



Universiteit  
Leiden  
The Netherlands

## Speech across species : on the mechanistic fundamentals of vocal production and perception

Ohms, V.R.

### Citation

Ohms, V. R. (2011, May 3). *Speech across species : on the mechanistic fundamentals of vocal production and perception*. Retrieved from <https://hdl.handle.net/1887/17608>

Version: Not Applicable (or Unknown)  
License: [Leiden University Non-exclusive license](#)  
Downloaded from: <https://hdl.handle.net/1887/17608>

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

# 4

## **Zebra finches exhibit speaker-independent phonetic perception of human speech**

Verena R. Ohms, Arike Gill, Caroline A. A. van Heijningen, Gabriël J. L. Beckers & Carel ten Cate

Humans readily distinguish spoken words that closely resemble each other in acoustic structure, irrespective of audible differences between individual voices or sex of the speakers. There is an ongoing debate about whether the ability to form phonetic categories that underlie such distinctions indicates the presence of uniquely evolved, speech-linked perceptual abilities or is based on more general ones shared with other species. We demonstrate that zebra finches (*Taeniopygia guttata*) can discriminate and categorize monosyllabic words that differ in their vowel and transfer this categorization to the same words spoken by novel speakers independent of the sex of the voices. Our analysis indicates that the birds, like humans, use intrinsic and extrinsic speaker normalization to make the categorization. This finding shows that there is no need to invoke special mechanisms, evolved together with language, to explain this feature of speech perception.

*Published in Proceedings of the Royal Society Series B-Biological Sciences (2010) 277: 1003-1009.*

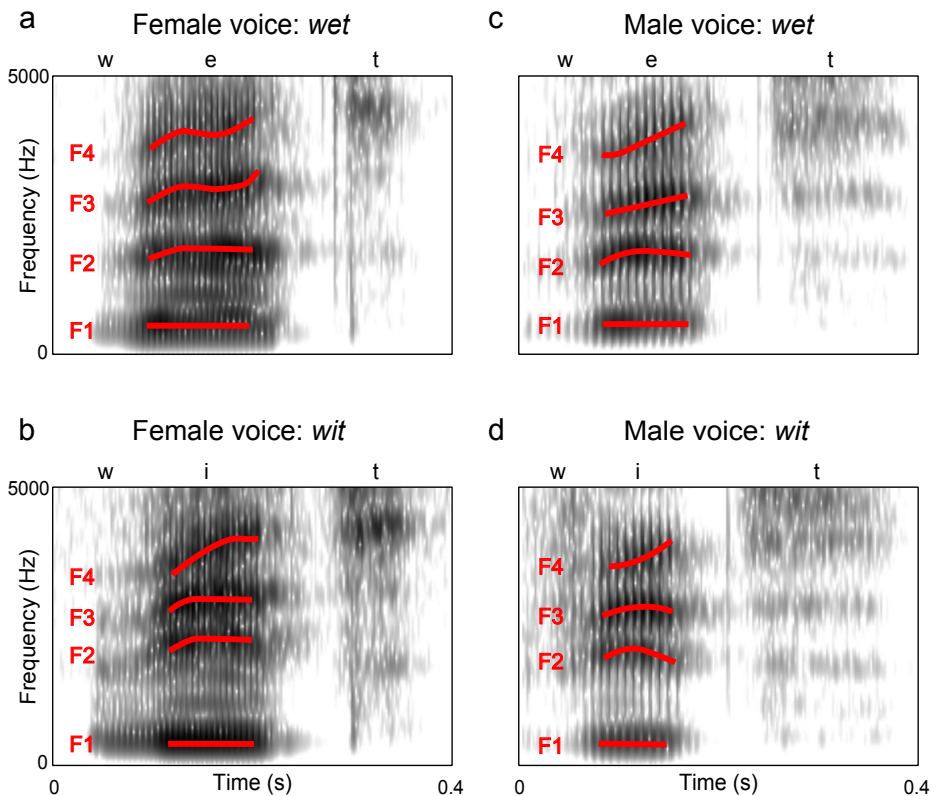
## Introduction

Human speech is a hierarchically organized coding system. A finite number of meaningless sounds, called phonemes, which are classes of speech sounds that are identified as the same sound by native speakers, are combined into an infinite set of larger units: morphemes or words. These larger units carry meaning and therefore allow linguistic communication (Yule 2006). An important role in the coding process is played by formants - vocal tract resonances that can be altered rapidly by changing the geometrical properties of the vocal tract using articulators such as tongue, lips and soft palate (Titze 2000). Changing the formant pattern of an articulation results in a different vowel produced (Fig. 4.1).

