



CrossMark  
click for updates

## Research

**Cite this article:** Spierings MJ, ten Cate C. 2014 Zebra finches are sensitive to prosodic features of human speech. *Proc. R. Soc. B* **281**: 20140480.

<http://dx.doi.org/10.1098/rspb.2014.0480>

Received: 27 February 2014

Accepted: 2 May 2014

### Subject Areas:

behaviour, cognition

### Keywords:

songbirds, prosody, vocal perception, zebra finches, speech

### Author for correspondence:

Michelle J. Spierings

e-mail: [m.j.spierings.2@biology.leidenuniv.nl](mailto:m.j.spierings.2@biology.leidenuniv.nl)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2014.0480> or via <http://rspb.royalsocietypublishing.org>.

# Zebra finches are sensitive to prosodic features of human speech

Michelle J. Spierings<sup>1,2</sup> and Carel ten Cate<sup>1,2</sup>

<sup>1</sup>Behavioural Biology, Institute Biology Leiden (IBL), Leiden University, PO Box 9505, Leiden 2300 RA, The Netherlands

<sup>2</sup>Leiden Institute for Brain and Cognition (LIBC), Leiden University, PO Box 9600, Leiden 2300 RC, The Netherlands

Variation in pitch, amplitude and rhythm adds crucial paralinguistic information to human speech. Such prosodic cues can reveal information about the meaning or emphasis of a sentence or the emotional state of the speaker. To examine the hypothesis that sensitivity to prosodic cues is language independent and not human specific, we tested prosody perception in a controlled experiment with zebra finches. Using a go/no-go procedure, subjects were trained to discriminate between speech syllables arranged in XYXY patterns with prosodic stress on the first syllable and XXYY patterns with prosodic stress on the final syllable. To systematically determine the salience of the various prosodic cues (pitch, duration and amplitude) to the zebra finches, they were subjected to five tests with different combinations of these cues. The zebra finches generalized the prosodic pattern to sequences that consisted of new syllables and used prosodic features over structural ones to discriminate between stimuli. This strong sensitivity to the prosodic pattern was maintained when only a single prosodic cue was available. The change in pitch was treated as more salient than changes in the other prosodic features. These results show that zebra finches are sensitive to the same prosodic cues known to affect human speech perception.

## 1. Introduction

Linguistic communication is made possible by our ability to produce and understand speech utterances. Essential to understanding the meaning of these utterances is sensitivity to paralinguistic information, which is provided by varying parameters like pitch, amplitude and duration of speech segments. These prosodic cues can alter the meaning and emphasis of a sentence and can reveal information about the emotional state of the speaker [1]. They are also important in the process of language acquisition and discrimination [2–6]. Interestingly, we can find variation in similar features when we look at animal vocalizations, like bird song. Zebra finches, for instance, alter the amplitude of their song depending on the distance to the receiver [7] and use different singing speeds for directed and undirected song [8]. Other songbird species can change prosody-like features in their songs to adapt to changing environments [9], or when displaying aggressive behaviour [10].

This shared use and importance of prosodic features in the vocalizations of humans and non-human animals might indicate that they also share the presence of general perceptual sensitivity to prosodic features. There is, indeed, evidence that some songbirds are sensitive to variation in pitch, amplitude and song syllable duration in their natural song [11]. Also, some mammal and bird species show a sensitivity to the general prosodic patterns in human speech [12–16]. These findings suggest a similar sensitivity for prosodic cues in humans and non-human animals. However, the experiments done thus far did not track which features of human prosody were used to make this discrimination. In this study, we used a songbird species, the zebra finch, to examine the sensitivity to various prosodic features in human speech in a more systematic way.

Songbirds are one of the most relevant groups in comparative language and speech research. Like speech, birdsong is characterized by a 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 [17,18]. Songbirds are known to exhibit sensitivities to frequency, amplitude and pauses in their own song repertoire [11,19,20]. An indication that they are also sensitive to stress patterns (that may involve similar parameters) in human speech comes from a study on Java sparrows [16]. However, which cues they rely on when listening to human speech, remains an open question.

For our experiments, we tested zebra finches, which are known to be able to discriminate and categorize monosyllabic words and are sensitive to differences in vowel formants in human speech [21–23]. This discrimination holds when the words are uttered by novel speakers or by speakers of a different sex [21]. We arranged human speech syllables to follow an XXYX or an XYYX structure, knowing that zebra finches are able to discriminate between these two structures [24]. Prosodic patterns were added and varied in salience (three, two, one, or no cues) depending on the experiment.

