biolinguistics

BRIEFS

Biolinguistics 14.SI: 130–144, 2020
www.biolinguistics.eu

Humans Discriminate Individual Zebra Finches by Their Song

Sabrina Schalz[1,*] & Thomas E. Dickins[1]

[1] Department of Psychology, Middlesex University, London, United Kingdom

[*] Corresponding author: ss3903@live.mdx.ac.uk

SS: https://orcid.org/0000-0003-0237-4367
TED: https://orcid.org/0000-0002-5788-0948

Comparative experiments have greatly advanced the field of biolinguistics in the 21st century, but so far very little research has focused on human perception of non-human animal vocalizations. Studies with zebra finch (*Taeniopygia guttata*) songs found that humans cannot perceive the full range of acoustic cues that zebra finches hear in their songs, although it remained unclear how much individual information is lost. Individual heterospecific discrimination by humans has only been shown with rhesus monkey (*Macaca mulatta*) voices. The present study examined whether human adults could discriminate two individual zebra finches by their songs, using a forced-choice Same-Different Paradigm. Results showed that adults can discriminate two individual zebra finches with high accuracy and without prior training. Discrimination mostly relied on differences in pitch contour, but discrimination was still possible with lower accuracy when pitch contour was removed. Future studies should expand these findings with more diverse non-human animal vocalizations.

Keywords: individual discrimination; zebra finch; pitch contour; human perception

1. Introduction

In the 21st century, the study of biolinguistics has made significant advances through comparative experiments with animal models. The majority of comparative studies so far have focused on non-human animals' perception of human language to draw inferences about which aspects of language are unique to humans (faculty of language in a narrow sense, or FLN) and which are not (faculty of language in a broad sense, or FLB; Hauser et al. 2002). For instance, java

Editors: Lluís Barceló-Coblijn, Universitat de les Illes Balears, Spain
Evelina Leivada, Universitat Rovira i Virgili, Spain

Received: 17 March 2020
Accepted: 11 May 2020
Published: 25 March 2021

ISSN 1450–3417

sparrows (*Padda oryzivora*) have been shown to discriminate patterns of emotional prosody in Japanese (Naoi et al. 2012). Java sparrows can also discriminate spoken English and Chinese (Watanabe et al. 2006), while the discrimination of Dutch and Japanese has been shown in cotton-top tamarin monkeys (*Saguinus Oedipus*; Ramus et al. 2000), rats (*Rattus norvegicus*; Toro et al. 2003), and large-billed crows (*Corvus macrorhynchos*; Schalz & Izawa 2020). Zebra finches (*Taeniopygia guttata*) can discriminate between familiar and unfamiliar song and speech in both English and Russian (Phillmore et al. 2017), perceive prosodic patterns in speech (Spierings & Cate 2014), as well as formant patterns in human speech and distinguish monosyllabic words despite speaker variation (Ohms et al. 2010).

Very few studies have tested human subjects' perception of non-human animal vocalizations. Presumably both directions could be possible with features that are considered part of the FLB. We may for instance argue that zebra finches perceive prosodic patterns in speech because prosody is not unique to language (FLB), but instead also found in birdsong. That gives us two equally intriguing possibilities: Either zebra finch prosody and speech prosody are fundamentally the same (although maybe superficially different) and can be perceived bi-directionally in their entirety by both species, or they overlap at best partially, and heterospecific perception is only feasible for one species but not the other. Vocal communication in non-human animals, although different from human language, are complex in their own right and human perception of acoustic details in these heterospecific vocalizations is far from trivial.

Studies with human infants have found that age plays a crucial role in the perception of heterospecific vocalizations. Both lemur (*Eulemur macaco flavifrons*) vocalizations and human speech but not backward speech support object categorization in three and four-months-old infants, but only speech promotes object categorization in six-months olds (Ferry et al. 2013). Exposure to primate vocalizations can extend this effect, while exposure to backward speech does not (Perszyk & Waxman 2016). Neonates initially show an equal preference for human speech and rhesus monkey (*Macaca mulatta*) vocalizations over synthetic sounds, while three-months-old infants prefer human speech over both rhesus monkey vocalizations and synthetic sounds (Vouloumanos et al. 2010). These results illustrate the effect of experience and age-related differences in human perception of heterospecific vocalizations and suggest that initial sensitivity to some heterospecific vocalizations is lost early in life due to lack of exposure and relevance. Despite this age-related decline in perception, studies with adults are nevertheless relevant and informative. Experiments with adults have shown that zebra finches are far more sensitive to temporal fine structure than humans. When presented with forwards and backwards repetitions of single periods taken from zebra finch contact calls, which differed only in the order of temporal fine structure cues, zebra finches, unlike humans, were still able to discriminate them (Dooling & Lohr 2006), which suggests that their songs may contain acoustic details that they can perceive but that we cannot (Dooling & Prior 2017). Further experiments on human perception of heterospecific vocalizations across multiple age groups are necessary to gain a more detailed understanding of the extent of the FLB. The goal of this study is therefore to further examine whether human adults perceive individual differences in zebra finch songs in a Same-Different task. Due to their intra-

individually stereotyped songs, zebra finches are a suitable model organism for this endeavour.

