The flexible listener: exploring zebra finch sensitivity to spectral and temporal sound features

Ning, Z.

Citation

Ning, Z. (2024, May 2). *The flexible listener: exploring zebra finch sensitivity to spectral and temporal sound features*. Retrieved from https://hdl.handle.net/1887/3750255

The Flexible Listener

Exploring zebra finch sensitivity to spectral and temporal sound features

Zhi-Yuan Ning
寧致遠

Ning, Zhi-Yuan

The Flexible Listener:
Exploring zebra finch sensitivity to spectral and temporal sound features

PhD thesis, Leiden University, the Netherlands

Cover design and lay-out by Zi-Yang Ning
Inspired by Zhi-Yuan Ning's thesis
Printed by Ridderprint, Ridderkerk

Dit onderzoek werd gefinancierd door de Chinese Scholarship Council

The Flexible Listener

Exploring zebra finch sensitivity to spectral and temporal sound features

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 2 mei 2024
klokke 11:15 uur

door

Zhi-Yuan Ning
geboren te Canton, China

Promotores

Prof. dr. C.J. ten Cate
Prof. dr. H. Honing

Promotiecommissie

Prof. dr. A.H. Meijer
Prof. dr. H.W. Slabbekoorn
Dr. M. Hoeschele (Universitat Wien, Austria)
Dr. S. Moorman (University of Groningen)
Prof. dr. K.R.L. Janmaat (Leiden University/ University of Amsterdam)

CONTENTS

General Introduction

Chapter 1

Chapter 1

In one of his monumental books, “The Descent of Man, and Selection in Relation to Sex”, Charles Darwin (1871) wrote that “the difference in mind between man and the higher animals, great as it is, is certainly one of degree and not of kind”. This perspective has served as a cornerstone in comprehending the intricate complexities of auditory perception. Darwin also noted: “The sounds uttered by birds offer in several respects the nearest analogy to language” (Darwin, 1871), referring to shared cognitive traits in communication among human beings and birds. Darwin’s idea that bird vocalizations offer an analogy to human language was based on his keen observations of the complexity and functionality of bird songs, the adaptability and diversity in vocalizations, and the similarity between the process of birdsong acquisition and human language development through exposure and imitation. This assertion aligns with the historical understanding that many animals rely on conspecific vocal signals for communication, which have evolved to convey specific information essential for, among others, mate attraction, individual identification, and resource defence. Such vocal signals are processed through species-specific auditory systems, enabling organisms to actively seek meaningful and relevant information from their environment, as envisioned by the concept of the “Umwelt” (von Uexküll, 1992). Noticeably, while each species possesses its own specific “umwelt”, there are also similarities in auditory perception between humans and non-human animals, even though the auditory capabilities of non-human animals may not fully equate to the auditory perceptual abilities as observed in the human recognition of acoustic variation in language or music.

Both vocal communications through language and music perception constitute two of the highest-level cognitive skills evident in humans. All humans (independent of their culture, region, preferences, etc), have a predisposition for music, just as we have for language. Historically, it is well known that many animals use vocal signals to communicate, some of which sound highly musical to humans (e.g., the sophisticated songs of humpback whales or melodious songs of songbirds). While animal vocalizations often share features of melodic and rhythmic characteristics of what we regard as *music*, it’s not easy to infer that animals possess the capacity for music. A more informative strategy is to identify which of the traits that enable humans to make or appreciate music we share with other animals, i.e., to examine their musicality. Musicality can be defined as a natural, spontaneously developing set of traits based on and constrained by our cognitive abilities and their underlying biology (Honing *et al.*, 2015). Similar to the language faculty, the human music faculty encompasses a suite of perceptual and cognitive abilities, some shared with nonhuman animals and some distinctively human (reviewed by Honing *et al*., 2015). Given that some animal vocalizations, in particular bird songs, are also characterized by spectral and temporal complexity as well as rhythmic patterns, some constituent cognitive components of musicality, such as relative pitch, tonal encoding, beat perception, and metrical encoding of rhythm, may be shared between humans and other species (Hoeschele *et al*., 2015). The study of musicality transcends the debate regarding whether animal vocalizations should be considered music, language, or both. Instead, it focuses on understanding the perceptual and cognitive mechanisms used to interpret sounds that may be deemed *musical* or *linguistic*. Drawing inspiration from Darwin's insights that the distinctions in mental traits between humans and animals exist on a continuum rather than being absolute, these subcomponents may have diverse evolutionary histories, and similar components or precursors may hence be present in other species. Conducting comparative research has the potential to reveal such similarities and thus provide insights into the evolutionary background of human language and musicality. Therefore, studying the mechanisms underlying the auditory perception of songbirds from a comparative perspective will be a valuable contribution for revealing shared perceptual abilities across species and offers insights into the evolution of human language and musicality.

Avian Model for Auditory Research

Songbirds are one of the most relevant groups for comparative language and speech research. Like speech, birdsong is characterized by the rapid production of acoustically varying syllables. Unlike the vocalizations in many other groups of animals, bird songs are learned from a tutor and, when acquiring their song, many songbird species go through similar phases to human infants learning language (Doupe & Kuhl, 1999; Bolhuis *et al*., 2010). As a widely used model species, studies on the zebra finch provide valuable insights into the intricate processes of vocal learning (e.g., Arnold, 1975; Boehner, 1983; Clayton, 1988; Zann, 1990; Mello, 2014; Hyland Bruno *et al*., 2021), rhythm detection (e.g., Nagel *et al*., 2010; van der Aa *et al*., 2015; Benichov *et al*., 2016; ten Cate *et al*., 2016; Lampen *et al*., 2019; Rouse *et al*., 2021), and the processing of complexly structured auditory stimuli (e.g., Okanoya & Dooling, 1990a; Okanoya & Dooling, 1990b; Uno *et al*., 1997; Lohr & Dooling, 1998; Dent *et al*., 2008; Osmanski *et al*., 2009; van Heijningen *et al*., 2009; Spierings & ten Cate, 2014; Spierings & ten Cate, 2016; Chen *et al*., 2016; Spierings *et al*., 2017). Like other social songbirds, zebra finches employ a diverse range of vocalizations, including song and other types, to communicate with conspecifics, with songs being particularly noteworthy for their acoustic complexity and consistent imitation within their vocal repertoire (Elie & Theunissen, 2020). A quantitative analysis of the zebra finch's complete vocal repertoire (Elie & Theunissen, 2016) revealed that vocalization types are primarily categorized based on the shape of the spectral envelope, attributed to formants produced by the syrinx and vocal tract, indicating that dynamic vocal tract shaping is not unique to humans or a few mammals. Zebra finch vocalization types were found to exhibit distinctions in spectral shape, pitch saliency, duration, and intensity, with their spectral shape primarily distinguishing vocalizations in various behavioural contexts, and pitch saliency further differentiating noisy calls from tonal or harmonic sounds (Elie & Theunissen, 2016). Surprisingly, despite being more distantly related to humans than other mammals, songbirds, including zebra finches, share striking similarities with humans in their perception of these acoustic features (ten Cate, 2018). Utilizing this model species in investigating both spectral and temporal features in auditory stimulus processing hence has the potential to significantly contribute to current theories regarding the biological foundations and origins of human speech processing and musicality, particularly in relation to spectral and rhythm perception. This kind of research aligns with the broader focus of comparative research in this context, which has primarily centred on songbirds' perception of two crucial aspects: spectral attributes (e.g., pitch, timbre/harmonics) and temporal attributes (e.g., tempo/rhythm).

Perception of Pitch & Spectral Attributes

Humans and birds share interesting similarities with regard to their auditory processing (Hoeschele, 2017). For example, humans and European starlings have similar frequency sensitivity, can perceive the pitch of the missing fundamental, and parse multiple pure-tone sequences into separate auditory streams (e.g., Hulse *et al.*, 1984; Hoeschele, 2017). Given these similarities, it is surprising to find a major difference in how humans and birds perceive sequences of tones. Humans readily recognize tone sequences that are shifted up or down in log frequency because the pattern of relative pitches is maintained (referred to as relative pitch). In contrast, birds were assumed to have a strong bias to rely on the absolute pitch for the recognition of tone sequences - a pitch-shifted melody seems to be perceived as an altogether different melody (Hulse *et al.*, 1984). However, starlings can maintain the discrimination between two songs shifted in frequency (Bregman *et al.*, 2012). It is unclear what causes this difference in responses between shifts in artificial tone sequences and in songs. Bregman *et al.*, (2016) propose that the perception of melodic sequences in songbirds relies on a perceptual representation that appears more closely tied to the spectral envelope rather than absolute pitch cues. Starlings, for instance, rely mainly on a perception of each tone based on its spectral envelope (the shape of the spectrum, formed by the relative amplitudes of the different frequency components) rather than abstracted features derived from the fundamental frequency (absolute pitch) or on the relative power in the harmonics (timbre) (Bregman *et al.*, 2016). This led to the hypothesis (Bregman *et al.*, 2016) that the spectral envelope governs avian tone sequence recognition: for pure tones, the spectral band envelope corresponds directly to pitch; for complex tones, the spectral band envelope contributes to both pitch and timbre perceptions. Noticeably, spectral envelope is not the only attribute that the birds attend to in auditory discrimination. A previous study from our own lab showed that zebra finches can discriminate artificial vowel-like elements differing in pitch and harmonic spectrum using either of these while ignoring the other, and it also showed that zebra finches can generalize a harmonic spectrum to a vocoded version (Burgering *et al.*, 2018). The findings of Burgering *et al.* (2019) and Bregman *et al.* (2016) suggest that while zebra finches demonstrate a capacity to detect pitch variations in songs, this sensitivity may not necessarily reflect an inherent pitch sensitivity but rather could be attributed to their sensitivity to the spectral envelope. Notably, this aspect has so far not been examined specifically in the context of song stimuli. Therefore, the perception of pitch and spectral attributes by zebra finches is one of the central topics addressed in this thesis.

Beat Detection & Perception of Temporal Regularities

Humans can easily detect the beat in music, perceive regularity in a series of pulses, and recognize melodies as being similar despite differences in the speed of performance. Although it was assumed that animals had similar abilities (Darwin, 1871), this has long remained untested. The question of whether animals can detect regularity in a stimulus got sudden attention with the discovery of Snowball, a Sulphur-crested cockatoo (*Cacatua galerita*) that could entrain head and body movements with the beat of several popular songs (Patel *et al.*, 2009). Parrots, such as Snowball, are vocal learners, and vocal learning is associated with evolutionary modifications to the basal ganglia, which play a key role in mediating a link between auditory input and motor output during learning (Petkov & Jarvis, 2014). However, other studies have revealed that this issue is also one with many open questions and have questioned the link between vocal learning and beat perception (ten Cate *et al.*, 2016; Wilson & Cook, 2016; Celma-Miralles & Toro, 2020). They suggest the presence of a graded scale for beat perception in avian species (ten Cate *et al.*, 2016), with some species (including the zebra finch) attending more strongly to local features of the individual stimuli rather than the overall regularity of the stimuli (which is a main feature human listeners attend to, e.g., van der Aa *et al*, 2015).

The perception of temporal regularity is one of the basic features of musicality. Yet, the current evidence for detecting pattern regularities in zebra finches seems ambiguous: a study using single-tone pulse strings as rhythmic stimuli showed that discrimination between isochronous and heterochronous stimuli disappeared with a 25% tempo change (van der Aa *et al*, 2015), and another study using more complex two-tone pulse strings as rhythmic stimuli found that some discrimination was maintained with a 25% shortening, but not with a 25% lengthening of element and pause durations (ten Cate *et al.*, 2016). In contrast to this sensitivity to tempo in tonal strings, zebra finches showed great tolerance for changes in song duration (61%~164%) in a song discrimination study by Nagel *et al.* (2010). The results of van der Aa *et al.* (2015) suggest the birds attended only to local temporal features (e.g., the exact duration of inter-onset intervals). By contrast, those of Nagel *et al.* (2010) indicate that birds do attend to somewhat global rhythmic features and might have used predominantly frequency or intensity cues during song categorization. The local feature bias hypothesis (ten Cate *et al.*, 2016) might provide a framework to interpret the discrepancies among those studies, which suggests a preference in birds for local temporal features in perception and discrimination tasks with simple stimuli and a lower sensitivity when these features are part of a spectrally more complex structure.

Cognitive Flexibility?

A possible explanation for the, at times, contradictory findings in avian cognition concerning the relevance of various spectral and temporal features for auditory discrimination might be the presence of cognitive flexibility. This cognitive flexibility involves the mental ability to adapt to changing conditions, switch between different tasks, or adjust one’s strategies in response to new information. Zebra finches appear to demonstrate this flexibility in their perceptual strategy, as demonstrated in the study by Burgering *et al.* (2018), where they could use either the pitch or the spectral envelope, depending on which is most relevant to succeed in the auditory task. What the different studies suggest is that birds can use pitch or other spectral features of acoustic stimuli (e.g., formant, spectral envelope, spectral centroid, etc.) to identify different stimuli. Despite the insights from Burgering *et al.* (2018), no study has systematically investigated how zebra finches invoke pitch or other spectral attributes, such as harmonicity,, spectral envelope, or other relevant spectral features, to differentiate between two tonal sequences or conspecific vocalizations in which these attributes have been manipulated In addition to the perceptual flexibility (i.e., the ability to perceive and interpret information from the environment in a flexible manner, including the capacity to attend to and process different sensory cues or modalities) that zebra finches demonstrate for spectral features, the apparently contradictory findings regarding the sensitivity to temporal features might also be explained by similar flexibility in songbirds' attention to various temporal features.

An open question is whether, and to what extent, the birds' attention to spectral attributes/features and temporal parameters/patterns depends on the differences between the training stimuli. If zebra finches can flexibly adjust their perceptual strategies to accommodate varying parameters or patterns in auditory stimuli, how do these parameters or patterns relate to each other? And are there biases in the attention and preference among zebra finches in employing cues such as spectro-temporal fine structure, temporal pattern, and spectral structure? Additionally, the presence of a graded scale for beat perception in avian species (ten Cate *et al.*, 2016) may indicate that the spectrally rich structure of songs, containing modulations of pitch and spectral contour, may interact with (or overshadow) the attention given to temporal changes. So, as for the perception in the frequency domain, a critical question for experimental studies on the interplay of spectral and temporal perception in zebra finches is how the nature of the stimuli affects the birds' discrimination. This focus leads to questions regarding how the perception of auditory stimuli is affected by stimulus complexity. This thesis has the ambition to fill critical gaps in our understanding of songbird auditory cognition by investigating the role of spectral and temporal features in zebra finches' discrimination of songs and other tonal sequences and spectrally complex stimuli. The research is expected to be useful for understanding the interplay between various spectral and temporal features in songbirds' auditory perception and to provide insights into the adaptability of songbirds' perceptual strategies across diverse acoustic contexts.

Thesis Outline

This thesis consists of 4 chapters of empirical research, addressing questions from song discrimination and song preferences to the perceptual interplay of specific acoustic features/patterns. These questions were addressed by using a well-established experimental paradigm, the Go-left/Go-right operant task, which was employed across three experimental chapters (**Chapter 2**, **Chapter 4**, and **Chapter 5**) to investigate the cues utilized by birds in discriminating natural songs and artificial stimuli. In this task, the birds have to learn to peck a central sensor to trigger a sound and then choose either the left or right sensor, with correct choices rewarded with food and incorrect ones resulting in a brief light-off as negative feedback. The birds are initially trained to discriminate between a pair of sounds without any constraints on the cues used for identification. Subsequently, they are tested with novel probe stimuli in which specific cues have been altered. Importantly, probe stimuli are presented without any reinforcement linked to the bird's choice, thus preventing the bird from learning a predetermined "correct" response through the reward/punishment pattern. Successful discrimination between probe stimuli derived from different training stimuli, with performance significantly above chance, serves as evidence of the bird's ability to recognize a modified version as being derived from a specific training stimulus. By offering sets of acoustic stimuli and letting the bird choose which to attend, this methodology enables the determination of the features that birds use to identify complex stimuli within a given context. Notably, this paradigm facilitates the presentation of multiple test stimuli within a short time frame during the test phase, can be applied to both male and female birds using a uniform approach, and allows for a clear distinction between incorrect responses (choosing the wrong answer), which may indicate that a test stimulus has modified a song feature in such a way that it now resembles the opposite training stimulus more than the original one or a failure to discriminate the probe sound, and a simple lack of response (no response), which could suggest factors such as confusion, fatigue, indifference, or inattention. **Chapter 3**, serving as a supplementary preference assessment following the operant task in **Chapter 2**, focuses on comparing the preferences for heterospecific songs and modified conspecific songs with those for natural conspecific songs.

Chapter 2 explores how zebra finches attend to spectral and temporal variation in recognizing conspecific song motifs. This chapter systematically examines the importance of spectral and temporal parameters when zebra finches have to discriminate two natural songs, which are either similar or different in their duration. Moreover, this chapter discusses the results from a discrimination task using an operant conditioning paradigm (the Go-left/Go-right task) in which birds are trained to respond to one type of sound for a food or water reward and to not, or to respond differently to another sound, with errors resulting in consequences such as brief periods of darkness or no reward, examining which sound features birds use for discrimination. The controlled experiment examined how the difference in song duration affects how zebra finches perceive and discriminate natural motifs of their conspecific songs, their noise-vocoded version in which the pitch was removed but the spectral envelope was maintained, as well as other modified versions varying in spectral (pitch, or frequency spectrum) and temporal features (duration/tempo). Through a comprehensive analysis, this chapter elucidates the types of parameters birds can leverage and hence the information that birds may extract from vocalization.

Chapter 3 examines the song preferences exhibited by both male and female zebra finches. Building upon the findings of **Chapter 2**, which explored the birds' perceptual sensitivity to various acoustic features in an operant discrimination task, **Chapter 3** deals with the question: does the low response rate to specific novel stimuli in the discrimination tasks arise from these stimuli being perceived as too different from the training songs or from a very low attractiveness of these stimuli? To investigate this, a 4-way operant choice test (referred to as the "carrousel") was used to measure birds' preferences. In this setup, birds could perch on different operant perches within the carrousel setup, thereby triggering the playback of four different songs. This chapter explores the birds' preference for different song types, including normal, duration-stretched, and vocoded conspecific songs, as well as heterospecific songs. This chapter also examines the presence or absence of a sex difference in song preference.

Chapter 4 addresses the question of how learning about the spectro-temporal structure and sequential order of song syllables relate to each other. The Go-left/Go-right task was once again used in this chapter to directly compare the birds' relative sensitivity to attend to spectro-temporal features and syllable sequence for song discrimination. Birds were either trained to discriminate between two song-syllable strings that consisted of identical syllables or trained to discriminate between two strings containing different syllables. This chapter examines whether zebra finches exhibit cognitive flexibility in their ability to attend to sequential and spectro-temporal features, depending on the salience of the differences between two auditory stimulus strings during discrimination learning.

Chapter 5 examines the zebra finches' sensitivity to pitch and formant patterns, two fundamental features crucial for human speech recognition and musical perception. It examines whether there is an interplay between pitch and formant when both vary between stimuli and which of these two spectral parameters is most salient to zebra finches. To investigate this, the stimuli used in this chapter were sound sequences consisting of five artificial elements, separated by brief pauses. These artificial elements featured either simultaneous pitch and formant contour ascending and descending in the same direction over a full sequence or opposite directions over a full sequence. By employing the Go-left/Go-right paradigm once more, this chapter examined which of these two spectral attributes is more important in recognition of artificial tone sequences and whether the way in which pitch and formant are combined in the training stimuli affects the ease of learning and discrimination of the test stimuli.

Chapter 6 contains a synthesis of the findings obtained across the four experimental chapters. It summarizes the findings of the previous chapters and, discusses them, offering overarching conclusions regarding the diverse cues involved in songbirds' auditory perception. Collectively, these findings offer a comprehensive perspective on auditory cognition in zebra finches. This chapter delves into the broader implications of the main conclusions of the thesis and explores how this research enhances our understanding of songbirds' perceptual flexibility with respect to the effect of the training context. Furthermore, this concluding chapter highlights potential research directions for future studies concerning the cognitive flexibility of songbirds.

References

Arnold A.P. (1975). The effects of castration on song development in zebra finches (*Poephila guttata*). *J. Exp. Zool.*, 191(2): 261–278. https://doi.org/10.1002/jez.1401910212

Benichov J.I., Benezra S.E., Vallentin D., Globerson E., Long M.A., Tchernichovski O. (2016). The Forebrain Song System Mediates Predictive Call Timing in Female and Male Zebra Finches. *Curr Biol.*, 26(3), 309–318. https://doi.org/10.1016/j.cub.2015.12.037

Boehner J. (1983). Song learning in the zebra finch (*taeniopygia guttata*): Selectivity in the choice of a tutor and accuracy of song copies. Anim. Behav., 31(1): 231-237. https://doi.org/10.1016/S0003-3472(83)80193-6

Bolhuis J.J., Okanoya K., Scharff C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.*, 11(11): 747–759. https://doi.org/10.1038/nrn2931

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. *Proc Natl Acad Sci.*, 113(6): 1666–1671. https://doi.org/10.1073/pnas.15153801

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

Celma-Miralles A., Toro J.M. (2020). Non-human animals detect the rhythmic structure of a familiar tune. *Psychon Bull Rev.*, 27(4): 694–699. https://doi.org/10.3758/s13423-020-01739-2

Chen J., Jansen N., ten Cate C. (2016). Zebra finches are able to learn affixation-like patterns. *Anim Cogn.*, 19(1): 65–73. https://doi.org/10.1007/s10071-015-0913-x

Clayton N.S. (1988). Song learning and mate choice in estrildid finches raised by two species. *Anim. Behav.*, 36(6), 1589-1600. https://doi.org/10.1016/S0003-3472(88)80101-5

Darwin C.R. (1871). The descent of man, and selection in relation to sex. London: John Murray. Volume 1. 1st edition

Dent M.L., Welch T.E., McClaine E.M., Shinn-Cunningham B.G. (2008). Species differences in the identification of acoustic stimuli by birds. *Behav. Processes*, 77(2): 184–190. https://doi.org/10.1016/j.beproc.2007.11.005

Doupe A.J., Kuhl P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.*, 22: 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567

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

Elie J.E., Theunissen F. (2020). The neuroethology of vocal communication in songbirds: Production and perception of a call repertoire. In The neuroethology of birdsong (pp. 175–209). New York, NY: Springer. Volume 71. https://doi.org/10.1007/978-3-030-34683-6_7

Hoeschele M., Merchant H., Kikuchi Y., Hattori Y., ten Cate C. (2015). Searching for the origins of musicality across species. *Philos. Trans. Royal Soc. B.*, 370(1664): 1–9. https://doi.org/10.1098/rstb.2014.0094

Hoeschele M. (2017). Animal Pitch Perception: Melodies and Harmonies. *Comp Cogn Behav Rev.*, 12: 5–18. https://doi.org/10.3819/CCBR.2017.120002

Honing H., ten Cate C., Peretz I., Trehub S.E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos Trans R Soc Lond B Biol Sci.*, 370(1664): 20140088. https://doi.org/10.1098/rstb.2014.0088

Hulse S.H., Cynx J., Humpal J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *J Exp Psychol Gen.*, 113(1): 38–54. https://doi.org/10.1037/0096-3445.113.1.38

Hyland Bruno J., Jarvis E.D., Liberman M., Tchernichovski O. (2021). Birdsong learning and culture: analogies with human spoken language. *Annu. Rev. Appl. Linguist.*, 7: 449-472. https://doi.org/10.1146/annurev-linguistics-090420-121034

Lampen J., McAuley J.D., Chang S.E., Wade J. (2019). Neural activity associated with rhythmicity of song in juvenile male and female zebra finches. *Behav Processes.*, 163: 45–52. https://doi.org/10.1016/j.beproc.2017.12.003

Lohr B., Dooling R.J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J Comp Psychol.*, 112(1): 36–47. https://doi.org/10.1037/0735-7036.112.1.36

Mello C.V. (2014). The zebra finch, Taeniopygia guttata: an avian model for investigating the neurobiological basis of vocal learning. *Cold Spring Harb Protoc.*, 2014(12): 1237–1242. https://doi.org/10.1101/pdb.emo084574

Nagel K.I.K., McLendon H.M., Doupe A.J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J Neurophysiol.*, 104(3): 1426–37. https://doi.org/10.1152/jn.00028.2010

Okanoya K., Dooling, R.J. (1990a). Detection of auditory sinusoids of fixed and uncertain frequency by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*). *Hear Res.*, 50(1-2): 175–183. https://doi.org/10.1016/0378-5955(90)90043-o

Okanoya K., Dooling, R.J. (1990b). Detection of gaps in noise by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*). *Hear Res.*, 50(1-2): 185–192. https://doi.org/10.1016/0378-5955(90)90044-p

Osmanski M.S., Marvit P., Depireux D.A., Dooling R.J. (2009). Discrimination of auditory gratings in birds. *Hear Res.*, 256(1-2): 11–20. https://doi.org/10.1016/j.heares.2009.04.020

Patel A.D., Iversen J.R., Bregman M.R., Schulz I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr Biol.*, 19(10): 827–830. https://doi.org/10.1016/j.cub.2009.03.038

Petkov C.I., Jarvis E.D. (2014). The basal ganglia within a cognitive system in birds and mammals. *Behav Brain Sci.*, 37(6): 568–604. https://doi.org/10.1017/S0140525X13004160

Rouse A.A., Patel A.D., Kao M.H. (2021). Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proc. Natl. Acad. Sci. U.S.A.*, 118(29): e2026130118. https://doi.org/10.1073/pnas.2026130118

Spierings M.J., ten Cate C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proc Biol Sci.*, 281(1787): 20140480. https://doi.org/10.1098/rspb.2014.0480

Spierings M.J., ten Cate C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proc. Natl. Acad. Sci. U.S.A.*, 113(27): E3977–E3984. https://doi.org/10.1073/pnas.1600483113

Spierings M.J., Hubert J., ten Cate C. (2017). Selective auditory grouping by zebra finches: testing the iambic-trochaic law. *Anim Cogn.*, 20(4): 665–675. https://doi.org/10.1007/s10071-017-1089-3

ten Cate C., Spierings M., Hubert J., Honing H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front Psychol.*, 7: 730. https://doi.org/10.3389/fpsyg.2016.00730

ten Cate C. (2018). The comparative study of grammar learning mechanisms: birds as models. *Curr. Opin. Behav. Sci.*, 21: 13-18. https://doi.org/10.1016/j.cobeha.2017.11.008

Uno H., Maekawa M., Kaneko H. (1997). Strategies for harmonic structure discrimination by zebra finches. *Behav. Brain Res.*, 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. *Behav Processes.*, 115: 37–45. https://doi.org/10.1016/j.beproc.2015.02.018

van Heijningen C.A., de Visser J., Zuidema W., ten Cate C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. U.S.A.*, 106(48): 20538–20543. https://doi.org/10.1073/pnas.0908113106

von Uexküll J. (1992). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, 89(4): 319–391. https://doi.org/10.1515/semi.1992.89.4.319

Wilson M., Cook P.F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon Bull Rev.*, 23(6): 1647–1659. https://doi.org/10.3758/s13423-016-1013-x

Zann R. (1990). Song and call learning in wild zebra finches in south-east Australia. *Anim. Behav.*, 40(5): 811–828. https://doi.org/10.1016/S0003-3472(83)80193-6

The Role of Spectral Features and Song Duration in Zebra Finch Song Recognition

Chapter 2

Zhi-Yuan Ning · Henkjan Honing · Harald van Mil · Carel ten Cate

This chapter is published in Animal Behaviour

(in press)

Chapter 2

ABSTRACT

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 either between 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.

INTRODUCTION

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., Nelson, 1989; Dabelsteen & Pedersen, 1992; Naugler & Ratcliffe, 1992). 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., Kreutzer et al., 1990; MacDougall-Shackleton & Hulse, 1996; Tu & Dooling, 2012; Neilans et al., 2010; Dooling & Prior, 2017). Such studies have provided important insights in 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 (e.g., Hulse et al., 1984; Bregman et al., 2012; Hoeschele, 2017; Dooling & Prior, 2017).

Over the years, the zebra finch (*Taeniopygia guttata*) 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; Lohr et al., 2006; Vernaleo et al., 2010; Vernaleo & Dooling, 2011; Lawson et al., 2018; Prior et al., 2018a; Prior et al., 2018b; Fishbein et al. 2021) examined the salience of various types of song changes on the identification of the target song. They showed that zebra finches are quite sensitive to changes in the spectro-temporal 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., Uno et al., 1997; Vignal & Mathevon, 2011; Geberzahn & Derégnaucourt, 2020; Mol et al., 2021) 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 (±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 discriminating among more complex tone sequences. Bregman et al. (2016) suggest that the spectral envelope governs avian tone sequence recognition. The importance of this feature may long have 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 aims 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 allows us to examine two factors. The first one 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 among 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 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 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 examine is the relevance of the spectral envelope *vs* 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. In this case we have no reason to expect a difference between the Equal-duration and the Unequal-duration groups. However, for each group we 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 to 30 individuals in aviaries (2m × 2m × 1.5m), 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) (70x30x45 cm) containing 3 pecking keys (sensors) with a red LED light at the top/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-h/10.5-h light/dark schedule. Sound stimuli were played through a speaker (Vifa MG10SD09–08; frequency range 100 – 15000 Hz) 1 meter 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.

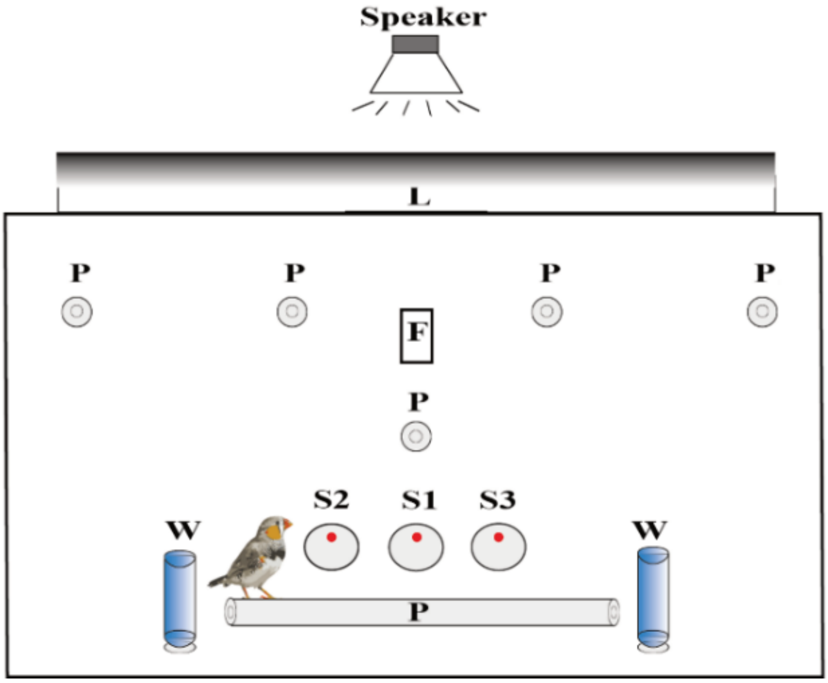


Figure 1. Schematic front view of the operant conditioning apparatus (Skinner box) used for the experiment. A speaker (top of figure) is suspended from the ceiling above the cage. Within the cage, there are several perches (P) for the bird to sit on, a food hatch (F) located in the upper middle of the back panel, 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.

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 will be indicated as 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 will be indicated as 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 380Hz and 22.5kHz. 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). The amplitude of each stimulus was adjusted by using the "Normalize" feature in Audacity (version 2.3.0).

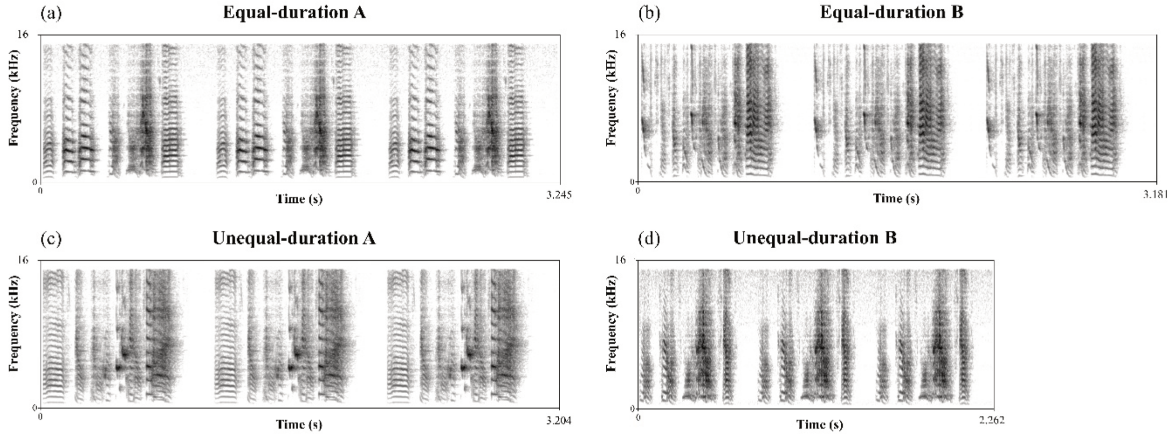


Figure 2. Spectrogram samples of training stimuli. Songs Equal-duration A (a) and Equal-duration B (b) form a pair of training stimuli used in the Equal-duration group, while songs Unequal-duration A (c) and Unequal-duration B (d) form a pair of training stimuli used in the Unequal-duration group.

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 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):

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	+1500Hz	+20%	-20%	+50%	-50%	Spectral envelope maintained but spectral contour averaged
2nd series	+500Hz	+8%	-8%	+20%	-20%	Spectral envelope & spectral contour maintained

Note: Two test series were used for the subgroups of both the Equal-duration and Unequal-duration experimental group of birds. The test stimuli differed in the degree to which pitch and duration were modified, which was 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).

- 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 subgroup of birds tested with series 1 this value was 1500 Hz and for subgroup of birds tested with series 2 this was 500 Hz.

- 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 remain 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.

- 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.

- 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 these two scripts were set to divide cut-off frequency bandwidths equally for 15 bands contiguous with smooth transitions (1000Hz bandwidth for one noise-vocoded band).

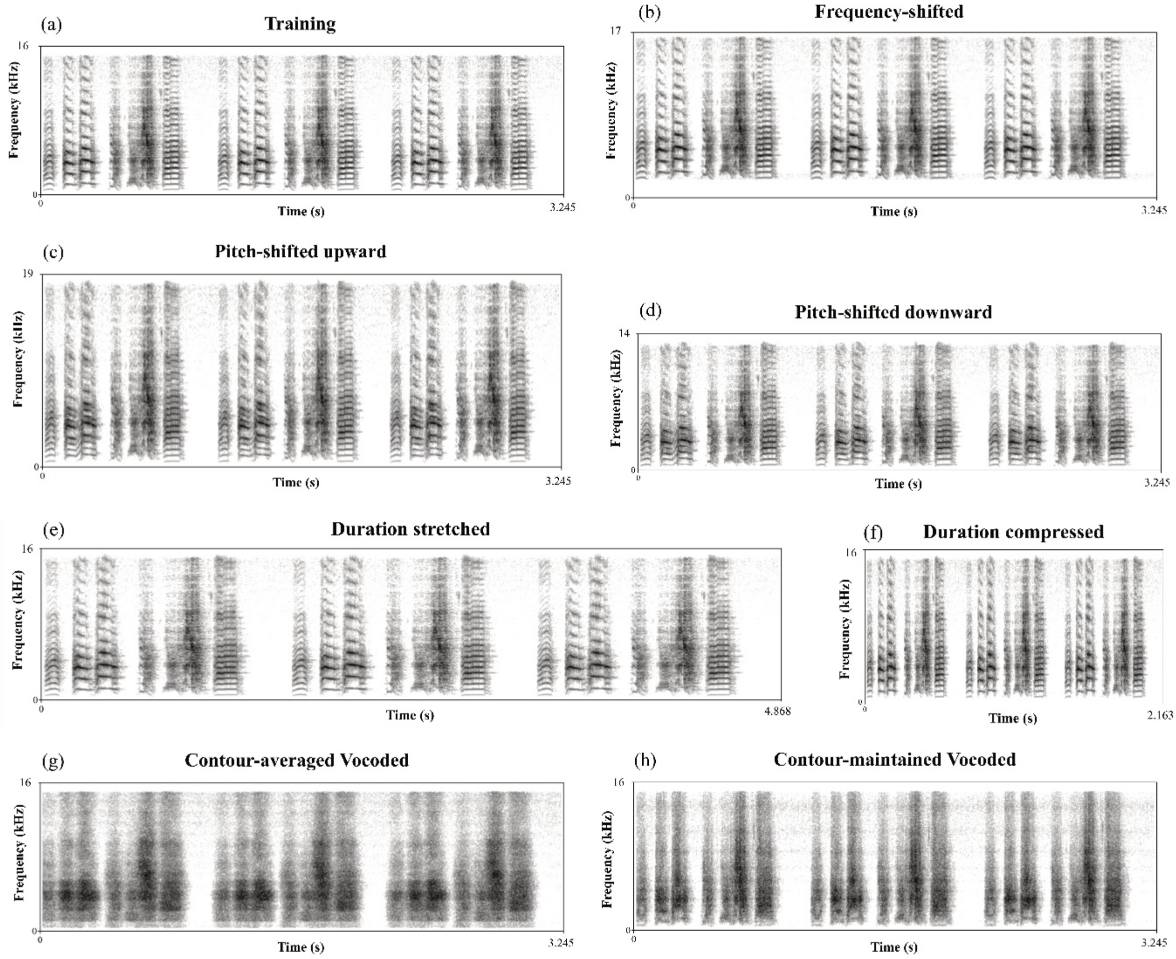


Figure 3. Examples of stimuli used in the test series, showing (a) the **Training stimulus**, and its modified versions. The whole frequency spectrum of **(b)** the **Frequency-shifted** version was shifted upwards by 1500 Hz. The frequency spectrum of the **Pitch-shifted** stimulus was either **(c)** 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 scrips – **(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, pre-training, 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 one day, with a median value of 26 (IQR 18-28) hours, the next phase started by closing the food hatch.

Pre-training 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 seconds), S3 (side sensor) = Sound B + Food hatch open (12 seconds). 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) hours. 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 one 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 sec access to food. If the wrong sensor was pecked, the light was off for three seconds. Before any sensor was pecked, only the S1 LED was on. For example, when playing song A, pecking sensor S2 caused the food hatch to open while pecking sensor S3 resulted in the pre-set dark time, and vice versa. If the bird did not respond within 25 seconds, 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 three to one seconds, the food acquisition time from the initial stage of 12 seconds to the later stage of 10 seconds. The duration of this phase varied from bird to bird, with a median value of 456 (IQR 278-655) hours. The proportion of correct responses out of all sounds that each bird responded to was calculated on a daily basis 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 three consecutive days (general discrimination score >0.75, the accuracy rate of each sensor pecking >0.60 for three 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 two days, the test phase began. The duration of the transition phase had a median value of 47 (IQR 46-50) hours.

Probe testing phase

In this phase, 16 test stimuli were introduced for 20% of the pecks on S1. Twelve of these were novel stimuli (belonging to either series 1 or series 2). The remaining 4 test stimuli were non-rewarded training sounds used as control and were presented twice as often as the other test stimuli. These test stimuli (non-rewarded 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 two to three weeks, with a median value of 394 (IQR 339-549) hours. 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-Wilcoxon 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 didn’t 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 interquartile range (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 Fig. 4 & Fig. 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%’ & ‘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’s methodologies, 2016), utilizing the ‘glmer’ function from the lme4 package (Bates et al., 2015). Model selection was carried out using the Wald Chi-Squared test. Finally, a post-hoc analysis was conducted on the chosen model, incorporating False Discovery Rate (FDR) correction using the emmeans package (Lenth, 2023).

Ethical Note

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 setup 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 (Ning Z) as well as the qualified animal caretaker (in possession of a so-called ‘art.13f2’ qualification – the qualification which the Dutch law requires), 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 hours (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 hours included the 10.5 hours 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 has been 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 setup 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

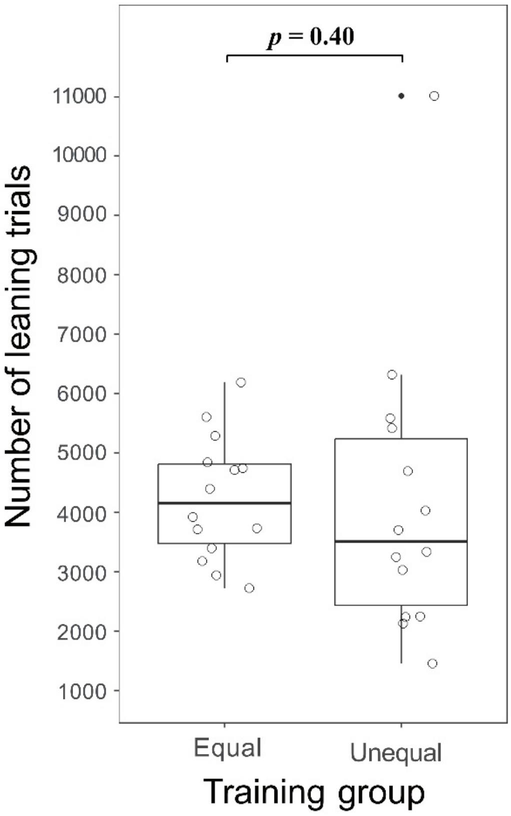


Figure 4. Number of learning trials needed to reach the learning criterion. Individual zebra finch results are shown with open dots. There is no significant difference between the Equal-duration group and the Unequal-duration group. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range.

The discrimination training lasted until a bird reached the learning criterion of over 75% correct responses to both sound A and sound B for consistent three days. All 28 birds finished the training and learned the discrimination on an average of 4209 (SD = 1840, N = 28) trials to reach the criterion. No significant difference ($p = 0.40$, $z = 0.87$; Fig. 4) was found between the Equal-duration group (M = 4243, SD = 1041) and the Unequal-duration group (M = 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 two training groups learn approximately equally fast.

The 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 address 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. However, 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. Finally, we use the same 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 among 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).

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

Variable	**Chisq.**	**Df**	**Pr(>Chisq.)**
① MODEL PC for spectrally-changed (Sound A & Sound B combined in Series 1)			
Training_Group	4.0689	1	0.0437 *
Test_Treatment	95.1957	4	< 2.2e-16 ***
Training_Group:Test_Treatment	11.4205	4	0.0222 *

TrainingTrails_scaled	0.0161	1	0.8990
Test_Treatment:TrainingTrails_scaled	14.1089	4	0.0070 **
② MODEL PC for spectrally-changed (Sound A & Sound B combined in Series 2)			
Training_Group	9.3598	1	0.0022 **
Test_Treatment	101.5756	4	< 2.2e-16 ***
Training_Group:Test_Treatment	9.7748	4	0.0444 *
③ MODEL PC for time-scaled (Sound A & Sound B separated in Series 1)			
Training_Group	0.0053	1	0.9421
Test_Treatment	48.2401	2	3.348e-11 ***
Training_Sound	0.0633	2	0.9689
Training_Group:Test_Treatment	9.0091	2	0.0111 *
Test_Treatment:Training_Sound	10.2016	4	0.0372 *
④ MODEL PC for time-scaled (Sound A & Sound B separated in Series 2)			
Training_Group	1.2867	1	0.25666
Test_Treatment	13.9044	2	0.00096 ***
Training_Sound	0.5413	2	0.76290
Training_Group:Test_Treatment	0.3930	2	0.82159
Test_Treatment:Training_Sound	2.4069	4	0.6613

Note: All variables shown here were used as fixed factors for corresponding models, no matter their p values were significant or not because these were the variables of our interest. 'Bird_ID' was used as the only random factor in all models. • $0.05 < P \leq 0.10$, * $0.01 < P \leq 0.05$, ** $0.001 < P \leq 0.01$, *** $P \leq 0.001$.