With these experiments we address not only whether zebra finches are capable of detecting and processing prosodic cues in human speech, but also whether the three parameters involved—pitch, amplitude and duration—are all contributing to the perceiving of the prosodic patterns.

## 2. Material and methods

### (a) Subjects

Eight zebra finches (four females and four males) from the Leiden University breeding stock were tested in all experiments sequentially. All birds were between 120 and 250 days post hatching and were naive to any other experiments. The birds lived in single sex groups on a 13.5 L : 10.5 D schedule at 20–22°C. During the experiment water, grit and cuttlebone were available ad libitum. Food was used as reinforcement and therefore only available after a correct trial. Food intake was monitored daily and the birds received additional food when necessary.

### (b) Apparatus

All zebra finch experiments were conducted in an operant conditioning cage (70 (l) × 30 (d) × 45 (h) cm). Each operant cage was in a separate sound attenuated chamber and illuminated by a fluorescent tube that emitted a daylight spectrum on a 13.5 L : 10.5 D schedule. A speaker (Vifa MG10SD109-08) was located 1 m above the cage. The cage was made from wire mesh except for the floor and a plywood back wall which supported two pecking keys with LED lights. A food hatch was located in between these two keys, easily accessible to the birds. Pecking the left key (sensor 1) elicited a stimulus and illuminated the LED light of the key on the right (sensor 2). Depending on the sound, the bird had to peck sensor 2 or had to withhold its response. A correct response resulted in access to food for 10 s and an incorrect response led to 15 s of darkness.

### (c) Stimuli

The stimuli were constructed from eight naturally spoken syllables which were all produced by a male as well as a female speaker. They were first equalized in pitch, amplitude and duration with PRAAT [25]. The syllables were chosen to contain the same number of heterorganic and homorganic sounds (mo, ka, pu, le, do, sa, nu, fi). Quadruplets were created that followed either an XXYX or an XYYX pattern and were consistent in voice type

**Table 1.** The parameters for the prosodic cues that were altered to create stressed and unstressed syllables. (The pitch contours describe the change in pitch over the stressed or unstressed syllable: the first number is the frequency at the start of the syllable, the last number is the frequency at the end of the syllable.)

|                                    | unstressed syllable                              | stressed syllable                                |
|------------------------------------|--|--|
| incline/decline pitch contour (Hz) | female voice<br>220–190<br>male voice<br>150–120 | female voice<br>215–250<br>male voice<br>145–180 |
| amplitude (dB)                     | 70   | 78   |
| duration (ms)                      | 250  | 325  |

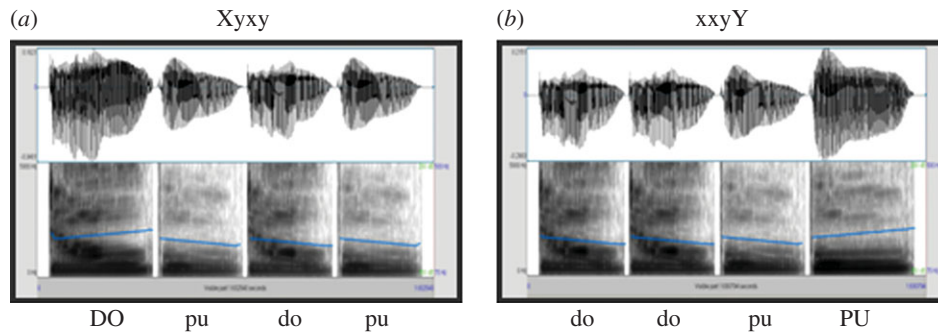
(male or female). Prosodic patterns, based on natural prosody in human speech [26], were added by modulating the pitch contour, amplitude and duration of the first or last syllable of a quadruplet, creating stressed (with all prosodic cues changed) and unstressed (no extra prosody added) syllables (table 1 and figure 1). The prosodic features were chosen in such a way that the prosodic pattern as well as the syntactic structure were of comparable salience to human adults, as assessed in pilot experiments. When trained with these stimuli, adults readily discriminated between the Xxyx and xxyY structures (capital letters indicate a stressed syllable, lower case an unstressed one). When subsequently asked to discriminate between structures in which the prosody was now switched to oppose the syntactic structure (xyxY and Xxyy), 12 out of the 32 participants based their responses on the prosodic pattern and 16 followed the syntactic structure (four participants did not choose a specific strategy). This was a clear indication that both the prosody and the structure were notable and salient cues for human adults.

