



## **Songbirds as Objective Listeners: Zebra Finches (*Taeniopygia guttata*) Can Discriminate Infant-directed Song and Speech in Two Languages**

**Leslie S. Phillmore<sup>1</sup>, Jordan Fisk<sup>1</sup>, Simone Falk<sup>2,3</sup>, and Christine D. Tsang<sup>4</sup>**

*<sup>1</sup>Dalhousie University, Canada*

*<sup>2</sup>Ludwig-Maximilian-Universität, Germany*

*<sup>3</sup>CNRS/Université Paris 3 - Sorbonne Nouvelle, France*

*<sup>4</sup>Huron University College at the University of Western Ontario, Canada*

Despite their acoustic similarities, human infants are able to discriminate between infant-directed song (as produced by human adults) and infant-directed speech in both English and Russian. However, experimenters are somewhat limited in what they can test using the preference paradigm with infants. As a complement to a previous infant study (Tsang, Falk, & Hessel, 2016), we asked whether a songbird, the zebra finch, could discriminate infant-directed song and speech in English and Russian, and tested responses to stimuli that humans could not categorize as either type. Male and female zebra finches learned to discriminate the stimuli in both languages equally well, although females were slightly faster at learning the discrimination, and generalized responses to untrained stimuli of the same categories. Bird responses to stimuli that humans could not categorize likewise did not follow a clear pattern. Our results show that infant-directed song and speech are discriminable as categories by non-humans, that song and speech are as easy to discriminate in both English and Russian, and that comparative studies together can provide more complete answers to research questions about auditory perception than using one species or one language alone.

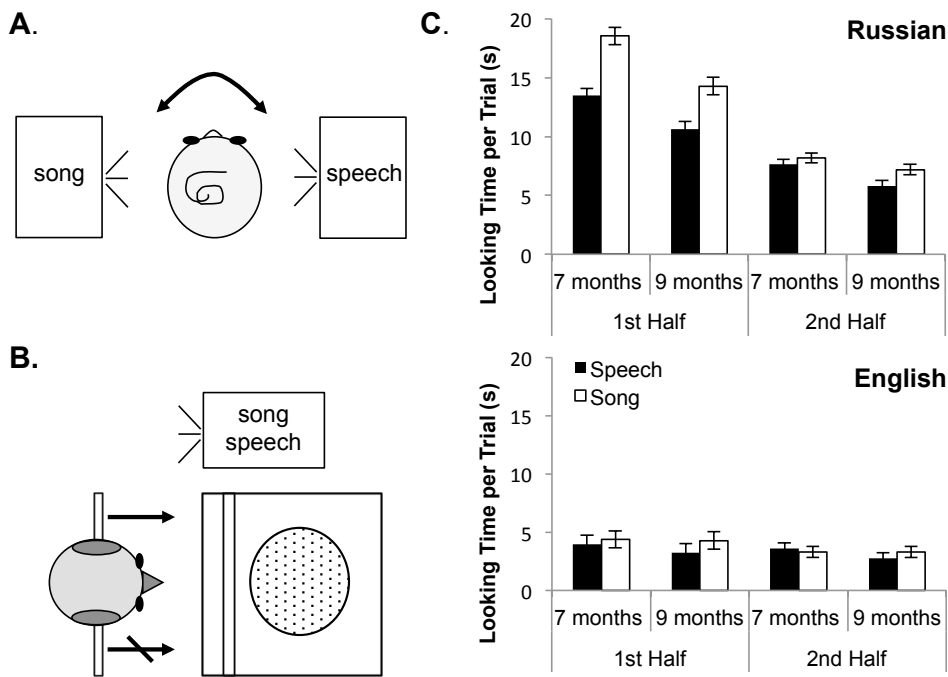
Human caregivers worldwide modify their pitch, intonation, rhythm and tempo when communicating with infants (Fernald, 1989). The specific acoustic features or paralinguistic features that differentiate infant-directed (ID) speech from adult-directed (AD) speech or non-ID speech are well established. ID speech is often characterized by several acoustic features that are related to changes in prosody; compared to non-ID speech, vocalizations directed at infants tend to have higher pitch, slower tempo, a loving voice timbre, shorter and more repetitive utterances, and longer pauses (Fernald & Kuhl, 1987). Several studies have documented infant preferences for ID communication over AD or non-ID communication (e.g. Fernald & Kuhl, 1987; Jusczyk, Frederici, Wessels, Svenkreund, & Jusczyk, 1993; Trainor, Clark, Huntley, & Adams, 1997).