Male zebra finches produce signature songs learned from a tutor early in life. These songs have two primary functions: courtship (Sossinka & Böhner 1980) and within-pair communication, for example to maintain the pair bond (D'Amelio et al. 2017) or to coordinate parental care (Boucaud et al. 2017). Songs follow an individualized, stereotyped pattern (the signature) and consist of an introduction followed by multiple motifs, which in turn consist of smaller elements. These motifs convey information about the identity of the individual, while introductory elements are fairly similar between males (Sossinka & Böhner 1980; Zann 1996). They contain both amplitude and temporal envelope cues than span up to multiple seconds, and fine structure cues of individual syllables, including amplitude, spectral and temporal cues (Dooling & Prior 2017). Recent studies have shown that zebra finches are primarily sensitive to the acoustic features contained within syllables as opposed to sequences of syllables (Lawson et al. 2018). These fine structure cues convey important information about the individual's identity, its sex and the specific call type (Prior et al. 2018).

Acoustic cues conveying information about the individual's identity are important components of the vocalizations of social animals, such as the zebra finch. Consequently, they can recognize conspecifics based on their song and any of their other calls using call-type specific signatures (Elie & Theunissen 2018). Humans, on the other hand, rely on passive voice cues and primarily discriminate each other based on fundamental frequency (perceived as pitch), followed by the frequency of the first formant (F1) for female voices and formant dispersion for male voices (Baumann & Belin 2010). There is a considerable sex difference in formant perception, as men are significantly better than women at using formant dispersion to assess the acoustic size of individual animals (Charlton et al. 2013), although it is unclear how far this difference extends into voice discrimination. Fundamental frequency also plays a major role in the voice systems of other animals, such as large-billed crows (Kondo et al. 2010). Individual discrimination and recognition is possible across species as well. Carrion crows (*Corvus corone*) have been found to discriminate familiar and unfamiliar human voices and jackdaw calls (Wascher et al. 2012). Captive cheetah (*Acinonyx jubatus*) can also discriminate between familiar and unfamiliar human voices (Leroux et al. 2018), while domestic dogs and domestic cats discriminate their owner's voice from that of an unfamiliar person (Adachi et al. 2007; Saito & Shinozuka 2013) and rhesus monkeys match a familiar human voice to the corresponding face (Sliwa et al. 2011). In turn, human infants (and to some degree, adults) can discriminate two individual rhesus monkeys by their voices (Friendly et al. 2014). At an age of six months, infants showed a more accurate discrimination compared to infants tested at 12 months, although with practice the 12 months old infants were able to outperform the six-month olds (Friendly et al. 2013).

The present experiment extends these findings by testing human adults' discrimination of two zebra finches by their song. As discussed above, human perception of zebra finch songs is likely far less detailed than that of zebra finches, at least with regards to temporal fine structure. The primary aim of this study is to examine whether humans are at all able to perceive individual differences in the

songs of zebra finches, and if so, to what degree. Further attention is given to explore which acoustic cues in zebra finch songs humans can use for this task, whether there is a correlation between the listener's sex and discrimination accuracy, and whether the discrimination improves with practice.

Results will extend findings on humans' perception of zebra finch songs, and more generally offer further insights into the commonalities between human and non-human vocalizations.

2. Material and Methods

The study was separated into condition 1 with natural zebra finch songs and condition 2 with manipulated songs as described in section 2.2. Condition 2 was designed as an extension to the previously conducted condition 1, which is reflected in its smaller sample size and analysis. Apparatus and procedure were the same for both conditions. The analysis was mostly the same unless stated otherwise for the respective aspect. Results were analysed in three parts to address the core questions: whether humans can discriminate individual zebra finches by their song and if so, how accurately, which acoustic cues play a role in this discrimination, and whether discrimination accuracy improves over time. Both conditions were approved by the Middlesex University Psychology Research Ethics Committee.

2.1. Participants

Participants were 50 adults (25 female) in condition 1, and 25 adults (14 female) in condition 2. All were students and staff at Middlesex University between the ages 18 to 50. Participants did not report hearing problems and gave informed consent. No participants were removed from the analysis.

2.2. Stimuli

Stimuli in condition 1 consisted of the natural song of two male zebra finches (3 and 4 months old) recorded at Bielefeld University. Animal housing and song recording were in compliance with all applicable national guidelines for the care and use of animals. The recordings were analysed in Praat version 6.0.49 (industry standard software for acoustic analysis; Boersma & Weenink 2019) and nine motifs per individual were selected. Selection was based on high similarity in pitch contour, intensity contour, duration and number of repeated elements. Each selected motif was then high-pass filtered at 500 Hz with Audacity version 2.3.0 (https://www.audacityteam.org) to reduce low-frequency background noise (e.g. perch clanging against the cage bar) without influencing the high-frequency song. Motifs of zebra finch *B* were shorter than those produced by zebra finch *A*, and so recordings from *A* had to be cut to remove total duration as a possible discrimination cue. Cuts were made at element boundaries for clean breaks, and as such stimuli differed in mean duration by 0.04 s, which we considered acceptable (see *Table 1* for mean values of acoustic features). In addition to differences in pitch and formant frequencies, the motifs also differed structurally as motifs *A*