### (d) Procedure

Zebra finches were first trained with a zebra finch song as the go stimulus and a pure tone as the no-go stimulus. In this shaping phase, they were familiarized with the go/no-go procedure without exposure to the experimental stimuli. The birds started with, on average, 150 trials a day. Within 3–6 days, this number increased to an average of 400 trials a day. During these days, their discrimination performance also increased significantly. When their performance reached our standard criterion (response to positive sounds more than 0.75, response to negative sounds less than 0.25 for 2 successive days), they progressed to the training phase.

During training all subjects were trained with Xxyx quadruplets (prosodic stress on first syllable) as the go stimuli and xxyY quadruplets (prosodic stress on final syllable) as the no-go stimuli. They were presented with three different go-stimuli; Abab, Cdcd and Efef, and three no-go stimuli; aabB, ccdD and eeff. The letters symbolize different syllables, which were also different for each bird. For example, sound A could be 'le' for one bird, but 'fi' for another. Capitalized letters represent a stressed syllable, and small letters an unstressed syllable. When a subject reached the standard criterion, it progressed to the test phase.

The test phase consisted of 80% reinforced training stimuli and 20% non-reinforced test stimuli. The test phase lasted until all the birds had completed 40 trials per stimulus. Experiment 1 tested whether the birds responded more to the prosodic pattern or the syntactic structure of the stimuli by changing these cues in the test phase. Experiment 2 tested whether the birds could generalize the prosodic pattern to test items constructed of new syllables. In experiment 3, we tested whether the birds would follow the



**Figure 1.** (a,b) Example of two training stimuli. 'DO pu do pu' follows the Xxyy structure and acts as a go-stimulus. The 'do do pu PU' stimulus follows the xxyY structure and is one of the no-go stimuli. In the sonogram, the change in pitch, amplitude and duration of the first or last syllable is visible. (Online version in colour.)

prosodic pattern when only a single prosodic cue was available. Experiment 4 tested the response of the birds to contradicting prosodic cues; one cue was on one edge of the quadruplet and the other two prosodic cues on the opposite edge. The fifth and final experiment tested whether the birds could also discriminate between the syntactic structures when there was no prosodic pattern added. The experiments were done in a stepwise manner, systematically varying the prosodic pattern of the test stimuli. Appendix A provides a full list of stimuli per experiment.

### (e) Analysis

For each experiment, we calculated the discrimination ratio (DR) for each stimulus pair and bird as the number of responses to the stimuli that was consistent with the positively reinforced prosodic pattern of the training, divided by the response total. The group data was then analysed with a linear mixed model (lmer) that took into account the repeated measurements per bird and the distribution of the pecking data. Random variables such as individual, sex and type of voice were checked for any influence on the differences in performance. The within-experiment analysis were conducted with a Wilcoxon signed-rank test with continuity correction on the response rates to the stimuli that followed the positively reinforced prosodic pattern from the training and the stimuli that followed the positively reinforced syntactic structure from the training ( $S^+$  = number of correct responses to prosodic pattern stimulus/total number of responses;  $S^-$  = number of incorrect responses to syntactic structure stimulus/total number of responses). Performance on an individual level was checked with a Clopper–Pearson binominal confidence interval measurement, using the responses to the positive stimulus and the responses to the negative stimulus per stimulus pair. These statistics are only presented when the birds differed from the group performance.

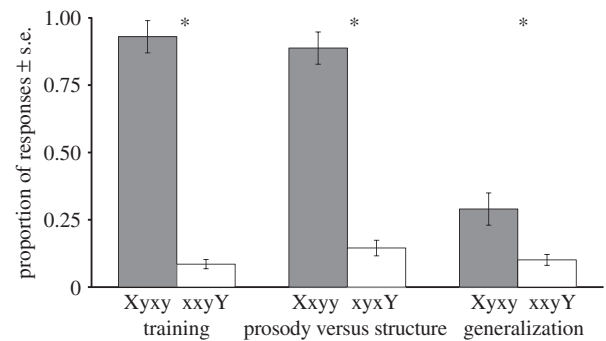
## 3. Experiment 1: prosody versus structure

### (a) Stimuli

The test stimuli for this experiment were similar to the training stimuli, but the prosodic pattern and syntactic structure were now switched. The test items were xxyY and Xxyy, while training items remained Xxyy and xxyY. All prosodic cues were kept as explained in §2.3.

