Cognitive mechanisms underlying speech sound discrimination: a comparative study on humans and zebra finches

Merel A. Burgering[1,2], *Carel ten Cate*[2,3] & *Jean Vroomen*[1]

[1]Department of Cognitive Neuropsychology, Tilburg University, Warandelaan 2, 5000 LE Tilburg, the Netherlands
[2] Institute Biology Leiden (IBL) Leiden University, Sylviusweg 72, 2300 RA Leiden, the Netherlands
[3] Leiden Institute for Brain and Cognition (LIBC), Leiden University, Leiden, the Netherlands
m.a.burgering@tilburguniversity.edu, c.j.tencate@biology.leidenuniv.nl, j.vroomenen@tilburguniversity.edu

Abstract

Speech sound discrimination in different species seems in many ways comparable to that of humans. Yet it is unclear what type of cognitive mechanisms are involved and whether these are shared among species.

To examine this, we trained human adults and birds (zebra finches) to discriminate two pairs of synthetic speech sounds that varied either along one dimension (vowel or sex of the speaker) or along two dimensions (vowel and speaker information needed to be integrated or combined). Subjects were assigned to one of the four stimulus-response mappings. Once training was completed, we tested generalization to new speech sounds that were either more extreme or more ambiguous than the trained sounds. Generalization to new sounds would reflect if they apply a rule or rely on an exemplar-based memory.

Humans learned the one-dimensional mappings faster than the two-dimensional mappings. Zebra finches learned all mappings equally fast, but showed the same tendency as humans. During the test, zebra finches performed in general higher on the trained sounds than on the extreme and ambiguous test-sounds, whereas humans performed higher on the extreme and trained test-sounds than on the ambiguous sounds. Humans had great difficulty with the task that required combining dimensions to form categories. These results demonstrate that birds rely on exemplar-based memory with some evidence for rule learning, whereas humans use a rule if possible.

Index Terms: categorization – information-integration – speech perception – comparative cognition – songbirds – zebra finches – human - XOR

1. Introduction

A variety of animal species can be trained to discriminate human speech sounds and form speech sound categories [1]. A recent study showed that zebra finches maintain discrimination between vowels when words were spoken by new speakers from the same sex or the other sex, which reveals the capability to generalize [2].

However, what type of cognitive mechanisms underlie this discrimination and generalization and whether animals and humans share these mechanisms is yet unclear. Learning to categorize sounds can be achieved via different mechanisms, such as exemplar-based memorization, prototype learning, rule-based learning or information-integration (II) [3].

To examine the cognitive mechanisms underlying auditory categorization, we developed a rule-based stimulus-response (SR) mapping, wherein the subject either had to discriminate the sounds based on the vowel (/i/ vs. /e/) or on the sex (male vs. female) of the speaker (hereafter: speaker). In addition, we developed two-dimensional SR-mappings: an II task and an exclusive-or (XOR) task that required the use of both dimensions to classify the stimuli.

Via a two-alternative forced-choice task with corrective feedback, we first trained birds and Dutch adults to categorize four sounds based on one or two dimension(s). Once training was completed, we tested generalization to new speech sounds from a matrix of sounds based on male-female and /e/-/i/ continua. These sounds were either more extreme, more ambiguous or intermediate between the trained sounds. For rule-based memory, we expected faster learning speed on one-dimensional mappings and generalization to new extreme and intermediate sounds. For exemplar-based memory, we expected no significant differences in learning speed between the various mappings, and similar generalization on ambiguous and extreme test-sounds.

2. Methods

2.1. Subjects & apparatus

Thirty-six adult zebra finches from the Leiden University breeding colony were individually housed in an operant conditioning chamber in a sound-attenuated room. Three horizontally aligned pecking sensors in the back wall of the cage, a fluorescent lamp, a food hatch, and a speaker were connected to an operant conditioning controller that registered all sensor pecks. Pecking the middle sensor elicited a sound. Depending on the sound, the bird had to peck the left or right sensor. A correct response resulted in temporary food access and an incorrect response led to a short period of darkness.

For humans, sixty students from Tilburg University were individually tested in a dimly lit sound-attenuated room. After a sound was presented through headphones, the participant responded by pressing one of two buttons on a response box after which they received immediate corrective feedback.

2.2 Stimulus material

Three stimulus matrices of morphed speech sounds were constructed with Tandem-STRAIGHT, each based on four different natural speech recordings from an earlier study [2] *wet* and *wit* spoken by a male and a female speaker. Sounds were decomposed into f0 trajectory, a time-frequency and an aperiodicity spectrogram, and next female-male continua for *wet* and *wit* were created by manually mapping time-frequency anchors of matching features in the spectrograms of the two sounds. Next, the female-male continua were matched in similar way to create *wet-wit* morphs. Four training-stimuli and twelve test-stimuli, including more extreme, ambiguous and intermediate sounds were used for all experiments.

2.3 Design & procedure

The subjects were randomly assigned to one of the four SR-mappings: based on vowel, speaker, XOR or II. Every task was completed by 15 humans and nine birds.

