



Universiteit  
Leiden  
The Netherlands

## **Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception**

Ohms, V.R.; Escudero, P.; Lammers, K.; Cate, C. ten

### **Citation**

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

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/76092>

**Note:** To cite this publication please use the final published version (if applicable).

# Zebra finches and Dutch adults exhibit the same cue weighting bias in vowel perception

Verena R. Ohms · Paola Escudero ·  
Karin Lammers · Carel ten Cate

Received: 8 April 2011 / Revised: 30 June 2011 / Accepted: 4 July 2011 / Published online: 15 July 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** Vocal tract resonances, called formants, are the most important parameters in human speech production and perception. They encode linguistic meaning and have been shown to be perceived by a wide range of species. Songbirds are also sensitive to different formant patterns in human speech. They can categorize words differing only in their vowels based on the formant patterns independent of speaker identity in a way comparable to humans. These results indicate that speech perception mechanisms are more similar between songbirds and humans than realized before. One of the major questions regarding formant perception concerns the weighting of different formants in the speech signal (“acoustic cue weighting”) and whether this process is unique to humans. Using an operant Go/NoGo design, we trained zebra finches to discriminate syllables, whose vowels differed in their first three formants. When subsequently tested with novel vowels, similar in either their first formant or their second and third formants to the familiar vowels, similarity in the higher

formants was weighted much more strongly than similarity in the lower formant. Thus, zebra finches indeed exhibit a cue weighting bias. Interestingly, we also found that Dutch speakers when tested with the same paradigm exhibit the same cue weighting bias. This, together with earlier findings, supports the hypothesis that human speech evolution might have exploited general properties of the vertebrate auditory system.

**Keywords** Speech evolution · Formants · Vowel perception · Zebra finches · Acoustic cue weighting

## Introduction

The evolution of speech and language is still a fiercely debated topic among scientists from various disciplines (Dooling et al. 1995; Hauser et al. 2002; Pinker and Jackendoff 2005; Anderson 2008; Fitch 2010). The original assumption that “speech is special” (Lieberman 1982) and that the mechanisms underlying speech perception are uniquely human (Lieberman 1975) has been challenged over the years by numerous studies indicating that the general ability of speech perception is widely shared with other species including both mammals (Kuhl and Miller 1975, 1978; Hienz et al. 1996; Eriksson and Villa 2006) and birds (Kluender et al. 1987; Dooling and Brown 1990; Ohms et al. 2010a).

The first evidence that human speech has likely exploited properties of the mammalian auditory system derives from experiments with chinchillas (Kuhl and Miller 1975, 1978), which have shown that these animals exhibit the same phonetic boundaries as humans do when categorizing different consonant–vowel syllables. Similar studies, however, revealed that the phonetic boundaries

---

V. R. Ohms (✉) · K. Lammers · C. ten Cate  
Behavioural Biology, Institute of Biology Leiden (IBL),  
Leiden University, Sylvius Laboratory, Sylviusweg 72,  
PO Box 9505, 2300 RA Leiden, The Netherlands  
e-mail: vrohms@indiana.edu

### *Present Address:*

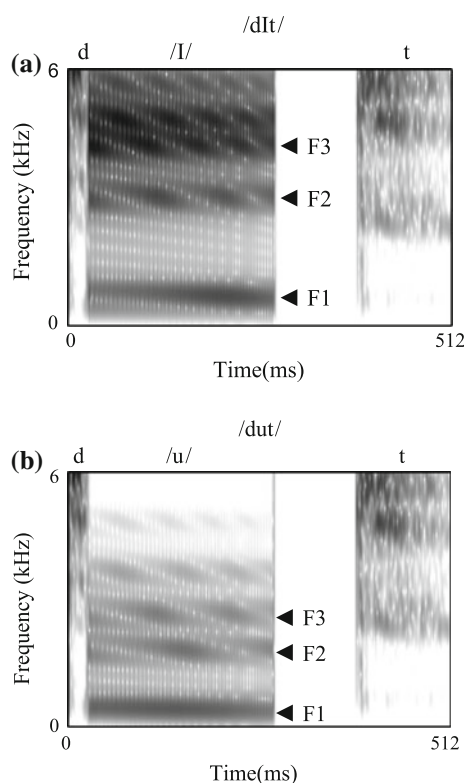
V. R. Ohms  
Medical Sciences, Indiana University Bloomington, Jordan Hall,  
1001 East Third Street, Bloomington, IN 47405, USA