### (b) Results and discussion

Results of experiment 1 showed that zebra finches responded more strongly to the prosodic pattern than to the syntactic structure (mean response Xxyy = 0.89, mean response xxyY = 0.14,  $Z = -2.53$ ,  $p = 0.011$ ; figure 2). These results were not influenced by sex of the bird or stimulus voice (sex ~ test \* DR:  $t = -4.32$ ,  $p = 0.14$ ; voice ~ test \* DR:  $t = -5.71$ ,  $p = 0.27$ ) and



**Figure 2.** Proportion of responses to the training and test items of experiment 1 and 2. The bars 'training' show the responses of the birds on the day before they started the test. In experiment 1, prosody versus structure, the grey bar indicates a stronger response to the prosodic pattern compared with the syntactic structure. The results of experiment 2, generalization, show responses to similar quadruplets as the training, but consisting of new speech syllables. We show the mean proportion of response of all birds  $\pm$  s.e. The asterisks indicate a significant difference between the responses to the two stimuli types.

were consistent for all individuals (from now on only reported when deviating). During training, the birds could have discriminated the quadruplets on the prosodic pattern, the syntactic structure or both. Our results indicate that in this experiment, all birds focused on the prosodic pattern.

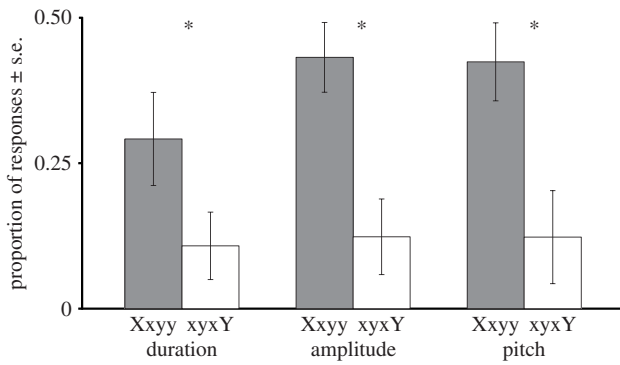
## 4. Experiment 2: generalization to unfamiliar syllables

### (a) Stimuli

Generalization was measured by changing the syllables from the training quadruplets to new syllables. During training, every bird heard six out of the eight possible syllables. The remaining two syllables were used in this experiment and were therefore different for each bird. The prosodic pattern and syntactic structure of the quadruplets remained the same as the training quadruplets, Xxyy and xxyY.

### (b) Results and discussion

Results showed that zebra finches still discriminated the stimuli (mean  $S^+ = 0.29$ , mean  $S^- = 0.1$ ,  $Z = -2.524$ ,  $p = 0.012$ ; figure 2). This could be obtained by either attending to the prosodic pattern or to the syntactic structure of the test items. From the strong response of the zebra finches to the prosodic pattern in experiment 1 and supported by the findings of the



**Figure 3.** Proportion of responses to test items which had one prosodic cue that was on the opposite position compared with the training stimuli. Shown are the mean proportions of response of all birds  $\pm$  s.e. The asterisks indicate a significant difference between the responses to the two stimuli types. Higher grey bars indicate a stronger response to the single prosodic cue compared with the syntactic structure.

subsequent experiments, we conclude that it is most likely that they generalized the prosodic pattern rather than the structural one.

## 5. Experiment 3: one prosodic cue

### (a) Stimuli

To understand the influence of single prosodic cues, the stimuli in this experiment had a single prosodic cue on the first or the last syllable, which could be pitch, amplitude or duration. The syntactic structure remained the same as during training, but the one prosodic cue was switched to the opposite position compared with the training stimuli. Training stimuli remained Xxyy and xxyY, the test stimuli were, for example, xyxY<sup>pitch</sup> and X<sup>pitch</sup>xxy.