Spectral changes affect the Equal- duration group most strongly

Figure 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 500Hz' 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.

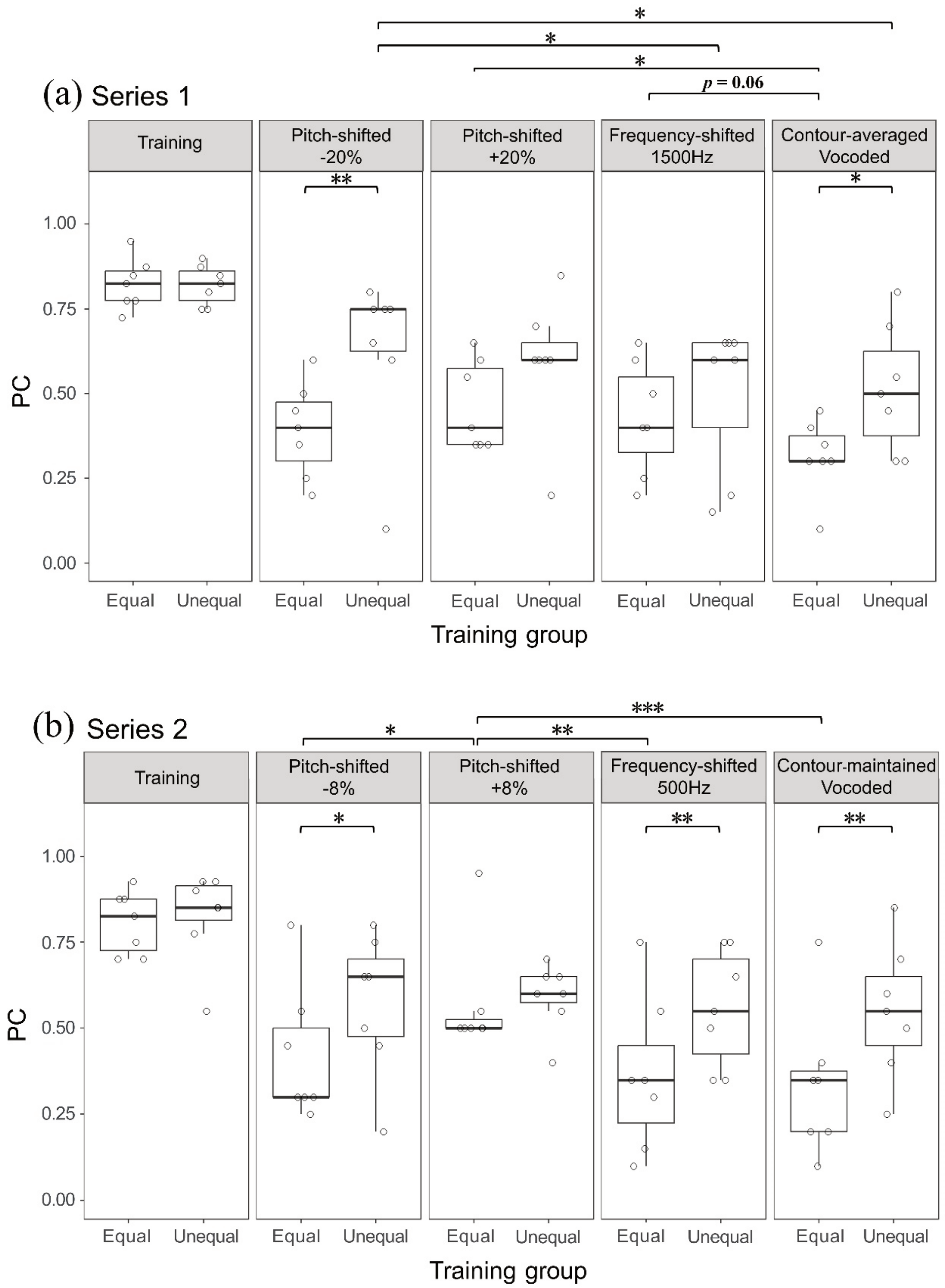


Figure 5. Proportion correct (PC) responses to spectrally-changed stimuli of series 1 (a) and series 2 (b). The significant between-group and within-group differences are indicated, with exception of the differences between the scores for the training stimuli and those for the other test stimuli - all spectrally-changed stimuli got a significantly lower PC than the training stimuli in both two series. *** refers to a significant difference of $P \leq 0.001$, ** refers to a significant difference of $0.001 < P \leq 0.01$, and * refers to a significant difference of $0.01 < P \leq 0.05$, for non-indicated comparisons P value is > 0.05. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range.

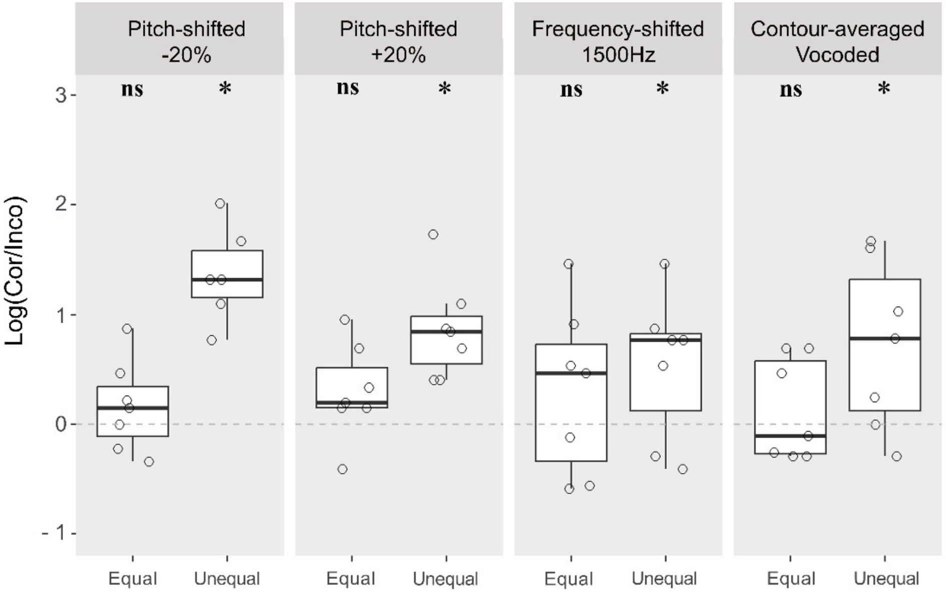
Differences in responses among the test stimuli

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 among 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 1500Hz' 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 1500Hz' (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 500Hz' ($P < 0.01$) & 'Contour-maintained Vocoded' ($P < 0.001$)). For the Unequal-duration group there is no significant difference in PC among the four spectrally-changed stimuli (Fig. 5b). On the whole, these results show a weak tendency that pitch-shifted versions have less impact on the PCs than vocoded versions. This implies that, if anything, the zebra finches are attending more strongly to precise spectral details of the song elements rather than to the spectral envelope. If they would attend more to the latter, vocoding would have 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 two 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 500Hz') 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 are more strongly attending to spectral features than the birds from the Unequal-duration group. In addition, the Unequal group showed a lower degree of recognition when the modifications were stronger (series 1) than when they were less strong (series 2).

(a) Series 1



(b) Series 2

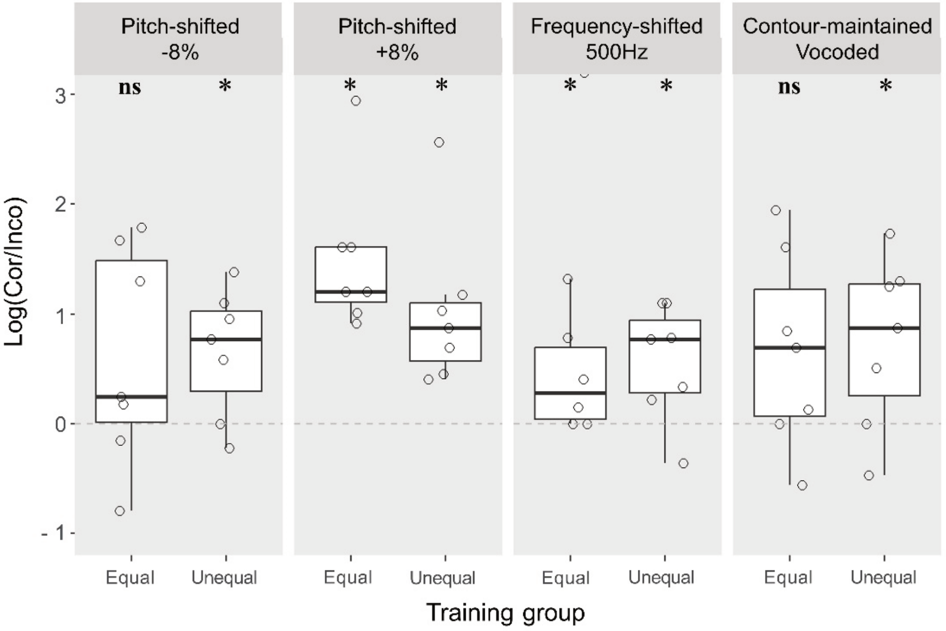


Figure 6. Visualisation of Log (Cor/Inco) for birds responding to spectrally-changed stimuli of series 1 (a) and series 2 (b). A * indicates that the Log(Cor/Inco) of a Test treatment is significantly different from 0; ns indicates that the Log(Cor/Inco) of a Test treatment is overlapping with 0. Box plots show median, 1st and 3rd quartile, and whiskers the 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(Cor/Inco)s was based on the counts of 'correct response' and 'incorrect response' from the same data set that was also used for Fig.5. Note that one bird's data point cannot be fully displayed on the panel (b) because it made no incorrect responses to the 'Frequency-shifted 500Hz' stimuli version, resulting in an infinitely large value after log-scaling.

The effect of duration changes

Tempo changes affect the Equal- and Unequal-duration group 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 if the results for the Time-scaled versions of training stimuli A and B are 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). Figure 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.

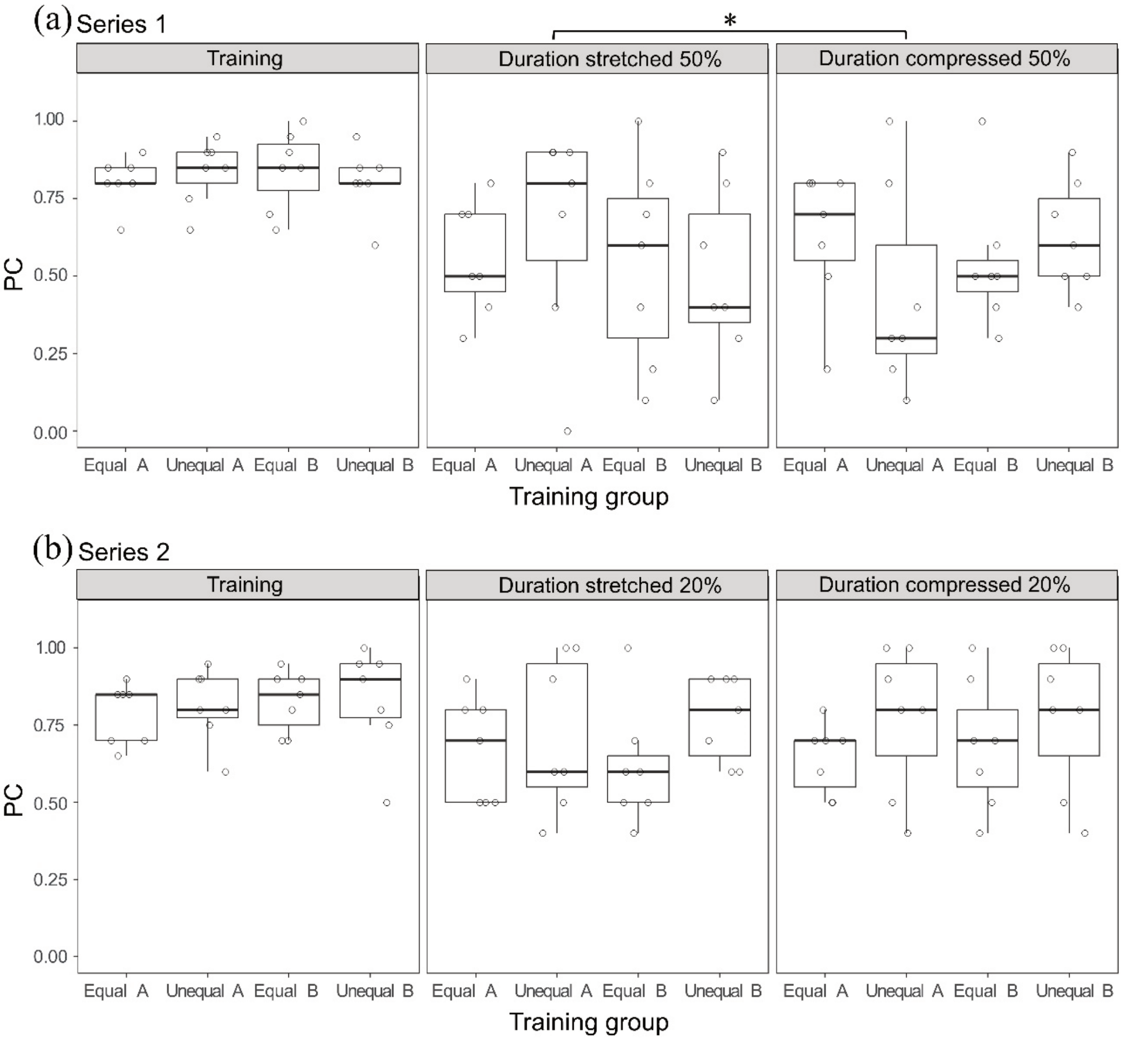


Figure 7. Proportion correct (PC) of responses to Time-scaled stimuli of series 1 (a) and series 2 (b).). For significant differences between Training and Duration-changed stimuli: see text. The only difference among the Duration-changed stimuli was present in series 1: * refers to a significant difference of $0.01 < P \leq 0.05$. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range.

Differences in responses among the test stimuli

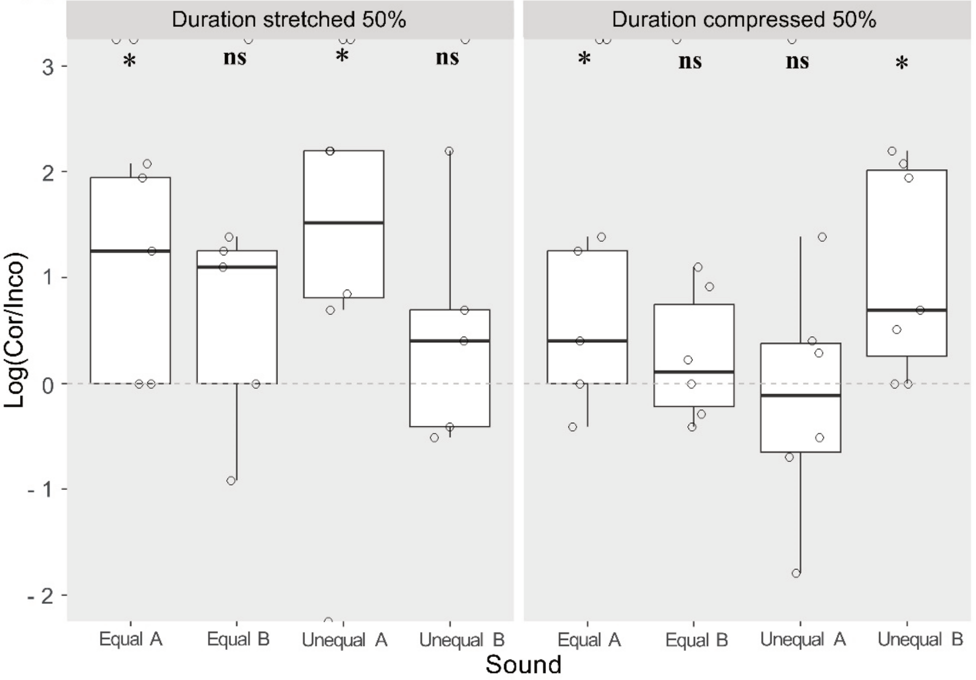
The pairwise comparisons of the PC for the Time-scaled versions of sound A and sound B for the Equal- and Unequal-duration groups are shown in Table A3. In both series 1 and series 2, there is no significant difference in the PC between all three sound versions (Training and two Time-scaled versions) between UnequalA and UnequalB, and between EqualA and EqualB. However, in series 1, the birds responded with a higher PC to the Training stimuli compared to the Time-scaled versions of soundA and soundB (in both Equal-duration and Unequal-duration groups). This difference is significant in all comparisons apart from the difference between the 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 birds of the Unequal-duration group 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 B than to 'Duration stretched 20%' B ($P < 0.01$), and a clear trend of difference between Training 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, and such an 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 attend to the song duration for song recognition. Training sound A was aways 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.

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 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 respond significantly above the chance to the 'Duration stretched 50%' sound A and 'Duration compressed 50%' sound B but respond 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 correct response (Fig. 8b).

(a) Series 1



(b) Series 2

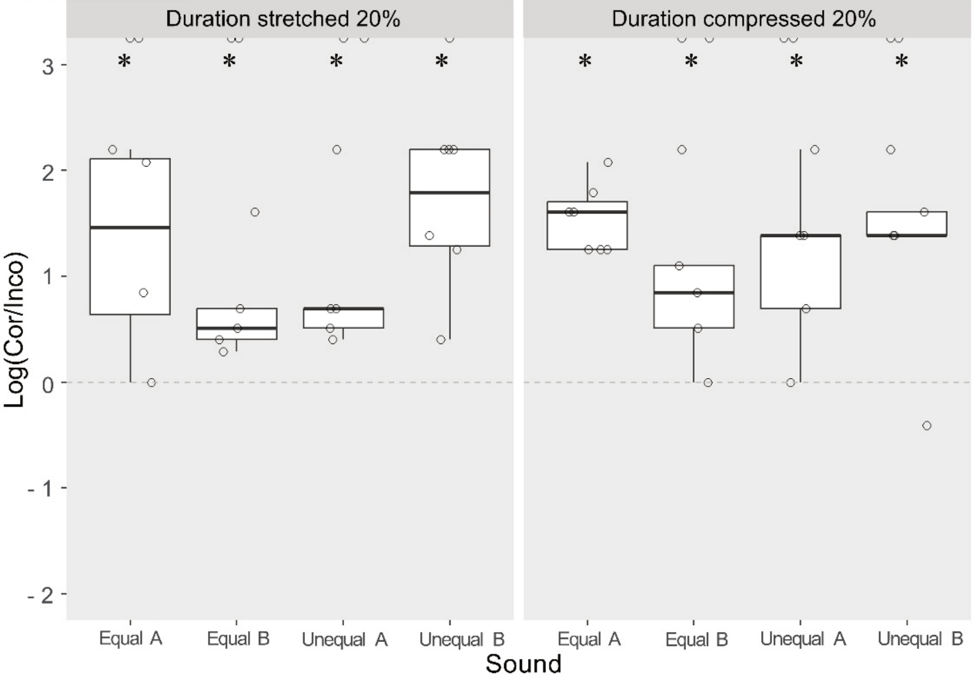


Figure 8. Visualisation of Log (Correct/Incorrect) for birds responding to time-scaled stimuli of series 1 (a) and series 2 (b). A * indicates that the Log(Cor/Inco) of a Test treatment is significantly different from 0; ns indicates that the Log(Cor/Inco) of a Test treatment is overlapping with 0. Box plots show median, 1st and 3rd quartile, and whiskers the 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(Cor/Inco)s was based on the counts of 'correct response' and 'incorrect response' from the same data set that was also used for Fig.7.

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 depended on whether the songs differed in duration or not, with spectral features having a less prominent role when duration is 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 the 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 the birds trained on songs of equal duration. Combined with our test results indicating that both experimental groups attend 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; Lohr et al., 2006; Vernaleo et al., 2010; Vernaleo & Dooling, 2011; Lawson et al., 2018; Prior et al., 2018a; Prior et al., 2018b; Geberzahn & Derégnaucourt, 2020; Fishbein et al. 2021; Mol et al., 2021) concluded that when zebra finches discriminate 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 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 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; Burgering et al., 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 is 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 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 zebra finches. The crucial test for being able to perceive a rhythmic pattern is whether humans or 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 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 thus is 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 that discrimination is 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 by 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 by 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 by their spectral envelope rather than pitch. Patel (2017) argued that this might be a common characteristic across birds, also for the discrimination of natural vocalizations. However, so far, no study 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 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) for distinguishing various vocalizations. Hence, one would expect spectral envelope to be important for discriminating 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 among 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. Attendance to such features has been demonstrated in a range of studies (e.g., Okanoya & Dooling, 1990; Uno et al., 1997; Dooling & Lohr, 2006; Lohr et al., 2006; Vignal & Mathevon, 2011; Prior et al., 2018a; Prior et al., 2018b).

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 demonstrate here, this could be spectral features, but also song duration. Similarly, in another study we showed that although zebra finches usually do not give much attention to the sequential order of the syllables when discriminating songs, they can use this sequence quite readily if needed (Ning et al., 2023), while Burgering et al. (2018, 2019) demonstrated that also the attendance to either the fundamental frequency (absolute pitch) or the energy distribution of a harmonic spectrum (spectral envelope) 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 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, as well as 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.

Acknowledgments

We thank Emmie Zeijlemaker for assistance with the experiment, Peter Snelderwaard for his technical assistance, and him and Michelle Geers for their animal care. Henkjan Honing was supported by an Open Competition grant from the Dutch Research Council NWO (406. 20. CW. 002). Zhi-Yuan Ning was supported by the CSC (China Scholarship Council) fellowship.

Appendix Figures & Tables

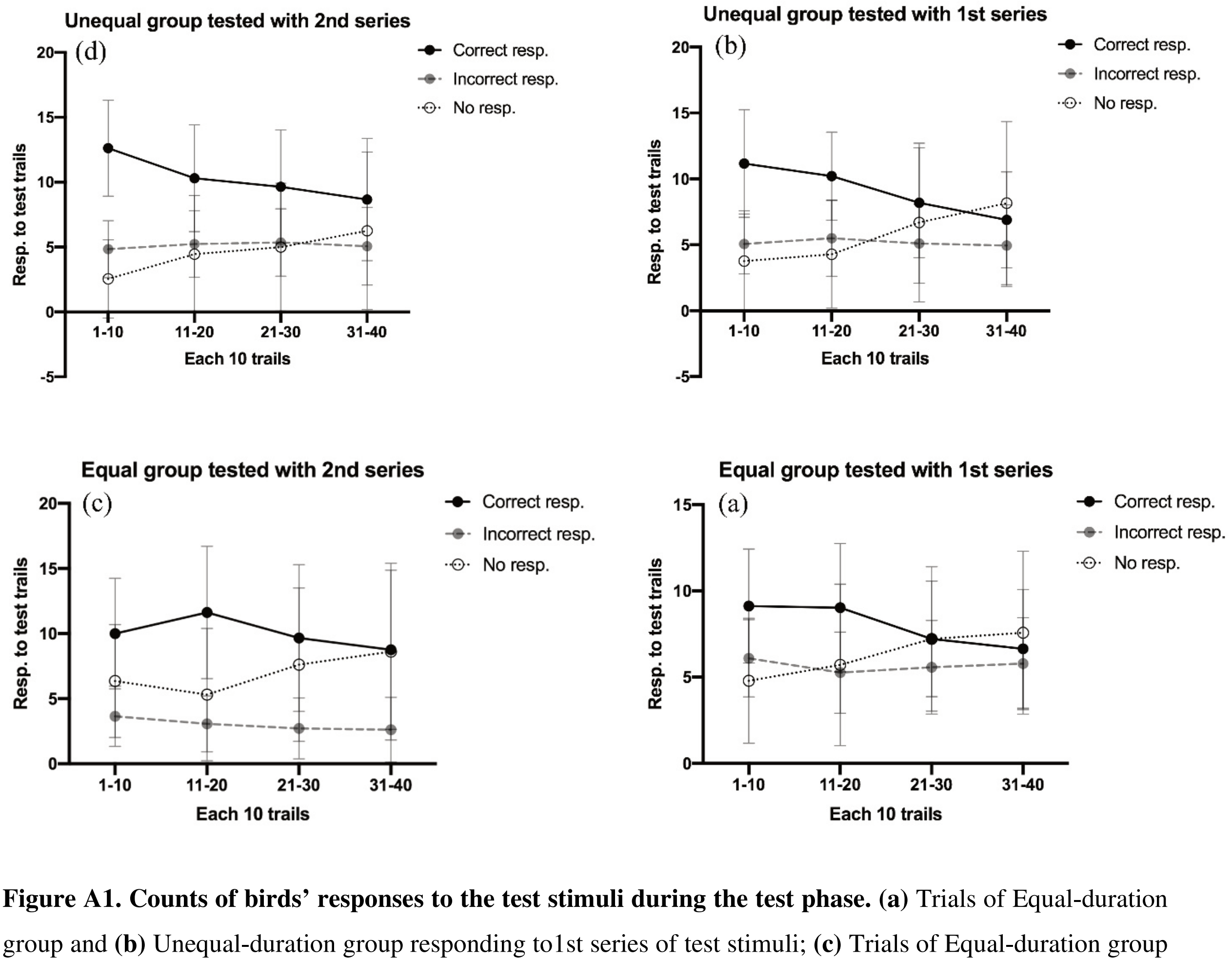


Figure A1. Counts of birds' responses to the test stimuli during the test phase. (a) Trials of Equal-duration group and **(b)** Unequal-duration group responding to1st series of test stimuli; **(c)** Trials of Equal-duration group and **(d)** Unequal-duration group responding to 2nd series of test stimuli. The total 40 test trials were divided into four 10-trials blocks. Lines across four 10-trials blocks refer to three categories of reaction to a sound: the 'correct response' (Black line), the 'incorrect response' (Black dash line), and a 'nonresponse' (Grey dash line).

Table A1 Post hoc test results of Binomial GLM for the interaction of spectrally-changed Test_Treatment & Training_Group

Stimuli	Training_Group	Series	estimate	SE	*Z*.ratio	*P*.value
a. PC (sound A+B in two training groups)						
Pair wise comparisons between training groups - Series 1						
Training	Equal - Unequal	1	0.0676	0.407	0.166	0.8680
Pitch-shifted -20%	**Equal - Unequal**	**1**	**-1.1345**	**0.359**	**-3.165**	**0.0080**
Pitch-shifted +20%	Equal - Unequal	1	-0.6006	0.351	-1.711	0.1450
Frequency-shifted 1500Hz	Equal - Unequal	1	-0.2342	0.349	-0.671	0.6281
Contour-averaged Vocoded	**Equal - Unequal**	**1**	**-0.8465**	**0.355**	**-2.387**	**0.0425**
Pair wise comparisons between test stimuli - Series 1						
Training - Pitch-shifted -20%	**Equal_Duration**	**1**	**2.1898**	**0.294**	**7.456**	**<.0001**
Training - Pitch-shifted +20%	**Equal_Duration**	**1**	**1.8516**	**0.290**	**6.384**	**<.0001**
Training - Frequency-shifted 1500Hz	**Equal_Duration**	**1**	**1.9422**	**0.290**	**6.688**	**<.0001**
Training – Contour-averaged Vocoded	**Equal_Duration**	**1**	**2.4698**	**0.298**	**8.301**	**<.0001**
Pitch-shifted -20% - Pitch-shifted +20%	Equal_Duration	1	-0.3382	0.250	-1.352	0.2519
Pitch-shifted -20% - Frequency-shifted 1500Hz	Equal_Duration	1	-0.2476	0.250	-0.989	0.3585
Pitch-shifted -20% – Contour-averaged Vocoded	Equal_Duration	1	0.2800	0.258	1.084	0.3478
Pitch-shifted +20% - Frequency-shifted 1500Hz	Equal_Duration	1	0.0906	0.246	0.368	0.7128
Pitch-shifted +20% – Contour-averaged Vocoded	**Equal_Duration**	**1**	**0.6182**	**0.254**	**2.431**	**0.0301**
Frequency-shifted 1500Hz – Contour-averaged Vocoded	Equal_Duration	1	0.5276	0.254	2.074	0.0635
Training - Pitch-shifted -20%	**Unequal_Duration**	**1**	**0.9877**	**0.296**	**3.341**	**0.0021**
Training - Pitch-shifted +20%	**Unequal_Duration**	**1**	**1.1834**	**0.292**	**4.059**	**0.0002**
Training - Frequency-shifted 1500Hz	**Unequal_Duration**	**1**	**1.6404**	**0.290**	**5.652**	**<.0001**
Training - Contour-averaged Vocoded	**Unequal_Duration**	**1**	**1.5557**	**0.289**	**5.374**	**<.0001**
Pitch-shifted -20% - Pitch-shifted +20%	Unequal_Duration	1	0.1957	0.259	0.755	0.5001
Pitch-shifted -20% - Frequency-shifted 1500Hz	**Unequal_Duration**	**1**	**0.6527**	**0.257**	**2.537**	**0.0224**
Pitch-shifted -20% – Contour-averaged Vocoded	**Unequal_Duration**	**1**	**0.5680**	**0.256**	**2.215**	**0.0446**
Pitch-shifted +20% - Frequency-shifted 1500Hz	Unequal_Duration	1	0.4570	0.253	1.810	0.1005

Pitch-shifted +20% – Contour-averaged Vocoded	Unequal_Duration	1	0.3724	0.252	1.480	0.1738
Frequency-shifted 1500Hz – Contour-averaged Vocoded	Unequal_Duration	1	-0.0846	0.249	-0.339	0.7344
Pair wise comparisons between training groups - Series 2						
Training	Equal - Unequal	2	-0.252	0.387	-0.652	0.5146
Pitch-shifted -8%	**Equal - Unequal**	**2**	**-0.889**	**0.339**	**-2.623**	**0.0145**
Pitch-shifted +8%	Equal - Unequal	2	-0.358	0.335	-1.069	0.3565
Frequency-shifted 500Hz	**Equal - Unequal**	**2**	**-1.095**	**0.344**	**-3.187**	**0.0035**
Contour-maintained Vocoded	**Equal - Unequal**	**2**	**-1.230**	**0.348**	**-3.539**	**0.0020**
Pair wise comparisons between test stimuli - Series 2						
Training - Pitch-shifted -8%	**Equal_Duration**	**2**	**1.9396**	**0.296**	**6.545**	**<.0001**
Training - Pitch-shifted +8%	**Equal_Duration**	**2**	**1.3160**	**0.291**	**4.518**	**<.0001**
Training - Frequency-shifted 500Hz	**Equal_Duration**	**2**	**2.2073**	**0.302**	**7.306**	**<.0001**
Training – Contour-maintained Vocoded	**Equal_Duration**	**2**	**2.3733**	**0.307**	**7.731**	**<.0001**
Pitch-shifted -8% - Pitch-shifted +8%	**Equal_Duration**	**2**	**-0.6236**	**0.265**	**-2.350**	**0.0268**
Pitch-shifted -8% - Frequency-shifted 500Hz	Equal_Duration	2	0.2677	0.277	0.966	0.3710
Pitch-shifted -8% – Contour-maintained Vocoded	Equal_Duration	2	0.4337	0.282	1.537	0.1555
Pitch-shifted +8% - Frequency-shifted 500Hz	**Equal_Duration**	**2**	**0.8912**	**0.272**	**3.279**	**0.0017**
Pitch-shifted +8% – Contour-maintained Vocoded	**Equal_Duration**	**2**	**1.0572**	**0.277**	**3.815**	**0.0003**
Frequency-shifted 500Hz – Contour-maintained Vocoded	Equal_Duration	2	0.1660	0.288	0.576	0.5647
Training - Pitch-shifted -8%	**Unequal_Duration**	**2**	**1.3031**	**0.286**	**4.548**	**<.0001**
Training - Pitch-shifted +8%	**Unequal_Duration**	**2**	**1.2099**	**0.287**	**4.214**	**0.0001**
Training - Frequency-shifted 500Hz	**Unequal_Duration**	**2**	**1.3646**	**0.286**	**4.768**	**<.0001**
Training – Contour-maintained Vocoded	**Unequal_Duration**	**2**	**1.3952**	**0.286**	**4.877**	**<.0001**
Pitch-shifted -8% - Pitch-shifted +8%	Unequal_Duration	2	-0.0932	0.249	-0.374	0.8876
Pitch-shifted -8% - Frequency-shifted 500Hz	Unequal_Duration	2	0.0615	0.248	0.248	0.8934
Pitch-shifted -8% – Contour-maintained Vocoded	Unequal_Duration	2	0.0921	0.248	0.372	0.8876
Pitch-shifted +8% – Frequency-shifted 500Hz	Unequal_Duration	2	0.1547	0.249	0.622	0.8876
Pitch-shifted +8% – Contour-maintained Vocoded	Unequal_Duration	2	0.1853	0.249	0.745	0.8876
Frequency-shifted 500Hz – Contour-maintained Vocoded	Unequal_Duration	2	0.0306	0.247	0.124	0.9016

Note: Response variables in GLMMs: the PC of 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)						
Pair wise 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 1500Hz	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 1500Hz**	**1**	**0.635**	**0.206**	**0.2314**	**1.0380**
Unequal_Duration	**Contour-averaged Vocoded**	**1**	**0.693**	**0.205**	**0.2900**	**1.0954**
Pair wise 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 500Hz**	**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 500Hz**	**2**	**0.564**	**0.202**	**0.1678**	**0.9606**
Unequal_Duration	**Contour-maintained Vocoded**	**2**	**0.727**	**0.213**	**0.3097**	**1.1443**

Note: Lower CL and Upper CL represent the lower and upper 95% confidence limits (CL) 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 toward correct responses. If they are both negative, then they are biased toward 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 GLM for the interaction of Time-scaled Test_Treatment & Training_Sound

Stimuli	Training_Sound	Series	Estimate	SE	*Z*.ratio	*P*.value
PC (sound A & B separated)						
Pair wise comparisons between training sounds - Series 1						
Duration stretched 50%	EqualA - EqualB	1	0.0594	0.345	0.172	0.8631
Duration stretched 50%	UnequalA - UnequalB	1	0.6801	0.355	1.915	0.1111
Duration compressed 50%	EqualA - EqualB	1	0.3645	0.350	1.043	0.3961
Duration compressed 50%	UnequalA - UnequalB	1	-0.7906	0.353	-2.238	0.1009
Pair wise comparisons between test stimuli - Series 1						
Training - Duration stretched 50%	**EqualA**	**1**	**1.10e+00**	**0.383**	**2.867**	**0.0124**
Training - Duration compressed 50%	EqualA	1	7.92e-01	0.387	2.047	0.0609
Duration compressed 50% - Duration stretched 50%	EqualA	1	3.05e-01	0.350	0.871	0.3835
Training - Duration stretched 50%	**EqualB**	**1**	**1.44e+00**	**0.402**	**3.573**	**0.0005**
Training - Duration compressed 50%	**EqualB**	**1**	**1.44e+00**	**0.402**	**3.573**	**0.0005**
Duration compressed 50% - Duration stretched 50%	EqualB	1	-4.05e+05	0.344	0.000	0.9999

Training - Duration stretched 50%	**UnequalA**	**1**	**9.58e-01**	**0.412**	**2.325**	**0.0201**
Training - Duration compressed 50%	**UnequalA**	**1**	**1.88e+00**	**0.407**	**4.620**	**<.0001**
Duration compressed 50% - Duration stretched 50%	**UnequalA**	**1**	**-9.20e-01**	**0.357**	**-2.582**	**0.0148**
Training - Duration stretched 50%	**UnequalB**	**1**	**1.44e+00**	**0.391**	**3.693**	**0.0007**
Training - Duration compressed 50%	**UnequalB**	**1**	**8.93e-01**	**0.395**	**2.259**	**0.0358**
Duration compressed 50% - Duration stretched 50%	UnequalB	1	5.50e-01	0.352	1.563	0.1179
Pair wise comparisons between training sounds - Series 2						
Duration stretched 20%	EqualA - EqualB	2	4.27e-01	0.379	1.127	0.8471
Duration stretched 20%	UnequalA - UnequalB	2	-3.22e-01	0.403	-0.800	0.8471
Duration compressed 20%	EqualA - EqualB	2	-7.40e-02	0.385	-0.192	1.0000
Duration compressed 20%	UnequalA - UnequalB	2	-9.20e-06	0.416	0.000	1.0000
Pair wise comparisons between test stimuli - Series 2						
Training - Duration stretched 20%	EqualA	2	4.88e-01	0.406	1.201	0.3444
Training - Duration compressed 20%	EqualA	2	5.62e-01	0.404	1.390	0.3444
Duration compressed 20% - Duration stretched 20%	EqualA	2	-7.39e-02	0.385	-0.192	0.8476
Training - Duration stretched 20%	**EqualB**	**2**	**1.32e+00**	**0.426**	**3.087**	**0.0061**
Training - Duration compressed 20%	EqualB	2	8.89e-01	0.433	2.053	0.0601
Duration compressed 20% - Duration stretched 20%	EqualB	2	4.27e-01	0.379	-1.127	0.2597
Training - Duration stretched 20%	UnequalA	2	7.05e-01	0.427	1.650	0.2967
Training - Duration compressed 20%	UnequalA	2	3.83e-01	0.440	0.870	0.4236
Duration compressed 20% - Duration stretched 20%	UnequalA	2	3.22e-01	0.403	0.800	0.4236
Training - Duration stretched 20%	UnequalB	2	6.11e-01	0.458	1.334	0.2734
Training - Duration compressed 20%	UnequalB	2	6.11e-01	0.458	1.334	0.2734
Duration compressed 20% - Duration stretched 20%	UnequalB	2	-1.27e-05	0.416	0.000	1.0000

Note: 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)						
	Pair wise 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**
	Pair wise 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**

Note: Lower CL and Upper CL represent the lower and upper 95% confidence limits (CL) 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 toward correct responses. If they are both negative, then they are more biased toward 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)						
	Pair wise 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**
	Pair wise 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**

Note: Lower CL and Upper CL represent the lower and upper 95% confidence limits (CL) 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 toward correct responses. If they are both negative, then they are more biased toward incorrect responses. Bold indicates significance.