P. Escudero  
MARCS Auditory Laboratories, University of Western Sydney,  
Locked Bag 1797, Penrith, NSW 2751, Australia

C. ten Cate  
Leiden Institute for Brain and Cognition, Postzone C2-S,  
PO Box 9600, 2300 RC Leiden, The Netherlands

involved in speech perception are not specific to humans or mammals in general, but that budgerigars also perceive consonants in a categorical fashion and that their phonetic boundaries roughly correspond to those of humans (Dooling et al. 1989). Furthermore, Japanese quail can correctly categorize /d/, /b/, and /g/ consonant–vowel syllables even if these consonants are presented in novel vowel contexts (Kluender et al. 1987).

Recent evidence suggests that zebra finches are also highly sensitive to formant patterns in human speech (Ohms et al. 2010a), which represent resonances of the vocal tract (Titze 2000). Different vowels have different formant frequencies. The frequency values of formants are dependent on the position of the tongue in the mouth cavity (Ladefoged 2006). For instance, the vowels in the syllables /dIt/ and /dut/ have different first (F1), second (F2), and third formant (F3) frequency values: /u/ has lower F1, F2, and F3 values than /I/ (see Fig. 1). Zebra finches, when trained to discriminate the produced words “wit” and “wet” differing only in their vowel sounds and hence in their formant frequencies, can categorize these words irrespective of sex



**Fig. 1** Spectrograms of two syllables differing only in their vowels. This figure shows spectrograms of two synthetic syllables: /dIt/ and /dut/. It is clearly visible that formant frequencies differ between the vowels with lower formant frequencies in /u/ compared to /I/. F1, first formant; F2, second formant; F3, third formant; kHz, kilohertz; ms, milliseconds

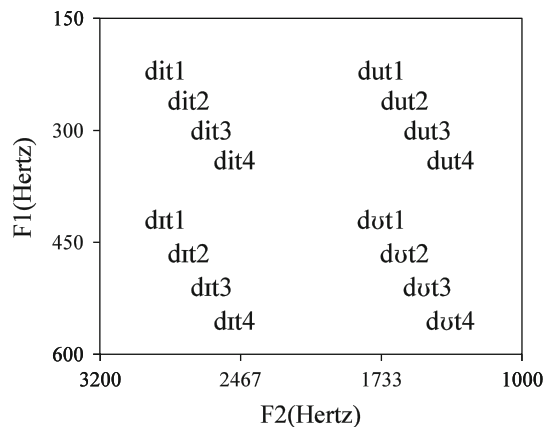
and identity of the speaker. The underlying mechanism seems to be, as in humans, a combination of extrinsic and intrinsic speaker normalization (Ohms et al. 2010a).

However, research in this direction treated speech sounds as unimodal entities and has not considered the fact that multiple acoustic features are involved in their production and perception. A major unsolved question in speech perception for humans as well as other species regards the relative contribution that those different acoustic features have in the perception of speech sound contrasts.

It has been shown for instance that when classifying the vowels in the Dutch words “taak” (task) and “tak” (branch), Dutch and German listeners prefer to use the vowels’ formant frequency differences to the vowels’ differences in duration, whereas Spanish L1 speakers who are learners of Dutch as a second language (L2) favor vowel duration (Escudero et al. 2009). Thus, listeners from different linguistic backgrounds weight temporal and spectral vowel parameters differently. However, spectral properties, namely the combination of first, second, and third formant frequencies, were used in the analysis, and the relative importance of individual frequencies was not considered. Therefore, the present study investigates whether different low (F1) and high (F2 and F3) formant frequencies are weighted differently in humans and in other species.