All subjects were trained to sort four training sounds into two categories (see figures 1 and 2). After performing at >0.75 for three days (birds) or one training-block of 32 trials (humans), the subject was tested on the trained and non-reinforced test-sounds.

2.5 Analyses

Learning speed was defined as the number of training trials (birds) or trainingblocks (humans) required to reach criterion of >0.75 correct.

For the test, the proportions 'correct' for different sound-groups were calculated by taking the average scores of the proportion of responses to a particular sound group on each side of the midline between the differentially reinforced stimuli (e.g. taking the average of the proportion of pecks to 'extreme *wit*' and 'extreme *wet*' for the vowel test). The proportions correct for the trained sounds included non-reinforced trials only.

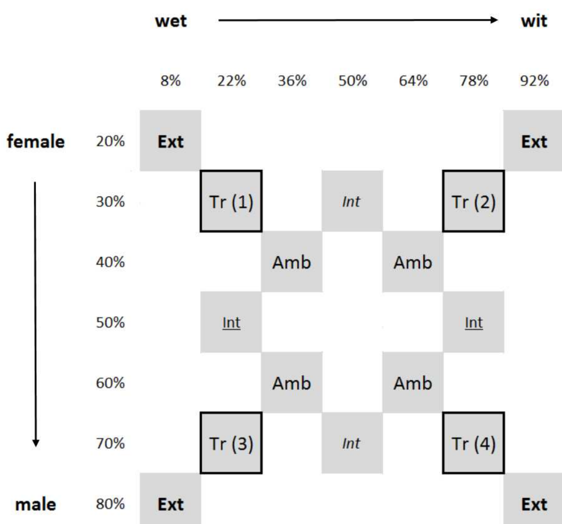

Figure 1: *Subjects were trained to sort four training sounds (Tr1, Tr2, Tr3, Tr4 for the vowel-, speaker- or XOR-task) into two categories. Upon reaching criterion they were tested on the trained and non-reinforced sounds, including intermediate sounds for the vowel (<u>Int</u>) and speaker task (Int). In the vowel task, Tr1 and Tr3 were assigned to one category and Tr2 and Tr4 to the other category. In the speaker task, Tr1 and Tr2 were assigned to one category and Tr3 and Tr4 were assigned to the other category. In the XOR training, Tr1 and Tr4 were assigned to one category and Tr2 and Tr3 to the other category.*

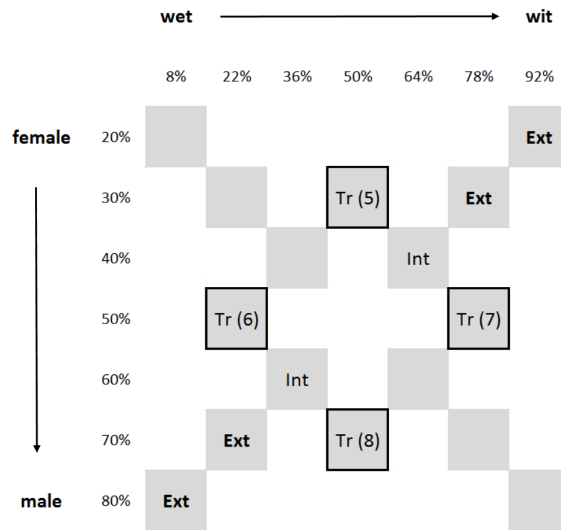

Figure 2: S*ubjects were trained to categorize four training sounds (Tr5, Tr6, Tr7 and Tr8 for the II-task) into two categories. Upon reaching criterion they were tested on trained and non-reinforced sounds. Here, Tr5 and Tr7, were assigned to one category and Tr6 and Tr8, were assigned to the other category.*

3. Results & conclusion

Humans learned the one-dimensional SR-mappings (categorization based on vowel or speaker) faster than the two-dimensional mappings (the II and XOR task). Zebra finches learned all mappings equally fast but showed the same tendency as humans. During the test phase, birds usually performed higher on the trained exemplars than on the extreme and ambiguous test-sounds whereas humans mostly performed higher on the extreme and trained test-sounds than on the ambiguous ones. These results reflect that birds rely more on exemplar-based memory than humans. In the rule-based task based on speaker, birds also show generalization for more extreme and intermediate sounds. Compared to birds, humans showed more generalization in both rule-based tasks. Humans had great difficulty with the XOR task, presumably because they confused the SR-mapping. These results demonstrate that birds rely on exemplar-based memory with weak evidence for rule learning, whereas humans prefer rule-based learning if possible.

4. References

[1] Dooling, R. J., & Brown, S. D. (1990). Speech-perception by budgerigars (Melopsittacus undulatus) - Spoken vowels *Perception & Psychophysics, 47*(6), pp 568-574.

[2] Ohms, V. R., Gill, A., Van Heijningen, C. A. A., Beckers, G. J. L., & ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society B-Biological Sciences, 277*(1684), pp. 1003-1009

[3] Smith, J. D., Zakrzewski, A. C., Johnson, J. M., Valleau, J. C., & Church, B. A. (2016). Categorization: The View from Animal Cognition. *Behavioral Sciences, 6*(2), 24.