Singing to infants (infant-directed song) is another form of vocal communication typically used in human caregiving contexts and shares with ID speech these same prosodic characteristics (Trainor, 1996; Trainor et al., 1997; Trehub, Trainor, & Unyk, 1993). However, ID speech and ID song do differ in one prosodic feature: pitch contour. Exaggeration and elongation of pitch contours are found in ID speech but not in ID song, likely because songs (as a musical form) have a very specific set of relative pitches that make up the melody, in turn restricting contour variation. These differences in pitch contour possibly contribute to human adult listeners' ability to readily distinguish between singing and speaking vocal registers.

Recently, Tsang, Falk, and Hessel (2016) showed that infant listeners in the second half of the first year of life are able to perceptually categorize between the two forms of infant-directed vocalizations. Using a preference test paradigm, 6- to 9-month-old infants discriminated between playback of ID speech and ID song

exemplars, and were able to do this across two different linguistic contexts—one familiar (English) and one unfamiliar (Russian). Interestingly, they were able to do this without pitch contour available as a cue, as the researchers ensured that pitch contour was matched across all stimulus groups. These findings are consistent with results showing infant listeners fixate attention towards maternal ID singing compared to maternal ID speaking when both audio and visual cues are available (Nakata & Trehub, 2004).

The attentional head-turn preference procedure (see Kemler-Nelson et al., 1995) is a commonly used infant behavioral paradigm that is often used to determine infant discrimination of acoustic sounds: stimuli are played in turn from one of two speakers, and presumably the infant will look longer at the stimuli it prefers (Figure 1A). A difference in the total length of time the infant looks at the two types of stimuli across all trials indicates a difference in preference, and thus discrimination. However, one major limitation of the head-turn preference paradigm is that it is impossible to separate preference and discrimination: the lack of an observed preference does not necessarily mean a lack of discrimination. A further complication in the past research with ID speech and ID song in particular is that we cannot always compare discriminability of ID song and ID speech stimuli independent of familiarity. For example, although infants with native English speaking mothers preferred ID song to ID speech regardless of whether it was in English or Russian, they also preferred the unfamiliar to the familiar language (Figure 1C; from Tsang et al., 2016).



**Figure 1. Experimental paradigms and previous results.** Panel A shows the experimental set up for testing infants in a preference test, measuring how long infants look towards a particular speaker. Panel B shows the experimental set up for testing birds in operant discrimination apparatus, measuring percentage of response to “go” (rewarded) or “no-go” (non-rewarded) stimuli. Panel C is data from Tsang et al. (2016) showing infants prefer ID song (white bars) to ID speech (black bars) regardless of language, and show longer attention to Russian (top panel) than to English stimuli (bottom panel).

To address these limitations, one can use objective listeners to determine discriminability of acoustic stimuli; one possibility is songbirds. Young songbird males must hear their father's song in order to learn to produce it accurately (Doupe & Kuhl, 1999), and both males and females must listen to song of conspecifics to make decisions about territorial defense or mate selection (Searcy & Beecher, 2009; Amy et al., 2008). Therefore, songbirds may be used as an unbiased and expert judge of the discriminability of acoustic stimuli, and an ideal comparator to infant listeners. Like infants, they do not produce the language they are discriminating. However, unlike infants, they are not learning the language they are being asked to discriminate, and thus must be doing the discrimination on the basis of acoustic characteristics as opposed to possible innate language mechanisms or processes that may be at play in infant listeners. In addition, birds may perform thousands of trials compared to fewer than 24 trials in infants (Tsang et al. used a 20 trial preference paradigm), we can measure the rate of acquisition of simultaneous discriminations, and we can test how birds respond to new, untrained stimuli that humans could not categorize as song or speech. In this study, we attempted to use birds as a tool to better understand the underpinnings of infant preferences for ID song and whether songbirds and humans may be using similar cues to perform the discrimination task.