Acoustic feature	Zebra finch *A* mean	Zebra finch *A* SD	Zebra finch *B* mean	Zebra finch *B* SD
Duration per motif (ms)	397.6	10	335.5	8
Intensity per motif (dB)	59.7	0.5	59.1	1.5
Pitch per motif (Hz)	3177.4	204	2888.8	309.3
Frequency of F1 (Hz)	3183.4	33.2	3368.5	73.5
Frequency of F2 (Hz)	4743.6	53.6	5177.4	91.9
F1-F2 dispersion (Hz)	1560.6	38.4	1807	54.7

Table 1: Acoustic features of the nine motifs of each zebra finch. Frequency range was set to a minimum 50 Hz and maximum 10,000 Hz for the pitch analysis (note that Praat measures pitch instead of F0) and to a maximum 10,000 Hz and 3 extracted formants for the formant analysis.

consisted of two elements while motifs *B* consisted of three. A silent 2 s interval was added at the end of each motif to create clear breaks between them. As indicated by the spectrogram of zebra finch *A*, three formants were initially extracted but since only two were reliably found for zebra finch *B*, F3 was not further analysed in this study (see Appendix, *Figures A1* and *A2*).

Stimuli for condition 2 were taken from condition 1 and then manipulated in Praat (Boersma & Weenink 2019). To test the influence of the signature encoded in the envelope of the song on the discrimination accuracy participants achieve, pitch contour (the pitch pattern across the entire motif) was removed from the recordings. All existing pitch points were removed, and new pitch points were added at the time points 0.0001 s, 0.1 s, 0.2 s, 0.3 s, and 0.4 s at the frequency of the mean pitch of the respective stimulus. This was done to continue to include mean pitch as possible discrimination cue (see *Table 2* for resulting acoustic features). After initial manipulation, each recording was then checked, and additional pitch points were added were necessary (see Appendix *Figures A3* and *A4* for natural and manipulated pitch contour).

2.3. Apparatus

The participant background questionnaire and the discrimination task were presented in the software PsychoPy version 3.2 (Peirce et al. 2019) on a desktop computer. The experiment was conducted in a quiet room, and stimuli were played using over-ear headphones.

Acoustic feature	Zebra finch *A* mean	Zebra finch *A* SD	Zebra finch *B* mean	Zebra finch *B* SD
Duration per motif (ms)	397.6	10	335.5	8
Intensity per motif (dB)	59.7	0.5	59.1	1.5
Pitch per motif (Hz)	3124	192.6	2872.5	304.7
Frequency of the first formant (Hz)	3190.2	171.4	2962.7	222
Frequency of the second formant (Hz)	5517.7	223.5	5563.1	286.6

Table 2: Acoustic features of the nine manipulated motifs of each zebra finch. Frequency range was set to a minimum 50 Hz and a maximum 10,000 Hz for the pitch analysis and to a maximum 10,000 Hz and 3 extracted formants (indicated by the spectrograms) for the formant analysis.

2.4. Procedure

Participants were tested with the forced-choice Same-Different Paradigm (Pisoni & Lazarus 1974). Each of 40 trials contained two vocalizations combined at random to avoid predictability, either produced by the same individual ("same"-trial) or two different individuals ("different"-trial). Before each experiment, participants received verbal instructions about the discrimination task emphasising that the choice would be between individuals of the same species, not two different species. After the verbal explanation, participants were shown the following instructions on the screen reiterating the verbal instructions: "You will now hear 40 sound pairs. A pair of sounds was either produced by the same animal or by two animals of the same species. After each pair, you will be asked to decide whether you heard the same animal or two different animals. Sounds are separated by a 2 s interval and only 0.3 s long.". Following the playback of each pair, participants were asked whether the song was sung by the same bird (keypress "y" for yes) or not ("n" for no). During the experiment, participants did not receive feedback on their discrimination accuracy.

2.5. Analysis

The analysis was conducted entirely in R (R Core Team 2019). The first part of the analysis focused on the degree of discrimination accuracy. Following the signal detection theory (Stanislaw & Todorov 1999), responses were divided into the four categories hit (y on a "same"-trial), miss (n on a "same"-trial), correct reject (n on a "different"-trial), and false alarm (y on a "different"-trial) to determine the hit rate (proportion of hit responses in same-trials) and the false alarm rate (false alarm responses in different-trials). These two values ranging from 0 to 1 were used to calculate the discrimination sensitivity index d' using the R package psyphy and the formula `dprime.SD(H, FA, method = "diff")`, (Knoblauch 2014). We chose d' scores over other success measures, such as the percentage of correct trials,