It has been argued in the past that many characteristics of speech are uniquely human (e.g. Lieberman 1975, 1984). Therefore it was a revolutionary finding when Kuhl and Miller (1975, 1978) who tested chinchillas on their ability to discriminate between /d/ and /t/ consonant-vowel syllables found that these animals have the same phonetic boundaries as humans and thereby challenged the view that the mechanisms underlying speech perception are uniquely human. A few years later the same phonetic boundary effect has been shown in macaques (Kuhl & Padden 1982). Nevertheless, there is still an ongoing debate about which parameters of human speech production and perception are unique to humans, with the implication that they evolved together with speech or language, and which are shared with other species (Lieberman & Mattingly 1985; Hauser *et al.* 2002; Trout 2003; Diehl *et al.* 2004; Pinker & Jackendoff 2005).

One of the most important phenomena in human speech concerns our ability to recognize words regardless of individual variation across speakers. Although human voices differ in acoustic parameters such as fundamental frequency and spectral distribution, we are able to distinguish closely similar words by using the relative formant frequencies in dependence of the fundamental frequency of an utterance. This feature enables the intelligibility of speech (Nearey 1989; Fitch 2000; Assmann & Nearey 2008). But does this mean that the human ability to perceive and normalize formant frequencies in order to develop an abstract formant percept evolved together with speech and language? Or has the evolution of language exploited a pre-existing perceptual property that allowed formant normalization? An important way to test this question is by examining whether this feature is present in other animals. If so, this suggests that it is not a uniquely evolved faculty.

Here we examined whether zebra finches trained to distinguish two words differing in one vowel only and produced by several same-sex speakers, generalize the distinction to a novel set of speakers of (1) the same sex and (2) the opposite sex. We chose for natural human voices instead of artificial stimuli to confront the animals with a situation humans have to deal with every day when vocally communicating: extracting the relevant sound features from irrelevant ones while listening and building up a percept that allows categorization of these words when originating from novel voices.



**Figure 4.1. Spectrograms of human voices**

(a) Female voice saying *wet*; (b) Female voice saying *wit*; (c) Male voice saying *wet*; (d) Male voice saying *wit*. Red lines indicate the formant frequencies. Note the difference in the distance of the first and second formant frequencies. In a this distance is smaller than in b and the same applies for c and d. F1, first formant; F2, second formant; F3, third formant; F4, fourth formant; s, seconds; Hz, Hertz.

## Material and Methods

### *Subjects*

We used three male and five female zebra finches (*Taeniopygia guttata*, aged 6 months to 2 years) from the Leiden University breeding colony. Prior to the experiment, birds were housed in groups of two or three animals and were kept on a 13.5 L:10.5 D schedule. Food, grit, and water were provided *ad libitum*. None of the birds had previous experience with psychophysical experiments. At the beginning of the study every animal was weighed to allow monitoring of the nutritional state. During the experiment the amount of food eaten by the birds was checked daily. If an animal ate less than necessary it was provided with additional food. In this case the bird was also weighed to ensure that it did not lose more than 20% of its initial body weight. All animal procedures were approved by the animal experimentation committee of Leiden University (DEC number 08054).

### *Stimuli*

We obtained naturally spoken Dutch words from second year students at Leiden University. A total of 10 female and 11 male native speakers of Dutch were recorded in the phonetics laboratory of the Faculty of Humanities, Leiden University using a Sennheiser RF Condenser Microphone MKH416T and Adobe Audition 1.5 software with 44.1 kilosamples/s, at a 16 bit resolution. Every speaker was asked to read a list of Dutch words in which the stimuli '*wit*' (wIt) and '*wet*' (wet) were embedded to prevent list-final intonation effects. The recordings were processed afterwards using the software Praat (version 4.6.09) freely available at [www.praat.org](http://www.praat.org) (Boersma, 2001) by cutting out the words *wit* and *wet* and saving both as separate wave files for each voice. To prevent intensity differences between stimuli from playing a role in the discrimination process, the average amplitude of all female and male voices respectively was normalized by using the root mean square of the average acoustic energy and equalizing it. During the experiment all stimuli were played back at approximately 70 dB SPL(A).

### *Apparatus*

The experiment was conducted in a Skinner box described earlier (Verzijden *et al.* 2007), which was placed in a sound attenuated chamber. Sounds were played through a Vifa MG10SD-09-08 broadband loudspeaker at approximately 70 dB SPL(A) attached one meter above the Skinner box. A fluorescent lamp (Lumilux De Luxe Daylight, 1150 lm,

L 18 W/ 965, Oscan, Capelle aan der IJssel, The Netherlands) served as light source and was placed on top of the Skinner box. It was switched on automatically every day from 7:00 a.m. to 8.30 p.m. whereby the light was gradually increasing and decreasing in a 15 minutes time window at the beginning and end of the light cycle respectively.