We used a Go/NoGo operant conditioning paradigm to test acoustic cue weighting in a species assumed to perceive vowel formants in similar ways as humans, namely the zebra finch (*Taeniopygia guttata*) (Dooling et al. 1995; Ohms et al. 2010a). In their own vocalizations, zebra finches show a variety of note types, covering a wide frequency range, which are produced using various articulators (Ohms et al. 2010b). Some song elements as well as zebra finch calls have harmonic structures comparable to human vowels, and it has been shown that these birds can detect rather subtle amplitude decreases in single harmonics (Lohr and Dooling 1998). Furthermore, when trained to detect inharmonicity, zebra finches outperform humans, which led to speculations that zebra finches might either use non-spectral cues or possess an enhanced ability to discriminate the same cues, which are otherwise also used by humans (Lohr and Dooling 1998).

In our experiment, stimuli varied in the first three formant frequencies, while other cues, temporal and spectral, were kept the same. We used four synthetic tokens of each of the vowels /i/, /I/, /u/, and /U/ (see Fig. 2; Table 1), which had similar F1 and F2 values to those reported earlier (Curtin et al. 2009). These vowels were embedded in the same natural d\_t frame resulting in the four syllables /dit/, /dIt/, /dut/, and /dUt/. Eight zebra finches were trained to discriminate two of the four syllables, which differed in all formant frequencies following a Go/NoGo paradigm. One



**Fig. 2** Stimuli. This figure shows a scatter plot of the first (F1) and second (F2) formant frequencies in Hertz of all 4 tokens used per word. /dit1/ and /dut1/, for example, have the same F1 but differ in F2, whereas /dit1/ and /dIt1/ have the same F2 but differ in F1. /dit1/ and /dUt1/ neither overlap in F1 nor in F2

syllable was associated to positive feedback, the other to negative feedback (see Table 2). After subjects had learned to reliably discriminate between the two syllables, the remaining two syllables were introduced as probe sounds. Probe sounds were never rewarded nor punished and either had the same F1 frequency as the positive stimulus and the same F2 and F3 frequencies as the negative stimulus or the other way around (see Fig. 2; Table 1). The responses of the birds to the probe sounds allowed us to draw conclusions about how these sounds were perceived by the subjects. Motivated by the clear categorization pattern found in the birds, we repeated the experiment with human listeners ( $n = 39$ , average 24.28 years, ranging from 19 to 34 years) using the same testing paradigm and stimuli to compare vowel formant cue weighting across species.

## Methods

### Stimuli

We used the software PRAAT (Boersma 2001) version 4.6.09 freely available at [www.praat.org](http://www.praat.org) to generate four

synthetic tokens of the syllables “deet” (/dit/), “dit” (/dIt/), “doot” (/dut/), and “dut” (/dUt/) based on an earlier study (Curtin et al. 2009). F1 and F2 values of these tokens are shown in Table 1. In order to compare the use of F1 and F2 differences in vowel perception, the tokens for the contrasts /dit/–/dIt/ and /dut/–/dUt/ differed in their F1 values, while the tokens for the contrasts /dit/–/dut/ and /dIt/–/dUt/ differed in their F2 and F3 values. In terms of F1, the fourth token of each syllable, namely /dit/4, /dIt/4, and /dut/4, had values that fell within one standard deviation of those reported earlier (Curtin et al. 2009), while those for /dUt/4 were identical to /dIt/4 for F1 and to /dut/4 for F2. Tokens 1–4 were generated in order to examine whether variation in F1 and F2/F3 values would lead to a different pattern in the use of these dimensions. Listeners heard only one set of tokens, e.g., /dit/1, /dIt/1, /dut/1, and /dUt/1. All synthesized vowel tokens were spliced in the middle of the same natural d\_t frame that was taken from a naturally produced /dut/ token to avoid anticipatory co-articulation, which refers to influences preceding segments can have on the target phoneme (Fowler and Saltzman 1993). The vowels had the same fundamental frequency (F0) and duration. They had a falling F0 contour that started at 350 Hz at the vowel onset and fell down to 250 Hz at the vowel offset, with both values being similar to those of a natural female voice (Curtin et al. 2009). The vowels had the same duration, namely 250 ms, in order for listeners to only use vowel formant differences when discriminating the vowels in the stimuli. The vowels also differed in their F3 values because, in English and Dutch, vowels with low F2 values, namely back or central vowels, are always produced with a low F3 value, which gives them their characteristic “rounding” feature. Thus, the script that was used to synthesize the vowels computed F3 values following the formula:  $F3 = F2 + 1,000 \text{ Hz}$  for /i/ and /I/ and the formula  $F3 = F2 + 400 \text{ Hz}$  for /u/ and /U/.