because they are less susceptible to participants' response biases (Stanislaw & Todorov 1999). If a participant answers "yes" in every trial (an extreme response bias), the hit rate and the false alarm rate will both be 1 and the d' score for equal rates is 0. This score reflects that the participant did not discriminate between "same"-trials and "different"-trials, whereas the percentage of correct trials depends entirely on how many "same"-trials were randomly chosen, resulting in a discrimination accuracy that could be anywhere between 0% and 100%. Since d' scores cannot be calculated with absolute values, the formula described by (Snodgrass & Corwin 1988) was used to correct absolute rates of 0 and 1 (see formula 1). Seven rates of 1 and 12 rates of 0 were corrected in condition 1, as well as 1 rate of 0 in condition 2. The lowest possible d' score of 0 was given for equal hit and false alarm rates (e.g. when a participant answers yes on every trial), and when the false alarm rate was higher than the hit rate, as d' scores cannot be negative. Consequently, d' scores ranged from 0 (no discrimination) to 5.94 (perfect discrimination) and indicate discrimination accuracy on a continuous scale rather than binary success or failure. Three single trials in condition 1 were missing and thus not included in the analysis.

(1) $$corrected\ rate = \frac{0.5 + response\ rate\ (either\ hit\ or\ false\ alarm)}{1 + number\ of\ trials\ (either\ same\ or\ different)}$$

Non-parametric statistical tests were chosen for data which were not normally distributed based on a Shapiro-Wilk normality test. This was the case with d' scores in both conditions, the responses types in condition 1, and the success trend in condition 2. Homogeneity of variance was confirmed with a Levene test for the one-way ANOVA, using the R package "car" (Fox & Weisberg 2019).

A one-sample Wilcoxon signed-rank test was used to assess whether d' scores were significantly above chance level (mu = 0). Additionally, a Mann-Whitney-U test was used to determine whether d' scores differed significantly between male and female participants. No participants were excluded from this part of the analysis.

The second part of the analysis focused on the relevance of different acoustic cues: mean pitch, mean F1, and formant dispersion (F1-F2) in Hz. These cues were chosen because they are the most important cues in human voice discrimination (Baumann & Belin 2010). As formant dispersion was very irregular in condition 2, this cue was only analysed for condition 1. If a given cue was relevant for the discrimination, "same" pairs with high differences should trigger the mistake "miss" more often, and "different" pairs with low differences should trigger the mistake "false alarm" more often. Stimuli pairs with a minimum occurrence per response type were chosen to focus on the most difficult combinations and to exclude those that only triggered the same response once or twice. For condition 1, pairs that triggered a "false alarm" or a "miss" response at least three times were selected. For condition 2, pairs that triggered a "false alarm" or a "miss" response at least four times, as well as pairs that triggered a "hit" or correct reject" at least five times were selected. These different thresholds were chosen in order to only include the most frequently occurring pairs while still including enough pairs for

analysis. "Hit" and "correct reject" responses from condition 1 were not included as the success rate was so high that the analysis of correctly categorized pairs would not be very insightful. The success rate in condition 2 was lower and the sample size smaller, which is why all four response types are included. A total of 23 "false alarm" and 22 "miss" pairs were selected for condition 1. For condition 2, 14 "false alarm", 16 "miss", 18 "hit", and 17 "miss" pairs were selected. Pairs with opposite stimuli order (e.g. a2b3 and b3a2) were treated as the same pair. Every selected pair was weighted once in the acoustic cue analysis. In condition 1, acoustic parameters were compared between "false alarm" and "miss" pairs using a Mann-Whitney U test. In condition 2, a one-way ANOVA was used to analyse all four response types. Participants with a d' score of 0 were excluded from this part of the analysis since they did not perceive any difference between stimuli (one excluded in condition 1, four in condition 2).

The third part of the analysis focused on the discrimination accuracy over time. A trend in discrimination success (measured as percentage of correct answers pooled from all participants per condition) was analysed with a linear regression model `lm(percentage correct ~ trial number)` for condition 1, and a Mann-Kendall trend test for condition 2 using the R package "Kendall" (McLeod 2011). No participants were excluded from this part of the analysis.

3. Results

In condition 1, the average d' score was 3.68 (SD = 1.54, 95% CI [3.24, 4.11]) with individual scores ranging from 0 to 5.94, the highest possible score. In condition 2, the average d' score was 1.3 (SD = 0.82, 95% CI [0.96, 1.63]) and individual scores ranged from 0 to 3.29. D' scores in both conditions were significantly above chance level ($p < 0.01$), and d' scores in condition 2 were significantly below scores from condition 1 ($p < 0.01$; see *Figure 1*). There was no significant difference in d' scores between female and male participants in either condition.

Neither mean pitch nor mean F1 or mean formant dispersion were significantly lower in "false alarm" responses than "miss" responses in condition 1.

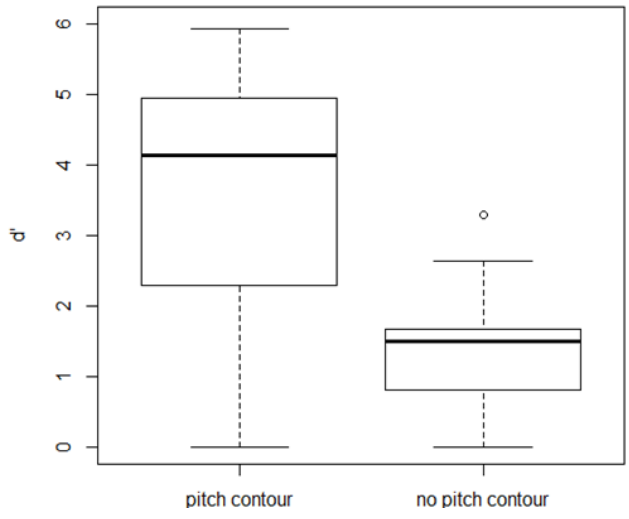

Figure 1: D' scores obtained in condition 1 (with pitch contour) and condition 2 (without pitch contour).