References

Bates D., Maechler M., Bolker B., Walker S. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw*., 67(1): 1-48. https://doi.org/10.18637/jss.v067.i01

Benjamini Y., Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*., 57(1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

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. *Philos Trans R Soc Lond B Biol Sci*., 376(1835): 20200335. https://doi.org/10.1098/rstb.2020.0335

Braaten R.F., Petzoldt M., Colbath A. (2006). Song perception during the sensitive period of song learning in zebra finches (Taeniopygia guttata). *J Comp Psychol*., 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. *Proc Natl Acad Sci*., 113(6): 1666–1671. https://doi.org/10.1073/pnas.15153801

Burgering M.A., ten Cate C., Vroomen J. (2018). Mechanisms underlying speech sound discrimination and categorization in humans and zebra finches. *Anim Cogn*., 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 both on the fundamental frequency ("Pitch") and spectral envelope. *J Comp Psychol.,* 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. https://www.jstor.org/stable/4535030

Dent M.L., Martin A.K., Flaherty M.M., Neilans E.G. (2016). Cues for auditory stream segregation of birdsong in budgerigars and zebra finches: Effects of location, timing, amplitude, and frequency. *J Acoust Soc Am.,* 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. *J Acoust Soc Am*., 112(2): 748-59. 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. *Ornithol Sci*., 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? *Anim Behav*., 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. *Anim Cogn.*, 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. *Nat Commun.*, 11(1): 2914. https://doi.org/10.1038/s41467-020-16398-z

Fishbein A.R., Prior N.H., Brown J.A., Ball G.F., Dooling R.J. (2021). Discrimination of natural acoustic variation in vocal signals. *Sci Rep.*, 11(1): 916. https://doi.org/10.1038/s41598-020-79641-z

Fox J., Weisberg S. (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Geberzahn N., Derégnaucourt S. (2020). Individual vocal recognition in zebra finches relies on song syllable structure rather than song syllable order. *J Exp Biol.*, 223(Pt 9): jeb220087. https://doi.org/10.1242/jeb.220087

Hulse S.H., Cynx J., Humpal J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *J Exp Psychol Gen.,* 113(1): 38–54. https://doi.org/10.1037/0096-3445.113.1.38

Hoeschele M. (2017). Animal pitch perception: Melodies and harmonies. *Comp Cogn Behav Rev.,* 12: 5-18. https://doi.org/10.3819/CCBR.2017.120002

Kreutzer M.L., Dooling R.J., Brown S.D., Okanoya K. (1990). A comparison of song syllable perception by five species of birds. *Int J Comp Psychol.*, 4(2): 141–155. https://doi.org/10.46867/C4H599

Lampen J., Jones K., McAuley J.D., Chang S.-E., 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): 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. *J Neuroendocrinol.*, 29(12): 10.1111/jne.12543. 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. *Anim Cogn.,* 21(4): 467–480. https://doi.org/10.1007/s10071-018-1182-2

Lenth R.V. (2016). Least-Squares Means: The R Package lsmeans. *J Stat Softw.*, 69(1): 1-33. https://doi.org/10.18637/jss.v069.i01

Lohr B., Dooling R.J., Bartone S. (2006). The discrimination of temporal fine structure in call-like harmonic sounds by birds. *J Comp Psychol.*, 120(3): 239-51. https://doi.org/10.1037/0735-7036.120.3.239

MacDougall-Shackleton S.A., Hulse S.H. (1996). Concurrent Absolute and Relative Pitch Processing by European Starlings (Sturnus vulgaris). *J Comp Psychol.*, 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. *Philos Trans RSoc Lond B Biol Sci*, 376(1836): 20200248. https://doi.org/10.1098/rstb.2020.0248

Nagel K.I.K., McLendon H.M., Doupe A.J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J Neurophysiol.*, 104(3): 1426–37. 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. https://www.jstor.org/stable/4535075

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

Nelson D.A. (1989). The Importance of Invariant and Distinctive Features in Species Recognition of Bird Song. *The Condor*, 91(1): 120–130. https://doi.org/10.2307/1368155

Ning Z.Y., Honing H., ten Cate C. (2023). Zebra finches (Taeniopygia guttata) demonstrate cognitive flexibility in using phonology and sequence of syllables in auditory discrimination. *Anim Cogn.,* 10.1007/s10071-023-01763-4. Advance online publication. https://doi.org/10.1007/s10071-023-01763-4

Okanoya K., Dooling R.J. (1990). Temporal integration in zebra finches (Poephila guttata). *J Acoust Soc Am.*, 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. *Comp Cogn Behav Rev.*, 12(1): 19–32. https://doi.org/10.3819/CCBR.2017.120003

Prior N.H., Smith E., Ball G.F., Dooling R.J. (2018a). Temporal fine structure and zebra finch vocalizations. *J Acoust Soc Am.*, 143(3): 1897-1897. https://doi.org/10.1121/1.5036170

Prior N.H., Smith E., Lawson S., Ball G.F., Dooling R.J. (2018b). Acoustic fine structure may encode biologically relevant information for zebra finches. *Sci Rep.*, 8(1): 6212. https://doi.org/10.1038/s41598-018-24307-0

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. 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. *Proc Natl Acad Sci.*, 118(29): e2026130118. 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. *Anim Behav.*, 203: 193-206. https://doi.org/10.1016/j.anbehav.2023.05.001

ten Cate C., Spierings M., Hubert J., Honing H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front Psychol.,* 7: 730. https://doi.org/10.3389/fpsyg.2016.00730

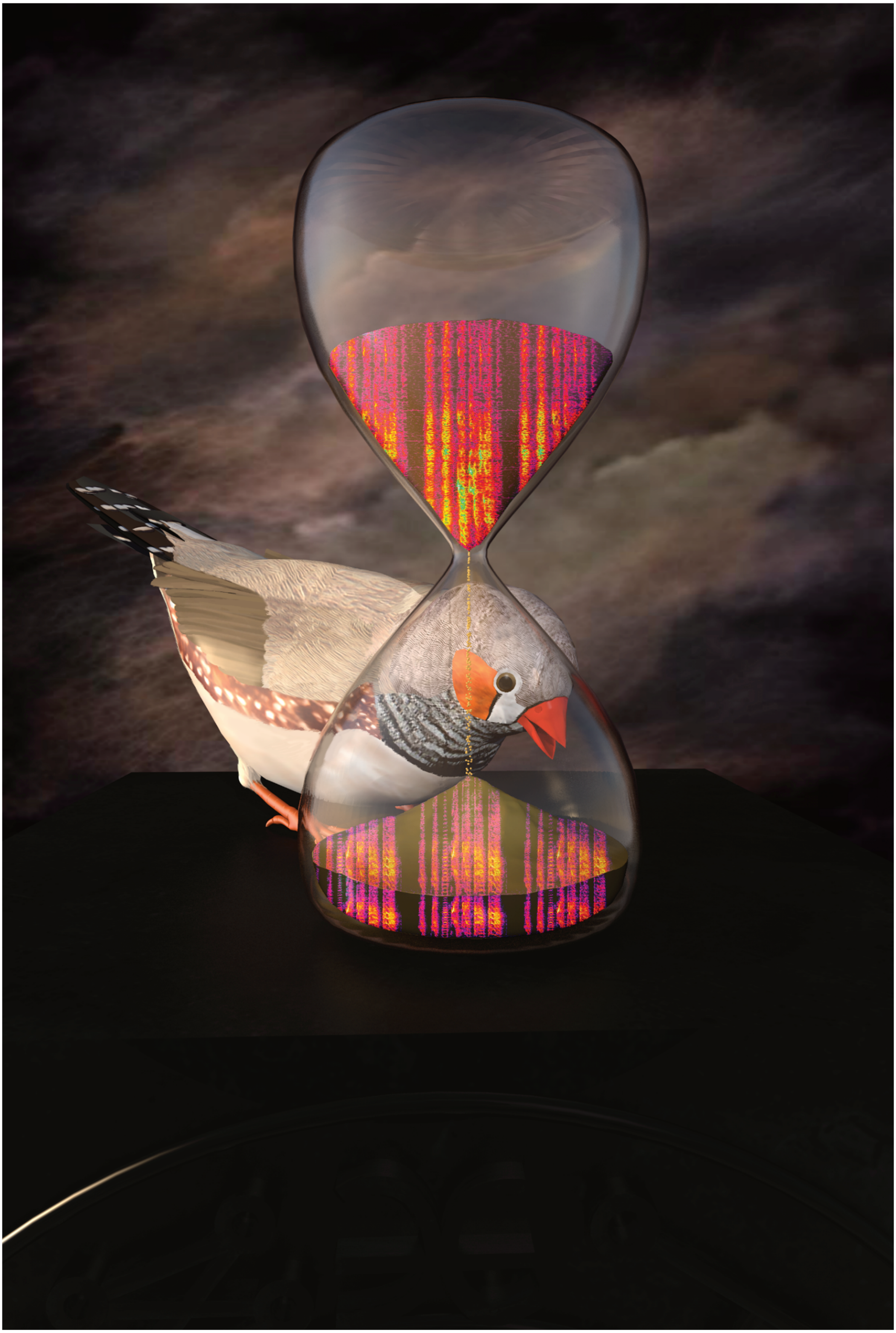
Tu H-W, Dooling R.J. (2012). Perception of warble song in budgerigars (Melopsittacus undulatus): Evidence for special processing. *Anim Cogn.*, 15(6): 1151-9. https://doi.org/10.1007/s10071-012-0539-1

Uno H., Maekawa M., Kaneko H. (1997). Strategies for harmonic structure discrimination by zebra finches. *Behav Brain Res.*, 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. *Behav Processes.,* 115: 37–45. https://doi.org/10.1016/j.beproc.2015.02.018

Vernaleo B.A., Dooling R.J., Leek M.R. (2010). Relative salience of acoustic features in zebra finch song. *J Acoust Soc Am.*, 127(3): 1863. https://doi.org/10.1121/1.3384460

Vernaleo B.A., Dooling R.J. (2011). Relative salience of envelope and fine structure cues in zebra finch song. *J Acoust Soc Am.*, 129(5): 3373–3383. https://doi.org/10.1121/1.3560121

Vignal C., Mathevon N. (2011). Effect of Acoustic Cue Modifications on Evoked Vocal Response to Calls in Zebra Finches (Taeniopygia guttata). *J Comp Psychol.*, 125(2): 150–161. https://doi.org/10.1037/a0020865



Female and Male Zebra Finches Preferences for Normal and Modified songs

Chapter 3

Zhi-Yuan Ning · Henkjan Honing · Peter Snelderwaard · Carel ten Cate

This chapter is prepared for submission

Chapter 3

ABSTRACT

The zebra finch is an established model species for song preference research in songbirds. In this study, a 4-way choice operant paradigm (the ‘carrousel’ cage) was used to measure the spontaneous preference of adult birds of both sexes for song stimuli of four different categories: a conspecific song, a duration-stretched version and a vocoded version of the conspecific song, and a heterospecific great tit song. An earlier operant song discrimination experiment showed that both male and female zebra finches treated duration stretched and vocoded song versions as being highly different from a normal song. The current results show that female song preference for normal and duration-stretched conspecific songs was stronger than male preference. Females also exhibited a gradient of preferences: they preferred the normal and the duration-stretched song over the vocoded version of the conspecific song, but the vocoded conspecific songs were still preferred over the heterospecific great tit songs. Males preferred duration-stretched conspecific songs over heterospecific songs, but otherwise did not differentiate among the stimuli. Overall, our results demonstrate that results about the importance of particular song features as obtained in the context of an operant discrimination task and of a preference test paradigm can differ. This shows that combining different approaches and testing paradigms may help to better understand the significance of various acoustical parameters in relation to the ecological functions and evolutionary processes of birdsong as relevant to both sexes.

INTRODUCTION

The study of song preferences and song recognition in songbirds is a crucial element for understanding their vocal communication. Song preferences may indicate the birds' preference for the singers and play a vital role in model selection for vocal learning during the juvenile phase, as well as in mate selection in adult songbirds (Fujii *et al*., 2022). In the context of vocal learning, research addresses whether conspecific song models are preferred over heterospecific ones (e.g., Marler & Peters, 1988) and which song features guide such preferences (e.g., Soha & Marler, 2000). Studies of song preferences in the context of mate choice examine which features make songs attractive (e.g., Vallet & Kruetzer, 1995; Gentner & Hulse, 2000) and whether these indicate specific qualities in the singers (e.g., Suthers *et al.* 2012; Wang *et al*., 2019). Preferences may also be shown when birds can choose between familiar conspecific songs, e.g., those of fathers or individuals of the same population, and unfamiliar ones (e.g., Fujii & Okanoya, 2022). Additionally, songbirds of both sexes often recognize the songs of mates or neighbours (e.g., Stoddard *et al*., 1990; Gentner *et al*., 2000). Song preferences and song recognition can thus be expressed in different contexts and the song features on which responses to particular songs are based will vary according to the context. This ability implies the presence of an auditory perceptual system that is capable of identifying, selecting, and using different vocal features in different contexts: choice of a song model, selecting a mate, or recognizing a neighbour may all be based on different song features. This makes the study of auditory perception and the cognitive processes involved in song preferences and song recognition an important area of research.

The zebra finch (*Taeniopygia guttata*) has become an established model species for song preference research in songbirds. Both male and female zebra finches produce different types of calls, but only males produce songs, which consist of a series of repetitions of an individual-specific motif composed of three to eight different syllables (Zann, 1996). Song preferences in zebra finches have been extensively studied using various experimental paradigms, mostly either operant conditioning or phonotaxis experiments. The operant paradigm involves training a subject bird to perform a specific response, such as pecking a key or hopping on a perch, in order to elicit song playback as a reward. The number of responses is then used as an indicator of a song preference. This paradigm is used extensively in song preference research, mostly in studies that offered bimodal choices (e.g., Braaten & Reynolds, 1999;

Neubauer, 1999; Houx & ten Cate 1999; Riebel, 2000; Riebel *et al.*, 2002; Riebel & Smallegange, 2003; Leadbeater *et al.*, 2005; Spencer *et al.*, 2005; Holveck & Riebel, 2007; Holveck & Riebel, 2010; Rodríguez-Saltos *et al.*, 2023; also see Riebel 2009 for a review) and occasionally also in studies that offered multiple choices (e.g., Ritschard *et al.*, 2010; Wei *et al.*, 2022). The phonotaxis paradigm involves presenting a subject bird with two or more different songs from speakers (or stimulus individuals) located on different sides of an experimental cage (Houtman, 1992; Neubauer, 1999; Adkins-Regan & Krakauer, 2000; Tomaszycki & Adkins-Regan, 2005; ter Haar *et al.*, 2014; Chen *et al.*, 2017; Fujii *et al.*, 2021; Mol *et al.*, 2021). The bird's approach to one or another speaker is taken as an indication of its preference for the corresponding song. This method is based on the natural behaviour of zebra finches to approach preferred songs (e.g., Loning *et al.*, 2023). In the operant conditioning paradigm, birds must work spontaneously to trigger a song playback, so operant conditioning is more closely related to the motivation to hear a preferred song than the phonoaxis paradigm for detecting song preferences.

The use of different paradigms and stimuli/sounds has allowed for a comprehensive investigation into zebra finch song preferences and has shed light on the features underlying these preferences and the role of experience in shaping them. This demonstrated, among others, that zebra finches of both sexes preferred conspecific song over either the songs of cross-fostered Bengalese finches (*Lonchura striata var. domestic*) (Clayton & Pröve, 1989; Campbell & Hauber, 2009), or European starling (*Stumus vulgaris*) (Braaten & Reynolds, 1999), and that young adult females preferred conspecific to canary (*Serinus canaria*) song in the two-choice preference test paradigm (Lauay *et al.*, 2004). Among conspecific songs of zebra finches some types of songs are preferred over others, such as longer songs over shorter ones (e.g., Clayton & Pröve, 1989; Neubauer, 1999), more complex songs over less complex ones (e.g., Woodgate *et al.* 2011), faster songs over slower ones (e.g., Chen *et al.*, 2017), and songs with better-defined harmonic elements over those with noisier elements (Tomaszycki & Adkins-Regan, 2005). Additionally, both males and females prefer the songs of their father or other model song over unfamiliar songs (e.g., Houx & ten Cate, 1999; Riebel *et al*, 2002) and over the songs of unfamiliar brothers (i.e., same parents but different clutch) (Riebel & Smallegange, 2003), and juvenile male zebra finches prefer the songs of their father over the songs of another familiar adult male during the auditory phase of song learning (Rodríguez-Saltos *et al.*, 2023) and song stimuli consisting of more common zebra finch song elements over stimuli consisting of less common elements (ter Haar *et al*, 2014). However, while the studies mentioned above indicate preferences for certain types of songs, they don't necessarily provide much insight into the nature of the song features on which these are based or on the perceptual and cognitive abilities that zebra finches have at their disposal to detect differences between songs. For instance, the finding that zebra finches prefer conspecific songs over heterospecific ones or familiar ones over unfamiliar ones shows that they can distinguish these, but it does not reveal on which acoustic parameters the discrimination is based. Also, the absence of a preference among various songs does not imply that zebra finches are insensitive to the acoustic differences between these songs. To address such issues, a different approach is required examining which features birds use, or are capable of using, for discriminating between songs or other auditory stimuli. This is done in experiments using operant discrimination paradigms such as a Go/No-go or Go-left/Go-right training, in which birds are trained to obtain a food or water reward for responding to one type of sound and not, or in a different way to another sound. Errors in responding can be followed by a signal such as a brief period of darkness or do not result in a reward. To address which features the birds use to discriminate between the training sounds they are given test sounds that systematically alter specific sound parameters and examine the impact on the ability to still recognize or distinguish the sounds. These discrimination experiments have explored the type of acoustic parameters to which zebra finches are able to attend in discriminating songs and other sounds, such as pitch, intensity (Nagel *et al*., 2010), spectro-temporal structure of elements (Dooling *et al*., 2002; Lohr *et al*., 2006; Vernaleo *et al*., 2010; Vernaleo & Dooling, 2011; Lawson *et al*., 2018; Prior *et al*., 2018; Fishbein *et al*. 2021), spectral envelope (Burgering *et al*., 2019), element sequences (Lawson *et al*., 2018; Geberzahn & Derégnaucourt, 2020; Ning *et al*., 2023) and duration of song (see Chapter 2).

It is important to note that operant discrimination experiments also have limitations in revealing the full extent of the cognitive processing involved. Specifically, while changes in a specific parameter may result in reduced recognition of a song, this does not reveal whether such a modified song is no longer recognized as a zebra finch song or is considered unattractive. As an example, our previous studies (Ning *et al*., 2023; Ning *et al*., submitted - Chapter 2, Chapter 4) showed that when song stimuli in an operant discrimination task were changed by vocoding – replacing a harmonic spectrum by noise in the same frequency range – they were treated as being highly different from the original training songs. This showed that the harmonic spectrum was a critical element for distinguishing the training songs. It may suggest that a vocoded song is no longer recognized as a zebra finch song at all and considered just as different from the normal song as heterospecific song or an arbitrary sound, but that need not be the case. A vocoded song version still contains some species-specific information, e.g., it maintains the temporal structure and frequency ranges of song syllables. This might still make such a stimulus interesting to zebra finches. Thus, while modifications made to training songs in the context of an operant discrimination task may result in song stimuli that are considered very different from the training songs it is an open question whether such manipulated stimuli still retain sufficient specificity to be recognized and potentially even be attractive to zebra finches. This can only be assessed by presenting such manipulated stimuli in a preference paradigm in which the preferences of zebra finches for normal songs and manipulated stimuli are compared.

In the current experiment we examine how the preference for non-manipulated songs (used as training stimuli in an earlier operant song discrimination study; see Ning *et al.*, submitted - Chapter 2) relates to the preference for some of the manipulated versions of these songs as well as to the preference for a heterospecific song. This allows for examining whether song modifications, that reduce the recognition of a familiar song in a discrimination task, also cause such stimuli to be less attractive in a preference test. Since zebra finches prefer conspecific songs over heterospecific ones (Clayton & Pröve, 1989; Braaten & Reynolds, 1999; Lauay *et al.*, 2004; Campbell & Hauber, 2009), we expect less attractive song variants to be treated more similar to heterospecific songs than to normal ones. Our song discrimination study showed no sex difference into which features and how strongly males and females responded, but this does not imply that males and females have similar song preferences. We therefore examined the song preferences of adult birds of both sexes in an identical context and with the same stimuli and method – something in itself of interest as this is not often done.

We tested zebra finch preferences in a four-way preference set up, using an octagonal choice cage in a multiple, interactive choice situation, or “carrousel” for short. This operant preference paradigm has been designed and successfully used in our laboratory’s previous preference research including preference tests on imprinting, courtship, and birdsong (ten Cate *et al.*, 2006; Holveck *et al.*, 2011; Wei *et al.*, 2022). Song preference tests in this set up demonstrated that zebra finches will voluntarily hop on an operant perch if this results in hearing a conspecific song, with this song exposure acting as a reinforcer for this behaviour (Wei *et al.*, 2022). The individual preferences were measured by the number of times the bird triggered playbacks of four different stimuli: a normal zebra finch song, its duration-stretched version, a vocoded version, and a heterospecific song. Zebra finches considered both the duration stretched, and the vocoded version of normal songs as being significantly different from the normal training songs in an operant discrimination task (Ning *et al.*, submitted).

METHODS

Subjects

Forty-five zebra finches (23 males and 22 females) were used in this experiment. They were tested at the age of 298 ± 50 days post-hatching (dph), (age males: M= 318, SD= 60, age females: M= 281, SD= 31). All birds originated from the in-house breeding colony at Leiden University. Before the experiment, the birds lived in single-sex groups of about 15 to 30 individuals in aviaries (2m × 2m × 1.5m), in which food and water were available ad libitum. The housing rooms were kept at 20–22°C and 40–60% humidity and illuminated with artificial lights (Philips Master TL5 HO 49W/830) from 07:00–20:30 (13.5h light : 10.5h dark) with a 15 min twilight phase with the light fading in and out at the beginning and the end of each day. A week before the preference test, birds were caught and transferred from the aviary to standard laboratory cages (two birds of equal sex in one cage) in order to acclimatize (cage size length × width × height: 80 × 40 × 40 cm) and reduce stress from catching in aviary.

Experimental setup

The birds were tested individually in a 4-way preference choice cage (carrousel cage) which was placed on a table (height: 75 cm) within a sound attenuated chamber (height: 250 cm, width × length irregular quadrilateral: 335 × 280 × 290 × 300 cm). The experimental setup consisted of a central octagon cage (height: 35 cm, diameter: 70 cm) containing 8 perches, of which 4 were sound perches (sensors), triggering playback of a sound when the bird hopped on them, and 4 were silent perches, not resulting in a sound being played. The cage had a wire-mesh covering (height: 35 cm, diameter: 70 cm) on its top and 8 wire mesh openings, one on each side closed off with white plastic partitions. Four sides had a loudspeaker (CB4500, Blaupunkt, Hildesheim, Germany) on the outside, broadcasting through a round opening (diameter = 3 cm) in the middle of the plastic partition (see Fig. 1). The birds triggered song playbacks by landing on small cross-shaped perches in front of the loudspeaker openings, which activated a microswitch connected to a computer outside the experimental room. The computer controlled the playbacks and recorded the moment the bird landed and departed

(Dell OptiPlex 3010 with RADEON HD soundcard and custom-written software in Visual Basic 6, by P.C. Snelderwaard). Hopping on a sound perch triggered a single playback sound, regardless of how long the bird stayed on a perch. However, if the bird left the perch before the end of the song playback and hopped to another sound perch, the previous playback was stopped and a new playback associated with the new perch was triggered. Prior to the experiment, playback levels on each perch were set to the maximum level of 68 dBA (re: 20 μPa) using a VOLTCRAFT SL-100 sound meter with the sound level set to high and A-weighting. A one-way mirror window in the door allowed observation of the room from outside but did not allow visibility of the birds everywhere in the carrousel. Therefore, a camera (Logitech HD 1080p, Lausanne, Switzerland) was mounted 1.5 meters above the carrousel and connected to the computer (software: Webcam7, v. 70.9.9.41) controlling the setup, allowing monitoring of individuals' behaviour during trials from outside the experimental chamber.

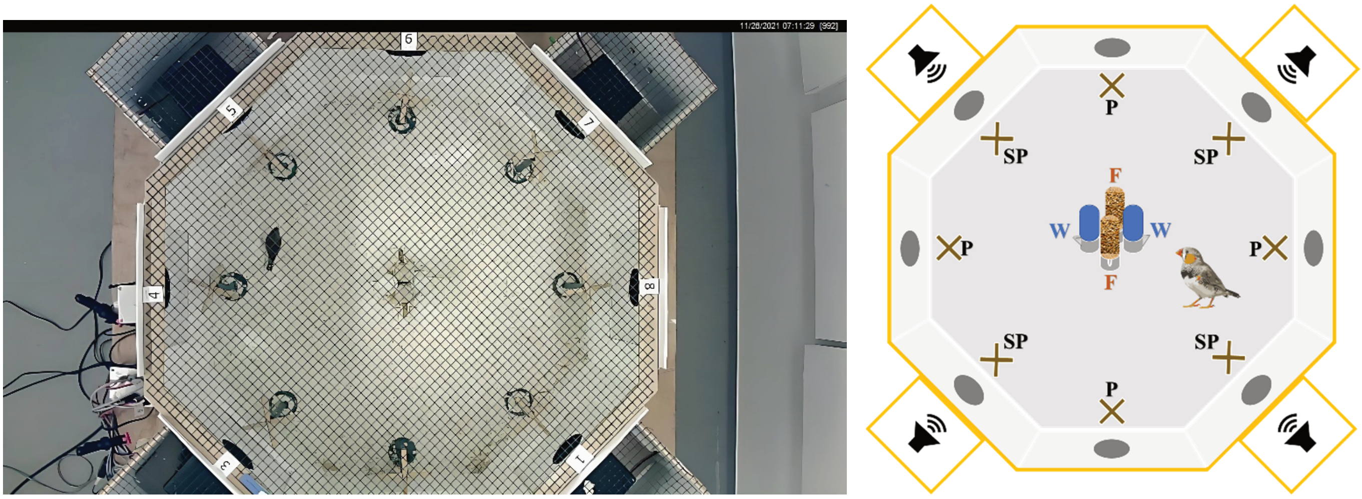


Figure 1. Experimental apparatus (Carrousel cage) used for the preference test. a) Photograph top view of the apparatus, and b) Schematic representation of the testing apparatus: Four **loudspeakers** are positioned symmetrically behind the plastic partitions on the diagonal side of the octagonal cage; eight grey ellipses represent the **round openings** through half of which the sounds of loudspeakers were broadcasted. Within the cage, there are eight cross-shape perches for the bird to sit on, with four of them being **sound perches** (SP) and the other four being **silent perches** (P). Two tubes with ad libitum **water** (W) and two tubes with ad libitum **food** (F) are placed in the middle of the floor panel. Landing on a sound perch triggered a single playback of the stimulus associated with it while landing on a silent perch resulted in no playback. The stimuli associated with the four sound perches were rotated one position clockwise per block, with each block lasting for two and a half hours.

Stimuli

Four types of stimuli were used in probing the preferences of the birds.

Normal zebra finch song

These songs consisted of motifs from the songs that also had been used as training stimuli for the Go-Left/Go-Right operant task in chapter 2. They consisted of a natural motif repeated several times without introduction notes. These motifs were extracted from representative recordings of adult males whose vocalizations had not been heard before by the experimental birds in this preference test. The same motif was repeated 3 to 6 times (resulting in a duration of 3.5 - 4.5 seconds). When played, the motifs within a song were normalized in root-mean-square (RMS) amplitude such that the average intensity was the same but the range of variation in volume recorded at the microphone was preserved. All zebra finch songs were filtered to a bandwidth between 380Hz and 15kHz (Fig. 2a). The songs were cut, synthesized, and filtered using Praat (version 6.0.54). The amplitude of each song was adjusted by using the “Normalize” feature in Audacity (version 2.3.0).

Duration-stretched zebra finch song

For the Duration-stretched manipulation, the duration of the whole song was stretched proportionally without any change in the frequency domain (Fig. 2c). The “change duration” script of Praat Vocal Toolkit was applied to obtain the 50% Duration-stretched versions (see the “duration +50%” stimuli version of test series 1 in chapter 2).

Vocoded zebra finch song

This song modification maintains the spectral and temporal envelopes of the elements within the motif, but averages the energy within specific frequency bands of the song elements, thus removing any harmonic structure (Fig. 2b). The vocoded version of songs was constructed by the Matt Winn’s Praat vocoded script (http://www.mattwinn.com/praat/vocode_all_selected_40.txt) which maintains the temporal envelope (see the Vocoded stimuli of test series 2 in chapter 2). This script was set to divide cut-off frequency bandwidths equally for 15 bands contiguous with smooth transitions (1000Hz bandwidth for each noise-vocoded band).

Heterospecific song

The heterospecific songs were selected from the field recordings of natural great tit (*Parus major*) songs. Great tit songs consist of repetitions of fixed syllables that deviate from the spectro-temporal structures of zebra finch conspecific songs. The total duration of a heterospecific song was selected to be similar to the duration of the zebra finch song in the stimuli set tested for a same individual (Fig. 2d). To obtain clear sounds from those field recording with which the environmental background noises were along, the “reduce noise” function of Praat Vocal Toolkit was applied to eliminate the background noise, and the denoised great tit songs were then high pass filtered at 2000 Hz by using Praat and were normalized as the same RMS amplitude as the zebra finch songs by using Audacity.

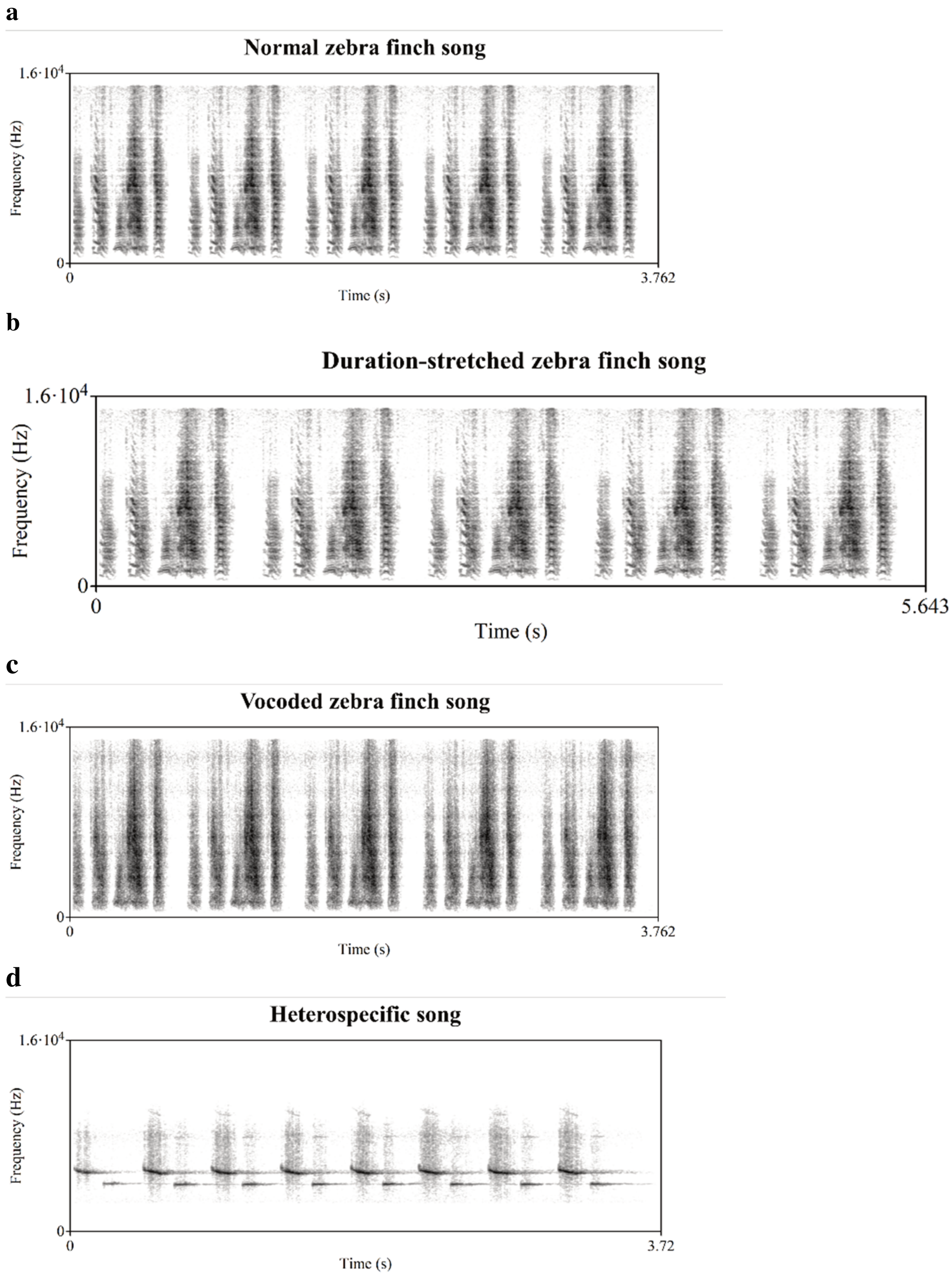


Figure 2. Spectrogram examples of four different stimuli used for the preference test. a - the original version of a **Normal zebra finch song.** The **Duration-stretched** version (b) stretched the normal song by 50%, and the **Vocoded** version (c) was produced by using the Matt Winn's Praat vocoded script. The **Heterospecific song** (d) was the song of a great tit of similar song duration but different in both spectral and temporal structure of elements compared to the original version of the zebra finch song.

Procedure

The experimental procedure consisted of two sessions: acclimatization and preference testing (Fig. 3). In the acclimatization session, each bird was placed into the carrousel at 17:00 the day before the preference testing and given ten hours overnight to acclimate. A five-seconds ambient sound recording of their aviary was triggered if the bird hopped on one of the sound perches before the light was turned off at 20:30. In this way, the birds were expected to learn that sound playback was associated with some of the perches before being transferred to the real testing session. The preference test lasted from 7:00 to 17:00 on the second day, and each of the four test stimuli was assigned to a specific sound perch. The stimuli were randomly assigned to the positions of sound perches, and the stimulus category assigned to a particular sound perch differed among birds. For each stimulus category, there were multiple songs, and each variant was only used for one or a few birds. These sound perch assignments were rotated one position clockwise every two and half hours, which constituted one "block", until each stimulus had been presented at each perch position once. Thus, the entire preference test was consisting of four blocks. Like the experiment by Wei *et al.* (2022), an additional loudspeaker (JBL Clip2) was placed on the floor directly under the carrousel broadcasting a continuous recording of ambient sound from the bird aviaries (45 dBA sound peak pressure level measured at the height of the perches inside the carrousel) to maintain the acclimation half hour before and during the session of preference testing. Ad libitum access to food and water was available throughout experiment, the birds were returned to their home aviaries after the testing was completed at 17:00.

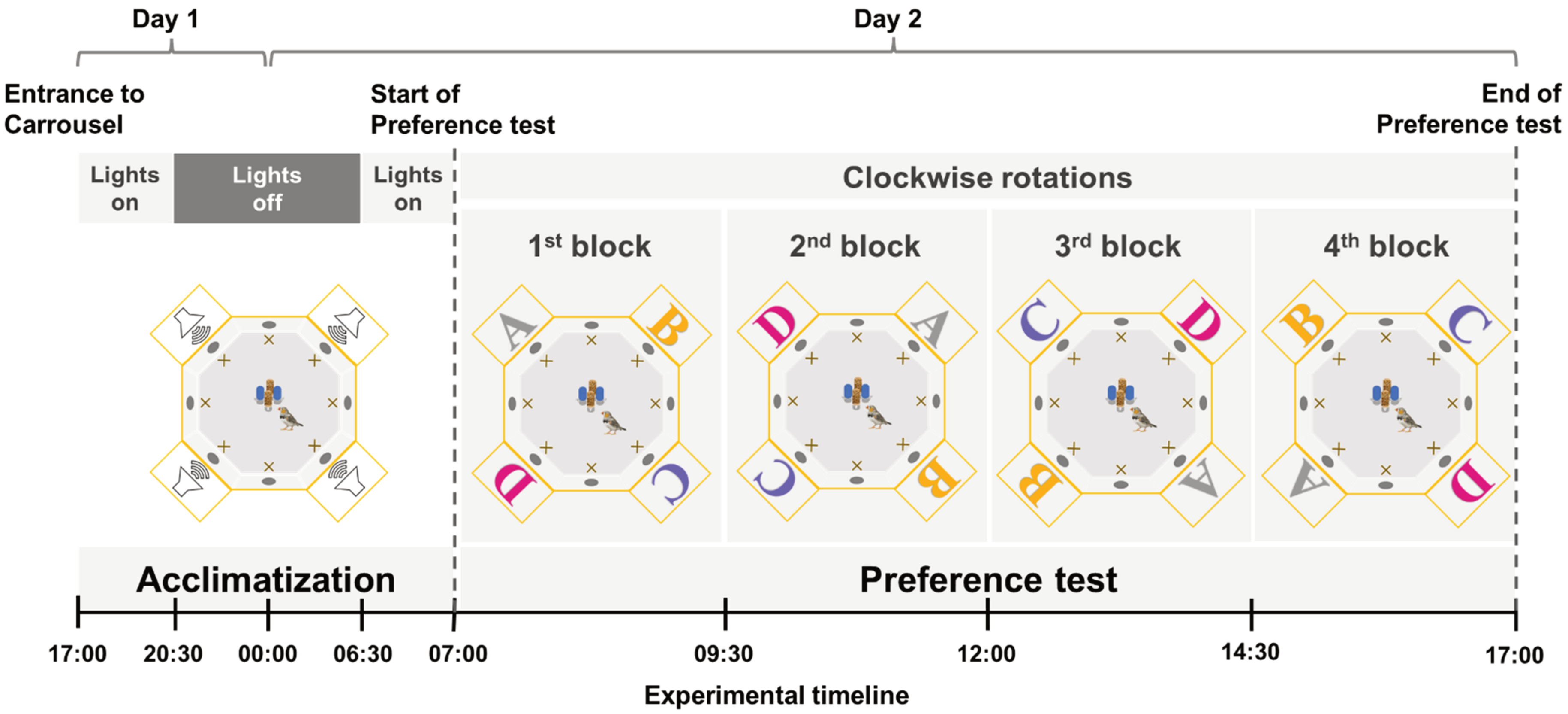


Figure 3. Experimental procedure. In the **Acclimatization session**, a bird was moved in the setup in the afternoon of the 1st day to acclimate the carrousel cage and the operant perches, and then to stay overnight. The **Preference test** session started in the morning of the 2nd day and lasted for ten hours; the stimuli (the labels A/B/C/D marked with different colors represent the different test sounds) associated with the four sound perches were rotated one position clockwise per each 2.5h block.

Analysis

The number of sound perch visits were chosen as a variable indicating a bird's preference toward a particular stimulus. For each bird, the number of visits on each sound perch was extracted from the data files registered individually by the logger software (original file format: text, transferred and saved into excel spreadsheets, v. 16.33). As reliable preferences can only be assessed in birds that visited all four perches giving them exposure to all the stimuli, our data analysis included only the counts of perch visits after a bird had hopped on all sound perches and had triggered all stimuli at least once (during the preference test session). Also, all stimuli had to have been visited at least once before 12:00 (at least underwent two full blocks) on day 2, and birds had to remain active and emit at least a total of 120 perch visits during the preference test session (average 30 perch visits per stimulus) to be included in the analysis. This resulted in 35 out of 45 birds reaching the criteria (with on average 851 ± 517 perch visits, range: 133–1942), which includes 16 males (with 670 ± 409 perch visits, range: 133–1509) and 19 females (with 1003 ± 547 perch visits, range: 156–1942).

A Generalized Linear Mixed-effects Model (GLMM) was used to assess the relative preferences that birds show among the 4 stimuli. This model analysis was conducted in Rstudio (R Core Team, 2016). We used the "Number of visits" for the various stimuli as response variable in GLMM. We used "Test_Stimulus" (Four versions of test stimuli), "Sex" (Male/Female), and the interaction between these two as covariates in the model with "Bird_ID" and "Stimulus_ID" as the random factors and a Poisson error structure of the "Number of visits". To assess whether any explanatory variables in a model were significant, we used Wald Chi-Squared tests to evaluate the fixed effects of the model (using the function car::Anova, R package car; Fox & Weisberg, 2019). A post-hoc analysis is performed on the final model with FDR correction (using the emmeans function, R package lsmeans; Lenth, 2016).

Ethics Statement

All animal housing, care, and use was performed with the permission of the Leiden Committee for Animal Experimentation (AVD number 1060020197507). None of 45 birds had any experience with this experimental setup or the stimuli preceding the experiment. Before being transferred to the setup, each experimental bird was co-housed with one other bird and underwent a physical examination. A brief removal from the social groups, which lasted no more than 24 hours, and behavioural observations in an open choice arena were considered non-procedural under the Experiments on Animals Act (Wod, 2014), which is the applicable legislation in the Netherlands, in accordance with the European guidelines (EU directive no. 2010/63/EU) for the protection of animals used for scientific purposes. The health and welfare of these birds was monitored by licensed and skilled personnel during the experiment, and all procedures were reviewed and monitored by the official Animal Welfare Body responsible for monitoring and implementing legal requirements. When showing signs of stress, deviating behaviour or illness a bird would be removed from the setup.

RESULTS

The ANOVA Type III showed that the number of perch visits differed significantly among the different test stimuli, and a trend towards a difference between males and females. Thus, the two factors “Test_Stimulus” and “Sex” as well as their interaction effects for the response variable “Number of visits” were selected as fixed factors for model analysis (see Table 1)

Table 1 ANOVA (Type III Wald chi square tests) table for selected GLMM

Variable MODEL: Number of visits	**Chisq.**	**Df**	**Pr(>Chisq.)**
Test_Stimulus	82.0014	3	< 2e-16 ***
Sex	3.4825	1	0.06202 .
Test_Stimulus:Sex	8.3438	3	0.03942 *

Note: GLMM with a Poisson distribution. All variables shown here were the variables of our interest and were thus used as fixed factors for the GLMM. “Bird_ID” and “Stimulus_ID” were included as two random intercepts. • $0.05 < p \leqslant 0.10$, * $0.01 < p \leqslant 0.05$, ** $0.001 < p \leqslant 0.01$, *** $p \leqslant 0.001$.

For the females, the perch visits to the Normal zebra finch song and its Duration-stretched version were significantly higher than to its Vocoded version and the Heterospecific song (both $p < 0.001$). In addition, the females visited the Vocoded zebra finch songs more often than the Heterospecific songs ($p < 0.05$). For the males, the perch visits to the Duration-stretched zebra finch song were significantly higher than to the Heterospecific song ($p < 0.01$), but there were no significant differences among the other stimuli. Overall females were more active than males and the pairwise comparisons per stimulus between males and females by the Post hoc Tukey's HSD tests showed a clear trend to a difference between two sexes’ towards more perch visits by females to both the Normal zebra finch song and its Duration-stretched version (both $p = 0.06$) (Fig. 4).

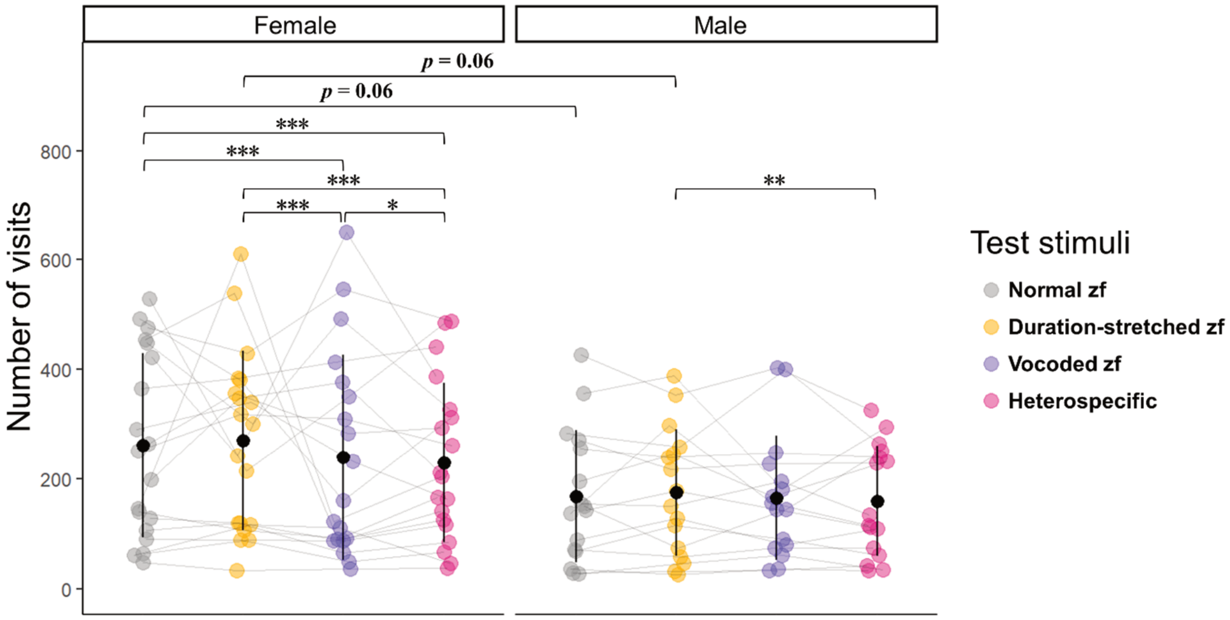


Figure 4. Number of sound perch visits for each sex group toward the test stimuli. *** refers to a significant difference of $p \leqslant 0.001$, ** refers to a significant difference of $0.001 < p \leqslant 0.01$, and * refers to a significant difference of $0.01 < p \leqslant 0.05$, for non-indicated comparisons p value is > 0.1. Error bars along with a black dot indicate Mean ± SD in perch visits to each test stimulus.

DISCUSSION

In this study, we examined the spontaneous preference of adult birds of both sexes for four different song stimuli. Both adult male and female zebra finches display a preference for normal or duration stretched conspecific songs over heterospecific ones. This finding is in line with other studies (Clayton & Pröve, 1989; Braaten & Reynolds, 1999; Lauay *et al*., 2004; Campbell & Hauber, 2009), confirming that song plays an important role in species/subspecies recognition. However, while other studies found no sexual differences in zebra finches' preference of conspecific over heterospecific songs (e.g., Braaten & Reynolds, 1999), we observed a strong trend towards a sex difference, with females showing stronger preferences than males for normal zebra finch songs and duration-stretched ones. The observed sex difference most likely reflects that sexual attraction plays a key role in female zebra finches' conspecific song preference, while the motivation of male zebra finches to respond to conspecific song may serve other social functions (e.g., Loning *et al*., 2023) different from the mate-choice driven motivation of females.

Our results further show that female zebra finches exhibited a gradient of preferences: they prefer conspecific songs that had not been manipulated in the spectral domain (nonmanipulated zebra finch song and duration-stretched zebra finch song), over the vocoded version of conspecific song, but vocoded song is preferred over the heterospecific great tit song. In contrast, the males don't differentiate among the stimuli, apart from a preference for the duration-stretched zebra finch song over heterospecific songs. The finding that males respond less than females and that their responses hardly differ towards the song variants makes it hard to conclude that males show a preference for a particular song variant over others. Overall, the results thus show that female zebra finch song preferences are more outspoken than male preferences, as well as more selective.

In our song discrimination study (see Ning *et al*., submitted - Chapter 2), we showed that both duration-stretched and vocoded song versions strongly reduced song recognition. However, in the current experiment, both manipulations differ substantially in their impact on the preference. Duration-stretched songs are at least as attractive as normal songs and preferred over vocoded ones by females and over heterospecific ones by males. Several earlier studies on song preferences showed that zebra finch females preferred longer or more complex songs (e.g., Clayton & Pröve, 1989; Neubauer, 1999), although this was not always the case (e.g., Holveck & Riebel, 2007; Riebel *et al*., 2009). Therefore Riebel (2009) argued that the duration of song is not a good explanatory variable for female song preference. However, in those studies longer songs differed in element composition and number from shorter ones, while in the current study, the duration-stretched manipulation on zebra finch songs didn't change the complexity in frequency domain or the number of syllables within the original song but stretched song duration proportionally, thus slowing down song tempo. This manipulation did not affect the preference for the song. The same manipulation did however affect the recognition of a training song when it was used in an operant discrimination task, indicating that it is a feature to which zebra finches can and do attend, but they apparently don't do this in the context of a preference test.