### Zebra finch testing

An extensive description of the testing procedure can be found elsewhere (Ohms et al. 2010a). Briefly, eight zebra finches were trained in a Go/NoGo operant conditioning

**Table 1** Formant values in Hertz of the synthesized stimuli

	deet (/dit/)			dit (/dIt/)			doot (/dut/)			dut (/dUt/)		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
T1	220	2,862	3,862	420	2,862	3,862	220	1,736	2,136	420	1,736	2,136
T2	260	2,742	3,742	465	2,742	3,742	260	1,616	2,016	465	1,616	2,016
T3	300	2,622	3,622	510	2,622	3,622	300	1,496	1,896	510	1,496	1,896
T4	340	2,502	3,502	555	2,502	3,502	340	1,376	1,776	555	1,376	1,776

This table gives the frequency values in Hertz of the first three formants of all synthesized stimuli used in this study. *T* token

**Table 2** Testing scheme

	Bird/group	S+	S–	Probes
This table shows which tokens of which stimuli were presented as either negative or positive stimulus to individual birds and groups of human participants. S+, positive stimulus; S–, negative stimulus	729/1	/dit/1	/dUt/1	/dIt/1 and /dut/1
	728/2	/dUt/2	/dit/2	/dIt/2 and /dut/2
	750/3	/dit/3	/dUt/3	/dIt/3 and /dut/3
	763/4	/dUt/4	/dit/4	/dIt/4 and /dut/4
	734/5	/dIt/1	/dut/1	/dit/1 and /dUt/1
	731/6	/dut/2	/dIt/2	/dit/2 and /dUt/2
	758/7	/dIt/3	/dut/3	/dit/3 and /dUt/3
	741/8	/dut/4	/dIt/4	/dit/4 and /dUt/4

chamber to discriminate between two syllables that differed in all formant frequencies from each other whereby every bird got a different set of stimuli (see Table 2). One of the syllables was associated to positive feedback, the other to negative feedback (see Table 2). Each trial was initiated by the birds pecking a report key, which resulted in the playback of either the positive or the negative stimulus. The birds had to peck a response key after hearing the positive stimulus (S+), e.g., /dit/, in order to get a food reward, while ignoring the negative stimulus (S–), e.g., /dUt/. Responding to the negative stimulus caused a 15 s time out in which the light in the experimental chamber went out. Playback of the positive and negative stimulus was randomized with no more than three consecutive positive or negative stimulus presentations. After each bird had reliably learned to discriminate between the two syllables, the remaining two syllables were introduced as probe sounds in 20% of the trials. Responses to probe sounds were never rewarded nor punished, and probes either had the same F1 frequency as the positive stimulus and the same F2 and F3 frequencies as the negative stimulus or the other way around (see Fig. 2; Table 1). The responses of the birds to the probe sounds allowed us to draw conclusions about how these sounds were perceived by the birds. All animal procedures were approved by the animal experimentation committee of Leiden University (DEC number 09058).

