EqualB	2	4.27e-01	0.379	1.127	0.847
Duration stretched 20%	UnequalA−UnequalB	2	−3.22e-01	0.403	−0.800	0.847
Duration compressed 20%	EqualA−EqualB	2	−7.40e-02	0.385	−0.192	1.000
Duration compressed 20%	UnequalA−UnequalB	2	−9.20e-06	0.416	0.000	1.000
Pairwise comparisons between test stimuli: series 2						
Training−Duration stretched 20%	EqualA	2	4.88e-01	0.406	1.201	0.344
Training−Duration compressed 20%	EqualA	2	5.62e-01	0.404	1.390	0.344
Duration compressed 20%−Duration stretched 20%	EqualA	2	−7.39e-02	0.385	−0.192	0.848
Training−Duration stretched 20%	**EqualB**	**2**	**1.32e+00**	**0.426**	**3.087**	**0.006**
Training−Duration compressed 20%	EqualB	2	8.89e-01	0.433	2.053	0.060
Duration compressed 20%−Duration stretched 20%	EqualB	2	−4.27e-01	0.379	−1.127	0.260
Training−Duration stretched 20%	UnequalA	2	7.05e-01	0.427	1.650	0.297
Training−Duration compressed 20%	UnequalA	2	3.83e-01	0.440	0.870	0.424
Duration compressed 20%−Duration stretched 20%	UnequalA	2	3.22e-01	0.403	0.800	0.424
Training−Duration stretched 20%	UnequalB	2	6.11e-01	0.458	1.334	0.273
Training−Duration compressed 20%	UnequalB	2	6.11e-01	0.458	1.334	0.273
Duration compressed 20%−Duration stretched 20%	UnequalB	2	−1.27e-05	0.416	0.000	1.000

Response variables in GLMMs: the proportion of trials that birds respond with pecking A or B. Each of the corrected pairwise multiple comparison tests is separated into clusters within the table. Bold indicates significance.

Table A4
Estimates and 95% confidence intervals for the recognition of Time-scaled stimuli

Training_Group	Stimuli	Series	emmean	SE	CL (95%)	
					Lower	Upper
Log(Cor/Inco)~Training_Group+Stimuli+Training_Group:Stimuli+Training_Sound+Training_Sound:Stimuli+(1|BirdID), (Sound A & Sound B combined)						
Pairwise comparisons: series 1						
Equal	**Duration stretched 50%**	**1**	**0.852**	**0.240**	**0.381**	**1.32**
Equal	**Duration compressed 50%**	**1**	**0.763**	**0.227**	**0.318**	**1.21**
Unequal	**Duration stretched 50%**	**1**	**1.050**	**0.259**	**0.541**	**1.56**
Unequal	**Duration compressed 50%**	**1**	**0.629**	**0.229**	**0.181**	**1.08**
Pairwise comparisons: series 2						
Equal	**Duration stretched 20%**	**2**	**1.21**	**0.277**	**0.668**	**1.75**
Equal	**Duration compressed 20%**	**2**	**1.41**	**0.286**	**0.849**	**1.97**
Unequal	**Duration stretched 20%**	**2**	**1.480**	**0.255**	**0.985**	**1.98**
Unequal	**Duration compressed 20%**	**2**	**1.520**	**0.253**	**1.023**	**2.02**

Lower CL and Upper CL represent the lower and upper 95% confidence limits of the confidence interval. If zero is part of that confidence interval, the treatment combination Training_Sound and Stimuli are not significantly different from 0. If both confidence levels are positive, then there is a bias towards correct responses. If they are both negative, then they are more biased towards incorrect responses. Each of the corrected pairwise multiple comparison tests is separated into clusters within the table. Bold indicates significance.

Table A5
Estimates and 95% confidence intervals for the correct identification of time-scaled stimuli

Training_Sound	Stimuli	Series	emmean	SE	CL (95%)	
					Lower	Upper
Log(Cor/Inco)~Training_Group+Stimuli+Training_Group:Stimuli+Training_Sound+Training_Sound:Stimuli+(1|BirdID), (Sound A & Sound B separated)						
Pairwise comparisons: series 1						
EqualA	**Duration stretched 50%**	**1**	**1.192**	**0.351**	**0.5049**	**1.8794**
EqualA	**Duration compressed 50%**	**1**	**1.024**	**0.314**	**0.4086**	**1.6400**
EqualB	Duration stretched 50%	1	0.511	0.288	−0.0539	1.0757
EqualB	Duration compressed 50%	1	0.501	0.288	−0.0640	1.0655
UnequalA	**Duration stretched 50%**	**1**	**1.804**	**0.404**	**1.0117**	**2.5962**
UnequalA	Duration compressed 50%	1	0.352	0.304	−0.2433	0.9474
UnequalB	Duration stretched 50%	1	0.295	0.285	−0.2627	0.8536
UnequalB	**Duration compressed 50%**	**1**	**0.906**	**0.304**	**0.3105**	**1.5023**
Pairwise comparisons: series 2						
EqualA	**Duration stretched 20%**	**2**	**1.535**	**0.396**	**0.7592**	**2.3108**
EqualA	**Duration compressed 20%**	**2**	**1.609**	**0.412**	**0.8019**	**2.4169**
EqualB	**Duration stretched 20%**	**2**	**0.886**	**0.345**	**0.2109**	**1.5619**
EqualB	**Duration compressed 20%**	**2**	**1.209**	**0.355**	**0.5127**	**1.9051**
UnequalA	**Duration stretched 20%**	**2**	**1.359**	**0.333**	**0.7059**	**2.0112**
UnequalA	**Duration compressed 20%**	**2**	**1.514**	**0.339**	**0.8483**	**2.1790**
UnequalB	**Duration stretched 20%**	**2**	**1.609**	**0.351**	**0.9199**	**2.2974**
UnequalB	**Duration compressed 20%**	**2**	**1.525**	**0.341**	**0.8578**	**2.1928**