### *Discrimination learning*

To train the birds to discriminate between acoustic stimuli we used a 'Go/NoGo' operant conditioning procedure (Verzijden *et al.* 2007). The positive ('Go') stimulus ( $S^+$ ) was an average zebra finch song whereas the negative ('NoGo') stimulus ( $S^-$ ) was a pure tone of 2 KHz constructed in Praat (Boersma 2001). During the training the birds had to learn that responding to  $S^+$  would lead to a 10 second food reward with access to a commercial seed mix, whereas responding to  $S^-$  would cause a 15 second punishment interval with the lights in the experimental chamber going out (Fig. A 4.1).

### *Experiment*

The actual experiment consisted of four successive phases. As soon as the birds reached the discrimination criterion ( $d'=1.34$ ) which we defined as a high response rate to the Go-stimulus (75% or more) and a low response rate to the NoGo-stimulus (25% or less) over three consecutive days they were transferred to the next stage. During the first stage of the experiment every bird had to learn to discriminate the words *wit* and *wet* of a single person (stage 1) whereby every bird started with a different voice. Four groups with two birds per group were formed (Fig. A 4.2). Two groups started with female voices and the other two groups with male voices. One of the groups that began the experiment with a female voice received *wit* as positive and *wet* as negative stimulus and vice versa for the other group. The birds that started with the male voices were treated accordingly. After the birds had reached the discrimination criterion they were switched to the next stage (stage 2) in which four new minimal pairs of the same sex as the first voice were added. After reaching the discrimination criterion birds were transferred to stage 3 in which the five voices used in stage 2 were replaced by five new voices of speakers of the same sex. In the final stage of the experiment (stage 4) the birds were confronted with five new voices of the opposite sex. The experiment was finished after the birds again fulfilled the discrimination criterion. To prevent pseudoreplication voices were randomly balanced over the four groups.

### *Performance evaluation*

To assess discrimination performance between *wit* and *wet* we calculated the  $d'$  and 95% confidence interval following the procedure used and described by others (Macmillan & Creelman 2005; Gentner *et al.* 2006) for every bird for the first 100, 200 and 300 trials directly after transition between the different phases. This is a sensitivity measure that subtracts the  $z$  score of the false-alarm rate (F), which is defined as the proportion of responses to a NoGo-stimulus divided by the total number of NoGo-stimulus presentations, from the  $z$  score of the hit rate (H), which is the proportion of responses to a Go-stimulus divided by the total number of Go-stimulus presentations. This measure allows evaluating how well two stimuli are discriminated from each other:  $d' = z(H) - z(F)$ . A  $d'$  of zero indicates no discrimination whereas a lower bound of the 95% confidence interval above zero can be considered to indicate significant discrimination (Macmillan & Creelman 2005; Gentner *et al.* 2006). Moreover this measurement is unaffected by a potential response bias (Macmillan & Creelman 2005).

### *Acoustic measurements*

In order to detect acoustic features that might have enabled distinction between *wit* and *wet* we measured word and vowel duration as well as fundamental frequency and the mean first (F1) and second (F2) formant frequencies of both words obtained by the different speakers using Praat software (Boersma 2001). We ran two-tailed Wilcoxon signed ranks tests separately for male and female voices to detect significant differences of the acoustic characteristics between *wit* and *wet*.

## **Results**

In the first phase of the experiment all birds learned to discriminate reliably between the two words *wit* and *wet* and fulfilled the discrimination criterion after an average of 41 blocks ( $40.72 \pm 3.41$  s.e.m.) with 100 trials per block.

However, this outcome does not imply generalized categorical discrimination as the birds might have learned the individual features of the training stimuli. In order to show that the birds had developed a generalized percept their performance should be independent of individual voices. In the next phase we therefore added four additional minimal pairs recorded by same-sex speakers to the first stimulus pair but maintained