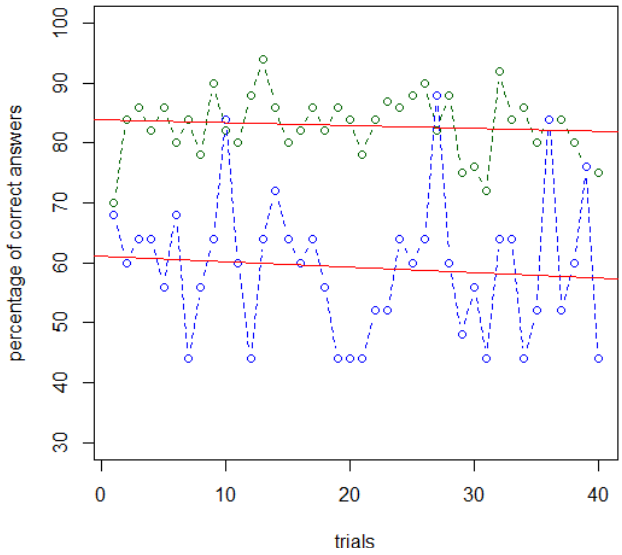

Figure 2: Percentage of correct answers (either "hit" or "correct reject") for each trial pooled from all participants per condition. Condition 1 is drawn in green, condition 2 in blue. The red linear regression lines indicate the overall trend for each condition.

There were also no significant differences in mean pitch or mean F1 frequency between response types in condition 2.

There was no significant trend across trials in either condition ($m = -0.04$ in condition 1 and $\tau = -0.12$ in condition 2; see *Figure 2*).

4. Discussion

These results show that humans can discriminate two individual zebra finches based on a short section of their song, even if pitch contour is not available as a discrimination cue. Overall, discrimination accuracy was very high with the majority of participants reaching either perfect or high d' scores, although success was highly variable inter-individually (see *Figure 1*).

This level of discrimination accuracy is especially remarkable since humans likely cannot perceive all details in zebra finch songs (Dooling & Prior 2017). While discrimination was far from perfect and there are surely some acoustic cues that participants did not perceive, this study shows that those cues that we do perceive are still enough for reliable individual discrimination. The most salient cue for this discrimination task seems to be pitch contour, a temporal envelope cue. Scores obtained in condition 2 without pitch contour were significantly lower and the percentage of correct answers across trials was more than 20 % lower in condition 2 compared to condition 1. The primary cue being part of the song envelope is in accordance with previous findings that humans are relatively insensitive to fine structure cues (Dooling & Lohr 2006).

This also suggests that our cue weighting of zebra finch songs differs considerably from that of zebra finches who are relatively insensitive to syllable sequences and instead focus on fine structure within syllables (Lawson et al. 2018). Since discrimination was still possible in condition 2 despite the removal of this cue, there must also be other, albeit less important cues that participants perceived

additionally. The analysis of stimuli pairs that triggered certain response types found that mean pitch, mean F1, and mean formant dispersion frequencies are unlikely to be contributing cues. This is contrary to findings that these three features are the most important cues for humans in voice discrimination (Baumann & Belin 2010). Mean pitch frequencies showed some variation intra-individually, but F1 and formant dispersion frequencies were fairly stereotyped between the two individuals (see *Tables 1* and *2*) and would have been available as useful cues.

Additionally, there was no difference in discrimination success between men and women, which has been observed for formant perception in acoustic size judgements (Charlton et al. 2013). Consequently, formants do not seem to be relevant for this task, although it is currently unclear why. Mean amplitude and overall duration were not available as cues, since they were standardized for all stimuli. By exclusion this leaves amplitude contour, timbre, and possibly, to some superficial degree, fine structure as possible cues in condition 2, and their potential relevance should be explored in future experiments. However, it is possible that the acoustic cues used by participants also vary inter-individually. Relevance of acoustic cues was analysed at the group level, but selective attention to certain cues over others and employed perceptual strategies could differ between individuals (Holt et al. 2018). Additionally, differences in participants' backgrounds (such as tonal languages or music training) may contribute to further attentional biases. Much more work is needed to narrow in on the acoustic cues that humans extract from zebra finch songs and how these may vary between different individuals and backgrounds.