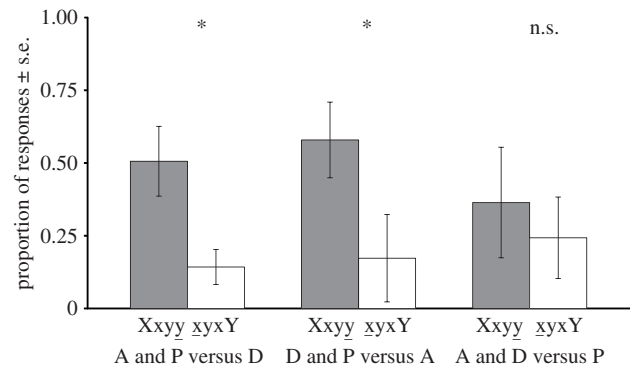
### (b) Results and discussion

Even with one prosodic cue, the birds responded more strongly to the prosodic pattern than to the syntactic structure of the test stimuli (duration: mean responses Xxyy = 0.29, mean responses xyxY = 0.11,  $Z = -2.524$ ,  $p = 0.012$ ; amplitude: mean responses Xxyy = 0.43, mean responses xyxY = 0.13,  $Z = -2.524$ ,  $p = 0.012$ ; pitch: mean responses Xxyy = 0.42, mean responses xyxY = 0.12,  $Z = -2.521$ ,  $p = 0.012$ ; figure 3). There was no difference in DR between the three test conditions (mean DR amplitude = 0.73, mean DR duration = 0.78, mean DR pitch = 0.79,  $t = -3.46$ ,  $p = 0.38$ ). This indicates that even a single and less notable prosodic cue is salient enough to outweigh the syntactic structure.

## 6. Experiment 4: two prosodic cues versus one prosodic cue

### (a) Stimuli

Stimuli for this experiment were constructed with all prosodic cues, but not on the same syllable. Either the first or the last syllable of a quadruplet had two prosodic cues stressed (e.g. pitch and duration); the other syllable was stressed by the one remaining prosodic cue (e.g. amplitude). The quadruplets were ordered such that the two prosodic cues were contradicting the syntactic structure from the training. The single cue



**Figure 4.** Proportions of responses to the test items in which two prosodic cues were not coinciding with the syntactic structure from the training. The remaining prosodic cue did correspond to the location of all prosodic features in the training. We show the mean proportion of response of all birds  $\pm$  s.e. The asterisks indicate a significant difference between the responses to the two stimuli types. Higher grey bars indicate a stronger response to the two prosodic cues. A, amplitude; D, duration; P, pitch.

was in line with the position that prosody had on the syntactic training structure. When trained on Xxyy and xxyY, a test quadruplet looked like X<sup>pitch,duration</sup>xy<sup>amplitude</sup> and x<sup>amplitude</sup>xyY<sup>pitch,duration</sup>. Here, the syllable that has one prosodic cue is indicated by a small, underlined letter.

### (b) Results and discussion

Results showed that when an increased pitch on a syllable was combined with any of the other two prosodic cues, this was salient enough for the birds to respond to it as if it were the training prosody (A and P versus D: mean responses Xxyy = 0.51, mean responses xyxY = 0.14,  $Z = -2.521$ ,  $p = 0.012$ ; D and P versus A: mean responses Xxyy = 0.58, mean responses xyxY = 0.17,  $Z = -2.380$ ,  $p = 0.017$ ; figure 4), even though a third cue was different from this pattern and corresponded to the syntactic structure. However, when an increased pitch was the single cue, the birds did not discriminate between the two structures (A and D versus P: mean responses Xxyy = 0.36, mean responses xyxY = 0.24,  $Z = -0.840$ ,  $p = 0.401$ ; figure 4). Five birds responded more to the two prosodic cues (amplitude and duration) and three birds showed the opposite response. For none of the birds did this lead to a significant difference in response to the stimuli. This indicates that, for the current set of stimuli, pitch is the most salient cue to the birds.

## 7. Experiment 5: no prosody

### (a) Stimuli

As a final test, we used stimuli that had the exact same syllables and syntactic structure as during training, but were composed of unstressed syllables only. This allowed us to learn whether the birds did discriminate the syntactic structures as well.

### (b) Results and discussion

When only syntactic structure is available to discriminate the stimuli, the birds still discriminated the ones that followed the go structure from the ones that followed the no-go structure (mean responses xyxy = 0.17, mean responses

$xyy = 0.09$ ,  $Z = -2.524$ ,  $p = 0.012$ ). This indicates that although the birds are very responsive even to single prosodic cues, they also learned about the syntactic structure of the stimuli. It also demonstrates that a single prosodic cue can overrule the structural information.

## 8. General discussion

Our experiments show that, when presented with three prosodic cues, zebra finches are sensitive to the same prosodic features of human speech as humans are. Moreover, the zebra finches respond stronger to the prosodic features of our stimuli than humans do. When required to either follow the prosodic cues or the syntactic structure, human participants are split equally in which of these they relied upon, whereas all of our zebra finches follow the prosodic pattern of the stimuli. The zebra finches are also able to generalize this prosodic pattern to quadruplets consisting of new syllables and to quadruplets that had a mismatch between the prosodic pattern, even with only a single prosodic cue present, and the syntactic structure. This is shown by an increased number of go-responses to stimuli in which the prosodic pattern matched with that of the go-stimuli of the training phase. Furthermore, the zebra finches do not differ in their responses to the stimuli in which only pitch, amplitude or duration is changed; they responded equally often to the stimuli in which the prosodic pattern followed the go-pattern of the training phase. When the cues are contradicting each other, pitch is the only cue that balances out the other two prosodic cues. However, it is notable in some experiments that the responses of the zebra finches to test items are in general lower than the responses to the training stimuli. This might have been owing to a novelty affect, causing the birds to respond in a risk aversive manner when confronted with new sounds.