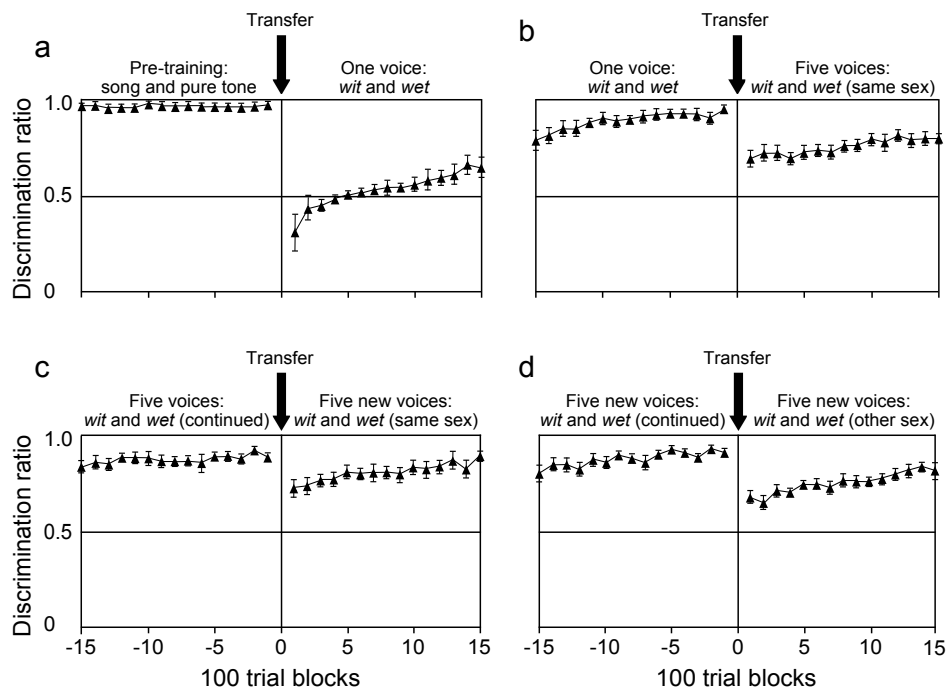
the same learning criterion. The mean  $d'$  (which is a measure of how well two stimuli are discriminated from each other) of the first 100 trial block after this transition was  $0.77 \pm 0.30$  ( $d' \pm$  s.e.m.) which is clearly above chance level ( $d'= 0$ ). After transition of stimulus sets (Fig. 2b) five out of eight birds immediately performed above chance level and all birds achieved a significant performance within the first three blocks after transition (mean  $d'= 0.94 \pm 0.17$  s.e.m. with the lower bound of the 95% confidence interval ranging from 0.14 to 0.94 ).

It could be argued that these results are biased through the incorporation of an already familiar voice in the stimuli sets. Hence, in the subsequent phase we switched to five completely unknown speakers of the same sex (Fig. 4.2c). Again, the average  $d'$  was already highly above chance level over the first 100 trials after transition ( $d'= 1.01 \pm 0.32$  s.e.m.) for six out of eight birds. Within 300 trials after transition all birds showed clear discrimination with a lower bound of the 95% confidence interval ranging from 0.2 to 1.57. Thus, the birds seem to have formed a generalized percept.

So far all voices were of the same sex and overlapped in several features. Therefore a more critical test is to check whether the birds are able to transfer the discrimination to the same words spoken by the opposite sex, i.e. whether the relevant acoustic features can be transferred to a context with larger differences in pitch and timbre compared to voices within the same sex. Consequently, we switched to five new voices of the opposite sex in the last phase of the experiment (Fig. 4.2d). This time all birds discriminated well above chance level (average  $d'= 0.9 \pm 0.59$  s.e.m.) within the first block after transition with the lower bound of the 95% confidence interval ranging from 0.02 to 0.59.

We measured various acoustic characteristics that may have allowed discrimination (Table A 4.1). It is possible that a consistent difference in either vowel or word duration between *wit* and *wet* enabled distinction, but neither vowel nor word duration differed regarding the male voices. There was a significant difference in vowel duration for the female voices (Wilcoxon signed ranks test,  $n = 10$ ,  $T+ = 47$ ,  $T- = 8$ ,  $p = 0.048$ ) with /I/ being shorter than /ε/, but as all birds showed a generally high selectivity irrespective of the sex of the voices it can be assumed that vowel duration was not involved in discrimination. Another cue that might have influenced discrimination is the fundamental frequency of the voices which is known to differ between vowels with /ε/ having a slightly lower fundamental frequency than /I/ (Peterson & Barney 1952). This observation complies with our measurements although the difference is only significant for the male voices (Wilcoxon signed ranks test,  $n = 11$ ,  $T+ = 59.5$ ,  $T- = 6.5$ ,  $p = 0.018$ ). However, the disparity in fundamental frequency between voices is much larger than