Vocoded songs are less attractive to females than normal songs. Such a response may be comparable to the finding that normal songs with noisier elements were found to be less attractive than those with a better-defined harmonic spectrum (Tomaszycki & Adkins-Regan, 2005). Nevertheless, vocoded songs arestill preferred over heterospecific songs. Noise elements with acoustic energy more or less uniformly distributed over all frequencies without any spectral structure and resembling a band of white noise have been reported to be natural element types present in zebra finch songs (Zann, 1993; Lachlan *et al*., 2016). Thus, although the noise-vocoded manipulation did result in substantial reduction of conspecific song recognition in both sexes in our discrimination studies (Chapter 2 & Chapter 4), natural songs modified in this way still retain sufficient specificity to be identified as a zebra finch song by female zebra finches. This may be based on the spectral envelop and temporal features (i.e., spectral contour) of the song syllables, which are still maintained after noise-vocoding. Nevertheless, it is clear that vocoded songs are less attractive than normal and duration stretched songs. So, while in our operant study both duration stretched and vocoded song versions were affecting song recognition, only vocoding had a negative impact on the preference.

To conclude, our study shows that modifying specific song features might have a different impact on song recognition than on song preference. Also, while no sex differences were observed in our operant song recognition study, they were present in the preference test. It demonstrates that results about the importance of particular song features as obtained in the context of an operant discrimination task and of a preference test paradigm can differ. Our results thus show that combining different approaches and testing paradigms may help to better understand the importance of various acoustical parameters in relation to the ecological functions and evolutionary processes of birdsong as relevant to males and females.

References

Adkins-Regan E., Krakauer A. (2000). Removal of adult males from the rearing environment increases preference for same-sex partners in the zebra finch. *Anim Behav.*, 60(1): 47–53. https://doi.org/10.1006/anbe.2000.1448

Amy M., Salvin P., Naguib M., Leboucher G. (2015). Female signalling to male song in the domestic canary, Serinus canaria. *R Soc Open Sci.*, 2(1): 140196. https://doi.org/10.1098/rsos.140196

Braaten R.F., Reynolds K. (1999). Auditory preference for conspecific song in isolation-reared zebra finches. *Anim Behav.*, 58(1): 105–111. https://doi.org/10.1006/anbe.1999.1134

Bregman M.R., Patel A.D., Gentner T.Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc Natl Acad Sci.*, 113(6): 1666–1671. https://doi.org/10.1073/pnas.1515380113

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. *J Comp Psychol.*, 133(1): 106-117. https://doi.org/10.1037/com0000143

Campbell D.L., Hauber M.E. (2009). Conspecific-only experience during development reduces the strength of heterospecific song discrimination in Zebra Finches (*Taeniopygia guttata*): a test of the optimal acceptance threshold hypothesis. *J Ornithol.*, 151: 379-389. https://doi.org/10.1007/s10336-009-0466-3

Chen Y., Clark O., Woolley S.C. (2017). Courtship song preferences in female zebra finches are shaped by developmental auditory experience. *Proceedings. Biological sciences*, 284(1855): 20170054. https://doi.org/10.1098/rspb.2017.0054

Clayton N., Pröve E. (1989). Song discrimination in female zebra finches and bengalese finches. *Anim Behav.*, 38(2): 352–354. https://doi.org/10.1016/s0003-3472(89)80096-x

Dooling R.J., Leek M.R., Gleich O., Dent M.L. (2002). Auditory temporal resolution in birds: Discrimination of harmonic complexes. *J. Acoust. Soc. Am.*, 112(2): 748-59. https://doi.org/10.1121/1.1494447

Elie J.E., Theunissen F.E. (2020). Author Correction: Zebra finches identify individuals using vocal signatures unique to each call type. *Nat. Commun.*, 11(1): 2914. https://doi.org/10.1038/s41467-020-16398-z

Fishbein A.R., Prior N.H., Brown J.A., Ball G.F., Dooling R.J. (2021). Discrimination of natural acoustic variation in vocal signals. *Sci. Rep.*, 11(1): 916. https://doi.org/10.1038/s41598-020-79641-z

Fox J., Weisberg S. (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fujii T.G., Coulter A., Lawley K.S., Prather J.F., Okanoya K. (2022). Song Preference in Female and Juvenile Songbirds: Proximate and Ultimate Questions. *Front Physiol.*, 13: 876205. https://doi.org/10.3389/fphys.2022.876205

Fujii T.G., Ikebuchi M., Okanoya K. (2021). Sex differences in the development and expression of a preference for familiar vocal signals in songbirds. *PloS one*, 16(1): e0243811. https://doi.org/10.1371/journal.pone.0243811

Fujii T.G., Okanoya K. (2022). Auditory and sexual preferences for a father's song can co-emerge in female Bengalese finches. *PloS one*, 17(3): e0254302. https://doi.org/10.1371/journal.pone.0254302

Geberzahn N., Derégnaucourt S. (2020). Individual vocal recognition in zebra finches relies on song syllable structure rather than song syllable order. *J Exp Biol.*, 223(9): jeb220087. https://doi.org/10.1242/jeb.220087

Gentner T.Q., Hulse S.H. (2000). Female European starling preference and choice for variation in conspecific male song. *Anim Behav.*, 59(2): 443–458. https://doi.org/10.1006/anbe.1999.1313

Gentner T.Q., Hulse S.H., Bentley G.E., Ball G.F. (2000). Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.*, 42(1): 117–133. https://doi.org/10.1002/(sici)1097-4695(200001)42:1<117::aid-neu11>3.0.co;2-m

Holveck M.J., Riebel K. (2007). Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. *Anim Behav.*, 74(2): 297-309. https://doi.org/10.1016/j.anbehav.2006.08.016

Holveck M.J., Riebel K. (2010). Low-quality females prefer low-quality males when choosing a mate. *Proc. R. Soc. B.*, 277(1678): 153–160. https://doi.org/10.1098/rspb.2009.1222

Holveck M.J., Geberzahn N., Riebel K. (2011). An experimental test of condition-dependent male and female mate choice in zebra finches. *PloS one*, 6(8): e23974. https://doi.org/10.1371/journal.pone.0023974

Houtman A.M. (1992). Female zebra finches choose extra-pair copulations with genetically attractive males. *Proc. R. Soc. B*, 249(1324): 3 - 6. https://doi.org/10.1098/rspb.1992.0075

Houx A.B., ten Cate C. (1999). Song learning from playback in zebra finches: Is there an effect of operant contingency? *Anim Behav.*, 57(4): 837–845. https://doi.org/10.1006/anbe.1998.1046

Lachlan R.F., van Heijningen C.A., Ter Haar S.M., ten Cate C. (2016). Zebra finch song phonology and syntactical structure across populations and continents—a computational comparison. *Front Psychol.*, 7:980. https://doi.org/10.3389/fpsyg.2016.00980

Lauay C., Gerlach N.M., Adkins-Regan E., Devoogd T.J. (2004). Female zebra finches require early song exposure to prefer high-quality song as adults. *Anim Behav.*, 68(6): 1249–1255. https://doi.org/10.1016/j.anbehav.2003.12.025

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. *Anim. Cogn.*, 21(4): 467–480. https://doi.org/10.1007/s10071-018-1182-2

Leadbeater E., Goller F., Riebel K. (2005). Unusual phonation, covarying song characteristics and song preferences in female zebra finches. *Anim Behav.*, 70(4): 909-919. https://doi.org/10.1016/j.anbehav.2005.02.007

Lenth R.V. (2016). Least-Squares Means: The R Package lsmeans. *J Stat Softw.*, 69(1): 1-33. https://doi.org/10.18637/jss.v069.i01

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): 239-51. https://doi.org/10.1037/0735-7036.120.3.239

Loning H., Verkade L., Griffith S.C., Naguib M. (2023). The social role of song in wild zebra finches. *Curr Biol.*, 33(2): 372–380.e3. https://doi.org/10.1016/j.cub.2022.11.047

Marler P., Peters S. (1988). The Role of Song Phonology and Syntax in Vocal Learning Preferences in the Song Sparrow, *Melospiza melodia*. *Ethology*, 77: 125-149. https://doi.org/10.1111/j.1439-0310.1988.tb00198.x

Mol C., Bolhuis J.J., Moorman S. (2021). Vocal learning in songbirds: the role of syllable order in song recognition. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 376(1836): 20200248. 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. *J Neurophysiol.*, 104(3): 1426–1437. https://doi.org/10.1152/jn.00028.2010

Neubauer R.L. (1999). Super-normal length song preferences of female zebra finches (*Taeniopygia guttata*) and a theory of the evolution of bird song. *Evol Ecol.*, 13(4): 365–380. https://doi.org/10.1023/A:1006708826432

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

Odom K.J., Hall M.L., Riebel K., Omland K.E., Langmore N.E. (2014). Female song is widespread and ancestral in songbirds. *Nat Commun.*, 5: 3379. https://doi.org/10.1038/ncomms4379

Prior N.H., Smith E., Lawson S., Ball G.F., Dooling R.J. (2018). Acoustic fine structure may encode biologically relevant information for zebra finches. *Sci. Rep.*, 8(1): 6212. https://doi.org/10.1038/s41598-018-24307-0

Riebel K. (2000). Early exposure leads to repeatable preferences for male song in female zebra finches. *Proc Biol Sci.*, 267(1461): 2553-8. https://doi.org/10.1098/rspb.2000.1320

Riebel K., Smallegange I.M., Terpstra N.J., Bolhuis J.J. (2002). Sexual equality in zebra finch song preference: evidence for a dissociation between song recognition and production learning. *Proc Biol Sci.*, 269(1492): 729–733. https://doi.org/10.1098/rspb.2001.1930

Riebel K., Smallegange I.M. (2003). Does zebra finch (*Taeniopygia guttata*) preference for the (familiar) father's song generalize to the songs of unfamiliar brothers?. *J. Comp. Psychol.*, 117(1): 61–66. https://doi.org/10.1037/0735-7036.117.1.61

Riebel K. (2009). Song and female mate choice in zebra finches: a review. *Adv. Stud. Behav.*, 40: 197-238. https://doi.org/10.1016/S0065-3454(09)40006-8

Riebel K. (2016). Understanding Sex Differences in Form and Function of Bird Song: The Importance of Studying Song Learning Processes. *Front. Ecol. Evol.*, 4: 62. https://doi.org/10.3389/fevo.2016.00062

Ritschard M., Riebel K., Brumm H. (2010). Female zebra finches prefer high-amplitude song. *Anim Behav.*, 79(4): 877-883. https://doi.org/10.1016/j.anbehav.2009.12.038

Rodríguez-Saltos C.A., Bhise A., Karur P., Khan R.N., Lee S., Ramsay G., Maney D.L. (2023). Song preferences predict the quality of vocal learning in zebra finches. *Sci. Rep.*, 13(1): 605. https://doi.org/10.1038/s41598-023-27708-y

Soha J.A., Marler P. (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Anim Behav.*, 60(3): 297–306. https://doi.org/10.1006/anbe.2000.1499

Spencer K.A., Wimpenny J.H., Buchanan K.L., Lovell P.G., Goldsmith A.R., Catchpole C.K. (2005). Developmental stress affects the attractiveness of male song and female choice in the zebra finch (*Taeniopygia guttata*). *Behav. Ecol. Sociobiol.*, 58(4): 423–428. https://doi.org/10.1007/s00265-005-0927-5

Stoddard P.K., Beecher M.D., Horning C.L., Willis M.S. (1990). Strong Neighbor-Stranger Discrimination in Song Sparrows. *Condor*, 92(4): 1051–1056. https://doi.org/10.2307/1368741

Suthers R.A., Vallet E., Kreutzer M. (2012). Bilateral coordination and the motor basis of female preference for sexual signals in canary song. *J. Exp. Biol.*, 215(17): 2950–2959. https://doi.org/10.1242/jeb.071944

ten Cate C., Verzijden M.N., Etman E. (2006). Sexual imprinting can induce sexual preferences for exaggerated parental traits. *Curr Biol.*, 16(11): 1128–1132. https://doi.org/10.1016/j.cub.2006.03.068

ter Haar S.M., Kaemper W., Stam K., Levelt C.C., ten Cate C. (2014). The interplay of within-species perceptual predispositions and experience during song ontogeny in zebra finches *(Taeniopygia guttata*). *Proc Biol Sci.*, 281(1796): 20141860. https://doi.org/10.1098/rspb.2014.1860

Tomaszycki M.L., Adkins-Regan E. (2005). Experimental alteration of male song quality and output affects female mate choice and pair bond formation in zebra finches. *Anim Behav.*, 70(4): 785–794. https://doi.org/10.1016/j.anbehav.2005.01.010

Vallet E., Kruetzer M. (1995). Female canaries are sexually responsive to special song phrases. *Anim Behav.*, 49(6): 1603–1610. https://doi.org/10.1016/0003-3472(95)90082-9

Vernaleo B.A., Dooling R.J., Leek M.R. (2010). Relative salience of acoustic features in zebra finch song. *J. Acoust. Soc. Am.*, 127(3): 1863. https://doi.org/10.1121/1.3384460

Vernaleo B.A., Dooling R.J. (2011). Relative salience of envelope and fine structure cues in zebra finch song. *J. Acoust. Soc. Am.*, 129(5): 3373–3383. https://doi.org/10.1121/1.3560121

Wang J., Liu J., Zhang Z., Ren H., Gao L., Hou J. (2019). Is male condition corrected with song features in Dusky Warblers (*Phylloscopus fuscatus*). *Avian Res.*, 10: 18. https://doi.org/10.1186/s40657-019-0158-5

Wei J., Liu Q., Riebel K. (2022). Generalisation of early learned tutor song preferences in female zebra finches (*Taeniopygia guttata*). *Behav Process*, 201: 104731. https://doi.org/10.1016/j.beproc.2022.104731

Wod. Dutch experiments on animals act, revised in 2014.

Woodgate J.L., Leitner S., Catchpole C.K., Berg M.L., Bennett A.T.D., Buchanan K.L. (2011). Developmental stressors that impair song learning in males do not appear to affect female preferences for song complexity in the zebra finch. *Behav. Ecol.*, 22(3): 566–573. https://doi.org/10.1093/beheco/arr006

Zann R.A. (1993). Structure, Sequence and Evolution of Song Elements in Wild Australian Zebra Finches, *The Auk*, 110(4): 702–715. https://doi.org/10.2307/4088626

Zann R.A. (1996). The Zebra Finch: a Synthesis of Field and Laboratory Studies. Oxford, Oxford University Press.

Zebra Finches Demonstrate Cognitive Flexibility in Using Phonology and Sequence of Syllables in Auditory Discrimination

Chapter 4

Zhi-Yuan Ning · Henkjan Honing · Carel ten Cate

This chapter is published in Animal Cognition
DOI: 10.1007/s10071-023-01763-4

ABSTRACT

Zebra finches rely mainly on syllable phonology rather than on syllable sequence when they discriminate between two songs. However, they can also learn to discriminate two strings containing the same set of syllables by their sequence. How learning about the phonological characteristics of syllables and their sequence relate to each other and to the composition of the stimuli is still an open question. We compared whether and how the zebra finches' relative sensitivity for syllable phonology and syllable sequence depends on the differences between syllable strings. Two groups of zebra finches were trained in a Go-Left/Go-Right task to discriminate either between two strings in which each string contained a unique set of song syllables ('Different-syllables group') or two strings in which both strings contained the same set of syllables, but in a different sequential order ('Same-syllables group'). We assessed to what extent the birds in the two experimental groups attend to the spectral characteristics and the sequence of the syllables by measuring the responses to test strings consisting of spectral modifications or sequence changes. Our results showed no difference in the number of trials needed to discriminate strings consisting of either different or identical sets of syllables. Both experimental groups attended to changes in spectral features in a similar way, but the group for which both training strings consisted of the same set of syllables responded more strongly to changes in sequence than the group for which the training strings consisted of different sets of syllables. This outcome suggests the presence of an additional learning process to learn about syllable sequence when learning about syllable phonology is not sufficient to discriminate two strings. Our study thus demonstrates that the relative importance of syllable phonology and sequence depends on how these features vary among stimuli. This indicates cognitive flexibility in the acoustic features that songbirds might use in their song recognition.

INTRODUCTION

Not only humans, but also songbirds learn their vocalizations early in life from their parents or other individuals. Vocal learning implies the presence of advanced auditory processing, including perception, memorization, and production of complex strings of sounds. Most emphasis in studies of vocal learning and auditory processing in birds is on the processes involved in learning the phonology, i.e., the spectro-temporal structure, of syllables, rather than on learning the syllable sequences (Vernes *et al*., 2021).

Songbird species show a large diversity in how syllables are arranged within songs. Some songbird species, such as the canary (*Serinus canaria*) (Lehongre *et al*., 2008), European starling (*Sturnus vulgaris*) (Eens, 1997), or willow warbler (*Phylloscopus trochilus*) (Gil & Slater, 2000) have a repertoire of syllables that are ordered in varying sequences to form phrases that together make up the song. The sequence of syllables sung within a given song is rarely an exact replicate of the previous song or of a sequence produced by the model from which the syllables are copied. This is in contrast to the vocalizations in species such as the white crowned sparrow (*Zonotrichia leucophrys*) (Soha & Marler, 2001), the chaffinch (*Fringilla coelebs*) (Riebel & Slater, 1999), song sparrow (*Melospiza melodia*) (Marler & Peters, 1987), or the zebra finch (*Taeniopygia guttata*) (Eales, 1985), in which songs consist of rather fixed sequences of syllables, and in which copied songs show limited element sequence divergence from the song models. The fact that these songbirds as well as others faithfully copy both the spectro-temporal structure of song syllables as well as their sequences, implies they have the ability to perceive and learn the phonology as well as the sequential order of conspecific syllables in great detail.

The zebra finch is an extensively used model species for comparative studies of vocal learning as well as auditory perception. With respect to sequence learning, despite the fact that zebra finches may have certain non-learned biases as to how different syllable types are distributed over a sequence (James & Sakata, 2017), there is ample evidence that syllable sequences are affected by learning (e.g., Eales, 1985).This is supported by the finding that zebra finch songs, both in captive and wild populations, show culturally transmitted differences in the position of specific syllable types, being more similar within than between colonies (Lachlan *et al*., 2016). Also, zebra finches first exposed to one set of syllables in a particular sequence and next exposed to a novel set, first acquire the phonological structure of the novel syllables and next adjust the sequence of these novel syllables, indicating the involvement of at least partially different learning processes (Lipkind *et al.*, 2013). Comparable evidence of a separation between learning the phonology of syllables and learning of their sequence can also be found on other songbirds, such as the white crowned sparrows (e.g., Soha & Marler, 2001; Plamondon *et al.*, 2010).

The finding that zebra finches attend to and learn about both phonology and syllable sequence demonstrates that both are perceived and suggests that they both are relevant for communication, for instance to distinguish between individuals. However, experiments addressing which song features zebra finches use to discriminate between songs suggest a striking imbalance between the role of syllable phonology and the role of syllables sequence. For instance, Braaten *et al.* (2006) used an operant discrimination task (Go/Nogo) to train adult and juvenile zebra finches to discriminate the natural forward song from its reversed version (i.e., a song played backwards). Tests in which a song was presented with syllables of non-reversed phonological structure in the reversed sequence and a song in which element sequence was maintained, but the syllables were reversed, showed that the original stimuli were discriminated on the phonological structure of the syllables and not by their sequence. A recent study, also using a Go/Nogo task, investigated the role of syllable sequences versus spectro-temporal fine structure of syllables for the process of individual recognition: zebra finches were trained to discriminate songs of one male conspecific from those of four others; thereafter they were exposed to hybrid stimuli combining the syllable sequences of one individual with the spectro-temporal features of another. The results demonstrated that zebra finches mainly rely on spectro-temporal details of syllables and pay less attention to syllable sequences (Geberzahn & Derégnaucourt, 2020). A laboratory playback experiment (Mol *et al.*, 2021) also suggested that syllable sequence is not an essential cue for recognition of familiar songs in zebra finches. In another study, Lawson *et al.* (2018) used a discrimination task to compare the ability of zebra finches to notice changes of syllable phonology and changes of syllable sequence in the motifs of natural songs. These results also showed that zebra finches could readily recognize the reversal of a single syllable in the motif, but largely ignore the change of syllable sequence in the motif. Similarly, zebra finches detect single syllable reversals more easily than a doubling of an inter-syllable interval (e.g., Dooling & Prior, 2017). Combined with evidence that zebra finches can detect differences between renditions of slightly different versions of the same song syllables (Fishbein *et al.*, 2021), demonstrating the attention to fine details of the spectro-temporal structure of syllables, such findings raised the question to what extent zebra finches attend to the sequences of syllables (Fishbein *et al.*, 2019).

Some studies have indicated that syllable sequence can play an additional role in song recognition. Lawson *et al.* (2018) showed that male zebra finches tested with their own songs or with those of familiar birds attended to sequences of syllables in addition to the spectro-temporal structure of these syllables. So, although zebra finches may thus show a strong bias to attend to spectro-temporal features of syllables to distinguish songs, they can also attend to syllable sequence. It suggests that more extensive experience with songs is needed before the birds acquire knowledge about syllable sequences. This was also suggested by an experiment by Braaten *et al.* (2006) showing that juvenile zebra finches could discriminate songs on the basis of syllable sequence alone, although this discrimination was more difficult to obtain than one based on syllable structure. However, in contrast to the studies indicating a marginal role of syllable sequences in song discrimination and suggesting that learning about sequences might be more difficult than about syllable phonology, a range of studies demonstrated that zebra finches can readily learn to distinguish strings consisting of identical syllables but differing in their sequence (e.g. van Heijningen *et al.*, 2013; Chen & ten Cate, 2015, 2017; Chen *et al.*, 2016; Spierings & ten Cate, 2016; Knowles *et al.*, 2018). In a study by van Heijningen *et al.* (2009), zebra finches were trained in a Go/Nogo task to discriminate between stimuli in which syllables were arranged in an ABAB or an AABB sequence. They readily acquired this discrimination. When next tested with stimuli of the same sequential structures but constructed of novel exemplars of the same type of syllables (and hence differing in fine spectro-temporal details), they generalized the discrimination to the novel exemplars based on the string structure. Evidence from a neural study (Cazala *et al.*, 2019) also using an AABB vs ABAB paradigm demonstrated that the caudomedial nidopallium (NCM) neurons encode the sequencing of syllables, which also supports the outcome of the behavioural studies described above in showing that zebra finches have no difficulty in distinguishing two strings by the sequence of their syllables. Zebra finches can thus readily use sequence information to distinguish strings differing in their sequence only.

The findings discussed above raise the question how learning about the spectro-temporal characteristics of syllables and about syllable sequences relate to each other and to the composition of the stimuli. The range of experiments mentioned above differ in methods and stimulus composition. So far, no experiment has directly compared the relative importance of spectral structure and sequence when zebra finches have to discriminate two syllable strings that either consist of different sets of syllables or consist of the same set of syllables, but different in the sequence, using similarly structured strings and identical training and testing procedures.

In the current study we use an operant discrimination paradigm - the Go-Left/Go-Right task - to examine the relative salience of syllable phonology and syllable sequence when zebra finches must distinguish two artificially constructed ‘song motifs’ that are either composed of different syllable types (the ‘Different-syllables group’), or two stimuli composed of the same set of syllables but differing in sequence (the ‘Same-syllables group’). We investigate whether the stimulus contrast in the training affects the ease of learning by examining the speed with which the discrimination is achieved. Next, we assess to what extent the birds in the two groups attend to the syllable phonology by assessing the responses to test strings consisting of reversed syllables or of vocoded versions of these syllables. To examine the importance of syllable sequence, we assessed the responses to test strings in which the sequences are shuffled.

METHODS

Subjects

Twenty-four zebra finches (12 males, and 12 females; ages 139 – 691 days post-hatching) were used in this experiment. All birds originated from the in-house breeding colony at Leiden University. Before the experiment, the birds lived in single-sex groups of about 15 to 30 individuals in aviaries (2m × 2m × 1.5m), in which food and water were available ad libitum.

The birds were divided randomly in two experimental groups; half of the birds were assigned to the Different-syllables group, and the other half of them to the Same-syllables group (6 males and 6 females in each group; age Different-syllables group: M=309, SD=184, age Same-syllables group: M=387, SD=246). Each group was trained to discriminate between two different strings consisting of five zebra finch syllables. Within each training group one half of the birds got training strings consisting of single-element syllables, and the other half another set of stimuli consisting of one complex syllable and four single-element syllables within a string.

Operant conditioning cage

The birds were trained and tested individually in an operant conditioning cage (Skinnerbox) (70×30×45 cm) using a Go-Left/Go-Right paradigm for training and testing. A cage contained 3 pecking keys (sensors) with a red LED light at the top/bottom of each sensor (Fig. 1a). 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-h/10.5-h light/dark schedule. Sound stimuli were played through a speaker (Vifa MG10SD09–08) 1 meter above the Skinnerbox. The volume of the speaker was adjusted to ensure that the sound amplitude in the Skinnerbox was approximately 65 dB (measured by an SPL meter, RION NL 15, RION). Sensors (S1, S2, S3), lamp, food hatch and speaker were connected to operant conditioning controller that also registered all sensor pecks.

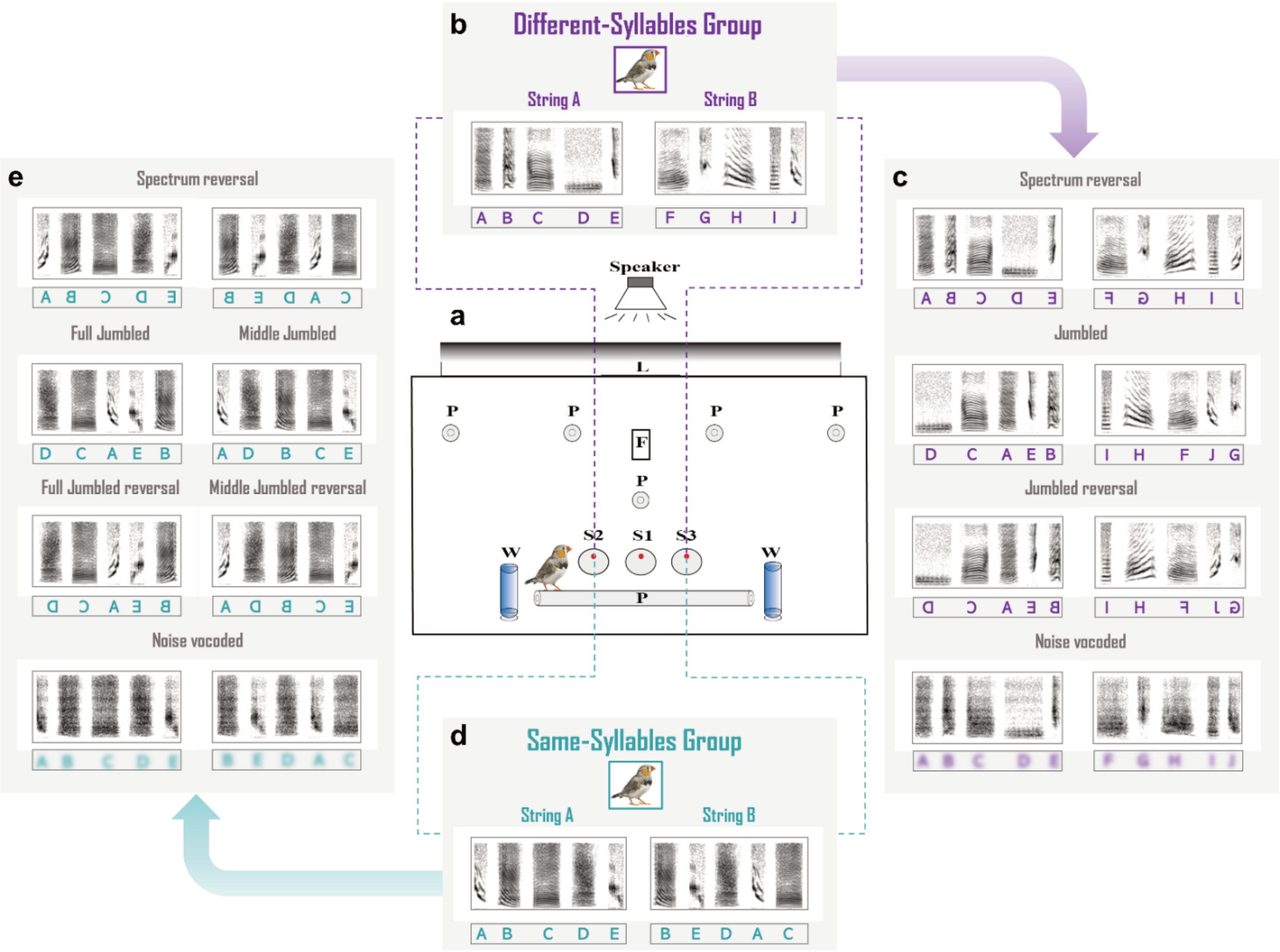


Figure 1. (a) Schematic view of the operant conditioning apparatus (Skinner box) used for the experiment. A speaker (S) is suspended from the ceiling above the cage. Within the cage, there are several perches (P) for the bird to sit on, a food hatch (F) is located in the upper middle of the back panel, a lamp (L) is placed at the top of the cage. Two tubes of ad libitum water (W) are placed symmetrically on two sides of the cage, three response keys (S1, S2, S3) with signal LEDs are lined horizontally in the lower middle of the back panel. **(b) An example of a pair of training strings for the Different-syllables group.** The birds of the Different-syllables group were trained with stimuli consisting of different syllable types: for instance, String A was the sequence of syllables A B C D E, while String B was the sequence of syllables F G H I J. **(c) Modified stimuli used in the testing phase for the Different-syllables group.** The birds of the Different-syllables group were tested with 4 modified versions of each training stimulus after completion of the training - see text for a description of these manipulations. **(d) A pair of training strings for the Same-syllables group.** For birds of the Same-syllables group, training stimuli consisted of the same syllables but arranged in different sequences: for instance, String A and String B consisted of the same five syllables A B C D E, but the sequences of these syllables were different between the two strings. **(e) Modified stimuli in the testing phase for the Same-syllables group.** These birds were also tested with 4 similarly modified versions of each training stimulus.

Stimuli

Training stimuli

Zebra finch syllables were selected from representative song recordings of adult males of the laboratory colony at Leiden University. The songs had not been heard before by the birds. Each string was composed of syllables belonging to different types, based on several distinctive acoustic features like the duration and spectral shape, mainly guided by the descriptions of syllable types in Lachlan *et al.* (2016). Each training string was thus consisting of five song units, each of which belonged to one of in total 13 types of single-element syllables and 8 types of complex syllables. Each bird got different combinations of syllable types as training stimuli.

The five syllables within one string were normalized in root-mean-square (RMS) amplitude and separated by a 30 ms silent interval between each two syllables to form a natural song-syllable string. The training stimuli in this experiment were 24 stimulus pairs (12 pairs for each training group), each consisting of two different strings. For the Different-syllables group, each bird was presented with a stimulus pair of which the two strings consisted of different syllable types (Fig. 1b). For the Same-syllables group, each bird was presented with a stimulus pair of which the two strings were consisting of a same set of syllables but arranged in a different sequence (Fig. 1d). To this end, we altered the syllable sequences of string A (indicated as “A-B-C-D-E”) into a different sequence “B-E-D-A-C” to construct the string B, which also avoids bigrams of syllables from string A.

When played, the strings were normalized such that the average intensity (RMS, calculated over the total duration of the stimulus) was the same for the two strings within a pair to avoid that amplitude differences affected the responses to the stimuli. The range of variation in volume recorded at the microphone was preserved. All stimuli were filtered to a bandwidth below 15kHz. All training stimuli were cut, synthesized, and filtered using Praat (version 6.1.12). The amplitude of each stimulus was adjusted by using the “Normalize” function in Audacity (version 2.3.0).

Test stimuli

To test the impact of spectral and sequential information that the birds used to discriminate the training strings, they were tested with modified versions of the training strings (Fig. 1b & 1d). We used Praat to modify each original training string to produce a version in which either the spectral features or the sequence of syllables was changed. For each training group modified stimuli were changed in an identical way (some examples of the training and test stimuli are provided as supplementary material):

- Spectrum reversal – The spectrum of each syllable in the string was reversed, but the sequence of the syllables was identical to the order in the training version. We used the “reverse selection” option in Praat to reverse the spectrum of each syllable of a training string without changing the initial order.

- Jumbled – The sequence of the syllables in the training strings of both training groups were altered from “A-B-C-D-E” to “D-C-A-E-B”. For instance, if the syllable sequence of the string A in the Different-syllables group is “A-B-C-D-E”, then the order manipulated version becomes “D-C-A-E-B”, and the manipulated version of string B (the original sequence “F-G-H-I-J”) becomes “I-H-F-J-G”. Thus the “Jumbling” was applied to both string A and string B in the Different-syllables group (Fig. 1c). Likewise, this modification was applied in the Same-syllables group, by which the sequence-manipulated version of string A became “D-C-A-E-B”, and the sequence manipulation of string B became “A-D-B-C-E”. Note that this means that the manipulated string B now has the same 1st and 5th syllables as present in training string A (“A-B-C-D-E”), since training string A and string B consisted of the same syllables.

Therefore, for the Same-syllables group, we distinguished in our analysis between the responses to “D-C-A-E-B”, which will be indicated as the “Full jumbled” test string and “A-D-B-C-E” which will be indicated as “Middle jumbled” test string, and we relate the responses to these test stimuli to the responses to training string A (Fig. 1e).

- Jumbled + Spectrum reversal – This manipulation was the combination of the above Jumbled alteration and Spectrum reversal. Both the spectrum of syllables and their sequence were changed (Fig. 1b and Fig. 1d).

- Vocoded – This modification maintains the spectral (and temporal) envelope of the syllables within the string, but averages the energy within specific frequency bands, thus removing any harmonic structure. To construct these stimuli, we used the Matt Winn's Praat vocoded script (http://www.mattwinn.com/praat/vocode_all_selected_v45.txt) to synthesize a vocoded morph of training strings. The script was set to divide cut-off frequency bandwidths equally for 15 bands contiguous with smooth transitions (1000Hz bandwidth for one noise-vocoded band).

Procedure

We used a Go-Left/Go-Right paradigm for training and testing (Fig. 1a). The training consisted of several phases.

Acclimation and pre-training

In the acclimation phase, the birds were moved to the Skinner boxes. The food hatch remained open, so food was freely accessible in a container behind the hatch. The LED lights on the pecking sensors were on. The goal of this phase was to acclimate the bird to the cage and to show it where to find food. The bird might also already learn to peck sensors spontaneously. If in this stage the central sensor, S1, was stimulated by pecking, it would play sound string A or sound string B with a 50% chance on each. The side sensor S2 produced one of the two training strings, and the other side sensor S3 produced the other string. The LEDs of all three sensors were illuminated to attract the attention from the bird. After a few hours to one night of acclimation, the pre-training phase started by closing the food hatch. In this phase, the food hatch was closed, and the bird had to learn 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. Once the bird started to peck all the sensors regularly 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, and then to peck S2 or S3, depending on the playback sound. If the bird pecked the sensor that was linked to the stimulus being played, this was rewarded with 12 sec access to food. If the wrong sensor was pecked the light was off for 1 sec. Before any sensor was pecked, only the S1 LED was on. If the bird did not respond within 15 seconds, a trial would end automatically without food reward or light-off penalty. The duration of this phase varied from bird to bird (range: 5 – 32 days). The proportion of correct responses (see ‘Analysis’ section below for calculation of the ‘Correct rate’) was calculated on a daily basis as the individual's discrimination rate among the training stimuli.

Transition phase

When a bird learned to associate the two training sounds with the corresponding sensors and had reached a Correct rate for the training stimuli greater than 0.75 for three 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, in which the reinforcement by food reward or darkness was reduced to occur randomly on 80% (instead of 100%) of trials. On the remaining 20% of trials, the responses were not reinforced, and the trial ended after 15 sec. If the bird kept the same level of discrimination for two days, the test phase began.

Probe testing phase

In this phase, 20% of the pecks on S1 resulted in presenting one of 10 test stimuli. These10 test stimuli were never reinforced and were randomly interspersed between training stimuli. Eight of these were modified versions of the training stimuli (four modified versions of stimulus A and four of stimulus B). The other two were non-reinforced training stimuli. The remaining 80% were training stimuli with reinforcement. Testing continued until each test stimulus had been presented 40 times to a bird. After reaching this, the bird was transferred back to its aviary. The order of stimulus presentation was random across subjects.

Analysis

For 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 two-tailed unpaired t-test (using the t-test function in GraphPad Prism 9.1.1) was used to detect differences between the two training groups.

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 a 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 'no-response' (not pecking a key). For the statistical analyses, we examined the proportion of 'correct responses' out of 'correct + incorrect responses' (Correct rate = Count_Correct / (Count_Correct + Count_Incorrect)), as well as the proportion of responses calculated as 'correct + incorrect responses' to modifications of sound A plus those to modification of sound B, as the proportion of the 40 presentations of each test stimulus (Response rate = (Count_Correct + Count_Incorrect) / (Count_Correct + Count_Incorrect + Count_NoResp)). In addition, we examined whether the individual test stimuli were discriminated above chance.

We used Generalized Linear Mixed-effects Models (GLMMs) to examine the discrimination of various test sounds by the birds. All model analyses were conducted in Rstudio (R Core Team, 2016). We calculated the 'Correct rate' and the 'Response rate' based on the counts of 'correct response', 'incorrect response', and 'no response', combining the response counts to (variants of) Training strings A and B, using the function cbind, R package mice; Van Buuren & Groothuis-Oudshoorn, 2011, and used these two rates as response variables in GLMMs in R (using the function glmer, R package lme4; Bates *et al.*, 2015). We used 'Training_Group' (Same or Different syllables), 'Test_Treatment', and the interaction between these two as covariates in the full model with 'Bird_ID', 'Age', 'Number_of_Training_Trials' as the random factors and a binomial error structure of the 'Correct rate' and the 'Response rate'. The best model was chosen based on corrected Akaike criterion (AICc) provided by dredge model selection (using the function Dredge, R package MuMIn; Bartoń, 2020). The model with the smallest value of AICc was considered to be the best model by default, but if 'Training_Group' was not part of the best model, we kept it in the final model anyway because this was a variable of our interest. To determine the effect and significance of the covariates, we ran the final models and, if applicable, used Post hoc Tukey's HSD tests to make pairwise comparisons of the test treatments (using the emmeans function, R package lsmeans; Lenth, 2016), with false discovery rate (FDR) correction of p-values (Benjamini & Hochberg, 1995) for multiple comparisons.

In the above model, the counts of the responses to (modifications of) both string A and string B were combined in all tests. This included the two test treatments 'Jumbled' and 'JumbledReversal' for both string A and B in the Same-syllables group. However, as outlined above, the jumbling of the syllables resulted in making the jumbled version of string B partly similar to training string A, and we therefore used string A as reference in this case. Because jumbling the strings for the Same-syllables group thus resulted in half of the jumbled strings being fully jumbled and the other half being middle jumbled, we also did a separate analysis for the data set of two Jumbled versions (MiddleJumbled/FullJumbled) in the Same syllables training group. In this analysis we compared the responses to training string A with those to the FullJumbled version of string A and those to the MiddleJumbled version in which the 1st and 5th syllables of the test string are the same as those of the training string A. In this analysis, 'Test_Treatment' was used as a fixed effect in the full model to gain insight into a possible comparison among three different stimuli versions (Training/MiddleJumbled/FullJumbled). The 'Bird_ID', 'Age', and 'Number_of_Training_Trials' were included as the random factors. Here we also used a model with binomial error structure of the Correct and the Response rates.

To examine whether the birds responded above chance (50%) to each of the testing stimuli, we applied a log(correct/incorrect) as the response variables against a log (Odds-ratio) = 0 in a GLM. If correct/incorrect = 1, then the probability of observing a correct response is as large as the probability of observing an incorrect response, representing both probabilities are 0.5, then log (Odds- ratio) = log (1) = 0. Therefore, comparing the outcomes of the Binomial GLM to 0 is comparing the results to the 50% chance for a correct response.

Ethics Statement

All animal housing, care, and use was approved by the national *Centrale Commissie voor Dierproeven* (CCD) of the Netherlands and the Leiden University Animal Welfare Body (AVD number 1060020197507). None of 24 birds had any experience with this experimental setup or the stimuli preceding the experiment. Each experimental bird underwent a physical examination before being transferred to the Skinnerboxes. During the experiment, the health and welfare of these birds was monitored daily. The food intake of the birds was monitored daily, and additional food was given when there were signs of a low food intake.

RESULTS

Learning speed

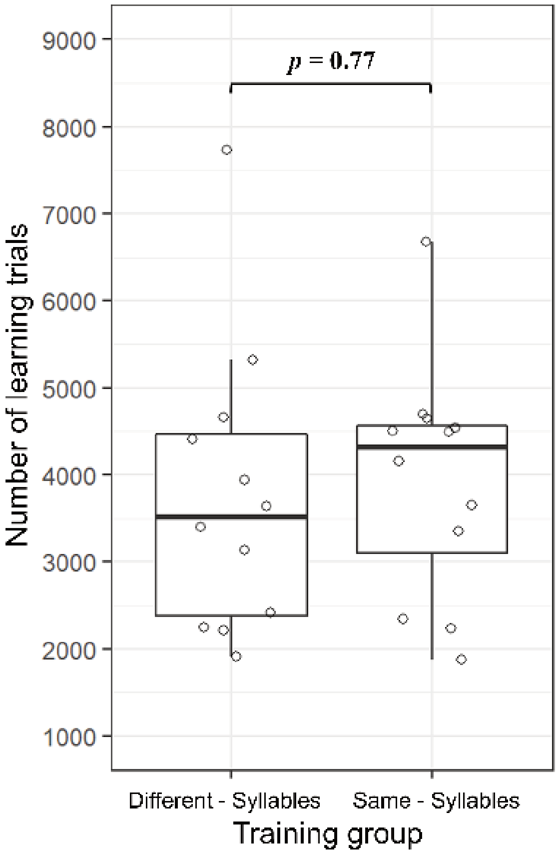


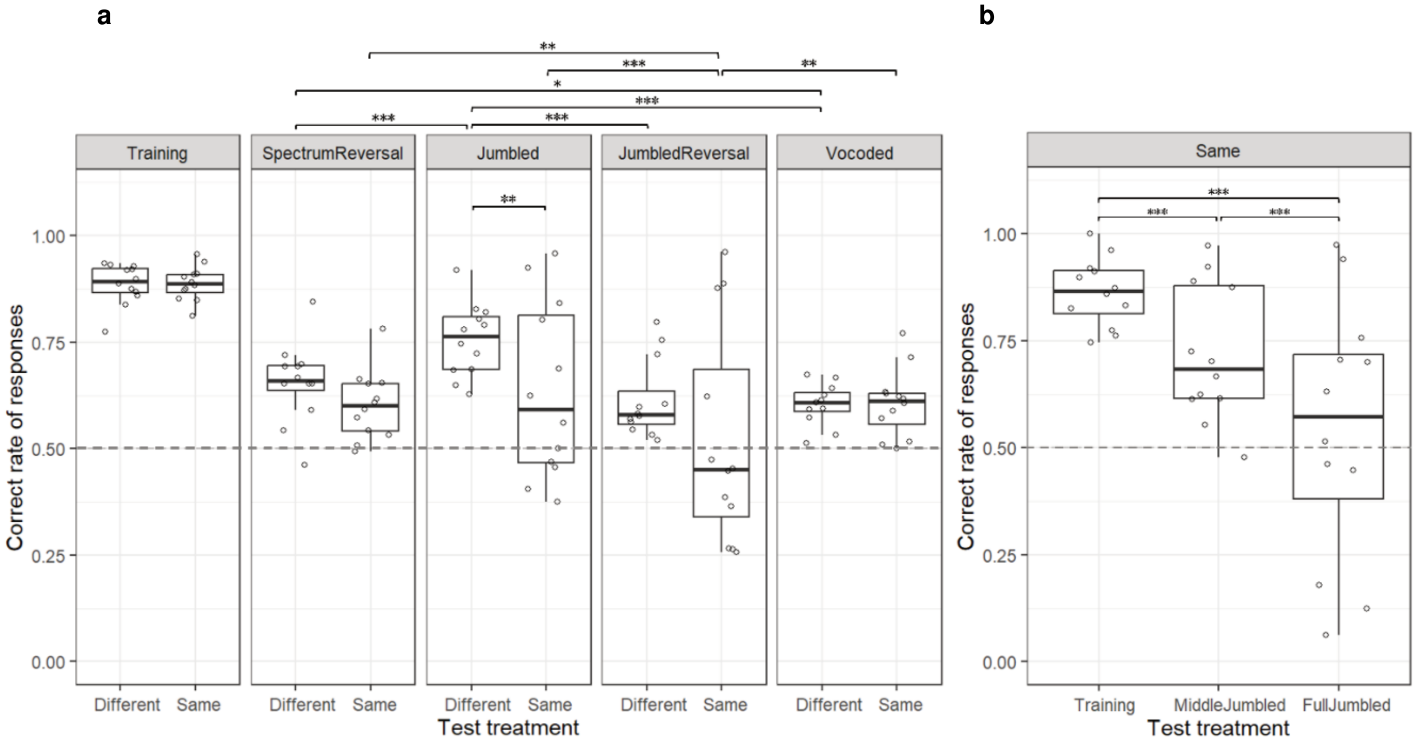
Figure 2. Number of learning trials needed to reach the learning criterion. Individual zebra finch results are shown with open circles. There is no significant difference between the Different-syllables group and the Same-syllables group in learning speed. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range.