The trend analysis (see *Figure 2*) shows that discrimination success is already high in the first trials without prior training. This is contrary to expectations based on previous findings on infants' sensitivity to non-human primate vocalizations that showed a rapid decrease in sensitivity with age and lack of exposure (Ferry et al. 2013; Perszyk & Waxman 2016; Vouloumanos et al. 2010). Even more so, it is contrary to the findings from the discrimination experiment with rhesus monkey voices in which adults only achieved an average d' score of 0.37 (Friendly et al. 2014), which is far below the mean d' scores of 3.68 and 1.3 observed here. To a large extent, this is likely due to the signature component of zebra finch songs, which is possibly easier to perceive than passive voice cues. Still, adults in condition 2 still outperformed those in the rhesus monkey study and it would be worth exploring how the discrimination of other animals would compare to these scores. The trend analysis also shows that participants' discrimination accuracy did not improve with practice, although accuracy could potentially increase with more extensive exposure exceeding 40 trials. However, the a priori high discrimination accuracy and lack of significant improvement show that this task does not require previous exposure or explicit training.

5. Conclusion

This study has shown that human adults are very sensitive to individual differences in zebra finch songs and predominantly use pitch contour to discriminate two individuals, although other acoustic cues play a role as well. Human participants do not seem to rely on mean pitch or mean formant frequencies in this

discrimination task. Discrimination accuracy is high without prior training and far exceeds the discrimination abilities observed for rhesus monkey voices in adults.

In the 21st century, the field of biolinguistics has made great advances in our understanding of shared features in human speech through comparative studies on non-human animals' perception of language, but the results obtained here show that we have not yet reached the limitations of our own perceptual capabilities with regards to heterospecific vocalizations. Going forward, more work should focus on exploring which components of non-human animal vocalizations humans of all age groups can perceive, which acoustic cues are used for this perception, and most intriguingly, why they can be perceived across species in the first place.

Data Availability

The data and code used for the analysis in this article are freely available from Figshare: https://doi.org/10.6084/m9.figshare.c.4998065.v1

Author Contributions

S.S. designed and performed the experiments, conducted the analysis, and led the writing of the manuscript. T.D. supervised the project, added to the argument, and contributed to the draft.

Acknowledgements

We would like to thank Barbara Caspers for enabling the recording of the zebra finch songs.

Appendix

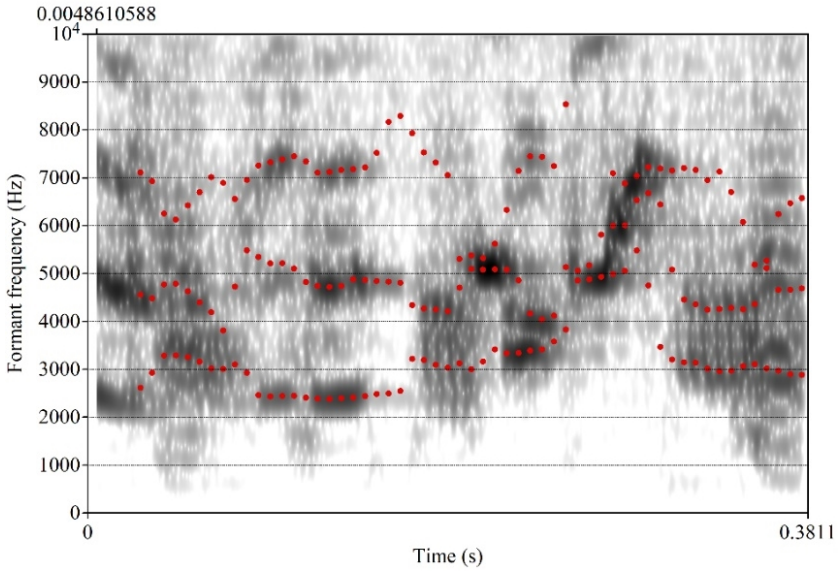

Figure A1: Sample spectrogram ranging from 0 Hz to 10,000 Hz showing one motif of zebra finch A with extracted formants drawn in (red dots) obtained in Praat (Boersma & Weenink 2019). Light area indicates pause between the two elements.

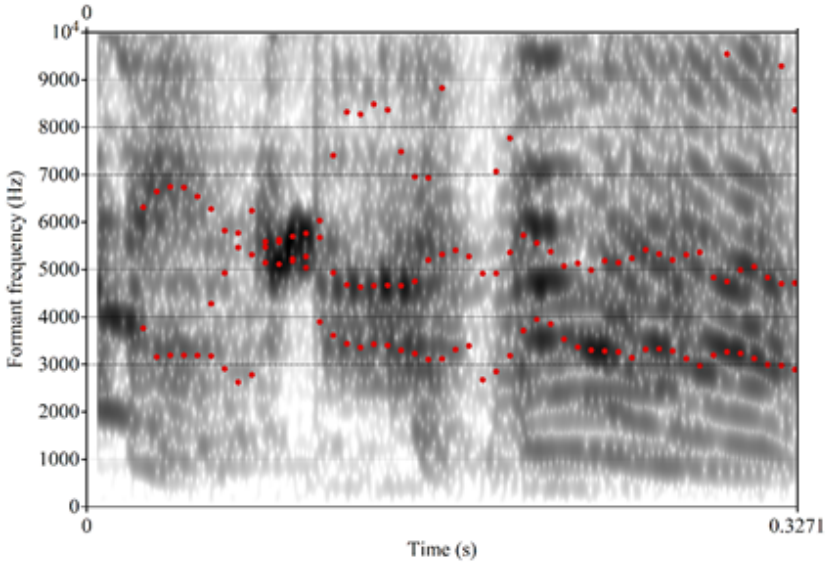

Figure A2: Sample spectrogram ranging from 0 Hz to 10,000 Hz showing one motif of zebra finch B with extracted formants drawn in (red dots) obtained in Praat (Boersma & Weenink 2019). Light areas indicate pauses between elements.