**Figure 4.2. Transitions between discrimination stages.**

Displayed is the discrimination ratio of the last fifteen 100 trial blocks before and after a transition between two stages. A discrimination ratio of 1.0 reflects perfect discrimination whereas a discrimination ratio of 0.5 indicates chance performance. The discrimination ratio is calculated as follows:  $(\text{Go S+} / \text{total S+}) / [(\text{Go S+} / \text{total S+}) + (\text{Go S-} / \text{total S-})]$ . **(a)** Transition between the pre-training phase in which all birds had to discriminate a zebra finch song from a 2 kHz pure tone and the training phase in which the animals were confronted with the first minimal pair. **(b)** shows the transition between the training phase and the subsequent experimental stage in which four additional minimal pairs of the same sex were added to the already familiar voice. **(c)** Transition between minimal pairs of now five familiar voices and five completely unknown voices of the same sex. **(d)** Transition from five voices to five new voices of the other sex. kHz, Kilohertz; Go S+, number of responses to a positive stimulus; total S+, number of positive stimulus presentations; Go S-, number of responses to a negative stimulus; total S-, number of negative stimulus presentations.

within voices, so that this feature alone cannot be sufficient for discrimination.

On the other hand we found a highly significant difference in the formant frequencies of the first (F1) and second (F2) formant between *wit* and *wet* as expected (Fig. 4.3a, Table A 4.1 and Table A 4.2). However, if the birds had only paid attention to the absolute frequency of F1 they should have treated the female *wit* as the male *wet* due to the overlap in F1 frequency (Fig. 4.3a, Table A 4.2) whereas in case they based their discrimination on F2 only they should have treated the male *wit* as the female *wet* as these words overlap in F2 frequency (Fig. 4.3a, Table A 4.2).

From phonetic research we know that humans do not discriminate vowels solely based on their absolute formant frequencies, but rather rely on relative formant ratios in dependence of the fundamental frequency (F0) of a speaker (Assmann & Nearey 2008). A common way of illustrating the relationship between formant frequencies and fundamental frequency as a method of intrinsic speaker normalization (Magnuson & Nusbaum 2007) is plotting the difference between F0 and F1 against the difference of F1 and F2 in Bark (Fig. 4.3b), which can be regarded as a two-dimensional perceptual similarity measure of different sounds. Applying this method to our stimuli, results in two clearly separate vowel categories despite an extensive overlap between the sexes (Fig. 4.3b).