The discrimination training lasted until the birds reached the learning criterion of over 75% correct responses to both sound A and sound B for three successive days. All twenty-four birds finished the training and reached the learning criterion in on average 3842 (SD = 1442, N = 24) trials. No significant difference ($p = 0.7733$, $t = 0.2916$, $df = 22$; Fig. 2) was found between the Different-syllables group (M = 3753, SD = 1579) and the Same-syllables group (M = 3932, SD = 1283). It suggests that birds from two training groups learned approximately equally fast.

Do training groups differ in responses to test stimuli?

We compared the Correct rates and Responses rates to the training and various test stimuli between both experimental groups (Fig. 3). For the Correct rate , the best model (model 1) was chosen based on AICc (Table 1). For the Response rate, we chose the model 3 with the same factors as model 1 for the Correct rate. It was not the most recommended model by the dredge model selection, but it contained the variables of our interest and was also close to the most recommend model (AICc = 723.1, delta = 7.41, Table 1).

The only significant difference between the two training groups concerns the Correct rate for the Jumbled version (Different – Same = 0.534 ± 0.173, *p* = 0.01, Table 2). There were no significant differences in the Correct rate for any of the other test stimuli between the two training groups (Fig. 3a). Note that the variation in Correct rate for the Jumbled test stimuli in the Same-syllables group is much larger than that for other test stimuli, which is caused by combining the responses to both the ‘Middle Jumbled’ and ‘Full Jumbled’ test stimuli (see below for the analysis separating among these stimuli). There were no significant differences in Response rates for any of the stimuli between two training groups (Fig. 3c).



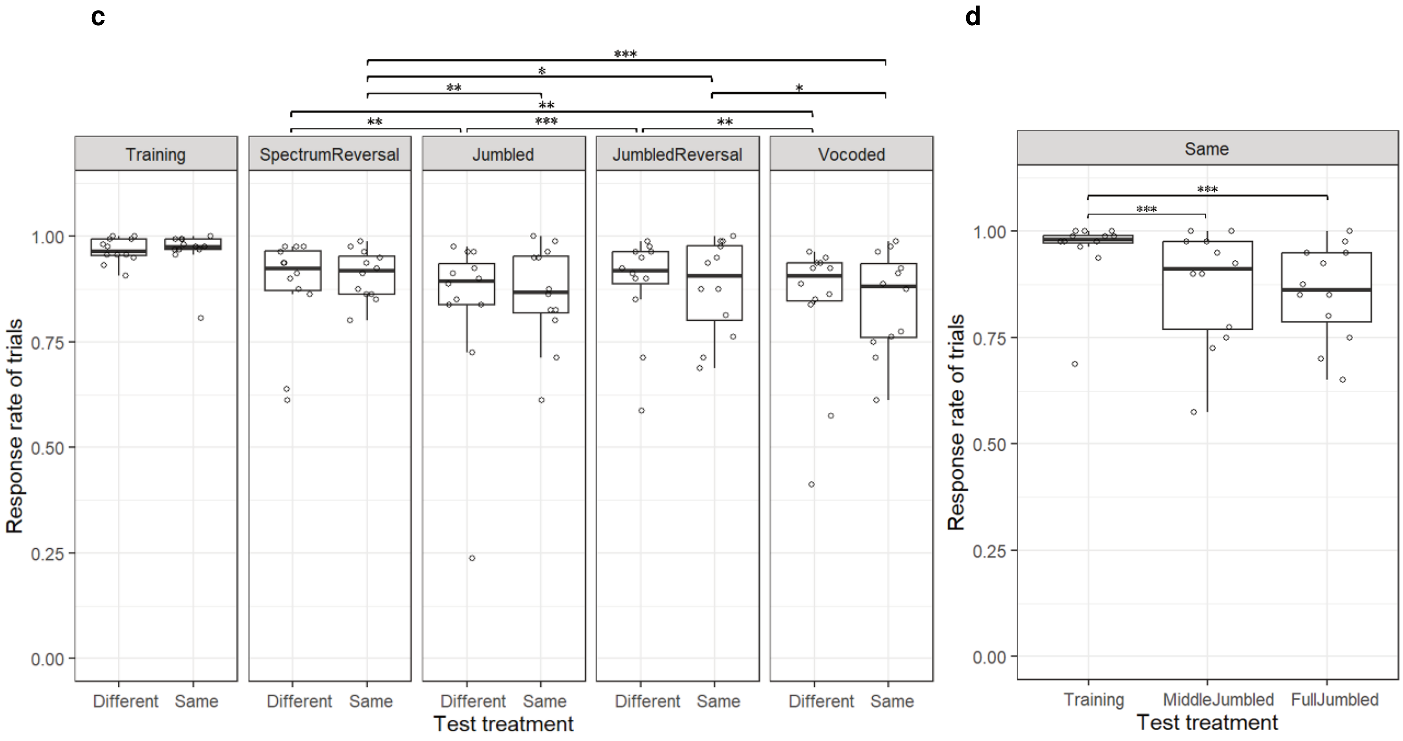


Figure 3. Correct rate of responses and Response rate of trials a) the proportion of correct responses (Correct rate) to the training and test stimuli for the two training groups ; b) the Correct rate of responses to the training stimulus and the two Jumbled versions for the 'Same-syllables' training group; c) the Response rates to the training and test stimuli for the two training groups; d) the Response rates to the training stimulus and the two Jumbled versions for the 'Same-syllables' training group. All test stimuli got significantly lower Correct rates and significantly lower Response rates than the training stimuli. Significant differences between the responses to the various test stimuli and between the training groups are indicated: *** refers to a significant difference of $p \leq 0.001$, ** refers to a significant difference of $0.001 < p \leq 0.01$, and * refers to a significant difference of $0.01 < p \leq 0.05$, for non-indicated comparisons p value is > 0.05. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range. The dashed line represents chance level, which was 50% for both tasks.

Table 1 Summary of the GLMs selection for (a) the proportion of correct responses if birds respond to one of two sounds; and **(b)** the proportion of trials that birds respond with pecking A or B

<table>
<tr><th>Model</th><th>df</th><th>logLik</th><th>AICc</th><th>Δi</th><th>wi</th></tr>
<tr><td>a. Correct rate of responses (sound A+B combined)</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1* Training_Group + Test_Treatment + Test_Treatment:Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>13</td><td>-481.009</td><td>991.5</td><td>0.00</td><td>0.964</td></tr>
<tr><td>2 Training_Group + Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_training_trials)</td><td>9</td><td>-489.805</td><td>999.2</td><td>7.79</td><td>0.020</td></tr>
<tr><td>3 Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>8</td><td>-491.141</td><td>999.6</td><td>8.13</td><td>0.017</td></tr>
<tr><td>4 Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>5</td><td>-789.401</td><td>1589.3</td><td>597.88</td><td>0.000</td></tr>
<tr><td>null (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>4</td><td>-790.736</td><td>1589.8</td><td>598.37</td><td>0.000</td></tr>
<tr><td>b. Response rate of trials (sound A+B combined)</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1 Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>8</td><td>-349.172</td><td>715.6</td><td>0.00</td><td>0.748</td></tr>
<tr><td>2 Training_Group + Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>9</td><td>-349.165</td><td>718.0</td><td>2.33</td><td>0.234</td></tr>
<tr><td>3* Training_Group + Test_Treatment + Test_Treatment:Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>13</td><td>-346.811</td><td>723.1</td><td>7.41</td><td>0.018</td></tr>
<tr><td>4 Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>5</td><td>-469.717</td><td>950.0</td><td>234.32</td><td>0.000</td></tr>
<tr><td>null (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)</td><td>4</td><td>-469.723</td><td>947.8</td><td>232.15</td><td>0.000</td></tr>
</table>

Note: Best four models of the model selection (ranked by AICc and logLik) and the null models. The Akaike weight (wi) indicates the probability of a better model in the model candidates set, and Delta AICc (Δi) was used to show the difference in AICc score between the best model and the model being compared. A * indicates the model we choose. Only information related to both sound A and sound B were shown here, the information about the two Jumbled versions in the 'Same-syllables' training group is not displayed in this table.

Do different test stimuli give rise to different responses?

The highest Correct and Response rates are present for the non-rewarded training stimuli. Thus, in both training groups all modifications affected the birds' responses (see Table 2). For the comparisons of responses to different test stimuli within each training group, Post hoc Tukey's HSD tests (Table 2) showed that the birds responded with a higher Correct rate and a higher Response rate to the training stimuli compared to all four testing stimuli in both training groups (Fig.3a and Fig.3c). The tests also showed that the birds of the Different-syllables training group responded with a significantly higher Correct rate to the Jumbled stimuli than to the JumbledReversal, the Vocoded stimuli and the SpectrumReversal stimuli (both $p < 0.0001$), and with a significantly higher Correct rate to the SpectrumReversal stimuli than to the Vocoded stimuli ($p < 0.05$), while the birds of the Same-syllables training group responded with a significantly lower Correct rate to the JumbledReversal stimuli than to the Jumbled stimuli ($p < 0.0001$), the Vocoded stimuli and the SpectrumReversal stimuli (both $p < 0.01$).

The birds of the Different-syllables training group had lower Response rate to the Jumbled stimuli and the Vocoded stimuli than to the JumbledReversal ($p < 0.01$), and had a significantly higher Response rate to the SpectrumReversal stimuli than to the Jumbled stimuli and the Vocoded stimuli (both $p < 0.01$), while the birds of the Same-syllables training group had significantly lower Response rate to the JumbledReversal ($p < 0.05$), the Vocoded ($p < 0.0001$) and the Jumbled stimuli ($p < 0.01$) than to the SpectrumReversal stimuli, and had a significantly higher Response rate to the JumbledReversal than to the Vocoded stimuli ($p < 0.05$).

To investigate the impact on discrimination of the two Jumbled versions in the Same-syllables training group, we split the data for the responses to the Jumbled version into responses to the MiddleJumbled version and FullJumbled version, comparing them with the responses given to training sound A. This showed that the birds responded with a higher Correct rate to Training sound A than to the MiddleJumbled test sound and with a higher Correct rate to the MiddleJumbled than to the FullJumbled test sound (Training – MiddleJumbled = 0.9071 ± 0.1812, MiddleJumbled – FullJumbled = 0.9094 ± 0.1603, both $p < 0.001$) (Fig 3b). There was no significant difference in the Response rate between these two Jumbled versions (MiddleJumbled – FullJumbled = 0.1404 ± 0.2004, $p = 0.76$), but both rates were lower than the Response rate to Training sound A (Training - MiddleJumbled = 1.3877 ± 0.2809, Training

– FullJumbled = 1.5281 ± 0.2783, both $p < 0.001$) (Fig 3d). These results (see Table S1 in the supplementary appendix) show that the birds of the ‘Same-syllables’ training group pay attention to the beginning and end, as well as to the middle syllables of the strings.

Table 2 Post hoc test results of Binomial GLMMs for the interaction of Test & Training_Group

Stimuli	**Training_Group**	**estimate**	**SE**	**z.ratio**	***p*.value**
a. Correct rate of responses (sound A+B in two training groups)					
Training	Different - Same	-0.007	0.198	-0.036	0.9715
SpectrumReversal	Different - Same	0.255	0.167	1.526	0.2117
Jumbled	**Different - Same**	**0.534**	**0.173**	**3.094**	**0.0100**
JumbledReversal	Different - Same	0.350	0.166	2.108	0.0878
Vocoded	Different - Same	-0.020	0.168	-0.119	0.9715
Training - SpectrumReversal	**Different**	**1.423**	**0.127**	**11.200**	**<.0001**
Training - Jumbled	**Different**	**0.958**	**0.133**	**7.217**	**<.0001**
Training - JumbledReversal	**Different**	**1.615**	**0.126**	**12.823**	**<.0001**
Training - Vocoded	**Different**	**1.673**	**0.127**	**13.188**	**<.0001**
SpectrumReversal - Jumbled	**Different**	**-0.465**	**0.110**	**-4.228**	**<.0001**
SpectrumReversal - JumbledReversal	Different	0.192	0.102	1.884	0.0662
SpectrumReversal - Vocoded	**Different**	**0.251**	**0.103**	**2.434**	**0.0186**
Jumbled - JumbledReversal	**Different**	**0.657**	**0.109**	**6.045**	**<.0001**
Jumbled - Vocoded	**Different**	**0.716**	**0.110**	**6.528**	**<.0001**
JumbledReversal - Vocoded	Different	0.059	0.102	0.578	0.5631
Training - SpectrumReversal	**Same**	**1.684**	**0.125**	**13.434**	**<.0001**
Training - Jumbled	**Same**	**1.499**	**0.127**	**11.781**	**<.0001**
Training - JumbledReversal	**Same**	**1.971**	**0.126**	**15.711**	**<.0001**
Training - Vocoded	**Same**	**1.660**	**0.127**	**13.076**	**<.0001**
SpectrumReversal - Jumbled	Same	-0.185	0.102	-1.821	0.0857
SpectrumReversal - JumbledReversal	**Same**	**0.287**	**0.099**	**2.889**	**0.0055**
SpectrumReversal - Vocoded	Same	-0.024	0.101	-0.238	0.8117
Jumbled - JumbledReversal	**Same**	**0.472**	**0.102**	**4.642**	**<.0001**
Jumbled - Vocoded	Same	0.161	0.104	1.555	0.1334
JumbledReversal - Vocoded	**Same**	**-0.311**	**0.101**	**-3.070**	**0.0036**
b. Response rate of trials (sound A+B in two training groups)					
Training	Different - Same	0.118	0.467	0.252	0.9724
SpectrumReversal	Different - Same	-0.221	0.420	-0.525	0.9724
Jumbled	Different - Same	-0.185	0.412	-0.448	0.9724

JumbledReversal	Different - Same	0.156	0.418	0.374	0.9724
Vocoded	Different - Same	0.014	0.411	0.035	0.9724
Training - SpectrumReversal	**Different**	**1.452**	**0.209**	**6.937**	**<.0001**
Training - Jumbled	**Different**	**1.896**	**0.204**	**9.291**	**<.0001**
Training - JumbledReversal	**Different**	**1.393**	**0.210**	**6.630**	**<.0001**
Training - Vocoded	**Different**	**1.859**	**0.204**	**9.097**	**<.0001**
SpectrumReversal - Jumbled	**Different**	**0.444**	**0.143**	**3.108**	**0.0027**
SpectrumReversal - JumbledReversal	Different	-0.059	0.152	-0.384	0.7790
SpectrumReversal - Vocoded	**Different**	**0.407**	**0.143**	**2.841**	**0.0056**
Jumbled - JumbledReversal	**Different**	**-0.502**	**0.144**	**-3.483**	**0.0010**
Jumbled - Vocoded	Different	-0.037	0.134	-0.272	0.7856
JumbledReversal - Vocoded	**Different**	**0.466**	**0.145**	**3.217**	**0.0022**
Training - SpectrumReversal	**Same**	**1.114**	**0.214**	**5.197**	**<.0001**
Training - Jumbled	**Same**	**1.594**	**0.206**	**7.746**	**<.0001**
Training - JumbledReversal	**Same**	**1.432**	**0.208**	**6.879**	**<.0001**
Training - Vocoded	**Same**	**1.756**	**0.204**	**8.621**	**<.0001**
SpectrumReversal - Jumbled	**Same**	**0.480**	**0.150**	**3.191**	**0.0024**
SpectrumReversal - JumbledReversal	**Same**	**0.319**	**0.154**	**2.071**	**0.0480**
SpectrumReversal - Vocoded	**Same**	**0.642**	**0.148**	**4.353**	**<.0001**
Jumbled - JumbledReversal	Same	-0.1616	0.141	-1.142	0.2534
Jumbled - Vocoded	Same	0.1623	0.135	1.206	0.2531
JumbledReversal - Vocoded	**Same**	**0.3239**	**0.138**	**2.341**	**0.0275**

Note: Response variables in GLMMs: (a) the proportion of correct responses if birds respond to one of two sounds; and (b) the proportion of trials that birds respond with pecking A or B. Only information related to both sound A and sound B were shown here, the information about the two Jumbled versions in the ‘Same-syllables’ training group weren't displayed in this table. Bold indicates significance.

Are modified stimuli still discriminated?

The above analyses concentrated on differences in the Correct rates between the groups and among the test stimuli. They don't test whether a low Correct rate also indicates that birds no longer discriminate between the modified version of training sound A and that of the similarly modified version of training sound B. If the birds are still capable of linking the modified stimuli to the respective training stimuli, the proportion of correct responses to the test stimuli should be higher than the proportion of incorrect responses. Table 3 and Fig. 4a show that for the Different-syllables group, all treatment combinations are significantly different from 0 in favour of a correct response. For the Same-syllables group, all treatments were also statistically different from 0 in favour of correct response, except the Test treatment JumbledReversal, which showed no significant difference from 0 (Fig. 4a).

For the data set of two Jumbled versions in Same-syllables group, MiddleJumbled is statistically different from 0 in favour of correct response, but FullJumbled is not significant different from 0 (Table 3), which is in line with the visualisation (Fig. 4b).

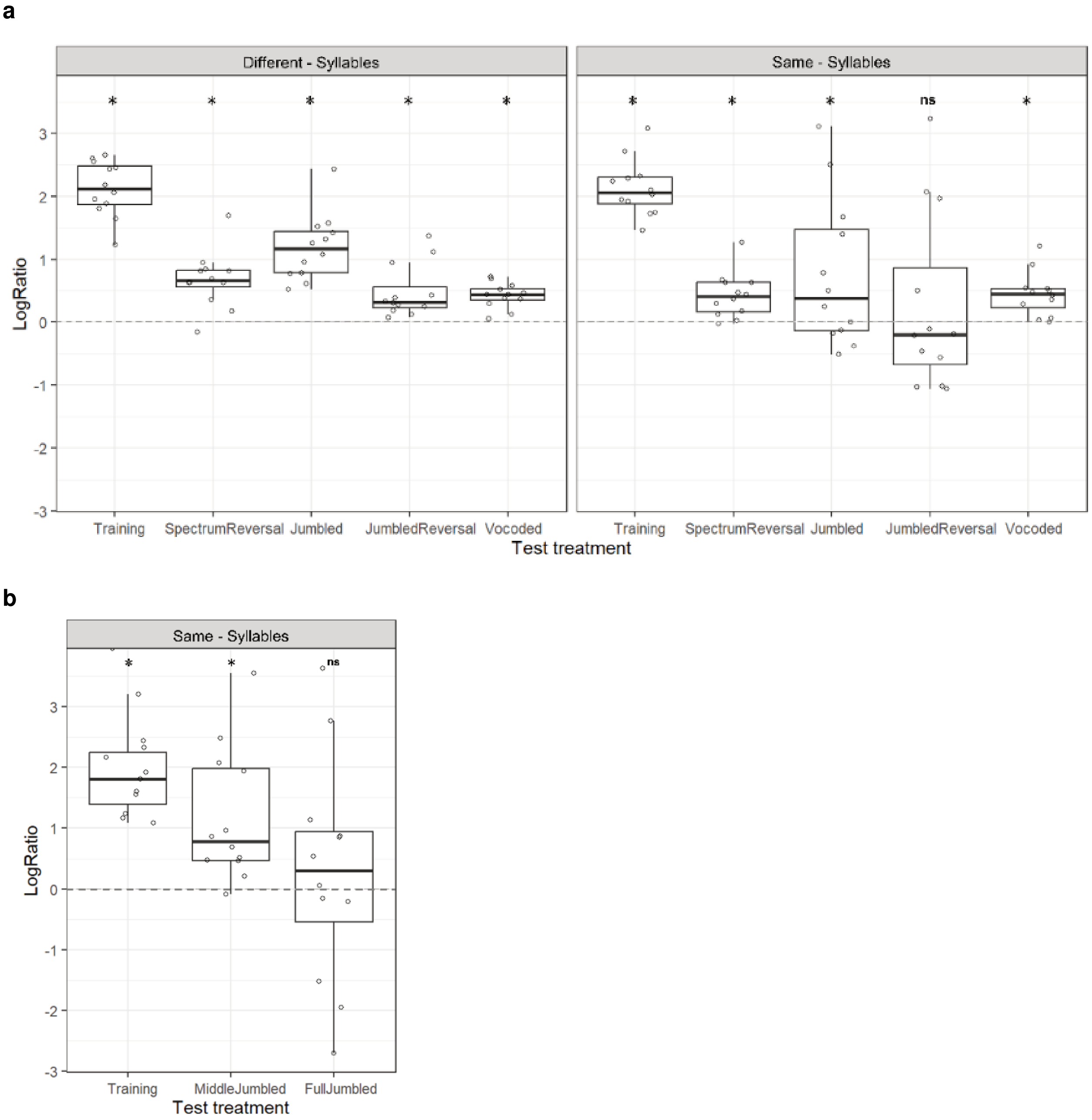


Figure 4. Visualisation of logRatios = log (Correct/Incorrect). a) For the Different-syllable group (left), all logRatios are statistically different from zero; for the Same-syllable group (right), the Test treatment JumbledReversal is not significantly different from 0; b) Results for the Jumbled test sounds of the Same-syllable group, split into Middle and Full Jumbled. For MiddleJumbled, there is a small overlap with zero; for Jumbled, it is statistically not different from 0. A * indicates that the logRatio of a Test treatment is significantly different from 0, 'ns' indicates that the logRatio of a Test treatment is overlapping with 0. Box plots show median, 1st and 3rd quartile, and whiskers the 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 logRatios was based on the counts of 'correct response' and 'incorrect response' from the same data set that was also used for Figure 3.

Table 3 Lower CL and Upper CL represent the lower and upper 95% confidence limits (CL) of the confidence interval

<table>
<tr><th rowspan="2">Training Group</th><th rowspan="2">Stimuli</th><th rowspan="2">estimate</th><th rowspan="2">SE</th><th colspan="2">CL (95%)</th></tr>
<tr><th>Lower</th><th>Upper</th></tr>
<tr><td colspan="6">LogRatio ~ Training_Group + Test_Treatment + Test_Treatment: Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials), data = sound A + sound B, n=24</td></tr>
<tr><td>Different syllables</td><td>Training</td><td>2.090</td><td>0.142</td><td>1.812</td><td>2.368</td></tr>
<tr><td>Different syllables</td><td>SpectrumReversal</td><td>0.667</td><td>0.121</td><td>0.430</td><td>0.904</td></tr>
<tr><td>Different syllables</td><td>Jumbled</td><td>1.132</td><td>0.127</td><td>0.883</td><td>1.381</td></tr>
<tr><td>Different syllables</td><td>JumbledReversal</td><td>0.475</td><td>0.120</td><td>0.241</td><td>0.710</td></tr>
<tr><td>Different syllables</td><td>Vocoded</td><td>0.417</td><td>0.121</td><td>0.180</td><td>0.653</td></tr>
<tr><td>Same syllables</td><td>Training</td><td>2.097</td><td>0.142</td><td>1.819</td><td>2.375</td></tr>
<tr><td>Same syllables</td><td>SpectrumReversal</td><td>0.412</td><td>0.119</td><td>0.179</td><td>0.646</td></tr>
<tr><td>Same syllables</td><td>Jumbled</td><td>0.597</td><td>0.121</td><td>0.360</td><td>0.835</td></tr>
<tr><td>Same syllables</td><td>JumbledReversal</td><td>0.125</td><td>0.119</td><td>-0.108</td><td>0.359</td></tr>
<tr><td>Same syllables</td><td>Vocoded</td><td>0.436</td><td>0.121</td><td>0.200</td><td>0.673</td></tr>
<tr><td colspan="6">LogRatio ~ Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials), data = sound A, n=12</td></tr>
<tr><td>Same syllables</td><td>Training</td><td>2.073</td><td>0.309</td><td>1.468</td><td>2.678</td></tr>
<tr><td>Same syllables</td><td>MiddleJumbled</td><td>1.166</td><td>0.298</td><td>0.581</td><td>1.751</td></tr>
<tr><td>Same syllables</td><td>FullJumbled</td><td>0.257</td><td>0.293</td><td>-0.319</td><td>0.832</td></tr>
</table>

Note: If zero is part of the confidence interval, the treatment combination Training Group and Stimuli are not significantly different from 0. If both confidence levels are positive, then there is a bias toward correct responses. If they are both negative, then they are more biased toward incorrect responses. Bold indicates significance.

DISCUSSION

Our results show that zebra finches are capable of using both spectral features and sequential information to discriminate strings consisting of conspecific song syllables. Confirming results obtained in earlier studies on zebra finches, our study also demonstrates that zebra finches will give higher priority to using spectral features than syllable/element sequence in discrimination when the syllables differ in phonology. When strings are composed of a same set of syllables, zebra finches learn about the syllable sequence in addition to the syllable phonology.

No effect of stimulus composition on learning speed

Various studies (Braaten *et al.*, 2006; Lawson *et al.*, 2018; Geberzahn & Derégnaucourt, 2020) demonstrated that when zebra finches learned to discriminate between two songs, they were very sensitive to changes in the spectral domain (syllable reversals) and hardly sensitive to sequential information (sequence reversals), similar to what we observed in our 'Different-syllables' training group. These studies indicated that the zebra finches ignored sequence cues in discrimination learning or that sequences were more difficult to learn than spectral features and might require more time. In line with this, some studies (Lawson *et al.*, 2018; Braaten *et al.*, 2006) indicated that if zebra finches used syllable sequences to distinguish songs, this occurred with songs to which the birds had been exposed more extensively. That learning to discriminate sequences consisting of the same sets of syllables might be more difficult than sequences consisting of different syllables was also suggested by a meta-analysis using data from 14 different acoustic Go/No-go experiments with zebra finches (Kriengwatana *et al.*, 2016), which indicated that stimuli (either zebra finch vocalizations or human speech syllables) differing in phonetic characteristics were learned faster than those differing in sequence only. However, in our experiment, allowing a direct comparison of learning speed of comparable stimuli in identical conditions, the learning speed of the training group relying only on sequence cues is not significantly lower than that of the group trained on stimuli with different syllables. This suggests that the Same-syllables group learned about the syllable sequence in parallel with learning about the syllable phonology, without requiring more extensive exposure or training.

Cognitive flexibility in processing syllable phonology and sequence

The comparison of the correct responses to the different test stimuli showed that both training groups were similarly strongly affected by changes of the spectro-temporal features of the syllables, thus noticing such changes equally well. It demonstrates that the Same-syllables group, which can only learn a sequence of syllables when they also learn the spectro-temporal features of these syllables, gives the same weight to the spectro-temporal features as the Different-syllables group does. The difference between the two training groups concerns their responses to the jumbled test sounds. Although the jumbled test stimuli received fewer correct responses and had a lower Response rate than the training stimuli in both groups, jumbling affected the Same-syllables group much more strongly than the Different-syllables group. For the Same-syllables group, the impact of jumbling is similar to that of spectral changes. Jumbling had a lesser impact than spectral modifications in the Different-syllables group, confirming that this group mainly (although not exclusively) relied on spectral features of the syllables to distinguish the training strings. Hence, the importance of syllable sequence increased when knowledge of the sequence is needed to correctly identify different strings. This finding indicates the presence of 'cognitive flexibility' in processing string information, in which sequence learning can be added to learning of spectro-temporal features of syllables when needed to distinguish strings.

No differences were observed between the responses of both groups to reversal of the syllables and vocoding them. Reversal of syllables reverses the within-syllable spectral and amplitude pattern (i.e., any frequency changes or increasing or decreasing amplitude over an element), while vocoding maintains these patterns, but removes pitch information. Apparently, all these dimensions are taken into account for identification of syllables. Nevertheless, both groups were capable of still discriminating reversed and vocoded versions of the training stimuli, indicating that the test stimuli still maintained sufficient gross spectral differences among the syllables of a string to allow for string identification.

That full jumbling strongly affected the Same-syllables group and resulted in absence of discrimination is no surprise, as full jumbling removed all information that might relate to the original syllable sequences. However, what is of interest is that middle-jumbled also got fewer correct responses than the training stimuli, indicating that the birds were not just relying on the first and last syllables of the syllable sequence (which was suggested by studies on zebra finches (Fishbein *et al*., 2019) and Bengalese finches (*Lonchura striata var. domestica*) (Mizuhara & Okanoya, 2020)) but also to the sequence of the middle syllables.

Vocal production learning and discrimination learning

Altogether the results indicate that sequence learning can be 'added to' learning about spectro-temporal features of syllables if these features alone are insufficient to distinguish two syllable strings. It indicates the presence of sequence learning as a separate, but nevertheless strongly connected or partially overlapping learning process, similar to what has been observed in several studies of song production learning (Liu *et al*., 2004; Braaten *et al*., 2006; Lipkind *et al*., 2013, 2017). This does not imply that song production learning and song discrimination learning rely on the same mechanisms. Song production learning occurs in male zebra finches only and only during a sensitive phase early in life, while discrimination learning can occur in both sexes and when adult. Also, vocal discrimination learning has been observed in vocal non-learning species, such as dove species (Beckers & ten Cate, 2001; Beckers *et al*., 2003), which give attention to both spectral and temporal structure of sound strings. Hence, vocal production learning and later occurring vocal discrimination or recognition learning are likely to rely at least partly on different mechanisms.

In conclusion, our study demonstrates that although zebra finches have a bias to attend to spectral features when recognizing or discriminating strings of syllables, they can also attend to the sequence when needed. Our study did not test whether the relative importance of syllable sequence might vary if the syllable similarity between strings also varies, e.g., when not all but only part of the syllables in a string are different, or when different strings contain different exemplars of the same syllable types. It is likely that such string modifications may affect the relative weight of spectro-temporal and sequence parameters in song discrimination. Such flexibility may explain why some studies on the cues that zebra finches use to distinguish songs demonstrated absence of any impact of changes in syllable sequences on discriminating strings (Lawson *et al*., 2018; Geberzahn & Derégnaucourt, 2020; Mol *et al*., 2021), while other studies (van Heijningen *et al*., 2009; Chen *et al*., 2016; Spierings & ten Cate, 2016) showed clear sequence learning. It shows that the use of particular cues within a specific experiment should not be taken as an inability to use other cues when such cues might be useful or needed to correctly identify different strings, although the importance of the ability to also learn about syllable sequences under natural conditions remains to be elucidated. A similar flexibility, in this case for using different spectral cues, was observed by Burgering *et al.* (2018; 2019), showing that depending on the differences among training sounds zebra finches used either pitch or spectral envelope to distinguish the training sounds. To what extend such a flexibility is also present for other song features awaits further exploration (see ten Cate & Honing, 2022). It is likely that zebra finches are not the only species that demonstrates such cognitive flexibility, although this remains to be tested. The benefit of such flexibility is that it may allow birds to adjust their perceptual tuning to those acoustic dimensions that are most relevant to distinguish songs of different individuals or other biologically relevant sounds.

Acknowledgments

We would like to thank Jeroen Hubert for help with the statistical analysis, Leda Aristidou for help with screening and synthesis of acoustic stimuli, Peter Snelderwaard for his technical assistance, and him and Michelle Geers for their animal care. Zhi-Yuan Ning was supported by the CSC (China Scholarship Council) fellowship.

References

Bartoń K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn

Bates D., Maechler M., Bolker B., Walker S. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw.*, 67(1):1-48. https://doi.org/10.18637/jss.v067.i01

Beckers G.J.L., ten Cate C. (2001). Perceptual relevance of species-specific differences in acoustic signal structure in Streptopelia doves. *Anim Behav.*, 62(3):511–518. https://doi.org/10.1006/anbe.2001.1768

Beckers G.J.L., Goossens B.M.A., ten Cate C. (2003.) Perceptual salience of acoustic differences between conspecific and allospecific vocalizations in African collared-doves. *Anim Behav.*, 65(3):605-614. https://doi.org/10.1006/anbe.2003.2080

Benjamini Y., Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.*, 57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Braaten R.F., Petzoldt M., Colbath A. (2006). Song Perception During the Sensitive Period of Song Learning in Zebra Finches (*Taeniopygia guttata*). *J Comp Psychol.*, 120(2):79–88. https://doi.org/10.1037/0735-7036.120.2.79

Burgering M.A., ten Cate C., Vroomen J. (2018). Mechanisms underlying speech sound discrimination and categorization in humans and zebra finches. *Anim Cogn.*, 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 both on the fundamental frequency ("Pitch") and spectral envelope. *J Comp Psychol.*, 133(1):106-117. https://doi.org/10.1037/com0000143

Cazala A., Giret N., Edeline J.M., Del Negro C. (2019). Neuronal encoding in a high-level auditory area: from sequence of elements to grammatical structure. *J Neurosci.*, 39(31):6150-6161. https://doi.org/10.1523/JNEUROSCI.2767-18.2019

Chen J., ten Cate C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behav Processes.*, 117:29-34. https://doi.org/10.1016/j.beproc.2014.09.004

Chen J., Jansen N., ten Cate C. (2016). Zebra finches are able to learn affixation-like patterns. *Anim Cogn.*, 19(1):65-73. https://doi.org/10.1007/s10071-015-0913-x

Chen J., ten Cate C. (2017). Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird. *J Exp Psychol Anim Learn Cogn.*, 43(3):295–302. https://doi.org/10.1037/xan0000145

Derégnaucourt S., Poirier C., Kant A.V., Linden A.V., Gahr M. (2013). Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *J Physiol Paris.*, 107(3):210–218. https://doi.org/10.1016/j.jphysparis.2012.08.003

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

Eales L.A. (1985). Song learning in zebra finches: Some effects of song model availability on what is learnt and when. *Anim Behav.*, 33(4):1293–1300. https://doi.org/10.1016/S0003-3472(85)80189-5

Eens M. (1997). Understanding the complex song of the European starling: an integrated ethological approach. *Adv Study Anim Behav*., 26:355–434. https://doi.org/10.1016/S0065-3454(08)60384-8

Fishbein A.R., Idsardi W.J., Ball G.F., Dooling R.J. (2019). Sound sequences in birdsong: how much do birds really care? *Philos Trans R Soc Lond B Biol Sci*., 375(1789):20190044. https://doi.org/10.1098/rstb.2019.0044

Fishbein A.R., Prior N.H., Brown J.A., Ball G.F., Dooling R.J. (2021). Discrimination of natural acoustic variation in vocal signals. *Sci Rep*., 11:916. https://doi.org/10.1038/s41598-020-79641-z

Geberzahn N., Derégnaucourt S. (2020). Individual vocal recognition in zebra finches relies on song syllable structure rather than song syllable order. *J Exp Biol*., 223(9):jeb220087. https://doi.org/10.1242/jeb.220087

Gil D., Slater P.J.B. (2000). Song organisation and singing patterns of the willow warbler, *Phylloscopus trochilus*. *Behaviour*, 137(6):759-782. https://doi.org/10.1163/156853900502330

James L.S., Sakata J.T. (2017). Learning biases underlie "universals" in avian vocal sequencing. *Curr Biol*., 27(23):3676-3682.e4. https://doi.org/10.1016/j.cub.2017.10.019

Knowles J.M., Doupe A.J., Brainard M.S. (2018). Zebra finches are sensitive to combinations of temporally distributed features in a model of word recognition. *J Acoust Soc Am*., 144(2):872. https://doi.org/10.1121/1.5050910

Kriengwatana B., Spierings M.J., ten Cate C. (2016). Auditory discrimination learning in zebra finches: Effects of sex, early life conditions and stimulus characteristics. *Anim Behav*., 116:99–112. https://doi.org/10.1016/j.anbehav.2016.03.028

Lachlan R.F., van Heijningen C.A., Ter Haar S.M., ten Cate C. (2016). Zebra finch song phonology and syntactical structure across populations and continents—a computational comparison. *Front Psychol*., 7:980. https://doi.org/10.3389/fpsyg.2016.00980

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. *Anim Cogn*., 21(4):467-480. https://doi.org/10.1007/s10071-018-1182-2

Lehongre K., Aubin T., Robin S., Del Negro C. (2008). Individual signature in canary songs: Contribution of multiple levels of song structure. *Ethology*, 114(5):425-435. https://doi.org/10.1111/j.1439-0310.2008.01486.x

Lenth R.V. (2016). Least-Squares Means: The R Package lsmeans. *J Stat Softw*., 69(1):1-33. https://doi.org/10.18637/jss.v069.i01

Lipkind D., Marcus G.F., Bemis D.K., Sasahara K., Jacoby N., Takahasi M., Suzuki K., Feher O., Ravbar P., Okanoya K., Tchernichovski O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, 498(7452):104–108. https://doi.org/10.1038/nature12173

Lipkind D., Zai A.T., Hanuschkin A., Marcus G.F., Tchernichovski O., Hahnloser R.H.R. (2017). Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nat Commun*., 8(1):1247. https://doi.org/10.1038/s41467-017-01436-0

Liu W.C., Gardner T.J., Nottebohm F. (2004). Juvenile zebra finches can use multiple strategies to learn the same song. *Proc Natl Acad Sci*., 101(52):18177–18182. https://doi.org/10.1073/pnas.0408065101

Marler P., Peters S. (1987). A sensitive period for song acquisition in the song sparrow, *Melospiza melodia*: A case of age-limited learning. *Ethology*, 76(2):89–100. https://doi.org/10.1111/j.1439-0310.1987.tb00675.x

Mizuhara T., Okanoya K. (2020). Do songbirds hear songs syllable by syllable? *Behav Processes*., 174:104089. https://doi.org/10.1016/j.beproc.2020.104089

Mol C., Bolhuis J.J., Moorman S. (2021). Vocal learning in songbirds: the role of syllable order in song recognition. *Philos Trans R Soc Lond B Biol Sci*., 376(1836):20200248. https://doi.org/10.1098/rstb.2020.0248

Plamondon S.L., Rose G.J., Goller F. (2010). Roles of syntax information in directing song development in white-crowned sparrows (*Zonotrichia leucophrys*). *J Comp Psychol*., 124(2):117-32. https://doi.org/10.1037/a0017229

Riebel K., Slater P.J.B. (1999). Song type switching in the chaffinch, *Fringilla coelebs*: timing or counting? *Anim Behav*., 57(3):655-661. https://doi.org/10.1006/anbe.1998.0984

Soha J.A., Marler P. (2001). Vocal syntax development in the white-crowned sparrow (*Zonotrichia leucophrys*). *J Comp Psychol*., 115(2):172-80. https://doi.org/10.1037/0735-7036.115.2.172

Spierings M.J., ten Cate C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proc Natl Acad Sci*., 113(27):E3977-84. https://doi.org/10.1073/pnas.1600483113

van Buuren S., Groothuis-Oudshoorn K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*., 45(3):1-67. https://doi.org/10.18637/jss.v045.i03

van Heijningen C.A., de Visser J., Zuidema W., ten Cate C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc Natl Acad Sci*., 106(48):20538–20543. https://doi.org/10.1073/pnas.0908113106

van Heijningen C.A., Chen J., van Laatum I., van der Hulst B., ten Cate C. (2013). Rule learning by zebra finches in an artificial grammar learning task: which rule? *Anim Cogn*., 16(2):165-175. https://doi.org/10.1007/s10071-012-0559-x

Vernes S.C., Kriengwatana B.P., Beeck V.C., Fischer J., Tyack P.L., ten Cate C., Janik V.M. (2021). The multi-dimensional nature of vocal learning. *Philos Trans R Soc Lond B Biol Sci*., 376(1836):20200236. https://doi.org/10.1098/rstb.2020.0236

Perceptual Interplay of Pitch and Formant Contours in Melody Recognition by Zebra Finch

Chapter 5

Zhi-Yuan Ning · Henkjan Honing · Carel ten Cate

This chapter is prepared for submission

Chapter 5

ABSTRACT

Auditory perception of complex acoustic sequences involves the integration of multiple perceptual attributes, such as pitch and formant contours. While both attributes contribute to speech and music perception, the relative importance of each and their potential interactions remain underexplored. Here, we investigate how zebra finches (*Taeniopygia guttata*) discriminate harmonic complex tone sequences, which were characterized by pitch and formant contours that were either both increasing or decreasing in the frequency domain or were going in the opposite direction, thus probing the interplay between pitch and formant contours, and evaluating the influence of training conditions. After being trained in a Go-left/Go-right paradigm, we next manipulate the pitch and formant contours of the tone sequences in test sounds to assess their role in sound sequence recognition and the presence of perceptual interactions. Zebra finches demonstrate remarkable sensitivity to both attributes, detecting variations across harmonic tones in pitch and formant contours. In most cases the responses of the two training groups to modified stimulus versions are the same, indicating training conditions have only a limited impact on the birds' attention given to pitch and formant contours. The current study adds to an expanding body of literature supporting cognitive flexibility in songbirds and highlights a holistic approach using harmonic complex tone sequences to provide a comprehensive perspective on auditory discrimination in zebra finches.

INTRODUCTION

Music and speech share certain characteristics: both consist of sequences of acoustic units that are systematically ordered, and the continuous acoustic dimension is partitioned by attending to perceptual attributes (e.g., pitch and formant) (Patel, 2008). The basic encoding of acoustic features underlying these attributes may involve largely overlapping subcortical circuits (Patel, 2011). The cognitive processing of a certain perceptual attribute can be quite different in speech and music, reflecting the different patterns and functions the attribute has in the two domains (Patel, 2008). However, recent studies have demonstrated that experience with a perceptual attribute (e.g., pitch) in one domain can affect the perception in another domain. For instance, lexical pitch perception may have an influence on musical pitch perception, and vice versa (e.g., Sadakata *et al.*, 2020; Choi W, 2021). One example of this concerns the perception of harmonic complex tones, i.e., tones that have a rich harmonic spectrum, that are present in both human music and speech. The perception of such complex tones has been shown to depend on whether they act as a musical tone (emphasizing pitch) or a speech syllable (emphasizing the formant structure) (Sadakata *et al.*, 2020; Albouy *et al.*, 2023). The perception of pitch and formant might also play a crucial role in the vocalization and communication of nonhuman animal (hereafter: animal) species (Hoeschele, 2017).