#### Human testing

Testing took place in a quiet room using a PC and the software E-Prime version 2.0. Stimuli were presented via headphones (Sennheiser HD595). Participants learned to discriminate between two of the syllables, following the same Go/NoGo procedure applied to the birds (see Table 2). Subjects were randomly allocated to the different test groups (1–8) until five persons per group had passed the training phase and could therefore proceed to the actual testing. The participants were asked to follow the instructions displayed in Dutch on the computer screen until a note appeared that announced the end of the experiment. Furthermore, it was pointed out that during the experiment something might

change, but that they were expected to just continue with the procedure. The Go/NoGo paradigm was not explained beforehand, so that the human subjects, just like the birds, had to figure out the correct procedure completely by themselves. The experiment started with the screen displaying the instruction: “Press ‘Q’ to start the trial”. After a subject pressed the button “Q”, either the positive or negative stimulus was played back, followed by the instruction: “Press ‘P’ after the positive stimulus”. A 2-s interval followed in which the subjects had time to press “P”. Pressing “P” after the positive stimulus resulted in the presentation of a smiley accompanied by a rewarding “ding” sound. Not pressing “P” during these 2 s resulted in the presentation of a sad face accompanied by a punishing “attack” sound. After playback of the negative stimulus, pressing “P” resulted in the presentation of the sad face accompanied by the “attack” sound, whereas not pressing “P” resulted in the presentation of the smiley and the “ding” sound. After this cycle had been completed, a new cycle started, again with the instruction: “Press ‘Q’ to start the trial”, until a total of 10 positive and 10 negative stimulus presentations had taken place. The order of stimulus presentations was random with no more than three positive or negative stimulus playbacks in a row. If a subject had at least 14 correct responses within the first 20 trials (70%), he or she automatically continued to the actual testing phase, which was announced by the note: “You are entering the actual testing procedure now”. If a subject did not reach the 70% correct responses criterion, he or she automatically underwent another training round, which was indicated by the sentence: “Your correct score is too low. You will enter another training round.” If a subject still did not achieve 70% correct responses in this second training, he or she did not continue to the testing phase, and the computer program was terminated with the note: “This is the end of the test. Thank you very much for your participation.” It should be noted at this point that from 52 participants only 40 figured out the Go/NoGo procedure and proceeded to the testing phase. During the testing phase, two probe sounds were presented next to the positive and negative stimulus. Each stimulus was presented 16 times in a random order with no more than three consecutive presentations of the same stimulus, resulting in a total of 64 trials. Contrary to the training phase, no feedback

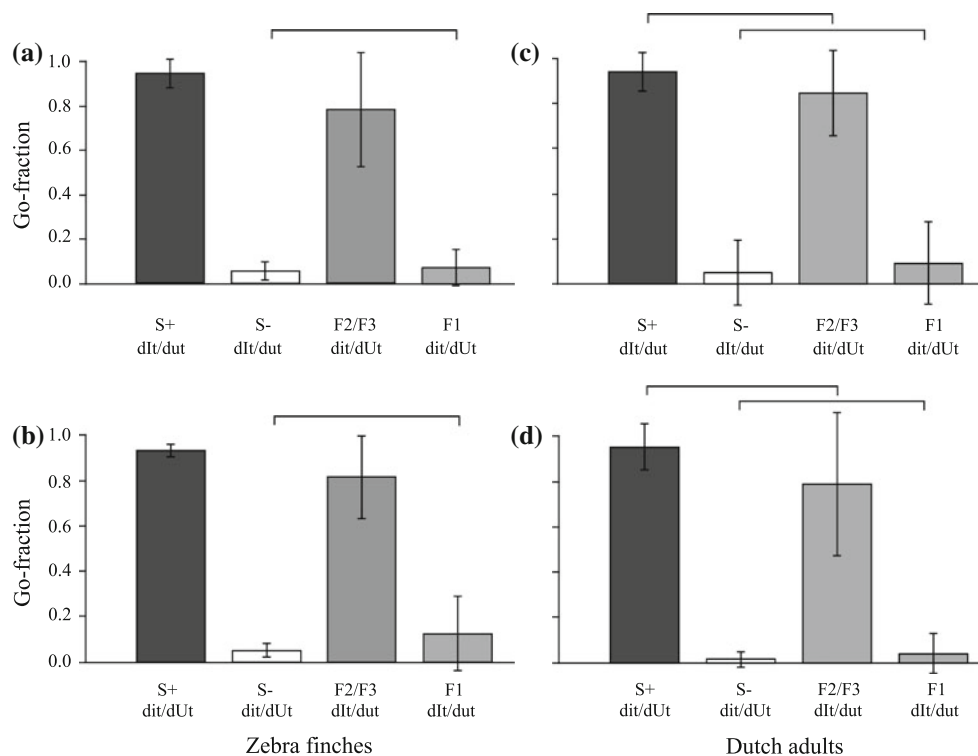
at all was provided in the testing phase. After the 64 trials, a note appeared announcing the end of the experiment and thanking the participants for their participation. The responses to all sounds were automatically saved in E-Prime. The results of one participant of group 7 were not included in the analysis, since this person reported to have forgotten which the original positive and negative stimulus was during the testing phase resulting in an “inverse response”, i.e., during testing, this person responded to the negative but not to the positive stimulus. Informed consent was obtained from the human participants after the nature of the experiment had been explained.