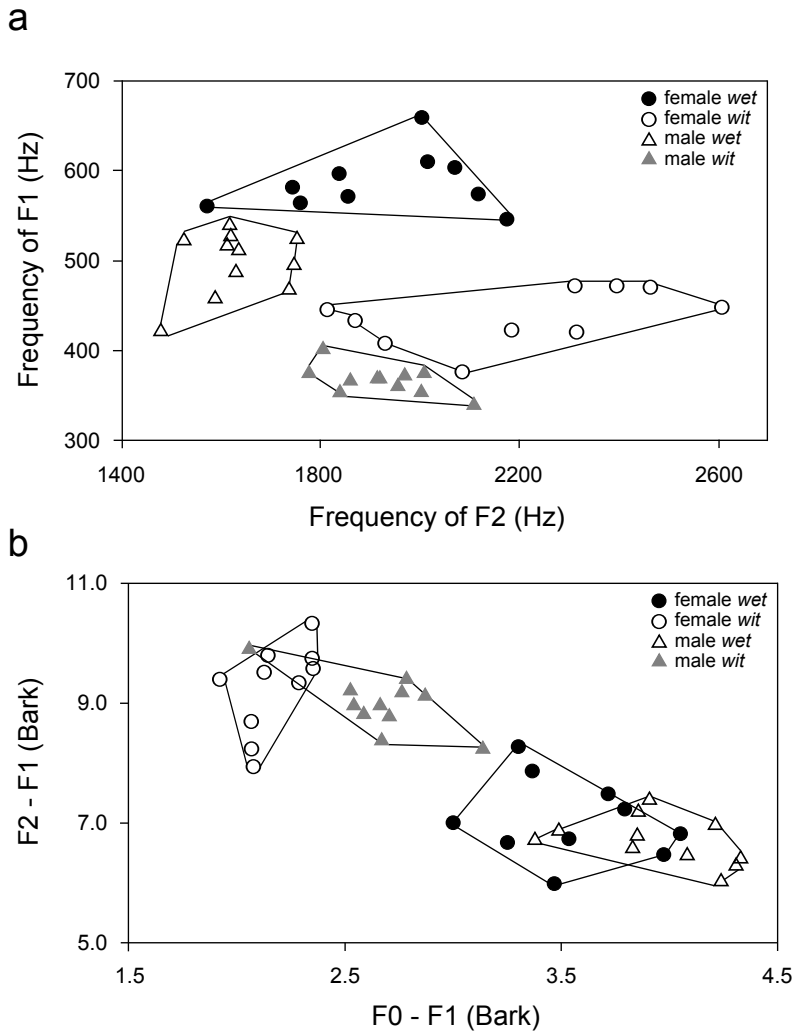
## Discussion

Previous studies on speech perception by non-human animals have suggested that the ability to discriminate human speech sounds based on their formant patterns, such as demonstrated in our study, is not unique to humans, but can be found in other taxa as well. Such studies have been carried out in several mammals, e.g. cats, chinchillas, monkeys and rats (Burdick & Miller 1975; Kuhl & Miller 1975, 1978; Kuhl 1981; Hienz & Brady 1988; Hienz *et al.* 1996; Eriksson & Villa 2006), and birds, such as budgerigars, pigeons, red-winged blackbirds and quail (Hienz *et al.* 1981; Kluender *et al.* 1987; Dooling *et al.* 1989; Dooling & Brown 1990; Dent *et al.* 1997). Most of these experiments used synthesized speech sounds lacking natural variation (Kuhl & Miller 1978; Hienz *et al.* 1981; Kuhl 1981; Hienz & Brady 1988; Dooling *et al.* 1989; Hienz *et al.* 1996; Dent *et al.* 1997; Eriksson & Villa 2006) to demonstrate that the way in which these were discriminated and categorized is equivalent to how humans do so. However, in order to show that animals do use the same mechanisms as humans do when categorizing speech

sounds it is crucial to work with natural and varying stimuli, which has been done only in a minority of studies (Burdick & Miller 1975; Kuhl & Miller 1975; Kluender *et al.* 1987; Dooling & Brown 1990). However, these studies either used isolated vowels or speech sounds from a small number of speakers. While definitely instructive none of these studies fulfilled the requirements of testing a phonemic contrast by employing different vowels embedded in a minimal pair of words. This might seem to be a minor detail when studying speech perception by animals, but yet is essential as humans do not simply make one-bit discriminations between single phonemes (Pinker & Jackendoff 2005), but have to extract relevant information from words that closely match each other in acoustic structure in other respects. Furthermore it is indispensable to use sufficiently different speakers (Magnuson & Nusbaum 2007).

Our experiment controlled for the above mentioned factors and our results strongly suggest that zebra finches use formants to make phonetically relevant discriminations and, similar to humans, abstract away from irrelevant variation between voices.

For humans, 'intrinsic normalization' theories (Nearey 1989) account for the phenomenon that sounds which are perceived as one phoneme can have several acoustic realizations (Liberman *et al.* 1967) by constituting that every speech sample can be categorized using a normalizing transformation. Our analyses indicate that zebra finches use a similar mechanism. However, these theories cannot explain the learning process also revealed by our data. Although the birds were able to immediately categorize *wit* and *wet* independent of speaker variability their performance dropped when confronted with new voices and then improved constantly (Fig. 4.2). Experiments with humans have also shown a clear speaker effect on speech discrimination. In a study by Magnuson and Nusbaum (2007) human subjects were presented with orthographic forms of a target vowel on a computer screen and asked to press the space bar when they heard the target vowel that they saw on the screen. Every subject had to do this task under different conditions, namely 'blocked-talker' condition, which means that all stimuli were from the same talker and 'mixed-talker' condition, which means that the stimuli were from two different talkers. In most cases the response time was significantly higher in the 'mixed-talker' condition compared to the 'blocked-talker' condition while the hit rate was significantly lower. The same speaker effect has been demonstrated by other studies in which the human ability to recognize whole words under varying conditions has been tested (Creelman 1957; Mullennix *et al.* 1989). In addition, also human subjects improve their discrimination performance over trial blocks (Mullennix *et al.* 1989) just as the



**Figure 4.3. Vowel diagrams.**

(a) The frequencies of the first and second formants of all individual voices saying *wit* and *wet* are plotted against each other. Especially with regard towards the second formant frequencies the male voices form denser clusters than the female voices, which show more variation. Nevertheless, the vowels /*l*/ and /*ɛ*/ can be clearly separated from each other. (b) The difference between the fundamental frequency and the first formant (in Bark) is plotted against the difference between the first and the second formant (in Bark) for all recordings used in the experiment. In contrast to the formant scatter plot in (a) this figure represents a two-dimensional perceptual concept in which male and female voices clearly overlap, whereas the two vowels /*l*/ and /*ɛ*/ are fully separated. F0, fundamental frequency; F1, first formant; F2, second formant.

zebra finches in the current study. This outcome indicates the presence of extrinsic normalization in humans and zebra finches, i.e. establishing a reference frame from the vowel distribution of the various speakers as a function of learned formant ranges (Magnuson & Nusbaum 2007; Nearey 1989).