Although animals can discriminate strings with different syntactic structures, the generalization of such structures to novel items has rarely been obtained in non-human animals [24,27]. Songbirds have only limited abilities to generalize different structures to novel sounds [24,27,28]. In contrast to this limited generalization based on structural cues, we show that zebra finches can readily generalize prosodic patterns to strings with novel syllables. This generalization suggests that the distinction learned during the training was not based on trivial cues related to the individual stimuli used for training, but on something more general that is shared by all stressed syllables, like the relative difference in pitch, amplitude or duration of the stressed syllable compared with the other ones. This ability for generalization or abstraction of the prosodic cues may also underlie the results obtained in a study on Java sparrows [16], where the sparrows were able to abstract the prosodic patterns of natural human speech sentences. Our results add to these findings by demonstrating that each of three different prosodic cues can be used to detect prosodic patterns. There was no detectable difference in the discrimination abilities of the birds when they heard a change in pitch, amplitude or duration as the only cue of the prosodic pattern.

When the zebra finches were presented with contradicting prosodic cues, we found a stronger effect of the pitch change compared with the other two prosodic features. In our experiment, pitch outweighed the other two prosodic

cues. It is known that birds in general, and also zebra finches, are better at detecting small differences in frequencies than proportionally similar changes in duration or intensity in their natural song [11,19,20]. Our results show that changes in pitch are also strong cues when songbirds listen to human speech syllables.

These results add to previous research showing that zebra finches are sensitive to phonetic cues by different speakers [21,22], by showing a sensitivity to other natural variations in human speech. They support the idea that such sensitivities either have a shared ancestral state or have evolved independently along highly similar lines in different groups of vocal communicators. The sensitivity of the birds to the features shown here could be related to the impact that these features might have in song interactions among conspecific birds. In the song of many songbirds, variations in these particular features can carry important information on, for example, the quality of the singing male [29]. When subjected to different environments or situations, songbirds are capable of modifying prosody-like features in their song [7,9]. For example, when food availability is decreased, an individual zebra finch often reduces its singing speed and song amplitude [30]. These song modifications show that at least some prosodic variation in birdsong is related to environmental or genetic parameters, indicating that these prosodic features could be of importance in vocal communication.

The subtle variation in vocal signals, which can be significant in communication, may also be the evolutionary origin of human sensitivity to such cues. One of the hypotheses concerning the evolution of human language is that it was preceded by a prosodic-protolanguage [31]. This protolanguage was based on a learned and generative vocal system with modifications in prosodic features as pitch, amplitude and duration. It might have its origins in a pre-existing sensitivity to meaningful variation in non-speech sounds present in our ancestors. The evolutionary process that shaped the sensitivity to relevant and sometimes subtle sound variations in birds, but most likely also that in other animals, might also have been at the base of the sensitivity to prosodic, paralinguistic features in human speech.

## 9. Conclusion

Results of our experiments showed that zebra finches are able to abstract and generalize prosodic patterns of human speech to new tokens. The zebra finches responded more strongly to the prosodic pattern than to the syntactic structure of the quadruplets. This response pattern persisted even when only one of three cues was present in a stimulus. Humans share this sensitivity to the prosodic pattern, although they also noted and responded to the syntactic structure of our stimuli. Our results show that sensitivity to prosodic cues is not linked to the possession of language and might have preceded language evolution, possibly originating from a pre-existing sensitivity to meaningful variation in pre-linguistic communicative sounds.

Research was approved by the Leiden Committee for animal experimentation DEC number 13007.

**Acknowledgements.** We thank Andreea Geambasu, Raquel G. Alhama, Clara Levelt and Willem Zuidema for useful comments on the experimental procedure and on the manuscript. We thank Anouk de Weger, Stefan Wever and Dexter Wessels for help with the practical work.