Pitch is conventionally defined as the perceptual correlate of a sound's fundamental frequency (f0) (Dowling & Harwood, 1986). However, what has been considered "pitch perception" in humans is mediated by several different mechanisms, not all of which involve estimating f0. A spectral-pattern tracking mechanism (irrespective of harmonic or inharmonic) that registers the direction of pitch shifts (i.e., contour) by tracking shifts in general spectral patterns, appears to operate for both musical tones and for speech (McDermott *et al.*, 2008; McPherson & McDermott, 2018). On the other hand, the f0-estimation mechanism (e.g., based on harmonic patterns in the spectrum of the sound) plays an important role in tasks that required judgments of pitch intervals (the magnitude of pitch shifts) or voice identity (McPherson & McDermott, 2018). Human listeners perceive harmonic or quasi-harmonic sounds as a coherent entity, rather than as a simultaneous collection of unrelated pure tones, which suggests that the human auditory system tends to "group" or "bind" together components that are presented simultaneously and are harmonically related (Micheyl & Oxenham, 2010). Noticeably, contour representations in other dimensions besides pitch (such as loudness and brightness contours) are also recognizable (by humans) (McDermott *et al.*, 2008). Nevertheless, the f0-based pitch quality is one of the most important characteristics of human auditory experience and plays a central role in human music cognition (Patel, 2003). Pitch perception is also a focus in animal auditory perception, as studying other species can provide a more comprehensive perspective to understand the evolutionary history of pitch perception (Honing *et al.*, 2015; Hoeschele *et al.*, 2015; Walker *et al.*, 2019).

Pitch perception has been studied in several mammal species, such as Japanese macaque (*Macaca fuscata*) (Izumi, 2001), crab-eating macaque (*Macaca fascicularis*) (Brosch *et al.*, 2006) and ferret (Walker *et al.*, 2009), as well as in many avian species, including European starling (*Sturnus vulgaris*) (Hulse & Cynx, 1985; MacDougall-Shackleton & Hulse, 1996), white-throated sparrow (*Zonotrichia albicollis*) (Hurly *et al.*, 1990), black-capped chickadee (*Parus atricapillus*) (Weary & Weisman, 1991), and zebra finch (*Taeniopygia guttata*) (Weisman *et al.*, 1994). Vocal learning avian species (especially songbirds) are more accurate than most of the mammals, including humans, when tested with absolute pitch (Weisman *et al.*, 1998; Weisman *et al.*, 2004; Friedrich *et al.*, 2007), and they do this more readily than identifying relative pitch relationships (Hulse *et al.*, 1984; Page *et al.*, 1989; Weisman *et al.*, 1994; MacDougall-Shackleton & Hulse, 1996; Bregman *et al.*, 2012). However, humans appear to rely on "octave equivalence" to solve an absolute pitch perception task (Hoeschele, 2017; See ten Cate & Honing, 2022 for a more elaborate discussion).

Nonetheless, pitch may need not to be the primary acoustic cue for the perception of sound patterns in animals. Rats (*Rattus norvegicus*) take variations in pitch (f0) to be less psychologically distant than changes in timbre (i.e., spectral quality or "sound color") in an operant conditioning task (Crespo-Bojorque *et al.*, 2022). In a series of operant conditioning tasks, Bregman *et al.* (2016) examined how a songbird, the European Starling (*Sturnus vulgaris*), perceives tone sequences in which one of three particular attributes (pitch, timbre, and spectral envelope) varies systematically over a sequence of four tones. Surprisingly, the starlings do not use pitch but the "acoustic spectral shape" (the overall distribution of the spectral energy for each tone: spectral envelope) to recognize successive tonal stimuli (Bregman *et al.*, 2016). The way starlings gravitate towards spectral envelope for complex tone sequence recognition contrasts the human bias to pitch in perceiving tone sequences. While the spectral envelope provides sufficient information for accurate tone sequence recognition in starlings, it's important to note that it is not the only cue they use for sound recognition. Starlings are capable of perceiving missing fundamentals in individual tones containing harmonic complexes (Cynx & Shapiro, 1986) and recognizing frequency-shifted conspecific songs in which the spectral envelope has been altered (Bregman *et al.*, 2012).

The “formant”, defined as “a characteristic peak in the spectral envelope” of vocal or musical sounds (e.g., the definition in the standards for acoustical terminology by Acoustical Society of America, 1994), has been studied in animals, suggesting that many species, including mammals and songbirds, can perceive formant information in acoustic signals. Rhesus macaques (*Macaca mulatta*) can spontaneously respond to a change in formant frequencies in their own species-typical vocalizations (Fitch & Fritz, 2006), or be trained to discriminate diverse sounds (human vowel, conspecific and heterospecific vocalizations, and artificial sounds) based on morphs and formants (Melchor *et al.*, 2021). Mice (*Mus musculus*) were found to share the same perceptual mechanism as humans, which combines specific formants and temporal patterns, for detecting auditory objects and sound event streams with biological communication functions (Geissler & Ehret, 2002). In birds, studies have shown that European starlings (*Sturnus vulgaris*) (Kluender *et al.*, 1998) and budgerigars (*Melopsittacus undulatus*) (Henry *et al.*, 2017) can attend to formant in vowel discrimination. Similarly, zebra finches (*Taeniopygia guttata*) were found sensitive to different formant patterns in human speech (Ohms *et al.*, 2012; Kriengwatana *et al.*, 2015; Burgering *et al.*, 2018).

Thus, to date, pitch and formant have been well documented as two of the most crucial perceptual attributes involved in speech and music. Nevertheless, the synchronous presence of both perceptual attributes may lead to perceptual interaction (concordance/competition), affecting how humans and other species organize sensory information and make perceptual judgments. Perceptual concordance refers to the state in which two or more perceptual attributes or cues in a sensory stimulus align or cooperate in the same direction or pattern, enhancing the perceptual grouping effect. On the other hand, perceptual competition denotes a situation where the same perceptual attributes or cues in a stimulus compete or work in different directions or patterns, potentially weakening the overall perceptual grouping effect. This competition doesn’t imply that one attribute be completely suppressed by another but both cues may have a noticeable impact on perception, even when in conflict. Until very recently, however, the majority of research on the topic of perceptual interaction has been influenced by cognitive studies of orienting responses in human perceptual systems. Much of the research on non-human animals has employed stimuli with relatively simple attribute patterns, such as single-tone strings, vowel-like units, or natural vocalizations. Consequently, the investigation of perceptual interaction has generally taken a secondary role in the field of auditory sensory perception research. Nevertheless, those interested in comprehending how the auditory perceptual system, typically in humans, handles and organizes more complex combinations of multiple attribute inputs into coherent perceptual objects, as well as how it interprets ambiguous inputs related to these attribute patterns, may find inspiration in the empirical research conducted in the domain of visual perception. Perceptual organization studies from human visual field show that the grouping effect is stronger and more stable when two cues concord in the same direction/pattern, while the grouping effect is weaker and more unstable when they compete in different directions/patterns (Kubovy & van den Berg, 2008; Luna & Montoro, 2011; Schmidt & Schmidt, 2013; Luna *et al.*, 2016; Montoro *et al.*, 2017; Villalba-García *et al.*, 2018). In fact, in situations of perceptual competition between two visual cues, the non-dominant cue is perceived to a certain extent, so it's not completely suppressed by the dominant cue (Luna *et al.*, 2016; Rashal *et al.*, 2017). Additionally, dominance dynamics between perceptual grouping cues were found in visual competition (Palmer & Beck, 2007; Luna *et al.*, 2016; Villalba-García *et al.*, 2021). While it's reasonable to consider that the perceptual mechanisms demonstrated to influence perceptual organization in the visual sensory domain when multiple attribute patterns are presented within a same stimulus might also apply to auditory sensory processing, caution is warranted when conducting such examinations. In the auditory domain, a recent study by McPherson & McDermott (2023) examined the effects of timbral differences on relative pitch judgments and suggested that relative pitch judgments are not completely invariant to timbre, even in naturalistic conditions, and even when such judgments are based on representations of the fundamental frequency (f0). However, the literature has paid less attention to the potential effect of perceptual interaction (concordance/competition) between auditory-perceptual attributes compared to visual-perceptual attributes so far, especially in cross-species studies. Hence, it is interesting to examine whether the above grouping mechanisms from visual perception are also present for auditory perception when there's concordance/competition between two acoustic attributes, in a cross-species comparison paradigm.

Previous studies with zebra finches (*Taeniopygia guttata*), a popular avianmodel species for investigating the cognitive basis of auditory perception, have demonstrated that they can perceive both pitch and formant in the sound discrimination task (e.g., Burgering *et al.*, 2019). The acoustic units used by Burgering *et al.*'s study (2019) resemble the structure of zebra finch distance calls (single syllable units with a harmonic structure). These vowel-like sounds can be regarded as a relatively simple acoustic stimulus in spectral structure, consisting of one unit with a single formant peak. Another study showed that zebra finches used a particular harmonic (the 2nd harmonic located between the frequency region from 2kHz to 3kHz) as the main discriminative cue and the fundamental frequency as a secondary discriminative cue when trained to perceive the harmonic structure from a male distance call (Uno *et al.*, 1997). It is likely that they are still quite sensitive to pitch (fundamental frequency), but pitch information needs to be prominent in their best hearing range. Perceptually, the pitch cues can be assessed from the harmonics, so the fundamental frequency need not be present to still be abstracted (Cynx & Shapiro, 1986). However, the potential interaction of both pitch and formant in zebra finch auditory perception has not been thoroughly examined. Specifically, the perception of changes over a series of units and the impact of Same-direction / Crossed-direction between these acoustic attributes remains unexplored.

In the current study, zebra finches were trained to perform an auditory discrimination task using a Go-left/Go-right paradigm with corrective feedback. The stimuli used consisted of sequences of five complex tones, with some small silent gaps between them. Each sequence was characterized by pitch and formant patterns (referred to as *pitch contour* and *formant contour*) that were either both increasing/decreasing in frequency or were going in the opposite direction.

In the *Same-direction condition*, the contours of pitch and formant are arranged in the same direction over the tone sequences (both ascending or both descending in frequency). In the *crossed-direction condition*, the contours of pitch and formant are in opposite directions (for example, when the pitch contour is ascending, the formant contour is descending), and the sound sequence can be perceived as two crossing contours. With this experimental paradigm, we were able to directly assess the relative contribution of pitch and formant pattern to a complex tone sequence, and determine how they contribute to the identification of the sequence.

METHODS

Subjects

Twenty-three zebra finches (12 males and 11 females) completed the task in this experiment. They were tested at the age of 334 ± 47 days post-hatching (dph), (age males: M= 366, SD= 37, age females: M= 299, SD= 26). All birds originated from the in-house breeding colony at Leiden University. Before the experiment, the birds lived in single-sex groups of about 15 to 30 individuals in aviaries (2m × 2m × 1.5m), in which food and water were available ad libitum. The housing rooms were kept at 20–22°C and 40–60% humidity and illuminated with artificial lights (Philips Master TL5 HO 49W/830) from 07:00–20:30 (13.5h light : 10.5h dark) with a 15 min twilight phase with the light fading in and out at the beginning and the end of each day. A week before the operant test, birds were caught and transferred from the aviary to standard laboratory cages (two birds of equal sex in one cage) in order to acclimatize (cage size: length × width × height = 80 × 40 × 40 cm) and reduce stress from catching in aviary. The birds were divided randomly in two experimental groups: twelve of the birds were assigned to the Same-direction group, and the other eleven birds to the Crossed-direction group (6 males and 6 females in Same-direction group, 6 males and 5 females in Crossed-direction group; age Same-direction group: M=316, SD=33, age Crossed-direction group: M=353, SD=51).

Operant conditioning cage

Zebra finches were trained and tested individually in an operant conditioning cage (Skinner Box) (70x30x45 cm). The cage was built from wire mesh walls and one foamed PVC back wall and was containing 3 pecking sensors with a red LED light at the top of each sensor (Fig. 1A). 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-h/10.5-h light/dark schedule. Sound stimuli were played through a speaker (Vifa MG10SD09–08, Vifa, Viborg, Denmark) 1 meter 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). Sensors (S1, S2, S3), lamp, food hatch and speaker were connected to operant conditioning controller that also registered all sensor pecks.

Stimuli

Training stimuli

Each experimental group was trained to discriminate between a pair of tonal sequences consisting of five complex tones, and within each group part of the birds got one pair of training stimuli (“sequence series 1”) and the rest another pair of training stimuli (“sequence series 2”) (see Table 1). The training stimuli were synthesized, normalized and filtered using Praat (version 6.1.12) and Audacity (version 2.3.0). The stimuli have not been heard before by the birds. To synthesize a sound unit, the first step was to construct a complex tone with a defined f0 value (e.g., 150Hz as the f0) by choosing the function “Create Sound as tone complex” (to create a sound combining the f0 and its constituent harmonics occurring at integer multiples of the f0) in Praat. Secondly, the complex tone was manipulated by the Effect “Fade in/Fade out” and its peak amplitude was Normalized to “-50 dB” by Audacity. Additionally, to create a formant for the unit, the amplitude located in a particular frequency range (e.g., the Formant peak located in 2.6kHz) was amplified (Width 1kHz, Gain 15dB x2) by applying the “Parametric EQ” function (an effect plugin) in Audacity. Afterwards these complex tones were combined into a single sequence using Praat, with each tone (duration: 0.21s) separated by a silence of 0.05s, resulting in a sequence with a duration of 1.25s. All stimuli were low-pass filtered (below 8kHz).

The training stimuli in this experiment were 4 stimulus pairs (2 sequence series/pairs for each training group, as shown in Table 1), each pair was consisting of two different tone sequences that differ in both pitch and formant direction. Two training groups were categorized by the training stimuli the birds trained with: for the Crossed-direction group, the training stimuli was a pair of complex tones that within which both the pitch contour and formant contour of a single training stimulus were going in the opposite direction: for instance, the pitch contour of training stimulus A was decreasing (in the frequency domain over an entire tone sequence), while the formant contour of training stimulus A was increasing (Fig. 1B). For the Same-direction group, the training stimuli was a pair of complex tones that within which both the pitch contour and formant contour of a single training stimulus were going in the same direction (both contours of two attributes were either increasing or decreasing) (Fig. 1D).

In each pair of training stimuli, one stimulus (e.g., training stimulus A) featured a formant-ascending pattern, for example, “2.6kHz- 3.1kHz- 3.6kHz- 4.1kHz- 4.6kHz”. This formant-ascending pattern increased the amplitude (+30 dB) within specific frequency bands of each complex tone. These amplitude-enhanced frequency bands fell within the zebra finches’ hearing threshold, as determined by Okanoya & Dooling (1987), typically peaking around 3.5-4.0 kHz. In contrast, the complex tones of the other stimulus (e.g., training stimulus B) followed a formant-descending pattern, for instance, “4.6kHz- 4.1kHz- 3.6kHz- 3.1kHz-2.6kHz”. Similarly, one sequence featured a pitch-ascending pattern, such as “150Hz- 220Hz-290Hz- 360Hz- 430Hz”, while the other exhibited a pitch-descending pattern, for instance, “430Hz- 360Hz- 290Hz- 220Hz- 150 Hz”. The stimuli from different sequence series of the same training group were arranged in the same way, but with different f0 and formant values. This was done to prevent the selected value set from accidentally coinciding with frequencies that might hold specific biological significance or relevance for the birds. When played, the sequences were normalized such that the average intensity (RMS, calculated over the total duration of the stimulus) was the same for the two sequences within a pair to avoid amplitude differences affecting the responses to the stimuli.

Table 1 Overview of the training groups

Training group	Sequence series	Subjects (n)	Sequence of tonal units
Crossed-direction	Series 1	3 ♂ + 3 ♀	430Hz(2.6kHz)-360Hz(3.1kHz)-290Hz(3.6kHz)-220Hz(4.1kHz)-150Hz(4.6kHz)
Crossed-direction	Series 2	3 ♂ + 2 ♀	405Hz(2.1kHz)-330Hz(2.7kHz)-255Hz(3.3kHz)-180Hz(3.9kHz)-105Hz(4.5kHz)
Same-direction	Series 1	3 ♂ + 3 ♀	150Hz(2.6kHz)-220Hz(3.1kHz)-290Hz(3.6kHz)-360Hz(4.1kHz)-430Hz(4.6kHz)
Same-direction	Series 2	3 ♂ + 3 ♀	105Hz(2.1kHz)-180Hz(2.7kHz)-255Hz(3.3kHz)-330Hz(3.9kHz)-405Hz(4.5kHz)

Note: In each training group, the tone sequences were arranged in the same way (the pitch and formant contours of a stimulus were either going in the same or crossed direction) but differed in f0 and Formant values of their complex tones. The numerical values between the "-" indicates the f0 and Formant values of each complex tone (e.g., "430Hz(2.6kHz)" stands for f0 = 430Hz and Formant peak = 2.6kHz). The values of tonal units for tone sequence are shown in this table, with training stimulus A presented as an exemplar. While training stimulus B is not displayed in this table, note that the values of tonal units for training stimulus B were identical to those of training stimulus A, with the only difference being that its tonal units were arranged in the reverse order of training stimulus A. See text for details.

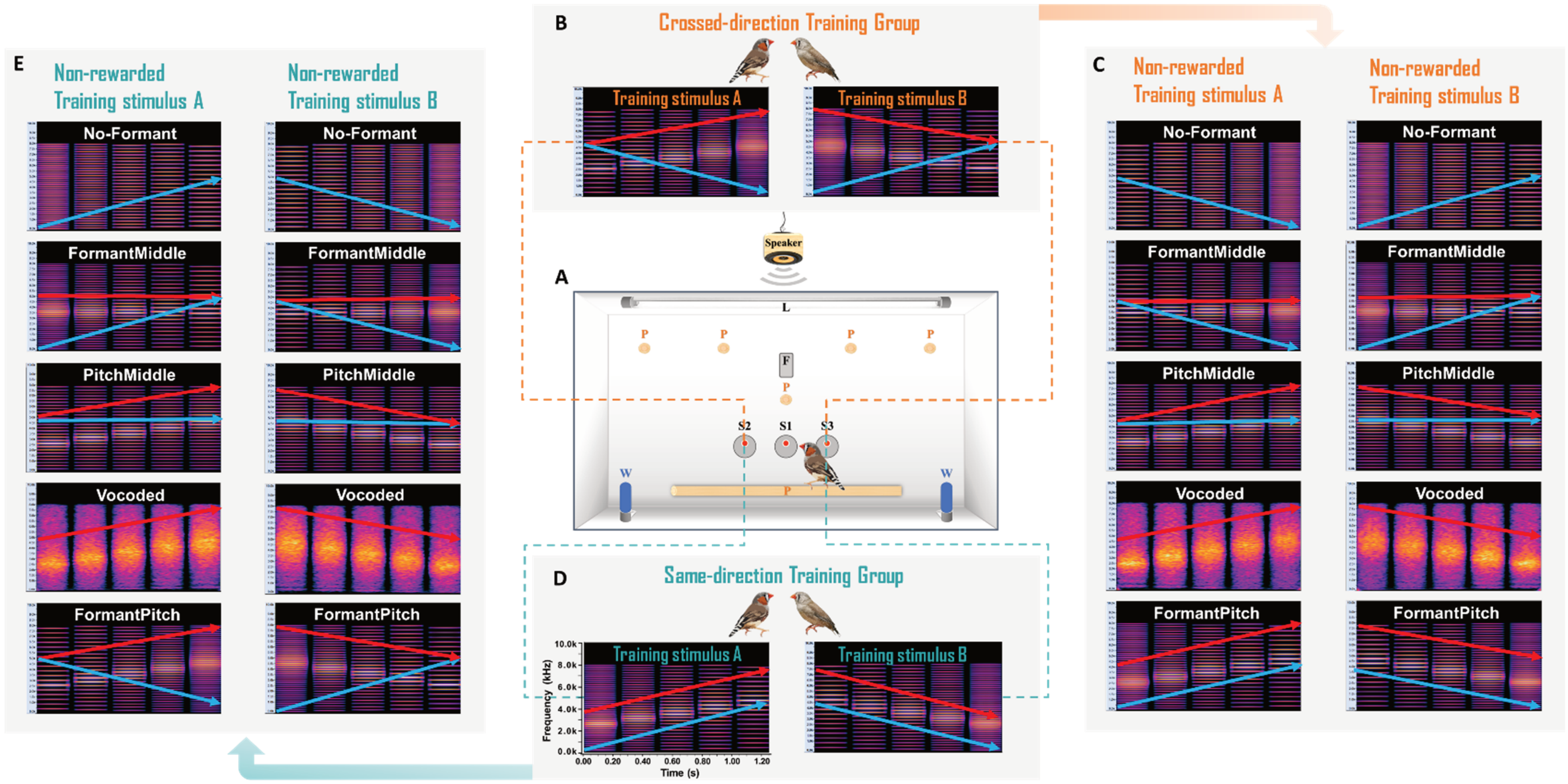


Figure 1. Operant conditioning apparatus (Skinner box) and stimuli used for the experiment: (A) Schematic front view of the Skinner box. A speaker (top of figure) is suspended from the ceiling above the cage. Within the cage, there are several perches (P) for the bird to sit on, a food hatch (F) located in the upper middle of the back panel, 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. **(B)** An example of a pair of training stimuli for Crossed-direction group. **(C)** Modified stimuli used in the testing phase for Crossed-direction group. The birds of Crossed-direction group were tested with 5 modified versions of each training stimulus after completion of the training — see text for a description of these manipulations. **(D)** A pair of training stimuli for Same-direction group. **(E)** Modified stimuli in the testing phase for Same-direction group. The birds of Same-direction group were also tested with 5 similarly modified versions of each training stimulus. In the spectrogram of each tone sequence, the direction (but not the exact value) of the pitch contour is indicated by a blue line, and formant contour is indicated by a red line.

Test stimuli

To test the relative importance of the pitch contour and formant contour in discrimination of the training stimuli, the birds were tested with modified versions of the training stimuli (Fig. 1C & 1E). Praat and Audacity were used to edit each original training stimulus to produce a version in which either pitch contour or the formant contour was changed. For both the Crossed-direction and the Same-direction training group, the test stimuli were always modified from the training stimuli in an identical way (some examples of the training and test stimuli are provided as supplementary material):

- No-Formant – The purpose of this manipulation was to create stimuli where all frequency bands have the same energy, while keeping the f0 values identical to training stimuli. The construction of the No-Formant version followed the same procedure as that of the training stimuli, except for omitting the Formant synthesis step. This ensured that the No-Formant version maintained the same characteristics (tone sequences with increasing/decreasing pitch contour) as the training stimuli, yet without any formant.

- FormantMiddle – In this stimulus a single, fixed formant was added to the training stimulus, using the formant value of the middle unit of the sequences. As such it preserved the pitch contour, while all elements have the same spectral envelope. This manipulation was accomplished by assigning a same "Frequency (Hz)" value (the "Parametric EQ" function in Audacity) to tonal units of different f0.

- PitchMiddle – In this manipulation, all pitches were equalized to the f0 of the middle unit, while preserving the formant contour of the initial training stimulus. For example, a sequence with five tonal units featuring "f0=290Hz" and arranged in a formants-ascending pattern of "2.6kHz- 3.1kHz- 3.6kHz- 4.1kHz- 4.6kHz". This adjustment was achieved by modifying the "Frequency (Hz)" parameter within the "Parametric EQ" function in Audacity.

- Vocoded – This modification maintains the spectral envelope (both its shape and position) of the training stimulus, but averages the energy within specific frequency bands, thus removing any harmonic structure. For this we used the Matt Winn's Praat vocoded script (http://www.mattwinn.com/praat/vocode_all_selected_v45.txt) to synthesize a vocoded version of training stimuli. The script was set to divide cut-off frequency bandwidths equally for 30 bands contiguous with smooth transitions (From lowCornerFreq 100Hz to highCornerFreq 8000Hz).

- FormantPitch – In this modification, the formant contour of the training stimulus was altered by adjusting the formant frequencies of each tonal unit and arranging them in a reversed order compared to the initial training stimulus. This modification aimed to combine the formant contour of one training stimulus with the pitch contour of the other training stimulus from the same training pair. For example, a modified sequence would share the pitch contour of training stimulus A while sharing the formant contour of training stimulus B. In the analysis, the responses to this ambiguous “FormantPitch” stimulus are compared to the training stimulus that shares the same formant contour (see Fig. 1C & 1E).

Procedure

A Go-Left/Go-Right paradigm was employed for both training and testing. The experimental procedure consisted of five phases, namely acclimation, pre-training, discrimination training, transition, and probe testing.

Acclimation phase

In the acclimation phase, each of the birds was introduced to a Skinner box, with the food hatch left open. The pecking sensors’ LED lights were illuminated to attract attention from the bird. The primary objective of this phase was to familiarize the birds with the cages and the location of the food source. Pecking the central sensor, S1, resulted in the playback of either sound A or sound B with a 50% probability for each. Pecking one of the side sensors, S2 or S3, triggered the playback of one of the two sounds. After a period of several hours or overnight, the food hatch was closed, marking the transition to the next phase.

Pre-training phase

This phase aimed to acquaint the birds with the training procedures. In this phase, the food hatch was closed so that the birds had to learn to peck all three sensors. Pecking each sensor had specific consequences: pecking S1 resulted in sound A or sound B playback without food reinforcement, pecking S2 led to sound A playback accompanied by a 15-seconds food hatch opening, and pecking S3 triggered sound B playback along with a 15-seconds food hatch opening. This training continued until the birds consistently pecked all sensors and associated sensor pecking with access to food. Some of the birds may also learn the association between specific sounds and the corresponding sensors in this phase. In cases when the birds did not spontaneously peck the sensors, the experimenter could activate or deactivate the LED lights to attract the bird's attention. Once the birds consistently pecked all sensors for several days, the discrimination training phase commenced.

Discrimination training

During the discrimination training phase, the birds were trained to peck the middle sensor (S1) to elicit sound playback, and then to subsequently peck either the left or right sensor, depending on the played sound triggered by the middle sensor. Correct responses, where the bird pecked the sensor associated with the played sound, were rewarded with a 15-second access to food hatch as the positive feedback. If an incorrect sensor was pecked, the light of the lamp was turned off for 1 second as a signal of negative feedback for the bird. Prior to any pecking, only the LED light for S1 was illuminated. For instance, when sound A was played, pecking S2 opened the food hatch, while pecking S3 resulted in a preset time of darkness, and vice versa. If a bird failed to respond within 15 seconds, the trial ended without a food reward or a light-off hint. The duration of this phase varied among individual birds. The discrimination rate for each bird, representing the proportion of correct responses out of all trials, was calculated on a daily basis. Once a bird achieved a discrimination score greater than 0.75 for the training stimuli for three consecutive days (with an accuracy rate of each sensor pecking exceeding 0.60 for three consecutive days), it was considered to have successfully discriminated the trained sequence pair, and the training transitioned to the next phase.

Transition phase

During the transition phase, the training stimuli remained the same as in the discrimination training phase, but the frequency of reinforcement by food or darkness was reduced to occur randomly on 80% of the trials (instead of 100% during the discrimination training phase). In the remaining 20% of trials (with stimuli identical to the training sounds), the subjects did not receive reinforcement in the form of food or darkness. If the birds maintained the same level of discrimination for two consecutive days during this phase, the test phase began.

Probe testing phase

In this phase, 20% of the pecks on S1 resulted in presenting one of twelve probe stimuli. These twelve probe stimuli were never reinforced and were randomly interspersed between training stimuli. Ten of these were modified versions of the training stimuli (five modified versions of training stimulus A and five of training stimulus B). The other two probe stimuli were non-reinforced training stimuli. The remaining 80% were training stimuli with reinforcement. Testing continued until each probe stimulus had been presented 40 times to a bird. After reaching this, the bird was transferred back to its aviary. The order of stimulus presentation was randomized across the subjects.

Analysis

To assess potential differences in the speed of discrimination learning between the two training groups, we analyzed the cumulative number of trials until reaching the learning criterion, including the day when the criterion was achieved. As the distribution of trial numbers did not conform to a normal distribution, a Mann-Whitney-Wilcoxon Test (R Core Team, 2016) was employed to examine any significant differences in learning speed (i.e., the number of required training trials) between the two training groups.

The reactions to the various test stimuli were classified 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 sensor). For the statistical analyses, we examined the proportion of "correct responses" out of "correct + incorrect responses" (Correct rate = Number_CorrectResp / (Number_CorrectResp + Number_IncorrectResp)), as well as the "response rate", calculated as "correct + incorrect responses" to modifications of sound A plus those to modification of sound B, as the proportion of the 40 presentations of each test stimulus (Response rate = (Number_CorrectResp + Number_IncorrectResp) / (Number_CorrectResp + Number_IncorrectResp + Number_NoResp)).

We used Generalized Linear Mixed-effects Models (GLMMs) to examine the discrimination of various test sounds by the birds. All model analyses were conducted in Rstudio (R Core Team, 2016 & lme4; Bates *et al.*, 2015). We calculated the "Correct rate" and the "Response rate" based on the counts of "correct response", "incorrect response", and "no response", combining the response counts to (variants of) Training stimuli A and B, (using the function cbind, R package mice; Van Buuren & Groothuis-Oudshoorn, 2011), and used these two proportions rates as response variables in GLMMs in R (using the function glmer, R package lme4; Bates *et al*., 2015). We used "Training_Group" (Same-direction or/ Crossed-direction training pairs), "Test_Treatment", "Sequence_Series" and the interaction between "Training_Group" and "Test_Treatment" as covariates in the full model with "Bird_ID", "Age", "Number_of_Training_Trials" as the random factors and a binomial error structure of the "Correct rate" and the "Response rate". The best model was chosen based on corrected Akaike criterion (AICc) provided by dredge model selection (using the function Dredge, R package MuMIn; Bartoń, 2020). The model with the smallest value of AICc was considered to be the best model by default, but if "Training_Group", "Test_Treatment" and the interaction between these two were not part of the best model, we kept them in the final model anyway because these were variables of our interest. To determine the effect and significance of the covariates, we ran the final models and, if applicable, used Post hoc Tukey's HSD tests to make pairwise comparisons of the test treatments (using the emmeans function, R package lsmeans; Lenth, 2016), with false discovery rate (FDR) correction of p-values (Benjamini & Hochberg, 1995) for multiple comparisons. In the above models, the counts of the responses to (modifications of) both sequence A and sequence B were combined in all tests. In the above models, the counts of the reactions to modifications of both sound A and sound B were combined.

Additionally, to determine whether the individual test stimuli were discriminated above chance (50%), the ratio of "Number_CorrectResp / Number_IncorrectResp" was assessed (specifically, whether this ratio differed from 1). We did so by applying the log (Number_CorrectResp / Number_IncorrectResp) (indicated as "Log (Cor/Inco)" from now on as the response variable against a log (Odds-ratio) = 0 in a GLM analysis. If correct/incorrect = 1, then the probability of observing a correct response is as large as the probability of observing an incorrect response, representing both probabilities are 0.5, then log (Odds- ratio) = log (1) = 0. Therefore, comparing the outcomes of the Binomial GLM to 0 is comparing the results to the 50% chance for a correct response.

RESULTS

Learning speed

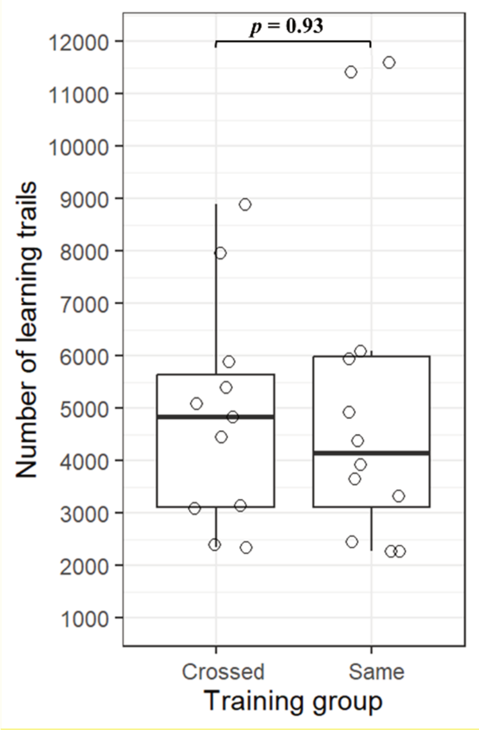


Figure 2. Number of learning trials needed to reach the learning criterion. Individual zebra finch results are shown with open circles. There is no significant difference between the Different-syllables group and the Same-syllables group in learning speed. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range.

The discrimination training lasted until the zebra finches reached the learning criterion of over 75% correct responses to both sound A and sound B for three days. All twenty-three birds finished the training and learned the discrimination on a median value of 4465 (IQR = 2857) trials to reach the criterion. No significant difference ($p = 0.93$, z = 0.12) was found between the Crossed-direction group (Median = 4841, IQR = 2808) and the Same-direction group (Median = 4158, IQR = 3126). It suggests that birds from two training groups learn approximately equally fast in both training conditions.

The impact of pitch and formant on stimulus classification

Do training groups differ in responses to test stimuli?

We compared the Correct rates and Response rates for both experimental groups in response to the training and various test stimuli (Fig. 1). We chose the two factors, “Training_Group” and “Test_Treatment”, along with their interaction effects, which were used as fixed factors in the statistical models for the response variables “Correct rates” and “Response rates” (see Table 2). Although the two selected models were not the most recommended ones based on the dredge model selection, they included the variables of interest and were still close to the most recommended models (based on AICc).

All modifications of the training stimuli resulted in a strong reduction of the correct rate indicating that both formant and pitch were used to distinguish the training stimuli, irrespective of the training group. Most test stimuli did not exhibit significant differences in the correct rate between the two training groups (see Fig. 3A), with the exception of the “No-Formant” version, which showed a significant distinction. In this case, the Crossed-direction group achieved a higher Correct rate compared to the Same-direction group (Crossed – Same = 0.347 ± 0.114, $p = 0.014$, as indicated in Table 3).

There were no significant differences in Response rates for any of the stimuli between the two training groups (Fig. 3B). Notably, the variation in Response rate for all five modified stimuli in the Same-direction group was more prominent compared to the Crossed-direction group. This suggests that the Same-direction training condition, rather than the Crossed-direction, might affect the consistency of individual responses to the modified stimuli, or that some individuals within the Same-direction group consistently respond more often than others.

Table 2 Summary of the GLMs selection for a) the proportion of correct responses if birds respond to one of two sounds; and b) the proportion of trials that birds respond with pecking A or B.

Model	df	logLik	AICc	Δi	wi
a. Correct rate of responses					
1 Training_Group + Test_Treatment + Test_Treatment:Training_Group + Sequence_Series + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	16	-408.692	853.9	2.80	0.073
2* Training_Group + Test_Treatment + Test_Treatment:Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	15	-408.694	851.3	0.24	0.263
3 Training_Group + Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	10	-415.021	851.8	0.69	0.210
4 Test_Treatment + Sequence_Series + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	10	-415.835	853.4	2.32	0.093
5 Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	9	-415.839	851.1	0.00	0.296
null (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	4	-764.165	1536.6	685.55	0.000
b. Response rate of trials					
1* Training_Group + Test_Treatment + Test_Treatment:Training_Group + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	15	-441.024	916.0	3.17	0.075
2 Training_Group + Test_Treatment + Sequence_Series + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	11	-445.658	915.4	2.60	0.100
3 Training_Group + Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	10	-446.210	914.2	1.34	0.187
4 Test_Treatment + Sequence_Series + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	10	-445.983	913.7	0.89	0.235
5 Test_Treatment + (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	9	-446.703	912.8	0.00	0.366
null (1|Bird_ID) + (1|Age) + (1|Number_of_Training_Trials)	4	-715.047	1438.4	525.58	0.000

Note: Best five models of the model selection (ranked by AICc and logLik) and the null models. The Akaike weight (wi) indicates the probability of a better model in the model candidates set, and Delta AICc (Δi) was used to show the difference in AICc score between the best model and the model being compared. A * indicates the model we choose.

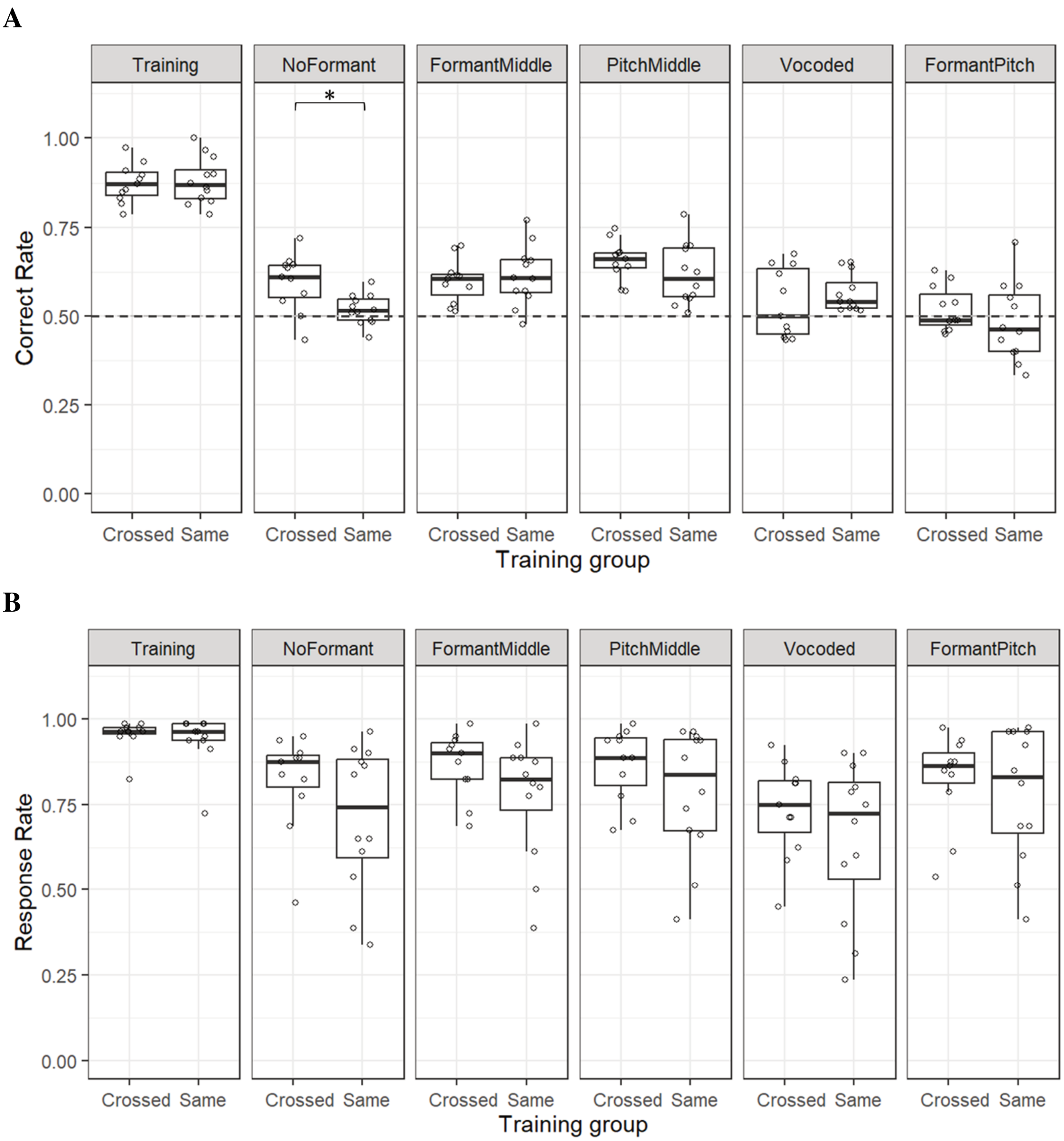


Figure 3. Correct rate of responses and Response rate of trials: (A) the proportion of correct responses (Correct rate) to the training and modified stimuli for the two training groups; **(B)** the Response rates to the training and modified stimuli for the two training groups. “Crossed” refers to the Crossed-direction training group, and “Same” refers to the Same-direction training group. Significant differences between the responses between the training groups are indicated: * refers to a significant difference of $0.01 < p \leq 0.05$, and for non-indicated comparisons p value is > 0.05. Box plots show median, 1st and 3rd quartile, and whiskers the 1.5 interquartile range. The dashed line represents chance level, which was 50% for both tasks.

Do different test stimuli give rise to different responses?

Post hoc Tukey's HSD tests (Table 3a, b) revealed significantly higher Correct rates and Response rates for the non-rewarded training stimuli compared to all five modified stimuli within each training group (both $p < 0.0001$).

Both the Crossed-direction and Same-direction groups exhibited higher correct rates in response to the "PitchMiddle" stimuli compared to the "No-Formant" (Crossed-direction group: $p < 0.05$; Same-direction group: $p = 0.0001$), "Vocoded" (Crossed-direction group: $p < 0.0001$; Same-direction group: $p < 0.05$), and "FormantPitch" (Crossed-direction group: $p < 0.0001$; Same-direction group: $p < 0.0001$) stimuli. Additionally, both groups showed higher correct rates in response to the "FormantMiddle" stimuli compared to "FormantPitch" stimuli (Crossed-direction group: $p < 0.01$; Same-direction group: $p < 0.0001$). Moreover, the Crossed-direction group responded with significantly higher correct rates to the "PitchMiddle" stimuli than to the "FormantMiddle" stimuli ($p < 0.05$) and responded with higher correct rates to the "No-Formant" stimuli than to the "Vocoded" ($p < 0.05$) and "FormantPitch" ($p < 0.01$) stimuli. The Same-direction group exhibited significantly higher correct rates when responding to the "FormantMiddle" stimuli than to the "No-Formant" ($p < 0.001$). Additionally, a discernible trend towards differentiation between the "No-Formant" and "Vocoded" stimuli, as well as between the "FormantMiddle" and "Vocoded" stimuli, was observed in the Same-direction group (both $p = 0.08$).

Both groups showed a higher response rate in responding to the "PitchMiddle" and the "FormantMiddle" stimuli than to the "No-Formant" (Crossed-direction group: both $p < 0.01$; Same-direction group: both $p < 0.001$) and the "Vocoded" (Crossed-direction group: both $p < 0.0001$; Same-direction group: both $p < 0.0001$). Additionally, both groups responded with a significantly lower response rate to the "Vocoded" stimuli than to other four modified stimuli (Crossed-direction group: both $p < 0.0001$; Same-direction group: both $p < 0.01$). Moreover, the Crossed-direction group responded with a significantly higher response rate to the "PitchMiddle" and the "FormantMiddle" stimuli than to the "FormantPitch" (both $p < 0.05$) stimuli. The Same-direction group responded with a significantly higher response rate to the "FormantPitch" stimuli than to the "No-Formant" ($p < 0.001$) and "Vocoded" ($p < 0.0001$) stimuli.

Overall, the response rates of birds in both groups shows a pattern that is somewhat similar to their correct rate for most of the modified stimuli. In both groups, the birds predominantly responded to the “FormantMiddle” and “PitchMiddle” stimuli, while responding least to the “Vocoded” stimuli. However, noteworthy is the relatively high response rate to the “FormantPitch” stimuli in both groups, even though the correct rate for this modified version was relatively low.