Many studies have examined songbirds' ability to discriminate various types of acoustic stimuli including pure tones (zebra finches, *Taeniopygia guttata*, ten Cate, Spierings, Hubert, & Honing, 2016), vocalizations of their own and other species (black-capped chickadees, *Poecile atricapillus*, Phillmore, Sturdy, Ramsay, & Weisman, 1998; European starlings, *Sturnus vulgaris*, Gentner, Hulse, Bentley, & Ball, 2000), and patterns in acoustic stimuli (Spierings & ten Cate, 2016). In fact, zebra finches are better than humans are at discriminating ranges of pure tones (Weisman, et al., 1998). Zebra finches can also discriminate between stimuli that differ in prosodic features (Spierings & ten Cate, 2014), meaning that prosodic features present in ID speech and song are discriminable to birds. In this study, we used ID song and ID speech stimuli used in the preference paradigm with infants (Tsang et al., 2016) in a study using a go/no-go operant discrimination paradigm with zebra finches (Figure 1B). Birds were required to respond to the rewarded (S+) stimuli and withhold response to non-rewarded (S-) stimuli: for some birds ID speech was rewarded and for others ID song was rewarded. The birds could learn to discriminate both Russian and English stimuli at the same time, allowing us to directly compare whether song and speech in one language was easier to discriminate than the other. We could also test generalization of the discrimination task to new stimuli they had not heard before to ensure birds were using the categories of song and speech (rather than memorizing individual stimuli) to perform the task. Finally, we could present birds with stimuli that humans had difficulty categorizing as song or speech. We could then analyze the response to these "ambiguous" stimuli to determine if birds could categorize them, or if they responded to them as did native speakers or non-native speakers. We used the birds as unbiased acoustic experts to extend our understanding of how ID song and ID speech may be discriminated by human infants and adults.

## Method

### Subjects, Housing and Food

Twelve adult zebra finches obtained from local pet stores (5 males and 7 females) were used in this study. Time between purchase and arrival at the lab was minimized and birds were closely monitored upon arrival for signs of stress. Dalhousie University's University Committee on Laboratory Animals approved all methods and procedures (Protocol # 14-067).

All new birds were housed separately for a one-week period following arrival at the lab in order to monitor for signs of illness. During habituation, birds were housed in individual cages (91 cm wide × 41 cm high × 46 cm deep) made of galvanized steel mesh with a stainless steel pan for collecting waste in an animal room. Birds were allowed free access to finch seed mix, water, cuttlebone and grit, and were maintained on a 12hr light:dark cycle. During the experiment, each bird lived and worked within an

operant chamber in a modified standard cage (37 cm wide × 31 cm high × 24 cm deep) with a plastic grating floor attached to the bottom of the cage to ensure that spilled food was not accessible. Each cage was contained in a ventilated, sound-attenuating enclosure (65 cm wide × 46 cm high × 42 cm deep) lighted with a 9W twin-tube fluorescent bulb, and contained two wooden perches, a water cup and a grit container. The perch in front of the feeder (the “request perch”) was equipped with infrared beams to monitor when the bird landed on it to start a trial. The other, plain perch was at the same height as the request perch but farther away from the feeder; there was approximately 20 cm between the two perches. The feeder also had infrared beams to monitor when the bird entered it (making a “go” choice) and had a motorized food cup which moved up to facilitate access to food and down to restrict access to food (finch mix). Water and grit were freely available.

The experiment was controlled by a microcomputer on each chamber and a desktop computer for each pair of chambers. Stimuli were stored on compact discs and played via disc drives within each desktop computer (one drive for each chamber). Stimuli were played via amplifier (Azur 350a, Cambridge Audio) to a speaker (Fostex FE108E) located inside the chamber beside the feeder and were broadcast at 70 dB (measured 20 cm from the loudspeaker; i.e., the distance between the perch and the loudspeaker). See Sturdy and Weisman (2006) for a complete description of the apparatus.