**Funding statement.** This research was supported by NWO-GW, grant no. 360.70.452.

## Appendix A

Training and test stimuli used for the experiments. A–H are the eight speech syllables that we used, these were different syllables for every individual. P, elevated pitch; D, increased duration and A, increased amplitude of the syllable.

| go-stimuli                          | no-go stimuli       |
|-------------------------------------|---------------------|
| training stimuli                    |                     |
| a <sup>PDA</sup> bab                | aabb <sup>PDA</sup> |
| c <sup>PDA</sup> dcd                | ccdd <sup>PDA</sup> |
| e <sup>PDA</sup> fef                | eeff <sup>PDA</sup> |
| test stimuli                        |                     |
| experiment 1: prosody versus syntax |                     |
| a <sup>PDA</sup> abb                | abab <sup>PDA</sup> |
| c <sup>PDA</sup> cdd                | ccdd <sup>PDA</sup> |
| e <sup>PDA</sup> eff                | efef <sup>PDA</sup> |
| experiment 2: generalization        |                     |
| h <sup>PDA</sup> ghg                | hhgg <sup>PDA</sup> |
| g <sup>PDA</sup> hgh                | gghh <sup>PDA</sup> |
| experiment 3: one prosodic cue      |                     |
| a <sup>D</sup> abb                  | abab <sup>D</sup>   |

(Continued.)

(Continued.)

| go-stimuli  | no-go stimuli                    |
|---|----------------------------------|
| c <sup>D</sup> cdd                                      | cdcd <sup>D</sup>                |
| e <sup>D</sup> eff                                      | efef <sup>D</sup>                |
| a <sup>A</sup> abb                                      | abab <sup>A</sup>                |
| c <sup>A</sup> cdd                                      | cdcd <sup>A</sup>                |
| e <sup>A</sup> eff                                      | efef <sup>A</sup>                |
| a <sup>P</sup> abb                                      | abab <sup>P</sup>                |
| c <sup>P</sup> cdd                                      | cdcd <sup>P</sup>                |
| e <sup>P</sup> eff                                      | efef <sup>P</sup>                |
| experiment 4: two prosodic cues versus one prosodic cue |                                  |
| a <sup>AP</sup> abb <sup>D</sup>                        | a <sup>D</sup> bab <sup>AP</sup> |
| c <sup>AP</sup> cdd <sup>D</sup>                        | c <sup>D</sup> dcd <sup>AP</sup> |
| e <sup>AP</sup> eff <sup>D</sup>                        | e <sup>D</sup> fef <sup>AP</sup> |
| a <sup>PD</sup> abb <sup>A</sup>                        | a <sup>A</sup> bab <sup>PD</sup> |
| c <sup>PD</sup> cdd <sup>A</sup>                        | c <sup>A</sup> dcd <sup>PD</sup> |
| e <sup>PD</sup> eff <sup>A</sup>                        | e <sup>A</sup> fef <sup>PD</sup> |
| a <sup>DA</sup> abb <sup>P</sup>                        | a <sup>P</sup> bab <sup>DA</sup> |
| c <sup>DA</sup> cdd <sup>P</sup>                        | c <sup>P</sup> dcd <sup>DA</sup> |
| e <sup>DA</sup> eff <sup>P</sup>                        | e <sup>P</sup> fef <sup>DA</sup> |
| experiment 5: no prosody                                |                                  |
| abab  | aabb                             |
| cdcd  | ccdd                             |
| efef  | eeff                             |