Table 3 Post hoc test results of Binomial GLM for the interaction of Test_Treatment & Training_Group

Stimuli	**Training_Group**	**estimate**	**SE**	**z.ratio**	***p*.value**
a. Correct rate of responses					
Training	Crossed - Same	-0.032	0.150	-0.214	0.8302
No-Formant	**Crossed - Same**	**0.347**	**0.114**	**3.039**	**0.0144**
FormantMiddle	Crossed - Same	-0.081	0.112	-0.721	0.5650
PitchMiddle	Crossed - Same	0.148	0.113	1.305	0.4869
Vocoded	Crossed - Same	-0.117	0.119	-0.985	0.4869
FormantPitch	Crossed - Same	0.128	0.111	1.155	0.4869
Training - No-Formant	**Crossed**	**1.522**	**0.129**	**11.807**	**<.0001**
Training - FormantMiddle	**Crossed**	**1.544**	**0.128**	**12.109**	**<.0001**
Training - PitchMiddle	**Crossed**	**1.281**	**0.129**	**9.929**	**<.0001**
Training - Vocoded	**Crossed**	**1.781**	**0.131**	**13.636**	**<.0001**
Training - FormantPitch	**Crossed**	**1.866**	**0.128**	**14.596**	**<.0001**
No-Formant - FormantMiddle	Crossed	0.022	0.106	0.210	0.8333
No-Formant - PitchMiddle	**Crossed**	**-0.241**	**0.108**	**-2.233**	**0.0320**
No-Formant - Vocoded	**Crossed**	**0.259**	**0.110**	**2.361**	**0.0249**
No-Formant - FormantPitch	**Crossed**	**0.344**	**0.106**	**3.236**	**0.0023**
FormantMiddle - PitchMiddle	**Crossed**	**-0.263**	**0.106**	**-2.477**	**0.0199**
FormantMiddle - Vocoded	**Crossed**	**0.237**	**0.108**	**2.190**	**0.0329**
FormantMiddle - FormantPitch	**Crossed**	**0.322**	**0.105**	**3.073**	**0.0035**
PitchMiddle - Vocoded	**Crossed**	**0.500**	**0.110**	**4.549**	**<.0001**
PitchMiddle - FormantPitch	**Crossed**	**0.585**	**0.107**	**5.489**	**<.0001**
Vocoded - FormantPitch	Crossed	0.085	0.108	0.786	0.4626
Training - No-Formant	**Same**	**1.901**	**0.127**	**14.935**	**<.0001**
Training - FormantMiddle	**Same**	**1.496**	**0.126**	**11.825**	**<.0001**
Training - PitchMiddle	**Same**	**1.461**	**0.126**	**11.562**	**<.0001**
Training - Vocoded	**Same**	**1.696**	**0.130**	**13.063**	**<.0001**

Training - FormantPitch	**Same**	**2.026**	**0.125**	**16.184**	**<.0001**
No-Formant - FormantMiddle	**Same**	**-0.405**	**0.108**	**-3.762**	**0.0003**
No-Formant - PitchMiddle	**Same**	**-0.440**	**0.107**	**-4.092**	**0.0001**
No-Formant - Vocoded	Same	-0.205	0.111	-1.837	0.0811
No-Formant - FormantPitch	Same	0.125	0.106	1.182	0.2543
FormantMiddle - PitchMiddle	Same	-0.035	0.107	-0.327	0.7440
FormantMiddle - Vocoded	Same	0.200	0.111	1.810	0.0811
FormantMiddle - FormantPitch	**Same**	**0.530**	**0.105**	**5.041**	**<.0001**
PitchMiddle - Vocoded	**Same**	**0.235**	**0.110**	**2.128**	**0.0455**
PitchMiddle - FormantPitch	**Same**	**0.565**	**0.105**	**5.381**	**<.0001**
Vocoded - FormantPitch	**Same**	**0.330**	**0.109**	**3.026**	**0.0037**
b. Response rate of trials					
Training	Crossed - Same	-0.036	0.365	-0.098	0.9964
No-Formant	Crossed - Same	0.525	0.316	1.663	0.5294
FormantMiddle	Crossed - Same	0.507	0.321	1.579	0.5294
PitchMiddle	Crossed - Same	0.416	0.321	1.296	0.5294
Vocoded	Crossed - Same	0.322	0.311	1.034	0.5576
FormantPitch	Crossed - Same	0.116	0.318	0.364	0.8542
Training - No-Formant	**Crossed**	**1.591**	**0.187**	**8.509**	**<.0001**
Training - FormantMiddle	**Crossed**	**1.202**	**0.193**	**6.239**	**<.0001**
Training - PitchMiddle	**Crossed**	**1.212**	**0.192**	**6.299**	**<.0001**
Training - Vocoded	**Crossed**	**2.130**	**0.182**	**11.694**	**<.0001**
Training - FormantPitch	**Crossed**	**1.557**	**0.187**	**8.313**	**<.0001**
No-Formant - FormantMiddle	**Crossed**	**-0.389**	**0.137**	**-2.844**	**0.0067**
No-Formant - PitchMiddle	**Crossed**	**-0.378**	**0.136**	**-2.773**	**0.0076**
No-Formant - Vocoded	**Crossed**	**0.539**	**0.121**	**4.459**	**<.0001**
No-Formant - FormantPitch	Crossed	-0.033	0.129	-0.259	0.8525
FormantMiddle - PitchMiddle	Crossed	0.010	0.144	0.072	0.9423
FormantMiddle - Vocoded	**Crossed**	**0.928**	**0.130**	**7.145**	**<.0001**
FormantMiddle - FormantPitch	**Crossed**	**0.355**	**0.137**	**2.589**	**0.0120**
PitchMiddle - Vocoded	**Crossed**	**0.917**	**0.130**	**7.080**	**<.0001**
PitchMiddle - FormantPitch	**Crossed**	**0.345**	**0.137**	**2.518**	**0.0136**
Vocoded - FormantPitch	**Crossed**	**-0.572**	**0.122**	**-4.709**	**<.0001**
Training - No-Formant	**Same**	**2.149**	**0.164**	**13.140**	**<.0001**
Training - FormantMiddle	**Same**	**1.743**	**0.166**	**10.516**	**<.0001**
Training - PitchMiddle	**Same**	**1.663**	**0.166**	**9.998**	**<.0001**

Training - Vocoded	**Same**	**2.484**	**0.163**	**15.273**	**<.0001**
Training - FormantPitch	**Same**	**1.707**	**0.166**	**10.282**	**<.0001**
No-Formant - FormantMiddle	**Same**	**-0.406**	**0.116**	**-3.509**	**0.0006**
No-Formant - PitchMiddle	**Same**	**-0.486**	**0.117**	**-4.162**	**0.0001**
No-Formant - Vocoded	**Same**	**0.335**	**0.110**	**3.060**	**0.0028**
No-Formant - FormantPitch	**Same**	**-0.442**	**0.116**	**-3.805**	**0.0002**
FormantMiddle - PitchMiddle	Same	-0.080	0.120	-0.663	0.5853
FormantMiddle - Vocoded	**Same**	**0.741**	**0.114**	**6.507**	**<.0001**
FormantMiddle - FormantPitch	Same	-0.036	0.120	-0.300	0.7641
PitchMiddle - Vocoded	**Same**	**0.821**	**0.115**	**7.140**	**<.0001**
PitchMiddle - FormantPitch	Same	0.044	0.121	0.363	0.7641
Vocoded - FormantPitch	**Same**	**-0.777**	**0.114**	**-6.794**	**<.0001**

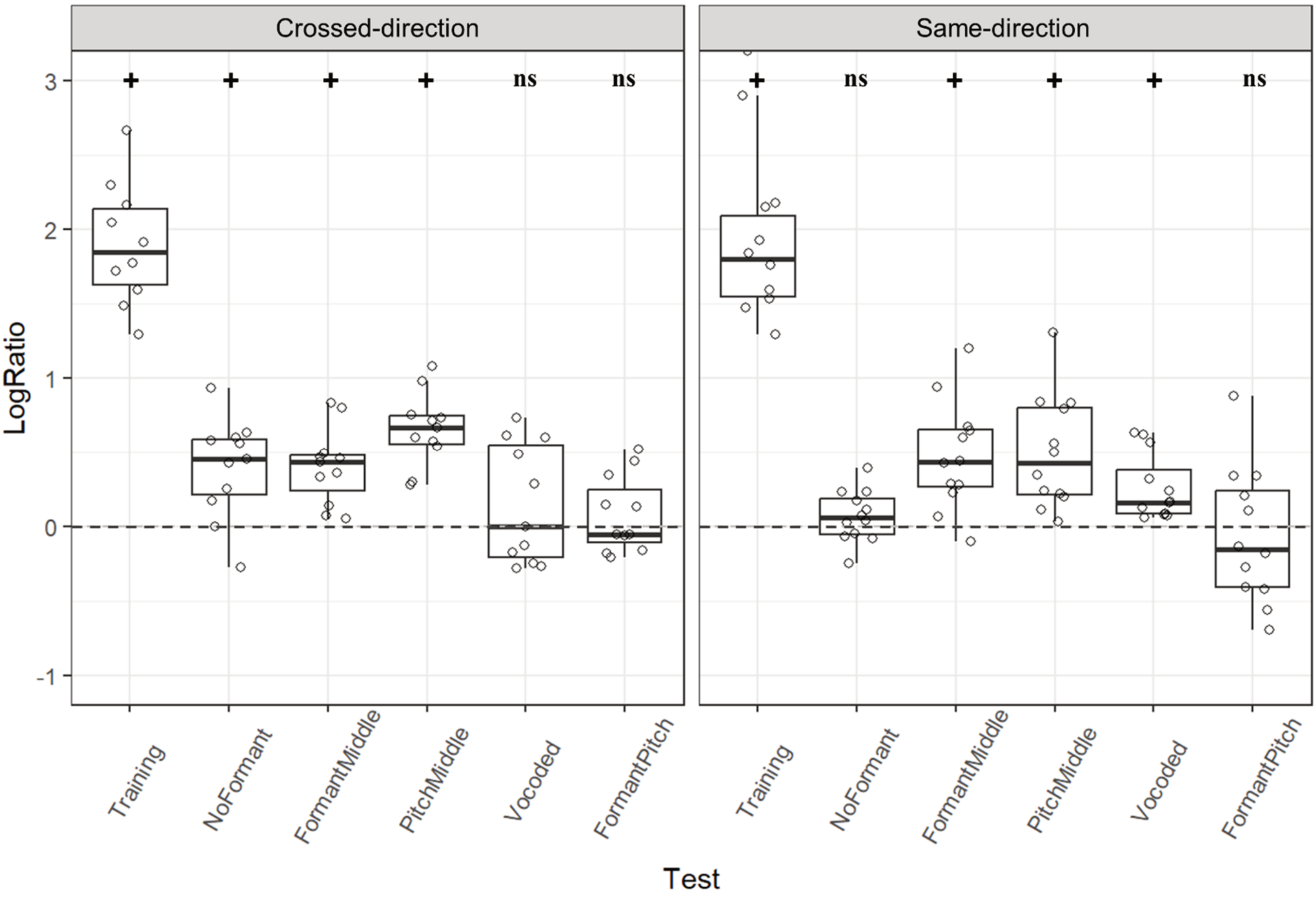
Note: Response variables in GLMMs: (**a**) the proportion of correct responses if birds respond to one of two sounds; and (**b**) the proportion of trials that birds respond with pecking A or B. “Crossed” refers to the Crossed-direction training group, and “Same” refers to the Same-direction training group. Each of the corrected pairwise multiple comparison tests is separated by borders within the table. Bold indicates significant differences <0.05.

Are modified stimuli still discriminated?

The previous analyses primarily focused on disparities in Correct rates among training groups and across test stimuli. However, these analyses did not show whether a low Correct rate means that birds are no longer able to discriminate between the modified versions of training sound A and training sound B. If the birds are still capable of associating the test stimuli with the respective training stimuli, the proportion of correct responses to the test stimuli should be higher than the proportion of incorrect responses. For the Crossed-direction group, two modified versions (the “Vocoded” and the “FormantPitch” versions) were statistically similar to 0, suggesting that the birds responded to these two modified versions by chance. In contrast, the rest test stimuli significantly differed from 0, indicating that these modified versions still showed resemblance to the training stimuli from which they were derived. In the Same-direction group, two modified versions (the “No-Formant” and the “FormantPitch” versions) were statistically similar to 0, with the remaining test versions showing a significant difference from 0, favouring correct responses (Table 4 & Fig. 4). In conclusion, both groups of birds maintained the ability to discriminate the “FormantMiddle” and “PitchMiddle” versions of the training stimuli, but their discrimination diminished for the “FormantPitch” version. Interestingly, the Crossed-direction group still differentiated the “NoFormant” version but lost discrimination for the "Vocoded" version, whereas the Same-direction group exhibited the opposite pattern for these two versions. These results suggest that different training conditions had some effect on birds' attention to pitch and formant in the training sequences.

Figure 4. Visualisation of logRatios = log (Correct/Incorrect). The Log (Cor/Inco) for two training groups responding to the various test stimuli. A + indicates that the Log (Cor/Inco) of a Test treatment is significantly above 0. A ns indicates that the Log (Cor/Inco) of a Test treatment is overlapping with 0. Box plots show median, 1st and 3rd quartile, and whiskers the 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 logRatios was based on the counts of "correct response" and "incorrect response" from the same data set that was also used for Fig.3.

Table 4 Estimates and 95% confidence intervals for the correct identification of test stimuli



Note: Lower CL and Upper CL represent the lower and upper 95% confidence limits (CL) 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 toward correct responses. If they are both negative, then they are biased toward incorrect responses. Bold indicates significance.

DISCUSSION

The present study examined the perceptual interaction between pitch and formant cues in zebra finches' auditory discrimination, employing a Go-left/Go-right paradigm. Through a systematic manipulation of the pitch and formant contours of tone sequences, our study investigated the relative contributions of these attributes to the recognition of sound sequences, as well as examining the presence of any perceptual interaction between them. Below, we discuss key findings concerning the effects of pitch and formant contour directions on birds' discrimination learning, the influence of training conditions on their discrimination of modified stimuli, and the interplay between pitch and formant contours in zebra finches' auditory discrimination.

Both training groups demonstrated similar learning speeds, suggesting that the perceptual interactions (if there were any) between pitch and formant contours, whether going in the same or opposite direction, did not affect the difficulty of acquiring discrimination. However, when analyzing Correct and Response rates for various test stimuli, distinctions between the training groups emerged. It then becomes evident that the relative importance of pitch and formant contours shows some effect of training conditions.

Among the modified versions tested, the "PitchMiddle" and "FormantMiddle" versions were consistently well-recognized by the birds, indicating that both formants and pitch, respectively, were attended for tone sequence recognition. In contrast, the responses to the "FormantPitch" version, despite its relatively high response rates, were at chance level. This suggests that the conflicting information presented by pitch and formant contours in the "FormantPitch" version led the birds to perceive it as ambiguous. On the other hand, among the five modified versions, the "Vocoded" version proved to be the most easily detected as differing from the training sounds, as the birds respond least to this version. The manipulation involving noise-vocoding not only disrupts the harmonic attributes (hence also removes pitch information) of the tones but also renders the spectrum of the stimulus "noise-like". Such "noise-like spectrum" alterations may likely capture the birds' attention, making the "Vocoded" stimuli distinguishable from the training stimuli when perceived by the birds. In addition, both groups showed a pronounced distinction in responding between training stimuli and their modified versions. This suggests that zebra finches excel in detecting spectral structures in either pitch or formant contours no matter the manipulation was on pitch or formant cues.

Remarkably, the "No-Formant" version had a higher impact on stimulus recognition within the same-direction group, which resulted in the same-direction group losing discrimination of this version, while remaining distinguishable in the crossed-direction group. This suggests that the crossed-direction group tends to focus slightly more on pitch contour than on formant contour for stimulus identification, although such a bias is not visible in the responses of both groups to the "FormantMiddle" and "PitchMiddle" versions. Zebra finches trained with crossed-direction sequences exhibited elevated correct rates for the "No-Formant" version, suggesting that the crossed-direction group paid relatively less attention to changes in formants compared to the same-direction group, although here also no difference is present between the groups in their responses towards the "FormantMiddle" and "PitchMiddle" versions. Such a difference would be expected if the groups really differed in their relative attention for pitch and formants.

A key question addressed in our study pertains to whether zebra finches when presented with stimuli containing two salient parameters prioritize one parameter for discrimination while disregarding the other, or whether they consider both parameters in their discrimination process. Our findings indicate that the latter strategy is adopted by these birds. Moreover, one might anticipate that differentiating between stimuli A and B could be more straightforward for the birds when both pitch and formant contours are oriented in the same direction, and as a result rely more on one parameter rather than taking both into account. However, based on our results, there is no evidence that this alignment had any impact on the birds' discrimination learning, apart from the small bias observed in the response to the "No-Formant" version in the crossed-direction group. These results together demonstrate that with these stimuli both parameters play comparable roles in zebra finches' tone sequence recognition. The variation in response to different modified versions can be explained by "additive effects" rather than more complex interactions between attribute contours or training conditions: both attributes are assessed and used to distinguish stimuli, and if one attribute remains constant throughout the stimulus sequence, the other suffices to keep discriminating the sequences.

With respect to the significance of our findings it is worthwhile to compare our study with the research of Bregman *et al.* (2016) and Burgering *et al.* (2019) on the role of various acoustic features in songbird's auditory perception. Bregman *et al.* (2016) investigated the ability of starlings to discriminate a sequence of synthetic harmonic tones. This investigation revealed that starlings were attending to spectral shape (i.e., spectral envelope) over absolute pitch in tonal sound discrimination tasks – they still showed recognition of vocoded tone sequences. This is in contrast with our finding that zebra finches could not identify vocoded versions of the tone sequences, which are maintaining the spectral shape, but lack pitch information. Our finding also differs from an earlier study by Burgering *et al.* (2019) on zebra finches' perception of vowel-like sounds, examining the roles of pitch and spectral envelope. This research revealed that in this case the zebra finches were responding to vocoded stimuli, hence attending to the spectral shape. The discrepancy between the results of our study and those of Bregman *et al.* (2016) and Burgering *et al.* (2019) may have different causes. It could be that starlings and zebra finches are sensitive to different vocal parameters. In relation to the study by Burgering *et al.* (2019), the finding that the zebra finches attend to other parameters (i.e., pitch and spectral envelope) than in the current study indicates that what gets attention may depend on the nature of the stimulus. It is still unclear what causes this disparity in zebra finches' responses between shifts in artificial tone sequences and in vowel-like sound elements. One possible reason could be that the nature of the training stimuli influences the birds' attention to specific acoustic features during discriminating tasks, as well as their future generalization of learning to novel stimuli. There might also be a methodological factor affecting the difference between the earlier studies and ours. In the current study, we utilize true "tone sequences" with silent gaps to partition the tone units. This stimuli design serves as a valuable complement to the starling experiment conducted by Bregman *et al.* (2016), which used similar "tone sequences" but without silent gaps. Previous studies, including those conducted by Bregman *et al.* (2016) and Burgering *et al.* (2019), have primarily focused on local features, whether in isolated units or in tone sequences lacking silent gaps. The absence of silent gaps raises the question whether birds perceive the entire "tone sequence" as a single acoustic object or as a sequence of tonal units processed sequentially. Our way of arranging stimuli prompted zebra finches to engage with the comprehensive contour of the overall tone sequence. This methodology is distinct from previous investigations that concentrated on localized features, such as the pitch or spectral attributes of individual acoustic units.

It is interesting to compare our results on zebra finches' attention to pitch and formant contours with a study on how humans attend to pitch and timbre (McPherson & McDermott, 2023). McPherson & McDermott (2023) demonstrated that judgments of harmonic sounds in humans relied on f0 representations, while relative pitch judgments were influenced by timbral differences, leading to biases in discrimination tasks. Comparatively, our findings highlight zebra finches' ability to integrate pitch and formant contours for discrimination. The fact that zebra finches are capable of attending to both attributes in sound recognition is noteworthy, as it differs from the human tendency to prioritize one attribute (i.e., pitch) over the other (i.e., timbre) in perceiving tonal sounds. Moreover, the stimuli used in our study, consisting of sequences of units, required broader attention to contour attributes rather than local features like the pitch or timbre of individual units. However, it's premature to determine the similarities or differences between our current study and these human studies. This is due to the limited number of studies on the perceptual interplay between attributes in human auditory perception, aside from a few studies (e.g., Shinn-Cunningham *et al.*, 2007; McPherson & McDermott, 2023). Conducting a similar experiment as presented here with human subjects would enable a direct comparison (Ning *et al.*, in prep).

CONCLUSION AND OUTLOOK

Our study investigated the interplay between pitch and formant attributes in zebra finches. The findings demonstrate that when tone sequences exhibit variations in both pitch and formant across a series of tones, zebra finches attend to both pitch and formant contours when distinguishing the series. This observation holds true regardless of whether the changes in pitch and formant across the tones occur in the same or opposite directions, indicating a limited impact of the direction of these changes on tone sequence discrimination. Furthermore, our study, in combination with earlier ones (e.g., Ning *et al.*, 2023) reaffirms the remarkable perceptual flexibility exhibited by zebra finches. This enhanced understanding of avian auditory perception prompts consideration of how the attention to acoustic attributes extends across species. It also indicates the relevance of future cross-species experiments to elucidate the differences between humans and songbirds in attending to pitch and formant cues. This line of inquiry holds promise for uncovering the underlying mechanisms of auditory perception and contributes to the broader field of cognitive research.

References

Albouy P., Mehr S.A., Hoyer R.S., Ginzburg J., Zatorre R.J. (2023). Spectro-temporal acoustical markers differentiate speech from song across cultures. *BioRxiv*, 2023.01.29.526133. Preprint. https://doi.org/10.1101/2023.01.29.526133

Bartoń K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. https://CRAN.R-project.org/package=MuMIn

Bates D., Maechler M., Bolker B., Walker S. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw.*, 67(1): 1-48. https://doi.org/10.18637/jss.v067.i01

Benjamini Y., Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.*, 57(1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

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 envelope, not pitch, for sound pattern recognition. *Proc Natl Acad Sci.*, 113(6): 1666–1671. https://doi.org/10.1073/pnas.1515380113

Brosch M., Selezneva E., Bucks C., Scheich H. (2004). Macaque monkeys discriminate pitch relationships. *Cognition*, 91(3): 259–272. http://dx.doi.org/10.1016/j.cognition.2003.09.005

Burgering M.A., Ten Cate C., Vroomen J. (2018). Mechanisms underlying speech sound discrimination and categorization in humans and zebra finches. *Anim Cogn.*, 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. *J Comp Psychol.*, 133(1): 106-117. https://doi.org/10.1037/com0000143

Choi W. (2021). Musicianship Influences Language Effect on Musical Pitch Perception. *Front Psychol.*, 12: 712753. https://doi.org/10.3389/fpsyg.2021.712753

Crespo-Bojorque P., Celma-Miralles A., Toro J.M. (2022). Detecting surface changes in a familiar tune: exploring pitch, tempo and timbre. *Anim Cogn.*, 25(4): 951-960. https://doi.org/10.1007/s10071-022-01604-w

Cynx J., Shapiro M. (1986). Perception of missing fundamental by a species of songbird (Sturnus vulgaris). *J Comp Psychol.*, 100(4): 356–360. https://doi.org/10.1037/0735-7036.100.4.356

Dowling W.J., Harwood D.L. (1986). Music cognition Cognitive processing. Orlando, FL: Academic Press.

Fitch W.T., Fritz J.B. (2006). Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J Acoust Soc Am.*, 120(4): 2132-2141. https://doi.org/10.1121/1.2258499

Friedrich A., Zentall T., Weisman R. (2007). Absolute pitch: frequency-range discriminations in pigeons (*Columba livia*): comparisons with zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*). *J Comp Psychol.*, 121(1): 95–105. https://doi.org/10.1037/0735-7036.121.1.95

Geissler D.B., Ehret G. (2002). Time-critical integration of formants for perception of communication calls in mice. *Proc Natl Acad Sci.*, 99(13): 9021-9025. https://doi.org/10.1073/pnas.122606499

Henry K.S., Amburgey K.N., Abrams K.S., Idrobo F., Carney L.H. (2017). Formant-frequency discrimination of synthesized vowels in budgerigars (Melopsittacus undulatus) and humans. *J Acoust Soc Am.*, 142(4): 2073. https://doi.org/10.1121/1.5006912

Hoeschele M., Merchant H., Kikuchi Y., Hattori Y., ten Cate C. (2015). Searching for the origins of musicality across species. *Philos Trans R Soc Lond B Biol Sci.*, 370(1664): 20140094. https://doi.org/10.1098/rstb.2014.0094

Hoeschele M. (2017). Animal Pitch Perception: Melodies and Harmonies. *Comp Cogn Behav Rev.*, 12: 5-18. https://doi.org/10.3819/CCBR.2017.120002

Honing H., ten Cate C., Peretz I., Trehub S.E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos Trans R Soc Lond B Biol Sci.*, 370(1664): 20140088. https://doi.org/10.1098/rstb.2014.0088

Hulse S.H., Cynx J. (1985). Relative pitch perception is constrained by absolute pitch in songbirds (*Mimus*, *Molothrus*, and *Sturnus*). *J Comp Psychol.*, 99(2): 176–196. http://dx.doi.org/10.1037/0735-7036.99.2.176

Hulse S.H., Cynx J., Humpal J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *J Exp Psychol Gen.*, 113(1): 38–54. https://doi.org/10.1037/0096-3445.113.1.38

Hurly T.A., Ratcliffe L., Weisman R. (1990). Relative pitch recognition in white-throated sparrows (Zonotrichia albicollis). *Anim Behav.*, 40(1): 176–181. http://dx.doi.org/10.1016/S0003-3472(05)80677-3

Izumi A. (2001). Relative pitch perception in Japanese monkeys (*Macaca fuscata*). *J Comp Psychol.*, 115(2): 127–131. http://dx.doi.org/10.1037/0735-7036.115.2.127

Kluender K.R., Lotto A.J., Holt L.L., Bloedel S.L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *J Acoust Soc Am.*, 104(6): 3568–3582. https://doi.org/10.1121/1.423939

Kriengwatana B., Escudero P., Kerkhoven A.H., ten Cate C. (2015). A general auditory bias for handling speaker variability in speech? Evidence in humans and songbirds. *Front Psychol.*, 6: 1243. https://doi.org/doi:10.3389/fpsyg.2015.01243

Kubovy M., van den Berg M. (2008). The whole is equal to the sum of its parts: a probabilistic model of grouping by proximity and similarity in regular patterns. *Psychol Rev.*, 115(1): 131-154. https://doi.org/10.1037/0033-295X.115.1.131

Lenth R.V. (2016). Least-Squares Means: The R Package lsmeans. *J Stat Softw.*, 69(1): 1-33. https://doi.org/10.18637/jss.v069.i01

Luna D., Montoro P.R. (2011). Interactions between intrinsic principles of similarity and proximity and extrinsic principle of common region in visual perception. *Perception*, 40(12): 1467-1477. https://doi.org/10.1068/p7086

Luna D., Villalba-García C., Montoro P.R., Hinojosa J.A. (2016). Dominance dynamics of competition between intrinsic and extrinsic grouping cues. *Acta Psychol (Amst).*, 170: 146-154. https://doi.org/10.1016/j.actpsy.2016.07.001

MacDougall-Shackleton S.A., Hulse S.H. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *J Comp Psychol.*, 110(2): 139–146. https://doi.org/10.1037/0735-7036.110.2.139

McDermott J.H., Lehr A.J., Oxenham A.J. (2008). Is relative pitch specific to pitch? *Psychol Sci.*, 19(12): 1263–1271. https://doi.org/10.1111/j.1467-9280.2008.0223

McPherson M.J., McDermott J.H. (2018). Diversity in pitch perception revealed by task dependence. *Nat Hum Behav.*, 2(1): 52-66. https://doi.org/10.1038/s41562-017-0261-8

Melchor J., Vergara J., Figueroa T., Morán I., Lemus L. (2021). Formant-Based Recognition of Words and Other Naturalistic Sounds in Rhesus Monkeys. *Front Neurosci.*, 15: 728686. https://doi.org/10.3389/fnins.2021.728686

Micheyl C., Oxenham A.J. (2010). Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. *Hear Res.*, 266(1-2): 36-51. https://doi.org/10.1016/j.heares.2009.09.012

Montoro P.R., Villalba-García C., Luna D., Hinojosa J.A. (2017). Common region wins the competition between extrinsic grouping cues: Evidence from a task without explicit attention to grouping. *Psychon Bull Rev.*, 24(6): 1856-1861. https://doi.org/10.3758/s13423-017-1254-3

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

Page S.C., Hulse S.H., Cynx J. (1989). Relative pitch perception in the European starling (Sturnus vulgaris): further evidence for an elusive phenomenon. *J Exp Psychol Anim Behav Process.*, 15(2): 137–146. https://doi.org/10.1037/0097-7403.15.2.137

Palmer S.E., Beck D.M. (2007). The repetition discrimination task: an objective method for studying perceptual grouping. *Percept Psychophys.*, 69(1): 68-78. https://doi.org/10.3758/bf03194454

Patel A.D. (2003). Language, music, syntax and the brain. *Nat Neurosci.*, 6(7): 674-681. https://doi.org/10.1038/nn1082

Patel A.D. (2008). Music, Language, and the Brain. New York, NY: Oxford University Press.

Patel A.D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Front Psychol.*, 2: 142. https://doi.org/10.3389/fpsyg.2011.00142

Ohms V.R., Escudero P., Lammers K., ten Cate C. (2012). Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception. *Anim Cogn.*, 15(2): 155-161. https://doi.org/10.1007/s10071-011-0441-2

Okanoya K., Dooling R.J. (1987). Hearing in passerine and psittacine birds: a comparative study of absolute and masked auditory thresholds. *J Comp Psychol.*, 101(1): 7-15. https://doi.org/10.1037/0735-7036.101.1.7

Rashal E., Yeshurun Y., Kimchi R. (2017). The time course of the competition between grouping organizations. *J Exp Psychol Hum Percept Perform.*, 43(3): 608-618. https://doi.org/10.1037/xhp0000334

Sadakata M., Weidema J.L., Roncaglia-Denissen M.P.M., Honing H. (2020). Parallel pitch processing in speech and melody: A study of the interference of musical melody on lexical pitch perception in speakers of Mandarin. *PLoS ONE*, 15(3): e0229109. https://doi.org/10.1371/journal.pone.0229109

Schmidt F., Schmidt T. (2013). Grouping principles in direct competition. Vision Res., 88: 9-21. https://doi.org/10.1016/j.visres.2013.06.002

Shinn-Cunningham B.G., Lee A.K., Oxenham A.J. (2007). A sound element gets lost in perceptual competition. *Proc Natl Acad Sci USA.*, 104(29): 12223–12227. https://doi.org/10.1073/pnas.0704641104

Standards Secretariat, Acoustical Society of America (1994). ANSI S1.1-1994 (R2004) American National Standard Acoustical Terminology, (12.41) Acoustical Society of America, Melville, NY.

ten Cate C., Honing H. (2022). Precursors of music and language in animals. *PsyArXiv*, Preprint. https://doi.org/10.31234/osf.io/4zxtr

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

van Buuren S., Groothuis-Oudshoorn K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.*, 45(3): 1-67. https://doi.org/10.18637/jss.v045.i03

Villalba-García C., Santaniello G., Luna D., Montoro P.R., Hinojosa J.A. (2018). Temporal brain dynamics of the competition between proximity and shape similarity grouping cues in vision. *Neuropsychologia.*, 121: 88-97. https://doi.org/10.1016/j.neuropsychologia.2018.10.022

Villalba-García C., Jimenez M., Luna D., Hinojosa J.A., Montoro P.R. (2021). Competition between perceptual grouping cues in an indirect objective task. *Q J Exp Psychol (Hove).*, 74(10): 1724-1736. https://doi.org/10.1177/17470218211010486

Walker K.M.M., Schnupp J.W.H., Hart-Schnupp S.M.B., King A.J., Bizley J.K. (2009). Pitch discrimination by ferrets for simple and complex sounds. *J Acoust Soc Am.*, 126(3): 1321-1335. http://dx.doi.org/10.1121/1.3179676

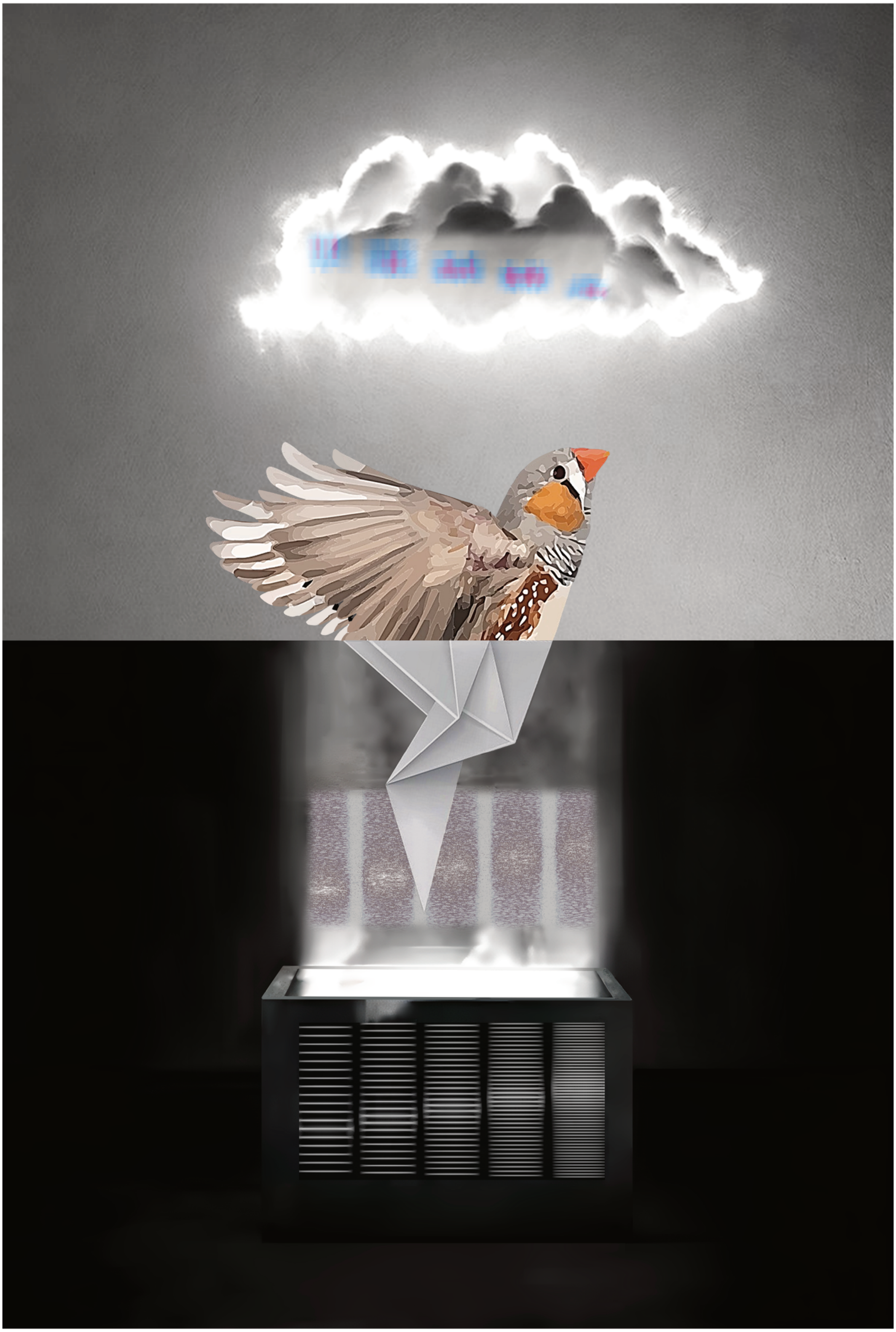
Walker K.M., Gonzalez R., Kang J.Z., McDermott J.H., King A.J. (2019). Acrossed-species differences in pitch perception are consistent with differences in cochlear filtering. *Elife*, 8: e41626. https://doi.org/10.7554/eLife.41626

Weary D.M., Weisman R.G. (1991). Operant discrimination of frequency ratio in the black-capped chickadee (*Parus atricapillus*). *J Comp Psychol.*, 105(3): 253–259. http://dx.doi.org/10.1037/0735-7036.105.3.253

Weisman N., Njegovan M., Ito S. (1994). Frequency ratio discrimination by zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*). *J Comp Psychol.*, 108(4): 363–372. https://doi.org/10.1037/0735-7036.108.4.363

Weisman R.G., Njegovan M.G., Sturdy C.B., Phillmore L., Coyle J., Mewhort D. (1998). Frequency-range discriminations: special and general abilities in zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*). *J Comp Psychol.*, 112(3): 244–258. https://doi.org/10.1037/0735-7036.112.3.244

Weisman R.G., Njegovan M.G., Williams M.T., Cohen J.S., Sturdy C.B. (2004). A behavior analysis of absolute pitch: sex, experience, and species. *Behav Process.*, 66(3): 289–307. https://doi.org/10.1016/j.beproc.2004.03.010



General Summary, Discussion & Conclusion

Chapter 6

Chapter 6

Summary of Findings from Different Chapters

Chapter 2 investigated the role of spectral characteristics and absolute duration of songs in zebra finch song recognition. The results indicated that zebra finches can attend to both spectral features and song duration in discriminating between two songs, and the relative salience of these acoustic parameters varies depending on whether these songs differ in absolute song duration or not in the training phase. Interestingly, when song duration serves as an additional distinguishing cue in song discrimination, spectral features play a less prominent role. In this chapter, I demonstrate that the acoustic features to which zebra finches attend when identifying or discriminating their conspecific songs may not be as fixed as previously assumed. Instead, their focus on these features appears to be influenced by the characteristics of the stimuli. This observed flexibility in auditory processing, coupled with the potential importance of variations in song duration for song perception, calls for systematic research into the acoustic parameters that zebra finches may utilize in the perception of their conspecific songs.

Chapter 3 expanded upon these insights by exploring the preferences of female and male zebra finches for normal and modified songs. By comparing the count of perch visits by which the birds could expose themselves to different stimuli among different sex groups, the study reveals an intriguing pattern: Females displayed a preference gradient, preferring normal and duration-stretched conspecific songs over vocoded versions, but still preferring vocoded conspecific songs over heterospecific great tit songs. In contrast, males only preferred duration-stretched conspecific songs over heterospecific songs and showed limited differentiation in their preference among the stimuli. In addition, female zebra finches display significantly greater activity in this task, underscoring the more pronounced song preferences observed in females compared to their male counterparts. Interestingly, while altering song duration did affect song discrimination in the operant discrimination task (Chapter 2), it did not significantly influence song preference in Chapter 3. As no sex difference was present in the discrimination studies in chapter 2, this chapter thus highlights that the attendance to, and importance of, various acoustic parameters is both task and sex dependent. It underscores the importance of employing multiple approaches and testing paradigms to avoid potential blind spots inherent in a single experimental paradigm.

Chapter 4 delved into the cognitive flexibility of zebra finches in using phonology and the sequence of syllables for auditory discrimination. Birds trained to discriminate between two song-syllable strings that consisted of identical syllables took an equal number of training trials to birds trained on strings containing different syllables. Test trials showed that birds trained on strings with identical syllables had learned sequential cues in addition to spectro-temporal ones. This outcome confirmed the cognitive flexibility of zebra finches, indicating their perceptual attention to the song features that offer the most prominent distinctions between two syllable strings, be it spectro-temporal or sequential. It demonstrates that the role of phonological and sequential features in zebra finches' song recognition depends on the degree of song differences in these features.

Chapter 5 investigated the interplay of pitch and formant contours in melody recognition by zebra finches. It explored how zebra finches perceive and recognize tonal strings with varying pitch and formant attributes, either simultaneously ascending or descending in the same direction or opposite directions over a full string. The results revealed that zebra finches possess remarkable sensitivity to both pitch and formant attributes, enabling them to detect variations in pitch and formant contours across harmonic elements. Interestingly, the impact of pitch and formant is almost identical for both training conditions, which might suggest no or limited interaction between the two attributes when learning to discriminate the different tonal strings. In culmination with the insights from previous chapters, Chapter 5 provides thus further evidence of the zebra finches' ability to attend to various acoustic parameters, not only in response to stimuli consisting of natural song elements but also in the context of artificial tonal stimuli.

Below I synthesize the findings from the preceding four chapters and discuss the details and broader context of auditory cognition. First, I discuss zebra finches' perceptual flexibility, their sensitivity to acoustic features in different contexts, and their ability to focus on global as well as local features. Next, I explore how the nature of stimuli impacts sound discrimination, comparing natural and artificial sounds' effects on zebra finches and starlings. After this, I investigate the influence of the training context on zebra finches' perceptual flexibility, particularly regarding spectral and temporal features in song discrimination. Subsequently, I discuss auditory perception and discrimination variations between zebra finches and starlings, comparing their responses to vocoded stimuli, spectral features, and sensitivity to noise disruptions in harmonic structures, emphasizing potential interspecies differences. I then propose that these findings on the context-dependent perceptual flexibility of zebra finches add to the zebra finch's significance as a model for auditory research. Lastly, I suggest future research directions, focusing on uncovering cognitive mechanisms of perceptual flexibility and advocating for cross-species comparative studies in this field.

Perceptual Flexibility of Zebra Finches

One overarching theme that emerges from the experimental chapters is the remarkable perceptual flexibility of zebra finches. Previous studies have demonstrated that zebra finches can perceive a range of acoustic features (e.g., Uno *et al*., 1997; Lohr & Dooling, 1998; Dooling *et al*., 2002; Lohr *et al*., 2006; Nagel *et al*., 2010; Vernaleo *et al*., 2010; Vernaleo & Dooling, 2011; Vignal & Mathevon, 2011; Lawson *et al*., 2018; Prior *et al*., 2018a; Prior *et al*., 2018b; Geberzahn & Derégnaucourt, 2020; Fishbein *et al*. 2021; Mol *et al*., 2021). However, while these studies revealed that zebra finches are capable of attending to various features, they generally did not investigate to what extent the choice of stimuli influenced the features to which they attended.

My findings suggest the perceptual flexibility of zebra finches in attending to either spectral features or duration of songs (Chapter 2) and to either spectral structure or sequence of syllables (Chapter 4) depending on the training context (i.e., the salient differences in sound features between the training stimuli). This observation aligns with the results of a study by Burgering *et al*. (2019), which demonstrated that zebra finches can use either pitch or spectral envelope as cues in discriminating vowel-like sounds, depending on the salient differences between the stimuli. This flexibility is also supported by behavioural evidence, such as a Go-Nogo operant task, where zebra finches were observed attending to specific features when differentiating between harmonic sounds (Uno *et al*., 1997). In Uno *et al*.' study (1997), when the only difference between two harmonic sounds was the second harmonic, zebra finches focused on this feature. Moreover, some zebra finches tended to rely on the fundamental frequency as a discriminative cue when the number of harmonics was limited. These results underline the zebra finches' ability to adapt perceptual strategies based on the contrasts between the stimuli they need to differentiate. My findings, along with those from other studies mentioned above, contribute to the understanding of perceptual flexibility in zebra finches' auditory processing, as different features appeared to be relevant in various contexts.

Moreover, the findings reveal that zebra finches' attention is not limited to spectral features but also extends to temporal features. These birds demonstrate a sensitivity to both local features, such as the spectro-temporal fine structure of syllables, and global features, including the sequential order of syllables within a song. The perceptual flexibility for global features is particularly evident in Chapter 4, where zebra finches showcased their ability to adapt and discriminate songs based on phonological and sequential aspects. This demonstrates their capacity to process and utilize the overall structure of songs rather than merely focusing on individual elements.