## Results

In the initial training procedure, all birds learned to discriminate between the two syllables that differed in all of the formant frequencies after 2,143 trials on average

( $2,143 \pm 258$  SEM,  $n = 8$ ) following the criterion described earlier (Ohms et al. 2010a).

After this initial discrimination stage, the two non-reinforced probe sounds were introduced in 20% of the stimulus presentations. The response pattern of the birds to these probe sounds compared to the training stimuli is given in Fig. 3. It is clear that the zebra finches utilized F2 and F3 differences to a greater extent than F1 differences, because they categorized stimuli primarily based on differences in F2 and F3 (see Fig. 3a, b) by responding to probe sounds that had the same F2 and F3 frequencies as the positive stimulus, while ignoring probe sounds with the same F2 and F3 frequencies as the negative stimulus. In other words, birds did not weight F1 differences between sounds as strong as F2 and F3 differences, because they responded similarly to stimuli and probe sounds which differed in F1. Thus, if a bird was trained to respond to, e.g., /dit/, it also responded to /dIt/, whereas if it was trained to respond to /dUt/, it also responded to /dut/.



**Fig. 3** Categorization patterns of training stimuli and probe sounds of both zebra finches and Dutch adults. This figure shows the average proportions including standard deviation of Go-responses of birds and humans to training and test stimuli. Every bird got between 50 and 100 probe sound presentations, whereas every human subject got 16 presentations per probe. Horizontal brackets indicate which Go-responses did not differ significantly from each other ( $P < 0.05$ ) analyzed with a simultaneous testing procedure based on G-tests of independence (Sokal and Rolf 1995). **a, c** Go-responses of zebra finches ( $n = 4$ ) and humans ( $n = 19$ ), respectively, that were first trained to discriminate /dit/ and /dUt/ and afterwards got /dit/ and

/dUt/ as probe sounds. F2/F3 beneath the bars indicates the Go-response to the probe sound that had the same F2 and F3 frequencies as the positive stimulus but the same F1 frequency as the negative stimulus, whereas F1 indicates the Go-response to the probe sound that had the same F1 frequency as the positive stimulus but the same F2 and F3 frequencies as the negative stimulus. **b, d** show the same information as panels **a** and **c** but for those birds ( $n = 4$ ) and humans ( $n = 20$ ) that were trained to discriminate /dIt/ and /dUt/ and got /dIt/ and /dut/ as probe sounds. S+, positively reinforced stimulus; S-, negatively reinforced stimulus



Therefore, zebra finches seem to weight frequencies above 1,000 Hz, i.e., those for which their auditory system is more sensitive, stronger (Dooling 2004).

Interestingly, human subjects gave responses that were highly similar compared to those of the zebra finches (see Fig. 3).