## References

- Kuhl PK. 2004 Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831–843. (doi:10.1038/nrn1533)
- Johnson EK, Seidl AH. 2009 At 11 months, prosody still outranks statistics. *Dev. Sci.* **12**, 131–141. (doi:10.1111/j.1467-7687.2008.00740.x)
- Jusczyk PW, Houston DM, Newsome M. 1999 The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* **39**, 159–207. (doi:10.1006/cogp.1999.0716)
- Seidl A, Johnson EK. 2006 Infant word segmentation revisited: edge alignment facilitates target extraction. *Dev. Sci.* **9**, 565–573. (doi:10.1111/j.1467-7687.2006.00534.x)
- Thiessen ED, Saffran JR. 2003 When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* **39**, 706–716. (doi:10.1037/0012-1649.39.4.706)
- Nazzi T, Jusczyk PW, Johnson EK. 2000 Language discrimination by English-learning 5-month-olds: effects of rhythm and familiarity. *J. Mem. Lang.* **43**, 1–19. (doi:10.1006/jmla.2000.2698)
- Brumm H, Slater PJB. 2006 Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Anim. Behav.* **72**, 699–705. (doi:10.1016/j.anbehav.2006.01.020)
- Sossinka R, Bohner J. 1980 Song types in the zebra finch (*Poephila guttata castanotis*). *Z. Tierpsychol.* **53**, 123–132. (doi:10.1111/j.1439-0310.1980.tb01044.x)
- Brumm H, Zollinger SA. 2011 The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* **148**, 1173–1198. (doi:10.1163/000579511X605759)
- Brumm H, Ritschard M. 2011 Song amplitude affects territorial aggression of male receivers in chaffinches. *Behav. Ecol.* **22**, 310–316. (doi:10.1093/beheco/arq205)
- Nagel KI, McLendon HM, Doupe AJ. 2010 Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *J. Neurophysiol.* **104**, 1426–1437. (doi:10.1152/jn.00028.2010)
- Ramus F, Hauser MD, Miller C, Morris D, Mehler J. 2000 Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* **288**, 349–351. (doi:10.1126/science.288.5464.349)
- Tincoff R, Hauser M, Tsao F, Spaepen G, Ramus F, Mehler J. 2005 The role of speech rhythm in language discrimination: further tests with a non-human primate. *Dev. Sci.* **8**, 26–35. (doi:10.1111/j.1467-7687.2005.00390.x)
- Toro JM, Trobalon JB, Sebastián-Gallés N. 2003 The use of prosodic cues in language discrimination tasks by rats. *Anim. Cogn.* **6**, 131–136. (doi:10.1007/s10071-003-0172-0)
- Toro JM, Trobalon JB, Sebastián-Gallés N. 2005 Effects of backward speech and speaker variability in language discrimination by rats. *J. Exp. Psychol. Anim. B* **31**, 95–100. (doi:10.1037/0097-7403.31.1.95)
- Naoi N, Watanabe S, Maekawa K, Hibiya J. 2012 Prosody discrimination by songbirds (*Padda oryzivora*). *PLoS ONE* **7**, e47446. (doi:10.1371/journal.pone.0047446)
- Bolhuis JJ, Okanoya K, Scharff C. 2010 Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759. (doi:10.1038/nrn2931)
- Doupe AJ, Kuhl PK. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
- Lee TTY, Charrier I, Bloomfield LL, Weisman RG, Sturdy CB. 2006 Frequency-range discriminations

- and absolute pitch in black-capped chickadees (*Poecile atricapillus*), mountain chickadees (*Poecile gambeli*), and zebra finches (*Taeniopygia guttata*). *J. Comp. Psychol.* **120**, 217–228. (doi:10.1037/0735-7036.120.3.217)
20. Weisman RG, Balkwill LL, Hoeschele M, Moscicki MK, Bloomfield LL, Sturdy CB. 2010 Absolute pitch in boreal chickadees and humans: exceptions that test a phylogenetic rule. *Learn. Motiv.* **41**, 156–173. (doi:10.1016/j.lmot.2010.04.002)
  21. Ohms VR, Gill A, Van Heijningen CAA, Beckers GJL, ten Cate C. 2010 Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proc. R. Soc. B* **277**, 1003–1009. (doi:10.1098/rspb.2009.1788)
  22. Ohms VR, 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**, 155–161. (doi:10.1007/s10071-011-0441-2)
  23. Dooling RJ, Best CT, Brown SD. 1995 Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *J. Acoust. Soc. Am.* **97**, 1839–1846. (doi:10.1121/1.412058)
  24. van Heijningen CAA, 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. USA* **106**, 20 538–20 543. (doi:10.1073/pnas.0908113106)
  25. Boersma P, Weenink D. 2014 PRAAT: doing phonetics by computer (computer program), v. 5.3.56. See <http://www.praat.org/> (accessed 24 January 2013).
  26. Koopmans van Beinum FJ. 1980 *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam, The Netherlands: Academische Pers B.V.
  27. ten Cate C, Okanoya K. 2012 Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Phil. Trans. R. Soc. B* **367**, 1984–1994. (doi:10.1098/rstb.2012.0055)
  28. van Heijningen CAA, 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**, 165–175. (doi:10.1007/s10071-012-0559-x)
  29. Spencer KA, MacDougall-Shackleton SA. 2011 Indicators of development as sexually selected traits: the developmental stress hypothesis in context. *Behav. Ecol.* **22**, 1–9. (doi:10.1093/beheco/arq068)
  30. Ritschard M, Brumm H. 2012 Zebra finch song reflects current food availability. *Evol. Ecol.* **26**, 801–812. (doi:10.1007/s10682-011-9541-3)
  31. Fitch T. 2005 The evolution of language: a comparative review. *Biol. Philos.* **20**, 193–230. (doi:10.1007/s10539-005-5597-1)