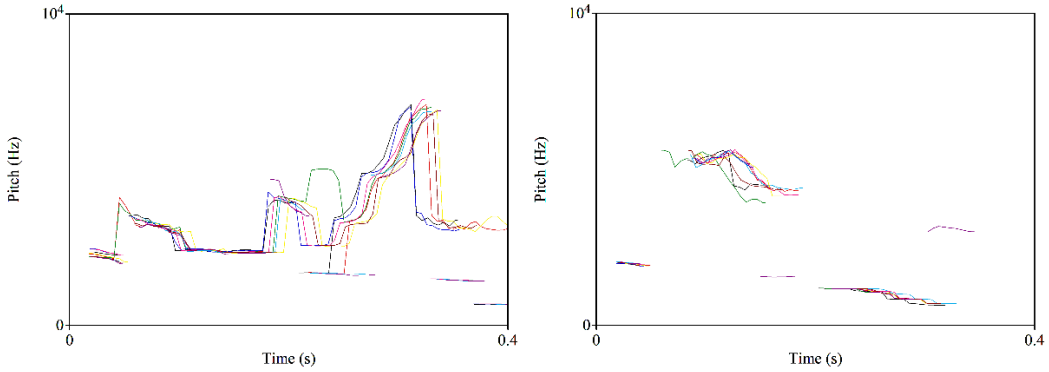

Figure A3: Natural pitch contour of the stimuli used in condition 1 produced by zebra finch A (left) and zebra finch B (right). Each colour corresponds to one motif per zebra finch.

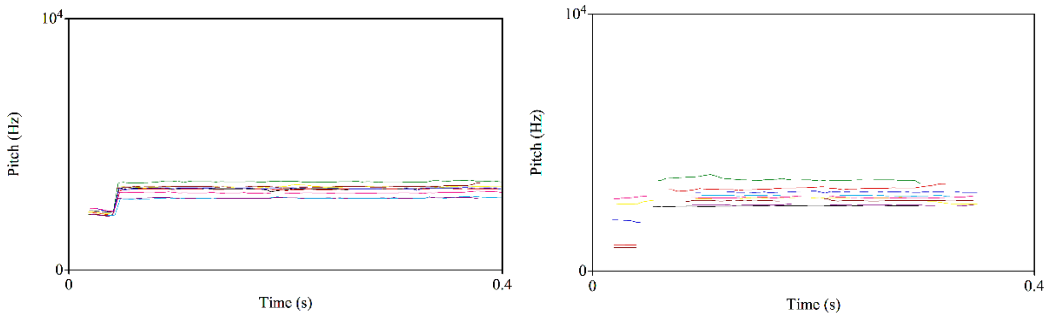

Figure A4: Manipulated pitch contour of the stimuli used in condition 2 produced by zebra finch A (left) and zebra finch B (right) where pitch contour was equalized. Each colour corresponds to one motif per zebra finch.

References

Adachi, Ikuma, Hiroko Kuwahata & Kazuo Fujita. 2007. Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition* 10(1), 17–21.

Baumann, Oliver & Pascal Belin. 2010. Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research* 74(1), 110–120.

Boersma, Paul & David Weenink. 2019. *Praat*: *Doing Phonetics by Computer*. http://www.fon.hum.uva.nl/praat.

Boucaud, Ingrid C.A., Emilie C. Perez, Lauriane S. Ramos, Simon C. Griffith & Clémentine Vignal. 2017. Acoustic communication in zebra finches signals when mates will take turns with parental duties. *Behavioral Ecology* 28(3), 645–656.

Charlton, Benjamin D., Anna M. Taylor & David Reby. 2013. Are men better than women at acoustic size judgements? *Biology Letters* 9(4): 20130270, doi: 10.1098/rsbl.2013.0270.

D'Amelio, Pietro B., Lisa Trost & Andries Ter Maat. 2017. Vocal exchanges during pair formation and maintenance in the zebra finch (taeniopygia guttata). *Frontiers in Zoology* 14: 13, doi: 10.1186/s12983-017-0197-x.

Dooling, Robert J. & Bernard Lohr. 2006. Auditory temporal resolution in the zebra finch (taeniopygia guttata): A model of enhanced temporal acuity. *Ornithological Science* 5(1), 15–22.

Dooling, Robert J. & Nora H. Prior. 2017. Do we hear what birds hear in birdsong? *Animal Behaviour* 124, 283–289.

Elie, Julie E. & Frédéric E. Theunissen. 2018. Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications* 9(1), 1–11.

Ferry, Alissa L., Susan J. Hespos & Sandra R. Waxman. 2013. Nonhuman primate vocalizations support categorization in very young human infants. *PNAS* 110(38), 15231–15235.