Repeated measures ANOVA revealed that the human subjects responded significantly slower to probe sounds compared to training stimuli ( $n = 39$ ,  $F = 7.519$ ,  $P < 0.01$ ) indicating that they did perceive a difference between the sounds but, nevertheless, treated them as tokens of the same category. For the birds on the other hand, no significant difference between reaction times was detected ( $n = 8$ ,  $F = 0.764$ ,  $P = 0.383$ ), although it is highly likely that they also perceived a difference between training and test stimuli, since they responded significantly less to the probe sound than they otherwise treated like the positive stimulus (see Fig 3a, b).

## Discussion

The results of our study are striking as they reveal a hitherto undiscovered parallel in speech perception between humans and birds. Up to now, differences in acoustic cue weighting strategies in speech perception have been attributed to developmental differences between ages (Curtin et al. 2009; Nittrouer 1996; Mayo et al. 2003; Mayo and Turk 2004) and linguistic background (Escudero et al. 2009; Ylinen et al. 2009). These previous studies have usually investigated cue weighting across acoustic dimensions, mostly temporal versus spectral cues, and have used different testing paradigms.

Based on their results, different hypotheses have been formulated to account for cue weighting differences between languages and ages. The “Developmental Weighting Shift” (Nittrouer 1996) hypothesis postulates that children apply optimal perceptual weighting strategies depending on their native language. Along this line, it has been hypothesized that early in life those features that are more salient might generally be weighted strongest in speech perception (Nittrouer and Lowenstein 2009). Several experimental outcomes favor this assumption. For instance, infants growing up in both Swedish and Canadian-English speaking environments prefer F1 differences, which have the most acoustic saliency, over F2 and F3 differences when distinguishing between vowels (Lacerda 1993, 1994; Curtin et al. 2009).

Our study adds a new perspective to acoustic cue weighting differences by including a non-related, but highly vocal, species. The discovery that both zebra finches and adult Dutch listeners exhibit the same cue weighting strategy for vowel perception might be explained by the

fact that both humans and birds show increased sensitivity in higher frequency regions between approximately 1 and 4 kHz. That means, this similarity across species might not be attributed to linguistic background, given that zebra finches obviously lack comparable experience with the Dutch language. The F2/F3 dimensions are in this region, and the absolute differences for the F2/F3 changes are larger than those in the F1 dimension. However, the perceptually more meaningful proportional changes in F2/F3 are in the same range as those of F1. Also, a computer model predicts no bias for any of the dimensions (Curtin et al. 2009), while human adults and zebra finches have a bias towards the F2/F3 dimension. At the same time, this is opposite to the bias found in human infants (Curtin et al. 2009). The outcome of our experiment might, therefore, be interpreted in favor of a general auditory processing hypothesis in mature organisms.

For normally raised adult zebra finches, which lack experience with human speech, the sensitivity matches the region with the most prominent frequency range of their natural songs. Whether this sensitivity arises from their exposure to a rich conspecific acoustic environment consisting, like human speech, of complex broad-band, amplitude- and frequency-modulated sounds (Lachlan et al. 2010) or whether their sensitivity is independent of such an acoustic experience remains an open question. Whatever the causes, our findings demonstrate that acoustic cue weighting underlying vowel perception in humans does not need to be a highly derived feature linked to the evolution of speech.

**Acknowledgments** We thank Dirk Vet from the Institute of Phonetic Sciences at the University of Amsterdam for helping us writing the E-Prime script to test human listeners and all subjects who volunteered to participate in this study. Gabriël J. L. Beckers from the Max-Planck-Institute for Ornithology helped extracting latencies from the data files of the birds. We also thank four anonymous referees for their comments on the manuscript. Funding was provided by the Netherlands Organization for Scientific Research (NWO) (grant number 815.02.011).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Anderson SR (2008) The logical structure of linguistic theory. *Language* 84:795–814
- Boersma P (2001) PRAAT, a system for doing phonetics by computer. *Glott Int* 5:341–345
- Curtin S, Fennell C, Escudero P (2009) Weighting of vowel cues explains patterns of word-object associative learning. *Dev Sci* 12:725–731