## Stimuli

The stimuli for this experiment were digital audio files consisting of either a speech or a song fragment in either English or Russian (see Supplementary Material for examples). These were made from stimuli originally used in an experiment with infants chosen to have matching pitch contours between stimulus sets: all the different forms of pitch contour typically found in ID speech (sinus, bell, u-shape) are represented in each of the stimulus categories (speech/song, English/Russian). The specific features used to match the Russian and English samples were the pitch range (in semitones), pitch variability (in semitones), percentage of vowels and tempo (number of syllables per second) using Praat Software Version 5.2.50 (Boersma & Weenink, 2013).

English- and Russian-speaking mothers were recorded singing and speaking to their infants in their first year of life and fragments of these recordings were used as stimuli for the infant study. However, we had to alter the stimuli to use with birds for several reasons. First, in the original study paradigm with infants, it was critical to equate the total amount of time of playback for each type of stimulus (English Song, English Speech, Russian Song, Russian Speech) between the two playback speakers. This meant that for Russian stimuli, because the speech samples ( $m = 2.06$  s,  $s = 0.87$ ) were on average shorter than the song samples ( $M = 2.93$  s,  $SD = 0.62$ ) they used six Russian speech and five Russian song stimuli, and matched them with six English speech and five English song stimuli. For the operant paradigm used with birds, we had to use equal numbers of exemplars for each type of stimulus for the discrimination and leave one for generalization tests. Because songbirds can discriminate the duration of stimuli (e.g., Weisman et al., 1999) and therefore could use it as a potential cue to solve the song and speech discrimination, we had to minimize the differences between categories of stimuli based on stimulus length. We also had to make sure the stimuli were short enough that the birds would sit on the perch and listen to each for their entire duration (between 1.5 and 3 s) before responding. Therefore we had to alter some of the stimuli used in the original experiment for use with birds to ensure minimal possible durations. We ensured as much of the original stimuli (i.e., pitch contour) was conserved as possible, and although they are not the exact stimuli used with infants, the manipulations did not affect the general acoustic differences between the Russian and the English sets (i.e., pitch stability, tempo).

Specifically, two Russian song stimuli were preserved as used in the infant study, two were shortened without altering contours (i.e., remove start and end of file when mother was not singing), and two had altered contours. Four English song stimuli were preserved, one was shortened without alteration, and one shortened with alteration. Three Russian speech stimuli were preserved, two were repeated (within the stimulus file), and one shortened with altered contour. Three English speech stimuli were preserved, and three were repeated. Overall the total number of stimuli that were altered in contour was only 4/24, and we ensured that files of different durations were distributed such that birds could not use duration to learn the discrimination. Finally, we normalized amplitude to ensure a consistent playback volume (approximately 70 dB as measured at the request perch), and applied a ramped on and off envelope to ensure there was no click at stimulus onset or offset.

**Training stimuli.** For Russian excerpts used in the infant study (Falk, 2011a,b), native English speakers listened to them and rated them on a Likert scale ranging from 1 (“Definitely Speech”) to 7 (“Definitely Song”). Stimuli that were rated on average below 1.5 were used as Speech stimuli, and above 6.5 as Song stimuli. For the present study with the birds, we used six of each of these to make up our training and generalization set. Each bird was trained with five of the six Russian song stimuli and five of the six Russian speech stimuli then tested with the remaining song stimulus and speech stimulus. We ensured that each stimulus in the set was used as a generalization test stimulus for at least one bird; most stimuli were used in a generalization test with two birds (except one set was tested with three). For English stimuli used in the infant study, excerpts were chosen from recordings of mothers singing and speaking to their infants in a controlled laboratory setting (S. Trehub, personal communication) and selected to match the acoustic features of the Russian song and speech stimuli: pitch range (in semitones), pitch variability (in semitones), percentage of vowels and tempo

(number of syllables per second). For the present study, we chose six each of song and speech, using five of each for training and one for generalization testing (as we did for the Russian stimuli). In probe testing, we presented eight “ambiguous” Russian stimuli not rated as definitely song or definitely speech by English speakers on the Likert scale: two stimuli rated at approximately 2.5, two at 3.5, two at 4.5, and two at 5.5.