Lower CL and Upper CL represent the lower and upper 95% confidence limits of the confidence interval. If zero is part of that confidence interval, the treatment combination Training_Sound and Stimuli are not significantly different from 0. If both confidence levels are positive, then there is a bias towards correct responses. If they are both negative, then they are more biased towards incorrect responses. Bold indicates significance.
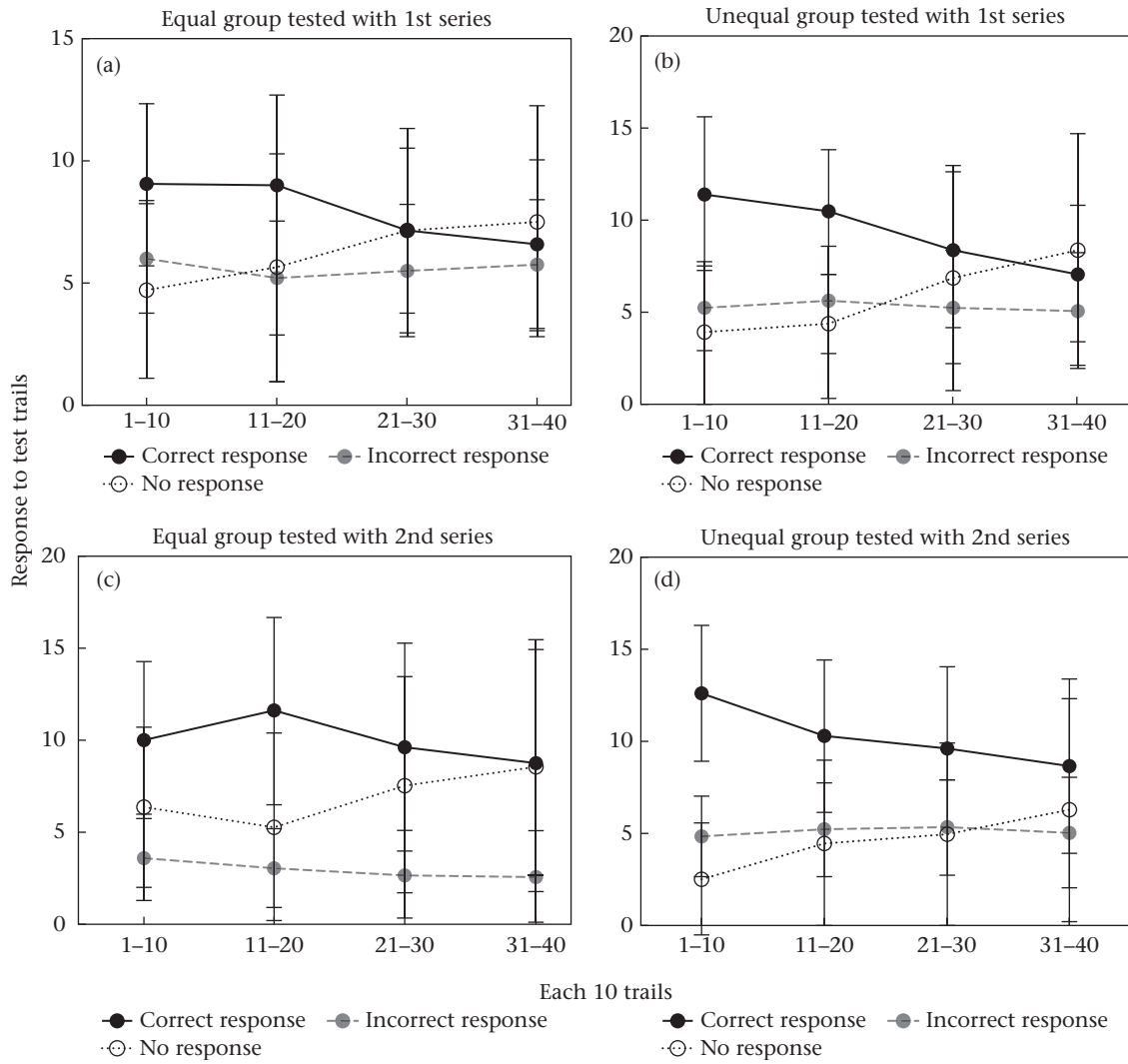
Figure A1. Counts of birds' responses to the test stimuli during the test phase. (a) Trials of Equal-duration group and (b) Unequal-duration group responding to the first series of test stimuli; (c) trials of Equal-duration group and (d) Unequal-duration group responding to the second series of test stimuli. The 40 test trials were divided into four 10-trial blocks. Lines across four 10-trial blocks refer to three categories of reaction to a sound: the 'correct response'; the 'incorrect response'; and a 'nonresponse'.