- Dooling RJ (2004) Audition: can birds hear everything they sing? In: Marler P, Slabbekoorn H (eds) *Nature's music. The science of birdsong*. Elsevier Academic Press, San Diego, pp 206–225
- Dooling RJ, Brown SD (1990) Speech perception by budgerigars (*Melopsittacus undulatus*): spoken vowels. *Percept Psychophys* 47:568–574
- Dooling RJ, Okanoya K, Brown SD (1989) Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced-voiceless distinction. *Percept Psychophys* 46:65–71
- 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
- Eriksson JL, Villa AEP (2006) Learning of auditory equivalence classes for vowels by rats. *Behav Process* 73:348–359
- Escudero P, Benders T, Lipski SC (2009) Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German and Spanish listeners. *J Phonetics* 37:452–465
- Fitch WT (2010) *The evolution of language*. Cambridge University Press, Cambridge
- Fowler CA, Saltzman E (1993) Coordination and coarticulation in speech production. *Lang Speech* 36:171–195
- Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: what is it, who has it and how did it evolve? *Science* 298:1569–1579
- Hienz RD, Aleszczyk CM, May BJ (1996) Vowel discrimination in cats: acquisition, effects of stimulus level and performance in noise. *J Acoust Soc Am* 99:3656–3668
- Kluender KR, Diehl RL, Killeen PR (1987) Japanese quail can learn phonetic categories. *Science* 237:1195–1197
- Kuhl PK, Miller JD (1975) Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosives. *Science* 190:69–72
- Kuhl PK, Miller JD (1978) Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *J Acoust Soc Am* 63:905–917
- Lacerda F (1993) Sonority contrasts dominate young infants' vowel perception. *J Acoust Soc Am* 93:2372
- Lacerda F (1994) The asymmetric structure of the infant's perceptual vowel space. *J Acoust Soc Am* 95:3016
- Lachlan RF, Peters S, Verhagen SL, ten Cate C (2010) Are there species-universal categories in bird song phonology and syntax? a comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taeniopygia guttata*), and swamp sparrows (*Melospiza georgiana*)? *J Comp Psychol* 124:92–108
- Ladefoged P (2006) *A course in phonetics*. Thomson Wadsworth, Boston
- Lieberman AM (1982) On finding that speech is special. *Am Psychol* 37:148–167
- Lieberman P (1975) *On the origins of language. An introduction to the evolution of human speech*. Macmillan, New York
- Lohr B, Dooling RJ (1998) Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J Comp Psychol* 112:36–47
- Mayo C, Turk A (2004) Adult-child differences in acoustic cue weighting are influenced by segmental context: children are not always perceptually biased toward transitions. *J Acoust Soc Am* 115:3184–3194
- Mayo C, Scobbie JM, Hewlett N, Waters D (2003) The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *J Speech Hear Res* 46:1184–1196
- Nittrouer S (1996) The relation between speech perception and phonemic awareness: evidence from low-SES children and children with chronic OM. *J Speech Hear Res* 39:1059–1070
- Nittrouer S, Lowenstein JH (2009) Does harmonicity explain children's cue weighting of fricative-vowel syllables? *J Acoust Soc Am* 125:1679–1692
- Ohms VR, Gill A, van Heijningen CAA, Beckers GJL, ten Cate C (2010a) Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proc R Soc B* 277:1003–1009
- Ohms VR, Snelderwaard PC, ten Cate C, Beckers GJL (2010b) Vocal tract articulation in zebra finches. *PLoS One* 5:e11923
- Pinker S, Jackendoff R (2005) *The faculty of language: what's special about it?* *Cognition* 95:201–236
- Sokal RR, Rolf FJ (1995) *Biometry: the principles and practice of statistics in biological research*. WH Freeman and Company, New York
- Titze IR (2000) *Principles of voice production*. National Centre for Voice and Speech, Iowa City
- Ylinen S, Uther M, Latvala A, Vepsäläinen S, Iverson P, Akahane-Yamada R, Näätänen R (2009) Training the brain to weight speech cues differently: a study of Finnish second-language users of english. *J Cogn Neurosci* 22:1319–1332