We also presented eight English probe stimuli that were chosen to match the pitch contours of the ambiguous Russian stimuli; mothers were instructed to sing or speak to their infants, but it was possible that their pitch contour did not match singing or speaking per se. In other words, for these stimuli, mothers may have intended to produce infant-directed song, but the pitch contour matched more closely infant-directed speaking (and vice versa).

## Procedure

**Preliminary training.** Each bird was first trained to use the perch and feeder in order to acquire food. Initially, the bird was familiarized with the location of food: the food cup remained in the up position and, at the beginning, additional food was scattered over the top of the food cup. During this stage, a red LED light at the back of the feeder was always on. Next, the bird was trained to activate the food cup and access food by flying into the feeder mechanism only when the red light was on, after which the food cup would become available for decreasing amounts of time. Following that, the bird was trained to sit on the perch across from the feeder mechanism for 1 s in order to activate the red light, at which point it could activate the food cup by entering the mechanism. The next stage was similar, but half the time the activated red light was accompanied by a 1-s 1000-Hz pure tone: the bird was rewarded with food access for responding to the light combined with the tone, and punished with a 30-s inter-trial interval (ITI) with the chamber lights out for responding to the light alone. In the final stage of feeder training, the red light was not used and the bird was rewarded for responding to the tone and punished (by turning lights out) for responding when the tone was absent. Responses were collected in 500-trial bins. Completion of each of the two stages of tone/no-tone discrimination occurred when the bird discriminated the tone from no-tone contexts with a discrimination ratio of 0.90 for a minimum two consecutive bins on one day. Discrimination ratio was calculated by dividing the percentage of response to the rewarded stimulus (tone or tone plus light) to the percentage of response to both rewarded and unrewarded (no tone or no tone, no light) stimuli; a discrimination ratio of 0.50 indicates chance performance and of 1.00 indicates perfect performance.

Finally, the bird was trained to remain on the perch until the end of the stimulus before responding to ensure the entire stimulus was heard before making a response. In this stage, the bird again had to sit on the perch in order to activate a tone, upon which it was rewarded with food for flying to feeder mechanism; unlike previous stages, the bird was punished for leaving its perch before the end of the tone with a 30-s ITI with the chamber lights out. Once the bird learned to sit on the perch reliably for the 1-s tone, we increased the length of the tone first to 2 s then to 3 s, to ensure birds would sit for the longest stimulus in the stimulus pool.

**Nondifferential training.** After learning to use the perch and feeder, each bird was exposed to the 20 training stimuli it would hear during discrimination training (see below), and trained to respond to all of stimuli at a high and uniform rate. As in the preceding stages, the bird was required to sit on its perch for 1 s in order to activate the auditory stimulus. If the bird responded by entering the feeder mechanism during the 1 s following completion of the stimulus, it was rewarded with 1 s of food access. A 30-s ITI followed. If it left the perch but did not enter the feeder mechanism, the trial ended after 1 s. If the bird did not leave the perch, the trial ended after 1 s and a 60-s ITI followed; this was intended to encourage the bird to leave the perch after each stimulus. If the bird left the perch prior to completion of the stimulus, the trial ended and the chamber lights were turned off during a 30-s ITI; this ensured that the bird was listening to the entire auditory stimulus before responding. Stimuli were selected in random order without replacement until all were played, at which point all stimuli were returned to the pool for selection; birds trained continuously for the entire 12-hr lights-on period and typically completed a total of about 1500 trials per day. Each bird remained on nondifferential training until they were responding to all 20 stimuli at a rate of at least 70% (number of times responding to stimulus/number of presentations of that stimulus) and were leaving the perch prior to stimulus offset on less than 10% of trials for a minimum of two consecutive 500-trial bins.