So, due to the design and the results of our study our evidence holds out against arguments that in the past allowed doubts about the universality of the auditory mechanisms underlying speech perception. With respect to speaker normalization our experiment therefore provides very strong evidence that non-human animals use the same perceptual principles as humans do when discriminating speech sounds by employing a combination of intrinsic and extrinsic speaker normalization and thereby suggests that the underlying mechanisms originally emerged in a context independent of speech.

It is mainly due to the lowering of the larynx that humans can produce so many distinct speech sounds (Lieberman *et al.* 1969). However, another effect of a lowered larynx is to increase the length of the vocal tract which causes a decrease of formant frequencies. This in turn can be used to exaggerate size, and playback experiments in red deer which possess a lowered larynx too, have shown that stags respond more to roars with lower formant frequencies compared to roars with higher formants (Reby *et al.* 2005). In humans, formant frequencies are used to correctly estimate age (Collins 2000) and they strongly influence the perceived height of a speaker (Smith & Patterson 2005) and hence can serve as indexical cues next to their function of coding linguistic information. Rhesus monkeys use formants in species-specific vocalizations as indexical cues as well (Ghazanfar *et al.* 2007) and although not many studies have investigated similar phenomena in bird vocalizations it has been shown that whooping cranes for example can perceive changes in formant frequencies in their own species calls and exhibit a different response pattern to calls with higher formants compared to lower formants (Fitch & Kelley 2000). These results led to the speculation that formant perception originally emerged in a wide range of species to assess information about the physical characteristics of conspecifics and that human speech has exploited the already existing sensitivity for formant perception (Ghazanfar *et al.* 2007; Fitch 2000).

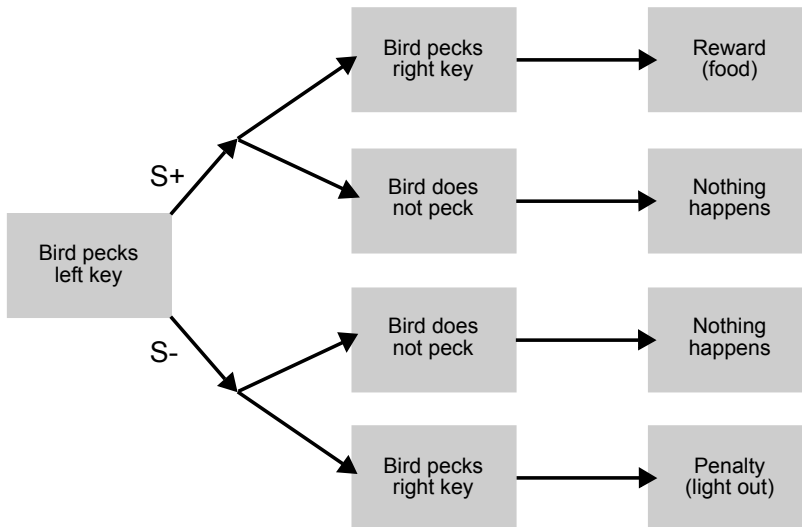
It can, of course, not be ruled out completely that unique perceptual abilities to facilitate speech perception did evolve in humans, or that the observed abilities evolved separately in birds and humans. In the latter case, this would indicate a remarkable convergence. However, our results, in combination with earlier findings, also support the hypothesis that the evolution of the variety of speech sounds in humans might have been

shaped by pre-existing perceptual abilities rather than being the result of co-evolution between the mechanisms underlying the production and perception of speech sounds.

### **Acknowledgements**

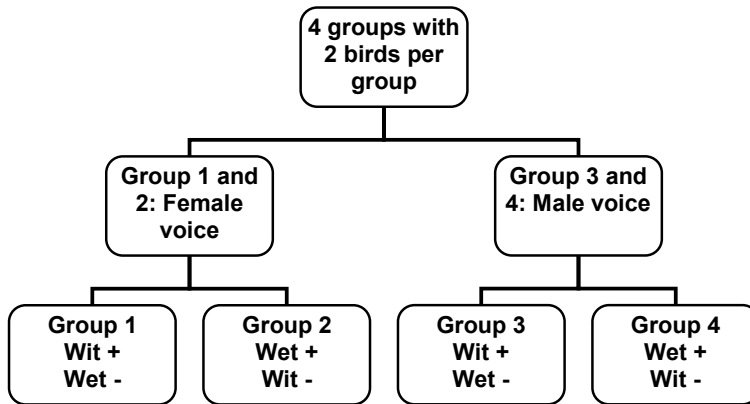
We thank Vincent J. van Heuven for advice considering the recording of the stimuli, for permission to use the phonetics laboratory and him, Katharina Riebel, Hans Slabbekoorn and Willem Zuidema as well as two anonymous referees for useful comments on the manuscript. Funding was provided by the Netherlands Organisation for Scientific Research (NWO). Grant number 815.02.011.