The observed perceptual flexibility enables birds to adapt their perceptual tuning to the most relevant acoustic dimensions for distinguishing between songs of different individuals or other biologically relevant sounds. It's likely that the zebra finch is not the unique species exhibiting this cognitive ability, though further crossed-species research is needed to confirm this. Also, many neurobiological studies on auditory perception in songbirds often treat perceptual sensitivities as static (e.g., Theunissen & Doupe, 1998; Sen *et al*., 2001; Meliza & Margoliash, 2012; Cazala *et al*., 2019; Inda *et al*., 2021). Our study suggests that researchers should consider the possibility that these sensitivities may vary based on the context (such as discriminating songs of different colony members or expressing a social or sexual preference) or the nature of the variation between the available stimuli. Investigating the regulation of this flexibility and identifying the brain areas crucial for specific contexts or tasks could be an intriguing topic for future neurobiological research.

Nature of Stimuli & Impact on Discrimination

Songbirds were assumed to have a strong bias to rely on the absolute pitch for the recognition of tone sequences - a pitch-shifted melody seems to be perceived as an altogether different melody (Hulse *et al*., 1984a). While it's commonly believed that songbirds primarily recognize tonal sounds based on absolute pitch rather than relative pitch (e.g., Weisman *et al*., 2004; Hoeschele *et al*., 2012b, 2013, 2014), it is interesting to note that starlings (Cynx *et al*., 1986; Braaten *et al*., 1990) and other songbirds (Hulse & Cynx, 1985; Hulse *et al*., 1992) did exhibit some degree of perceptual sensitivity for relative pitch. When these birds were tested with artificial stimuli, they could only transfer their discrimination abilities between different tone sequences to novel patterns within the range of frequencies used to construct the patterns they initially learned to discriminate (Page *et al.*, 1989). However, starlings can maintain the discrimination between two songs shifted in frequency (Bregman *et al.*, 2012). Some other songbirds, like the white-throated sparrow (Hurly *et al.*, 1990) and black-capped chickadee (Weisman & Ratcliffe, 1989; Hoeschele *et al.*, 2012a), were also found to be sensitive to the relative pitch features in conspecific song. It is still unclear what causes this difference in responses between shifts in artificial tone sequences and in conspecific songs. A possible explanation is that the nature of the training stimuli impacts the birds' attention to specific acoustic features during discrimination tasks and the subsequent generalization of their learning to novel stimuli. This stimulus-dependent sensitivity in zebra finches and other songbirds implies that perceptual flexibility in sound discrimination may also be expected to exist across songbird species.

Equally noteworthy is the influence of the stimulus nature on the perceptual salience of different acoustic features that zebra finches may prioritize. Given the paramount role of song learning in zebra finches, encompassing abilities such as individual identification, mate attraction, rival repulsion, and the discrimination between conspecific and heterospecific songs, one would expect these birds to focus on a spectrum of acoustic parameters, including temporal aspects (e.g., duration), source-related characteristics (F0-based pitch), and filter-related attributes (such as formants) inherent in their conspecific songs. While one of our chapters delved into artificial stimuli, specifically tonal strings featuring pitch and formant contours (Chapter 5), others focused their investigations on conspecific songs (Chapter 2 and Chapter 4).

Patel (2017) suggested a hierarchy of cues for tone sequence recognition in songbirds, emphasizing spectral shape as a primary feature and pitch saliency as a significant secondary feature distinguishing tonal from noisy sounds. According to this, pitch perception might be a vital cue in distinguishing call types based on pitch saliency, especially if harmonic tones are less acoustically complex than real musical instrument sounds or if the spectral structure of complex harmonic tones remains consistent across tones (e.g., Bregman *et al.*, 2012). Surprisingly, while starlings are capable of relational processing (which entails the cognitive analysis of relationships, patterns, or dependencies among various features of a specific acoustic attribute within a given auditory context, emphasizing the interaction and contribution of its components to the overall perception of sound) for acoustic parameters like loudness (Bernard & Hulse, 1992) and rhythm (Hulse *et al.*, 1984b) in tone sequences, they do not exhibit the same ability for relational pitch processing. This implies that relative pitch may carry less weight than other parameters such as loudness and rhythm if a strict “hierarchy of cues” (as proposed by Patel, 2017) is employed by songbirds for sound recognition. However, it's worth noting that Patel also acknowledged that this hierarchy is not absolute and may vary depending on the complexity of the acoustic stimuli. In fact, my findings align with this notion that these distinct strategies might result from specific biases — in this context, the term “biases” is more precise than “hierarchy” — reflecting the weight or priority that certain birds give to certain features of an acoustic signals.

Perceptual flexibility emphasizes the adaptability of sensitivities to various acoustic features or perceptual attributes. The interplay between the nature of the training stimuli and the birds’ sensitivity to specific acoustic features during discrimination tasks can lead to a more nuanced understanding of sound perception in songbirds. This suggests that while certain acoustic features or perceptual attributes, such as pitch, may play a crucial role in song discrimination, the flexibility and adaptability of these birds’ perceptual systems are influenced by the context and stimuli they encounter during training and testing. Therefore, it’s essential to consider the context-dependent nature of their cognitive processes in auditory discrimination research.

Impact of Training Context

The findings of this research highlight the value of the training context in modulating the perceptual flexibility of zebra finches. As elucidated in Chapter 2 and Chapter 4, the training context plays a pivotal role in determining the extent to which zebra finches rely on spectral and temporal features for discriminating songs. This highlights the need to account for training conditions when interpreting their auditory discrimination capabilities. Furthermore, Chapter 5 investigates whether the relative influence of pitch and formant in discriminating tone sequences differs depending on whether they change in the same or opposite directions across the sequences. Unlike the findings from Chapter 2 and Chapter 4, where conspecific sounds were used, Chapter 5 employed two distinct artificial stimulus sets. Remarkably, the result in Chapter 5 does not significantly vary with changes in the training context regardless of whether formant and pitch changes were congruent or opposite suggesting that both attributes are important but don’t seem to interact with each other. Hence, the training context is clearly influential, but it’s not the determining factor.

Sexual Bias on Song Preferences, as well as Song Discrimination?

The zebra finches display perceptual flexibility in song discrimination, yet they still have preferred song features they focus on. The research of Chapter 3 added another dimension to song discrimination by investigating potential sexual biases in song preferences. Chapter 3 brought to light that both female and male zebra finches exhibited preferences for duration-stretched conspecific songs over heterospecific songs, with females showing a gradient of preferences on the four song types (normal, duration-stretched, vocoded conspecific songs, and heterospecific songs) and displaying more pronounced preferences than males. The preference for conspecific songs over heterospecific ones in zebra finches is not surprising, as it has been supported by both behavioural and neural experimental evidence (Stenstrom *et al.*, 2022).

These findings prompt consideration of the distinct questions addressed by the two experimental paradigms and their respective limitations. Discrimination tasks primarily assess a songbird's ability to differentiate and recognize specific auditory cues, while preference tests focus on the bird's behavioural responses and subjective preferences for auditory stimuli. Comparing the results of both paradigms using the same stimulus set contributes to a more comprehensive understanding of songbirds' auditory perception of specific stimuli. However, it is worth noting that our results did not reveal any sexual differences in discrimination tasks, as evidenced in Chapter 2, Chapter 4, and Chapter 5. This suggests that the sexual biases observed in preference tests might be absent or comparatively subtle when birds are subjected to training and testing within the operant discrimination paradigm. This disparity implies that the specific experimental paradigm employed could influence the extent to which sexual differences are discernible. Furthermore, the variation in results between preference tests and discrimination tasks is not due to preference tests being more sensitive to sexual biases, but rather because these two experimental paradigms measure different cognitive traits. Moreover, the synthesis of these findings underscores the divergence in cognitive processes underpinning song preferences and auditory discrimination. This highlights the importance of tailoring the choice of a behavioural paradigm based on its specific focus on distinct cognitive functions.

Do Zebra Finches Attend to the Same Features as Starlings Do in Sound Discrimination?

The results from Chapter 2 of this thesis and those presented by Nagel *et al.* (2010) indicate that zebra finches are more sensitive to pitch changes (in the sense that pitch-shifted versions are considered different) of songs than starlings are, which can still show discrimination of songs with pitch shifts up to ±40% (Bregman *et al.*, 2012). Moreover, starlings appear to exhibit a greater inclination to focus on specific spectral features compared to zebra finches, which seem to attend to a wider variety of features. A comparable pattern emerges concerning specific features in starlings' attention to temporal characteristics. Starlings were more sensitive to changes in tone length (i.e., sound duration) than changes in the inter-onset interval (Hulse *et al.*, 1984b), while zebra finches appeared to assign equal importance to both temporal aspects (ten Cate *et al.*, 2016).

The findings in this thesis with regard to how zebra finches process vocoded stimuli, including conspecific songs and artificial tones, differ from earlier studies with starlings. Bregman *et al.* (2016) demonstrated that when presenting starlings with vocoded versions of sequences of tones (varying pitch and timbre), they maintained discrimination between these sequences. However, pitch-shifted versions did not, indicating that spectral envelope rather than pitch was used for discrimination by starlings. While Patel (2017) suggested that this might be a common strategy among birds, the reduced discrimination of noise-vocoded songs in zebra finches in my studies (Chapter 2 and Chapter 4) was surprisingly not in line with the studies on starlings' discrimination of artificial tone sequences. Even when tested with artificial tone stimuli (Chapter 5), zebra finches from both training conditions responded the least to the vocoded version among the five modified versions. The finding that zebra finches hardly discriminate noise-vocoded stimuli, whether they are conspecific songs or artificial sounds, suggests a distinction between songbird species in perceptual strategies regarding the utilization of spectral envelope in sound discrimination.

However, Burgering *et al.*'s study (2019) found support for the idea that zebra finches use spectral envelope for discriminating vowel-like sounds. The effectiveness of the stimulation of zebra finches' calls does not appear to rely on their fine acoustic structure, as these calls still evoke a strong behavioural response even when their acoustic features have been significantly modified (Vignal & Mathevon, 2011). This might suggest that spectral envelope is important for discriminating among calls or other shorter sounds, whereas discriminating songs relies more on attending to other spectral features, including pitch and harmonics, which has been demonstrated in various studies (e.g., Okanoya & Dooling, 1990; Uno *et al.*, 1997; Dooling & Lohr, 2006; Lohr *et al.*, 2006; Vignal & Mathevon, 2011; Prior *et al.*, 2018a; Prior *et al.*, 2018b).

In my experiments, the vocoded version proved to be the most easily detected as differing from the training sounds when compared to other stimulus modifications. This may be due to the fact that noise-vocoding disrupts not only pitch and the harmonic attributes of songs and sequence of tones but also introduces a but also introduces a noise-like, non-harmonic spectrum into the stimulus. These alterations, apparently, make the vocoded versions perceptually distinct from the training stimuli. Nonetheless, the zebra finches might still be capable of using spectral envelope as an acoustic cue in sound discrimination (supported by discrimination above chance in some cases of Chapter 4 and Chapter 5). If this is indeed the case, it suggests that zebra finches are more sensitive to the disruption of pitch and harmonic structures than starlings are. However, so far, there have been no studies examining how starlings respond to vocoded versions of natural songs (the stimuli Bregman *et al.* (2016) used were artificially synthesized sounds), leaving it to be explored whether such a stimulus would yield similar outcomes compared to testing with artificial sounds.

Lastly, more research is needed to test these claims more generally across avian species. Future research should explore cognitive flexibility in a broader range of avian species, including both songbirds and non-songbirds, using identical experimental paradigms and stimuli. This will aid in identifying the neural, ecological, and evolutionary factors associated with the cognitive flexibility.

An Ideal Avian Model for Auditory Research

This thesis has contributed to establishing the zebra finch as an avian model for research on auditory perception and processing, further contributing to the value attributed to these birds in the field of comparative cognition research. Their extensively studied neurobiological and behavioural characteristics, along with the observed perceptual flexibility in auditory discrimination, emphasize their significance as model species for understanding auditory processing in comparative studies. By using this avian model, I presented insights into the auditory abilities of zebra finches, showcasing their perceptual flexibility across spectral, temporal, and sequential attributes. It is evident that the birds' auditory sensitivity extends beyond the realm of local features, such as the fine-grained spectro-temporal structure of individual syllables, to encompass the ability to attend to global features, including the sequential order of syllables. Most notably, this sensitivity for global features is demonstrated when the birds successfully discriminate between two artificial tonal strings based on pitch contour and formant contour (as unveiled in Chapter 4). The notion of context-dependent cognitive flexibility and these birds' sensitivity beyond local features together suggest that there may still be other unexplored cognitive traits within this avian model.

Open Questions of Songbird Auditory Perception & Avenues for Future Research

While this thesis has yielded valuable insights into zebra finches' auditory perception, there remain open questions that elude our understanding. For example, further exploration is still needed to determine whether and how the interplay among various acoustic attributes impacts cognitive flexibility in avian auditory perception, and how these impacts depend on stimulus nature and ecological context. Additionally, there is still a lack of studies investigating which other species exhibit similar or contrasting levels of cognitive flexibility in auditory perception. As I conclude this study, it is imperative to delineate promising avenues for future research, capitalizing on the knowledge gleaned from our findings.

Future research endeavours should set their sights on unravelling the neural mechanisms that underpin the observed perceptual flexibility in zebra finches. Furthermore, delving into cross-species comparisons holds the promise of providing a broader perspective on auditory cognition (e.g., the impact of vocal learning on perceptual flexibility). Such comparative studies, encompassing both songbirds and non-songbirds, can offer valuable insights into the evolutionary aspects of auditory perception.

In addition, future research directions should take into account the ecological relevance of zebra finch auditory discrimination in their natural habitats. Investigating how their perceptual flexibility translates to real-world scenarios and influences social interactions can provide a holistic understanding of their song recognition abilities. This ecological perspective will bridge the gap between laboratory findings and the practical application of these perceptual abilities in the wild, offering a comprehensive view of zebra finches' auditory cognition. In conclusion, our research highlights the extraordinary perceptual flexibility of zebra finches, revealing their capacity to integrate both spectral and temporal attributes. This adaptability results from a dynamic interplay among factors such as training conditions, the nature of stimuli, and ecological relevance. The birds' adaptability for auditory perception, as a sort of multifaceted cognitive procedure influenced by various factors, does not stand alone but rather arises from the interplay among various acoustic attributes. Understanding the complex cognitive processes governing avian perceptual flexibility not only enriches our understanding of auditory perception but also provides valuable insights into cognitive abilities across diverse species.

References

Bernard D.J., Hulse S.H. (1992). Transfer of serial stimulus relations by European starlings (Sturnus vulgaris): Loudness. *J Exp Psychol Anim Behav Process.*, 18(4): 323–334. https://doi.org/10.1037/0097-7403.18.4.323

Braaten R.F., Hulse S.H., Page S.C. (1990). Discrimination and classification of rising and nonrising pitch patterns by the European starling (Sturnus vulgaris). *Anim learn Behav.*, 18(4), 352–364. https://doi.org/10.3758/BF03205317

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. *Proc Natl Acad Sci.*, 113(6): 1666–1671. https://doi.org/10.1073/pnas.15153801

Burgering M.A., Vroomen J., ten Cate C. (2019). Zebra finches (Taeniopygia guttata) can categorize vowel-like sounds both on the fundamental frequency ("Pitch") and spectral envelope. *J Comp Psychol.*, 133(1): 106-117. https://doi.org/10.1037/com0000143

Cazala A., Giret N., Edeline J.M., Del Negro C. (2019). Neuronal Encoding in a High-Level Auditory Area: From Sequential Order of Elements to Grammatical Structure. *J Neurophysiol.*, 39(31): 6150–6161. https://doi.org/10.1523/JNEUROSCI.2767-18.2019

Cynx J., Hulse S.H., Polyzois S. (1986). A psychophysical measure of pitch discrimination loss resulting from a frequency range constraint in European starlings (Sturnus vulgaris). *J Exp Psychol Anim Behav Process.*, 12(4): 394–402. https://doi.org/10.1037/0097-7403.12.4.394

Dooling R.J., Leek M.R., Gleich O., Dent M.L. (2002). Auditory temporal resolution in birds: discrimination of harmonic complexes. *J Acoust Soc Am.*, 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. *Ornithol Sci.*, 5(1): 15-22. https://doi.org/10.2326/osj.5.15

Fishbein A.R., Prior N.H., Brown J.A., Ball G.F., Dooling R.J. (2021). Discrimination of natural acoustic variation in vocal signals. *Sci Rep.*, 11(1): 916. https://doi.org/10.1038/s41598-020-79641-z

Geberzahn N., Derégnaucourt S. (2020). Individual vocal recognition in zebra finches relies on song syllable structure rather than song syllable order. *J Exp Biol.*, 223(Pt 9): jeb220087. https://doi.org/10.1242/jeb.220087

Hoeschele M., Cook R.G., Guillette L.M., Brooks D.I., Sturdy C.B. (2012a). Black-capped chickadee (Poecile atricapillus) and human (Homo sapiens) chord discrimination. *J Comp Psychol.*, 126(1): 57–67. https://doi.org/10.1037/a0024627

Hoeschele M., Guillette L.M., Sturdy C.B. (2012b). Biological relevance of acoustic signal affects discrimination performance in a songbird. Animal cognition, 15(4): 677–688. https://doi.org/10.1007/s10071-012-0496-8

Hoeschele M., Weisman R.G., Guillette L.M., Hahn A.H., Sturdy C.B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Anim Cogn.*, 16(4): 599–609. https://doi.org/10.1007/s10071-013-0597-z

Hoeschele M., Cook R.G., Guillette L.M., Hahn A.H., Sturdy C.B. (2014). Timbre influences chord discrimination in black-capped chickadees (Poecile atricapillus) but not humans (Homo sapiens). *J Comp Psychol.*, 128(4): 387–401. https://doi.org/10.1037/a0037159

Hulse S.H., Cynx J., Humpal J. (1984a). Absolute and relative pitch discrimination in serial pitch perception by birds. *J Exp Psychol Gen.*, 113(1): 38–54. https://doi.org/10.1037/0096-3445.113.1.38

Hulse S.H., Humpal J., Cynx J. (1984b). Discrimination and generalization of rhythmic and arrhythmic sound patterns by European starlings (Sturnus vulgaris). *Music Percept.*, 1(4): 442–464. https://doi.org/10.2307/40285272

Hulse S.H., Cynx J. (1985). Relative pitch perception is constrained by absolute pitch in songbirds (Mimus, Molothrus, and Sturnus). *J Comp Psychol.*, 99(2): 176–196. https://doi.org/10.1037/0735-7036.99.2.176

Hulse S.H., Takeuchi A.H., Braaten R.F. (1992). Perceptual invariances in the comparative psychology of music. *Music Percept.*, 10(2): 151–184. https://doi.org/10.2307/40285605

Hurly T.A., Ratcliffe L., Weisman R. (1990). Relative pitch recognition in white-throated sparrows, Zonotrichia albicollis.. *Anim Behav.*, 40(1): 176–181. https://doi.org/10.1016/S0003-3472(05)80677-3

Inda M., Hotta K., Oka K. (2021). High responsiveness of auditory neurons to specific combination of acoustic features in female songbirds. *Eur J Neurosci.*, 53(5): 1412–1427. https://doi.org/10.1111/ejn.15047

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. *Anim Cogn.*, 21(4): 467–480. https://doi.org/10.1007/s10071-018-1182-2

Lohr B., Dooling R.J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J Comp Psychol.*, 112(1): 36–47. https://doi.org/10.1037/0735-7036.112.1.36

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

Meliza C.D., Margoliash D. (2012). Emergence of selectivity and tolerance in the avian auditory cortex. *J Neurophysiol.*, 32(43): 15158–15168. https://doi.org/10.1523/JNEUROSCI.0845-12.2012

Mol C., Bolhuis J.J., Moorman S. (2021) Vocal learning in songbirds: the role of syllable order in song recognition. *Philos Trans RSoc Lond B Biol Sci.*, 376(1836): 20200248. https://doi.org/10.1098/rstb.2020.0248

Nagel K.I.K., McLendon H.M., Doupe A.J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J Neurophysiol.*, 104(3): 1426–37. https://doi.org/10.1152/jn.00028.2010

Okanoya K., Dooling R.J. (1990). Temporal integration in zebra finches (Poephila guttata). *J Acoust Soc Am*., 87(6): 2782–2784. https://doi.org/10.1121/1.399069

Page S.C., Hulse S.H., Cynx J. (1989). Relative pitch perception in the European starling (Sturnus vulgaris): Further evidence for an elusive phenomenon. *J Exp Psychol Anim Behav Process*., 15(2): 137–146. https://doi.org/10.1037/0097-7403.15.2.137

Patel A.D. (2017). Why Doesn't a Songbird (the European Starling) Use Pitch to Recognize Tone Sequences? The Informational Independence Hypothesis. *Comp Cogn Behav Rev*., 12(1): 19–32. https://doi.org/10.3819/CCBR.2017.120003

Prior N.H., Smith E., Ball G.F., Dooling R.J. (2018a). Temporal fine structure and zebra finch vocalizations. *J Acoust Soc Am*., 143(3): 1897-1897. https://doi.org/10.1121/1.5036170

Prior N.H., Smith E., Lawson S., Ball G.F., Dooling R.J. (2018b). Acoustic fine structure may encode biologically relevant information for zebra finches. *Sci Rep*., 8(1): 6212. https://doi.org/10.1038/s41598-018-24307-0

Sen K., Theunissen F.E., Doupe A.J. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol*., 86(3): 1445–1458. https://doi.org/10.1152/jn.2001.86.3.1445

Stenstrom K., Voss H.U., Tokarev K., Phan M.L., Hauber M.E. (2022). The Direction of response selectivity between conspecific and heterospecific auditory stimuli varies with response metric. *Behav Brain Res*., 416: 113534. https://doi.org/10.1016/j.bbr.2021.113534

ten Cate C., Spierings M., Hubert J., Honing H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front Psychol*., 7: 730. https://doi.org/10.3389/fpsyg.2016.00730

Theunissen F.E., Doupe A.J. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *J. Neurosci*., 18(10): 3786–3802. https://doi.org/10.1523/JNEUROSCI.18-10-03786.1998

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

Vernaleo B.A., Dooling R.J., Leek M.R. (2010). Relative salience of acoustic features in zebra finch song. *J Acoust Soc Am*., 127(3): 1863. https://doi.org/10.1121/1.3384460

Vernaleo B.A., Dooling R.J. (2011). Relative salience of envelope and fine structure cues in zebra finch song. *J Acoust Soc Am*., 129(5): 3373–3383. https://doi.org/10.1121/1.3560121

Vignal C., Mathevon N. (2011). Effect of Acoustic Cue Modifications on Evoked Vocal Response to Calls in Zebra Finches (Taeniopygia guttata). *J Comp Psychol*., 125(2): 150–161. https://doi.org/10.1037/a0020865

Weisman R.G., Njegovan M.G., Williams M.T., Cohen J.S., Sturdy C.B. (2004). A behavior analysis of absolute pitch: Sex, experience, and species. *Behav Processes*., 66(3): 289–307. https://doi.org/10.1016/j.beproc.2004.03.010

Weisman R., Ratcliffe L. (1989). Absolute and relative pitch processing in black-capped chickadees, Parus atricapillus. *Anim Behav*., 38(4): 685–692. https://doi.org/10.1016/S0003-3472(89)80014-4

Nederlandse samenvatting

De flexibele luisteraar: onderzoek naar de gevoeligheid van zebravinken voor spectrale en temporele geluidskenmerken

Nederlandse samenvatting

Vocale communicatie door taal en de perceptie van muziek zijn twee van de meest geavanceerde cognitieve vaardigheden bij mensen, waarbij een universele aanleg voor beide wordt waargenomen. Zowel de muziekvaardigheid als de taalvaardigheid van de mens omvat verschillende perceptuele en cognitieve capaciteiten. Vergelijkend onderzoek naar de auditieve perceptie van taal en muziek omvat het begrijpen van mogelijke gedeelde cognitieve eigenschappen tussen mensen en niet-menselijke dieren, zoals het herkennen van relatieve toonhoogte, het coderen van tooninformatie, het waarnemen van ritmes, metrische codering van ritme en het verwerken van opeenvolgende geluidseenheden.

Onderzoek tussen soorten heeft inzichten opgeleverd in de al lang bestaande vraag of muzikaliteit een inherent onderdeel is van de menselijke natuur of een culturele uitvinding (sommige onderzoekers gaan er bijvoorbeeld van uit dat het puur een culturele uitvinding is die voortbouwt op hersencircuits die om andere redenen zijn geëvolueerd) . Hebben mensen unieke hersenkenmerken voor het verwerken van muziek, of is ons muzikale vermogen volledig afhankelijk van hersencircuits die voor andere doeleinden zijn geëvolueerd? Volgens de hypothese van vocaal leren en ritmische synchronisatie kan ons vermogen om nauwkeurig op muziek te bewegen bijvoorbeeld voortkomen uit neurale circuits die verband houden met complex vocaal leren. Papegaaien, zoals de beroemde ‘Snowball’, vertonen net als mensen ritmische synchronisatie, gekoppeld aan veranderingen in de basale ganglia. Sommige onderzoeken trokken echter het veronderstelde verband tussen het vocale leren en de beatperceptie bij vogelsoorten in twijfel, en stelden een graduele schaal voor de beatperceptie voor. Bovendien hebben recente onderzoeken de aannames over de manier waarop zangvogels tonen herkennen in twijfel getrokken. Die wijzen er op dat vogel zich eerder baseren op de spectrale envelop dan op de absolute toonhoogte.

Zangvogels leren, net als mensen, complexe vocalisaties en delen opvallende overeenkomsten met mensen in hun perceptie van verschillende akoestische kenmerken, ondanks dat ze minder verwant zijn aan mensen dan andere zoogdieren, waardoor ze een uitstekend model zijn voor het bestuderen van auditieve perceptie. Zebravinken leveren als modelsoort inzichten in vocaal leren, ritmedetectie en complexe auditieve stimulusverwerking, en dragen daarmee bij aan ons diepere begrip van de biologische grondslagen van menselijke spraakverwerking en muzikaliteit. Zebravinken richten zich meer op lokale kenmerken dan op algemene regelmaat, en vertonen een variërende gevoeligheid voor tempoveranderingen in ritmische stimuli. Bovendien suggereert de ‘local feature bias’ hypothese een voorkeur voor eenvoudigere stimuli, wat wijst op een verminderde gevoeligheid voor complexere structuren. De soms tegenstrijdige bevindingen op het gebied van de auditieve perceptie van vogels, met name wat betreft de rol van spectrale en temporele kenmerken bij auditieve discriminatie, evenals het onderzoek naar de vraag of vogels bij auditieve discriminatie prioriteit geven aan lokale kenmerken boven globale kenmerken, en de tegenstrijdige bevindingen over de gevoeligheid voor temporele kenmerken en elementreeksen, kunnen verklaard worden door cognitieve flexibiliteit. Zoals in eerdere onderzoeken is waargenomen, vertonen zebravinken deze flexibiliteit door bij het onderscheiden van geluiden gebruik te maken van de toonhoogte of van de spectrale envelop, afhankelijk van de taak die moet worden uitgevoerd. Hoewel verschillende onderzoeken suggereren dat zangvogels, waaronder zebravinken, verschillende spectrale kenmerken kunnen gebruiken voor het identificeren/differentiëren van verschillende stimuli bij verschillende taken, ontbreekt een systematisch onderzoek naar hoe zebravinken deze kenmerken gebruiken om toonreeksen of gemanipuleerde vocalisaties te discrimineren. Vragen over hoe vogels strategieën aanpassen op basis van trainingsstimuli of vooroordelen in de aandacht voor kenmerken zoals spectro-temporele detailstructuur zijn nog onbeantwoord. In grote lijnen is mijn onderzoek gericht op het opvullen van lacunes in het begrijpen van de auditieve cognitie van zangvogels, door inzicht te verschaffen in de flexibiliteit in perceptuele strategieën in verschillende akoestische contexten, en daarnaast te onderzoeken hoe de complexiteit van stimuli de discriminatie van zebravinken beïnvloedt in het samenspel van spectrale en temporele perceptie.

In **hoofdstuk 2** heb ik onderzocht hoe zebravinken onderscheid maken tussen twee liedjes met dezelfde of verschillende duur. De focus lag op het begrijpen van de relatieve bijdrage van spectrale kenmerken en de duur van liedjes aan de herkenning van liedjes. In een operant conditioneringsparadigma dat bekend staat als de Go-left/Go-right-taak, werden vogels getraind met een voedselbeloning om op de ene toets te pikken als reactie op het ene liedje en op een andere toets te pikken als reactie op een ander liedje. Fouten resulteerden in korte perioden van duisternis. Het experiment onderzoekt hoe verschillen in de duur van de zang de discriminatie van zebravinkliedjes beïnvloedt, inclusief aangepaste versies met variaties in spectrale (toonhoogte of frequentiespectrum) en temporele kenmerken (duur/tempo). Ik demonstreer dat zebravinken rekening kunnen houden met zowel spectrale kenmerken als de duur van de zang tijdens discriminatie, en het belang van deze parameters hangt af van het feit of de zang tijdens de trainingsfase verschilt in absolute duur. Met name wanneer de duur van een liedje als extra onderscheidend kenmerk dient, spelen spectrale kenmerken een minder prominente rol. Mijn bevinding betwist de opvatting dat de aandacht van zebravinken voor specifieke akoestische kenmerken vaststaat bij het identificeren of onderscheiden van hun soortgenoten. In plaats daarvan lijkt hun focus te worden beïnvloed door stimuluskenmerken, wat wijst op flexibiliteit in auditieve verwerking.

In **hoofdstuk 3** onderzocht ik de zangvoorkeuren van zowel mannelijke als vrouwelijke zebravinken, voortbouwend op de resultaten uit hoofdstuk 2. De centrale vraag was of het lage responspercentage op bepaalde nieuwe stimuli bij discriminatie te wijten was aan het feit dat ze als te verschillend van trainingsliedjes werden ervaren of omdat ze een zeer lage aantrekkelijkheid hadden. Om dit te onderzoeken hebben vogels met een ‘carrousel test’ de mogelijkheid gekregen om op verschillende zitstokjes te gaan zitten, waaraan vier soorten liedjes waren gekoppeld: normale liedjes, liedjes met verlengde duur, gevocodeerde liedjes, en heterospecifieke liedjes. De resultaten van deze voorkeurstest brengen intrigerende patronen aan het licht, waarbij vrouwtjes een voorkeursgradiënt vertonen met een voorkeur voor normale en verlengde duur liedjes boven vocoded-versies, en een voorkeur voor vocoded versies boven heterospecifieke koolmees liedjes. Mannetjes hadden een beperkte differentiatie in hun voorkeuren. Ondanks dat de verandering in de duur van het lied de discriminatie in Hoofdstuk 2 beïnvloedde, had dit geen significante invloed op het liedvoorkeur in Hoofdstuk 3. Hoewel er geen sekseverschil was in het discriminatieonderzoek, vertoonden vrouwen sterkere reacties vergeleken met mannen in het voorkeursonderzoek. Dit geeft aan dat de aandacht van de vogels voor verschillende zangkenmerken zowel taak- als geslachtsafhankelijk is. Daarom is het aanbevolen om verschillende methoden en testbenaderingen te gebruiken voor een uitgebreider begrip.

In **Hoofdstuk 4** heb ik onderzocht of zebravinken flexibele aandacht kunnen besteden aan sequentiële en spectro-temporele kenmerken, afhankelijk van de saillante verschillen tussen liedjes tijdens het leren van discriminatie. Met behulp van de Go-left/Go-right-taak heb ik de gevoeligheid van de vogels voor spectro-temporele kenmerken en elementvolgorde voor het onderscheidende van zang vergeleken. Het onderzoek omvatte het trainen van vogels om onderscheid te maken tussen twee reeksen zangelementen van liedjes met identieke of verschillende elementen. Uit testproeven bleek dat vogels die getraind waren op kunstmatige liedjes met identieke elementen zowel de elementvolgorde als spectro-temporele kenmerken leerden, terwijl vogels die getraind waren op stimuli met verschillende elementen voornamelijk spectro-temporele signalen gebruikten om liedjes te onderscheiden. Dit demonstreert de cognitieve flexibiliteit van zebravinken, wat hun vermogen aangeeft om zich te concentreren op de zangkenmerken die de meest opvallende verschillen bieden tussen twee liedjes, zowel spectro-temporeel als sequentieel. Vergelijkbaar met de studie in Hoofdstuk 2, waar vogels aandacht besteden aan de duur van de zang naast spectrale kenmerken wanneer de liedjes in duur verschilden, bevestigen de resultaten van Hoofdstuk 4 opnieuw dat de zangherkenning van zebravinken afhangt van de aard van de kenmerken waarin liedjes van elkaar verschillen.

In **hoofdstuk 5** heb ik de melodieherkenning van deze vogels onderzocht, waarbij ik heb onderzocht of zebravinken prioriteit geven aan toonhoogte of het harmonische spectrum wanneer beide variëren over een reeks tonen, en welke van deze kenmerken voor hen belangrijker zijn. De studie draagt bij aan ons begrip van hoe zangvogels toonhoogte- en formantpatronen herkennen, die essentieel zijn voor de menselijke spraak- en muziekperceptie. Om dit te onderzoeken gebruikte ik stimuli die bestonden uit geluidssequenties bestaande uit vijf kunstmatige elementen, gescheiden door korte pauzes. Deze kunstmatige elementen hadden toonhoogte- en formantcontouren die ofwel in dezelfde richting over een volledige reeks omhoog en omlaag gingen, of in tegengestelde richtingen over een volledige reeks. Het Go-left/Go-right-paradigma werd in dit onderzoek opnieuw gebruikt om de betekenis van deze spectrale kenmerken te onderzoeken bij het herkennen van kunstmatige toonsequenties en hoe hun combinatie het leren en discriminatie beïnvloedde. De resultaten laten zien dat zebravinken variaties in zowel toonhoogte als formant over harmonische elementen kunnen detecteren, wat wijst op de opmerkelijke gevoeligheid van deze vogels voor zowel toonhoogte als formant. Interessant is dat de impact van toonhoogte en formant vrijwel gelijk was in beide trainingsomstandigheden, wat duidt op een beperkte interactie tussen de twee eigenschappen tijdens het leren van discriminatie.

Zebravinken komen naar voren als een uistekend vogelmodel voor auditief onderzoek, met veel perceptuele flexibiliteit. Mijn onderzoek benadrukt hun vermogen om aandacht te besteden aan verschillende akoestische kenmerken, met een focus op spectrale of temporale aspecten, afhankelijk van de trainingscontext. Dit aanpassingsvermogen komt overeen met breder onderzoek dat de genuanceerde discriminatiemogelijkheden van zebravinken liet zien. De resultaten van mijn studies naar zebravinken onthullen soortspecifieke verschillen in perceptuele strategieën in vergelijking met studies aan spreeuwen. Terwijl zebravinken letten op specifieke spectrale kenmerken, letten spreeuwen meer op de spectrale envelop.

Nederlandse samenvatting

De context waarin zebravinken worden getraind, speelt een cruciale rol bij hun perceptuele flexibiliteit. Dit inzicht benadrukt de noodzaak om trainingsomstandigheden in overweging te nemen bij het interpreteren van auditieve discriminatiecapaciteiten. Bovendien heeft de aard van de stimuli een aanzienlijke invloed op de discriminatie van zangvogels. Mijn onderzoek betrof niet alleen zang, maar ook kunstmatige stimuli, waarbij ik de wisselwerking tussen de aard van de stimuli en perceptuele salientie laat zien.

Hoewel mijn proefschrift waardevolle inzichten biedt in de auditieve perceptie van zangvogels, laat het vragen open. Verder onderzoek naar de wisselwerking tussen verschillende akoestische kenmerken en de impact ervan op de cognitieve flexibiliteit in de auditieve perceptie van vogels is nodig. Toekomstige studies zouden zich verder moeten uitstrekken dan zangvogels (bijvoorbeeld studies over niet-zangvogels) om overeenkomsten of verschillen in cognitieve flexibiliteit tussen soorten te identificeren. Afgezien van het gedragsaspect is het van cruciaal belang om prioriteit te geven aan het begrip van neurale mechanismen achter perceptuele flexibiliteit, samen met vergelijkingen tussen soorten, voor een alomvattend begrip van auditieve cognitie. Gezien de ecologische relevantie zal auditieve discriminatie in natuurlijke habitats helpen de kloof tussen laboratoriumbevindingen en relevantie in het veld te overbruggen, en inzicht bieden in cognitieve vaardigheden van verschillende soorten. Concluderend toont mijn onderzoek de perceptuele flexibiliteit van zebravinken aan, gevormd door factoren als trainingsomstandigheden, de aard van de stimulus en ecologische relevantie, waardoor ons begrip van auditieve perceptie wordt verrijkt en inzicht wordt geboden in cognitieve vaardigheden van verschillende soorten.

Acknowledgments

Acknowledgments

This thesis signifies more than two decades of collaborative work involving researchers from both the music cognition and animal behaviour fields in the Netherlands. Since my arrival on September 13th, 2018, I have felt at home. Numerous individuals have generously offered their help and guidance throughout my PhD journey, providing me with unique opportunities that I have embraced with gratitude, enabling me to present such in-depth, interdisciplinary studies in this thesis. Below, I express my immense gratitude to a few of those who played a pivotal role in supporting me on this journey.

Foremost, I extend my sincere thanks to my two supervisors. Carel ten Cate deserves my gratitude for his enlightened and generous support to the younger generation in academia. Two years before his retirement, he took the risk of considering an unfamiliar foreigner's PhD application, and I am thankful for his excellent supervision and commitment to my research project. Most importantly, I am grateful for his patience in addressing my "naive" questions and providing swift feedback, significantly aiding in advancing my research skills. I also want to express my appreciation to Henkjan Honing for his creativity, inspiration, and enthusiasm as a supervisor, making academic discussions feel comfortable. I vividly recall asking him about the potential usage of research on animal musicality, and he responded, "The motivation to spend time focusing on useless topics is human curiosity; that's why we're human beings". This reminder helped clarify my purpose in pursuing a PhD in the Netherlands. Throughout the years, both Carel and Henkjan have provided unwavering support, not only by offering the opportunity to focus on this project systematically over five years but also academically and emotionally during the challenging journey to complete this thesis and to overcome the difficulties caused by the COVID-19 pandemic.

I've been a part of Leiden's Behavioural Biology research group, a place that has always felt like home. This sentiment is thanks to my wonderful colleagues, including Annebelle, Christian, Danielle, Eira, Errol, Fabian, Fien, Gabrielle, Inge, James, Jeroen, Jiangnan, Jing, Judith, Kees, Meike, Merel, Michelle Spierings, Michelle Geers, Nick, Peter, Quanxiao, Ralph, and Roy. Our inspiring coffee break conversations covered science, birds, fishes, mussels, and intriguing "Weird Fun Facts", showcasing the diverse scientific and non-scientific interests that still amaze me. Special thanks to Jeroen for his invaluable advice on tackling statistical issues in R. I express gratitude to Peter Snelderwaard for technical support and Michelle Geers and Roy van Swetselaar for their assistance with animal care. They've been an invaluable support pillar in our research group, always ready to answer questions on bird welfare, lab administration, and equipment operation. Peter's dedicated help for over three months on the preference test, starting early in the morning, deserves special mention.

Acknowledgments extend to our senior staff in the Behavioural Biology group, Hans and Katharina, for inspiring science discussions and career advice. Their guidance kept me grounded and reminded me of the essence of our work. I am deeply grateful to the dedicated students, Emmie and Leda, for their crucial assistance with experiments. Much appreciation to our collaborator Harald van Mil for his crucial contributions to designing and clarifying the statistical analyses in my projects, as well as to Yizhen Dai for his assistance with statistical analyses.

A separate acknowledgment goes to the Music Cognition group from the University of Amsterdam. Special thanks to Jiaxin for her dedicated work on the project and the enjoyable collaboration. I express gratitude to the other members of the Music Cognition group—Atser, Ashley, Bas, Berit, David, Evert, Fleur, Makiko, Marcel, Mariëlle, and Xuan—for their relaxed and helpful demeanour, engaging brainstorming sessions, strong ideas, and insightful discussions about the project. A particular shoutout to Berit and Evert for their programming contributions to the project.

I also engaged in activities beyond the lab, where I had the pleasure of meeting amazing individuals. I extend my heartfelt thanks to my friends - Bradley, Chengjian, Roberta, Ruochen, Thomas, Thijs, and Xiangyu. I am grateful for the dedication and energy each of them contributed to the happy time we had in the Netherlands. Their reminders that this work can be both enjoyable and captivating, along with their convincing encouragement that I possessed the qualities for an academic path, meant a lot.

Lastly, and most importantly, I am fortunate to have unwavering support from my typical and unique Chinese family, conforming to the stereotype of an ideal family in some aspects. I express gratitude to my parents, brother, and other family members for their steadfast support and encouragement. They celebrated every significant moment with me and stood by me whenever I needed assistance, including my brother's professional drawing for the thesis cover and the support of my uncle and cousin.

Curriculum Vitae

Curriculum Vitae

Zhi-Yuan Ning was born in Canton province, People's Republic of China, on August 21st, 1986. He earned a Bachelor's degree in Biological Sciences, with a major in Biochemistry & Molecular Biology and a specialization in Secondary Metabolites and Structural Anatomy of Medicinal Plants, from South China Agricultural University in 2010. After completing his bachelor's degree, he worked for almost four years at the Guangdong Entomological Institute (CAS) as a Research Assistant. During this period, he collaborated with other members of the research program, conducting biological experiments on the biological control of citrus insect pests. He also independently undertook the technical development and finalization of a crop protection product—an environmental-friendly insecticide spraying glue. During this time, he was granted a National Invention Patent of China for an ecological control method for citrus pests based on the utilization of predatory mites.

In 2014, he started a master's project on animal ecology and ethology at the Chinese Academy of Sciences. For his master's thesis, he worked together with Prof. dr. Huijian Hu and Dr. Daoying Lan on the PAE behaviour Coding System on Semi-captive Hamadryas baboon. During this period, Zhi-Yuan developed a keen interest in animal behaviour and cognition. His master's study on Hamadryas baboon behaviour led to the publication of a peer-reviewed scientific article.

After completing his master's at the Chinese Academy of Sciences in 2017, he was awarded a grant from the Chinese Scholarship Council to pursue his PhD study abroad in animal behaviour and cognition. In 2018, he moved to Leiden to embark on his PhD project, investigating diverse cues involved in zebra finches' auditory perception under the supervision of Prof. dr. Carel ten Cate (Leiden University) and Prof. dr. Henkjan Honing (University of Amsterdam). Throughout his PhD, he conducted several empirical studies using both natural songs and artificial stimuli, addressing questions ranging from song discrimination and preferences to the perceptual interplay of specific acoustic features/patterns. Over the years, Zhi-Yuan has developed a growing interest in comparative cognition studies on non-human animals, and his academic journey in this scientific field is still unfolding.

Publications

Peer-reviewed articles

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.

Ning, Z., Dong, G., Tang, H., Lan, D., & Hu, H. (2017). Behavioral diversity of semi-captive hamadryas baboons (Papio hamadryas) on the basis of the PAE coding system. *Biodiversity Science*, 25(9), 1008-1018.

Ning, Z., Honing, H., van Mil, H., & ten Cate, C. (in press) The role of spectral features and song duration in zebra finch (Taeniopygia guttata) song recognition. *Animal Behaviour*.

Manuscripts for publication

Ning, Z., Honing, H., Snelderwaard, P., & ten Cate, C. (in preparation) Female and male zebra finches (Taeniopygia guttata) preferences for normal and modified songs.

Ning, Z., Honing, H., & ten Cate, C. (in preparation) Perceptual interplay of pitch and formant contours in melody recognition by zebra finch (Taeniopygia guttata).