Fox, John & Sanford Weisberg. 2019. *car: Companion to Applied Regression.* R package version 3.0-5, https://cran.r-project.org/package=car.

Friendly, Rayna H., Drew Rendall & Laurel J. Trainor. 2014. Learning to differentiate individuals by their voices: Infants' individuation of native- and foreign-species voices. *Developmental Psychobiology* 56(2), 228–237.

Hauser, Marc D., Noam Chomsky & W. T. Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598), 1569–1579.

Holt, Lori L., Adam T. Tierney, Giada Guerra, Aeron Laffere & Frederic Dick. 2018. Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research* 366, 50–64.

Knoblauch, Kenneth. 2014. *psyphy: Functions for analyzing psychophysical data in R*. R package version 0.1-9, https://cran.r-project.org/package=psyphy.

Kondo, Noriko, Ei-Ichi Izawa & Shigeru Watanabe. 2010. Perceptual mechanism for vocal individual recognition in jungle crows (Corvus macrorhynchos): contact call signature and discrimination. *Behaviour* 147(8), 1051–1072.

Lawson, Shelby L., Adam R. Fishbein, Nora H. Prior, Gregory F. Ball & Robert J. Dooling. 2018. Relative salience of syllable structure and syllable order in

zebra finch song. *Animal Cognition* 21(4), 467–480.
Leroux, Maël, Robyn S. Hetem, Martine Hausberger & Alban Lemasson. 2018. Cheetahs discriminate familiar and unfamiliar human voices. *Scientific Reports* 8(1), 1–6.
McLeod, A. I. 2011. *Kendall: Kendall rank correlation and Mann-Kendall trend test*. R package version, 2.2 https://cran.r-project.org/package=Kendall.
Naoi, Nozomi, Shigeru Watanabe, Kikuo Maekawa & Junko Hibiya. 2012. Prosody discrimination by songbirds (Padda oryzivora). *PLoS ONE* 7(10): e47446.
Ohms, Verena R., Arike Gill, Caroline A. A. van Heijningen, Gabriel J. L. Beckers & Carel ten Cate. 2010. Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society B* 277(1684), 1003–1009.
Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas K. Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1), 195–203.
Perszyk, Danielle R. & Sandra R. Waxman. 2016. Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition* 153, 175–181.
Phillmore, Leslie S., Jordan Fisk, Simone Falk & Christine D. Tsang. 2017. Songbirds as objective listeners: Zebra finches (Taeniopygia guttata) can discriminate infant-directed song and speech in two languages. *International Journal of Comparative Psychology* 30: 32722.
Pisoni, David B. & Joan H. Lazarus. 1974. Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America* 55(2), 328–333.
Prior, Nora H., Edward Smith, Shelby Lawson, Gregory F. Ball & Robert J. Dooling. 2018. Acoustic fine structure may encode biologically relevant information for zebra finches. *Scientific Reports* 8(1): 6212.
R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Ramus, F., M. D. Hauser, C. Miller, D. Morris & J. Mehler. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288(5464), 349–351.
Saito, Atsuko & Kazutaka Shinozuka. 2013. Vocal recognition of owners by domestic cats (felis catus). *Animal Cognition* 16(4), 685–690.
Schalz, Sabrina & Ei-Ichi Izawa. 2020. Language discrimination by large-billed crows. *The Evolution of Language (EvoLang13): Proceedings of the 13th International Conference*, 380–389. https://brussels.evolang.org/proceedings/evolang13_proceedings.pdf.
Sliwa, Julia, Jean-René Duhamel, Olivier Pascalis & Sylvia Wirth. 2011. Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *PNAS* 108(4), 1735–1740.
Snodgrass, Joan G. & June Corwin. 1988. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General* 117(1), 34–50.
Sossinka, Roland & Jörg Böhner. 1980. Song types in the zebra finch poephila

guttata castanotis1. *Zeitschrift für Tierpsychologie* 53(2), 123–132.
Spierings, Michelle J. & Carel ten Cate. 2014. Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society B* 281(1787): 20140280.
Stanislaw, H. & N. Todorov. 1999. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* 31(1), 137–149.
Toro, Juan M., Josep B. Trobalon & Núria Sebastián-Gallés. 2003. The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition* 6(2), 131–136.
Vouloumanos, Athena, Marc D. Hauser, Janet F. Werker & Alia Martin. 2010. The tuning of human neonates' preference for speech. *Child Development* 81(2), 517–527.
Wascher, Claudia A. F., Georgine Szipl, Markus Boeckle & Anna Wilkinson. 2012. You sound familiar: Carrion crows can differentiate between the calls of known and unknown heterospecifics. *Animal Cognition* 15(5), 1015–1019.
Watanabe, Shigeru, Erico Yamamoto & Midori Uozumi. 2006. Language discrimination by Java sparrows. *Behavioural Processes* 73(1), 114–116.
Zann, Richard A. 1996. *The Zebra Finch*: *A Synthesis of Field and Laboratory Studies* (Oxford ornithology series 5). Oxford: Oxford University Press.