## Appendix



**Figure A 4.1. Schematic overview of the Go/No-Go procedure.**

In the Go/NoGo operant conditioning task the bird elicits every trial by pecking on the left pecking key and subsequently hears either a positive or a negative stimulus. If the bird hears a positive stimulus and pecks on the right key it receives a 10 second food reward. On the other hand if the bird responds to a negative stimulus the lights in the experimental chamber will go out for 15 seconds which serves as a punishment. If the bird does respond neither to a positive nor a negative stimulus within 6 seconds after it heard the sound nothing happens and a new trial can be elicited by the bird by pecking the left key again. S+. positive stimulus; S-, negative stimulus.



**Figure A 4.2. Overview over the 4 different treatment groups.**

Groups 1 and 2 started the discrimination experiment with a female voice, whereas groups 3 and 4 started with a male voice. For groups 1 and 3 *wit* was the positive stimulus and *wet* the negative stimulus, for groups 2 and 4 *wet* was the positive and *wit* the negative stimulus.



**Table A 4.1. Results of the statistical voice analysis.**

Male voices	Parameter	<i>n</i>	' <i>wit</i> '	' <i>wet</i> '	<i>T+</i>	<i>T-</i>	<i>p</i>
			<i>x ± s.d.</i>	<i>x ± s.d.</i>			
	Word duration (ms)	11	415 ± 33.5	399 ± 43.4	47	19	0.240
	Vowel duration (ms)	11	117 ± 15.6	121 ± 14.4	21.5	33.5	0.541
	F0 (Hz)	11	123 ± 14.8	115 ± 10.4	59.5	6.5	<b>0.018</b>
	F1 (Hz)	11	369 ± 15.7	500 ± 35.8	66	0	<b>0.001</b>
	F2 (Hz)	11	1923 ± 99.1	1631 ± 87.9	66	0	<b>0.001</b>
	F1/F2 ratio	11	0.193 ± 0.016	0.307 ± 0.023	0	66	<b>0.001</b>
<hr/>							
Female voices	Parameter	<i>n</i>	<i>x ± s.d.</i>	<i>x ± s.d.</i>	<i>T+</i>	<i>T-</i>	<i>p</i>
	Word duration (ms)	10	427 ± 59.8	439 ± 87.0	24	31	0.721
	Vowel duration (ms)	10	109 ± 19.6	114 ± 22.0	47	8	<b>0.048</b>
	F0 (Hz)	10	220 ± 19.8	213 ± 20.1	46	9	0.064
	F1 (Hz)	10	437 ± 31.2	587 ± 32.1	0	55	<b>0.002</b>
	F2 (Hz)	10	2199 ± 267.8	1916 ± 192.1	54	1	<b>0.004</b>
	F1/F2 ratio	10	0.201 ± 0.023	0.309 ± 0.031	0	55	<b>0.002</b>

Wilcoxon signed ranks test (two-tailed) were conducted to calculate differences in various acoustic parameters between the words *wit* and *wet* whereby male and female voices were compared separately. Significant p-values are printed bold. ms, milliseconds; Hz Hertz; F0, fundamental frequency; F1 first formant; F2 second formant.

Table A 4.2. Formant values of all used voices.

Female						Male					
'wit'			'wet'			'wit'			'wet'		
F1	F2	ratio	F1	F2	ratio	F1	F2	ratio	F1	F2	ratio
(Hz)	(Hz)		(Hz)	(Hz)		(Hz)	(Hz)		(Hz)	(Hz)	
470	2466	0.191	659	2006	0.329	341	2108	0.162	497	1748	0.284
422	2186	0.193	596	1839	0.324	363	1955	0.186	528	1620	0.326
448	2608	0.172	546	2176	0.251	369	1860	0.198	519	1612	0.322
376	2086	0.180	582	1744	0.334	356	1839	0.194	489	1629	0.300
472	2397	0.197	609	2017	0.302	374	1969	0.190	526	1753	0.300
471	2314	0.204	574	2119	0.271	356	2003	0.178	470	1737	0.271
408	1929	0.212	561	1573	0.357	403	1804	0.223	525	1524	0.344
420	2316	0.181	603	2070	0.291	377	2008	0.188	424	1478	0.287
445	1816	0.245	571	1858	0.307	370	1912	0.194	542	1616	0.335
433	1870	0.232	565	1759	0.321	371	1916	0.194	514	1634	0.315
						377	1775	0.212	460	1588	0.290

In this table the frequencies of the first and second formant and the values of the first formant divided by the second formant are listed for all individual voices. F1, first formant; F2, second formant; Hz, Hertz.