**Discrimination training.** In discrimination training, entering the feeder following an S+ stimulus was rewarded with 1 s of food access. Entering the feeder after an S- stimulus was punished with a 30-s ITI with the chamber lights out, therefore the correct response after hearing an S- was to leave the request perch and then return to start a new trial. The other contingencies that were in place during nondifferential training (e.g. lights out for leaving perch before stimulus offset) also applied during discrimination training.

Each bird had a training set of five out of the six exemplars of each stimulus: English Song, English Speech, Russian Song, and Russian Speech. The remaining exemplar of each was used as probe stimuli to test generalization. One male and one female (with the exception of one female because there were seven) had the same stimulus and probe sets such that each of the six stimuli of each type was tested as a probe stimulus. In addition, for half the birds (four females and two males) song (both English and Russian) was

the S+ stimulus and speech was the S-, and for the others (three females and three males) speech was the S+ and song was the S-. This time discrimination ratios were calculated separately for each language. For example, we divided the number of feeder visits in response to English S+ stimuli (song or speech, depending on group) by the total number of visits in response to both English S+ and English S- stimuli. Discrimination training continued until a bird performed with an overall discrimination ratio of at least 0.80 for three consecutive 500-trial blocks.

**Probe testing.** After discrimination training was complete, birds were given two days of free feed before moving on to the next stage. We then presented birds with two days of modified discrimination training, during which S+ stimuli were rewarded on 85% of correct visits to the feeder rather than 100% in order to reduce the birds ability to discriminate probe stimuli from training stimuli based on how often they were rewarded. The next day, we added the probe stimuli to the stimulus pool. All of the new probe stimuli were rewarded on 15% of visits (on the rest of these trials nothing happened, regardless of a go or no-go response). We did this so that birds could not discriminate the new stimuli based on reward frequency; birds' responses to the probe stimuli were due to transfer of rules learned during original discrimination training rather than new learning during probe testing. In addition, having the probe stimuli rewarded at a low and uniform rate encouraged birds to continue responding to the new songs. This procedure has been used previous in the literature using the same apparatus and design (e.g. Phillmore, Sturdy, & Weisman, 2003; Bloomfield, Farrell, & Sturdy, 2008). The remaining contingencies in place for S+ and S- stimuli during modified discrimination training continued during probe testing. Probe training lasted for the day, but we only used the first 500 trials in our analyses.

In the first probe test (generalization) we added four stimuli to the pool: one novel exemplar each of English Song, English Speech, Russian Song, and Russian Speech; these were the four S+ and S- stimuli not initially included in each particular bird's training set. In the second probe test we added eight ambiguous stimuli: four Russian and four English, and for the last probe we added the remaining four Russian and four English ambiguous stimuli. We alternated probe sessions with one day of modified discrimination training (S+ rewarded at 85%).

## Response Measures and Analyses

In addition to discrimination ratios, we also calculated how many of the five S+ stimuli were being successfully discriminated from the S- stimuli in order to make sure that birds were attending to and learning all S+ stimuli rather than generating a high discrimination ratio by responding to only one or two S+ stimuli. To do this, we calculated the response rate (number of visits to the feeder divided by the total number of times leaving the perch) for each stimulus and then used those numbers to determine how many of the S+ stimuli had a response rate higher than the upper bound of the 95% confidence interval (i.e., mean response rate plus two standard deviations) of the response rates for the S- stimuli.

Both discrimination ratio and the number of S+ stimuli successfully discriminated were calculated for each of the first four bins (2000 trials), as well as the final bin during discrimination training (the third of the three consecutive bins of 500 trials). For generalization probe trials, we compared average percentage of response to the probe stimuli during the first 500-trial bin of the day of probe testing to the average percentage of responses to training stimuli on the last day of discrimination training. For ambiguous stimuli, we looked at percentage of response to those stimuli compared to generalization probe stimuli and to percentage response to training stimuli at the end of discrimination training.

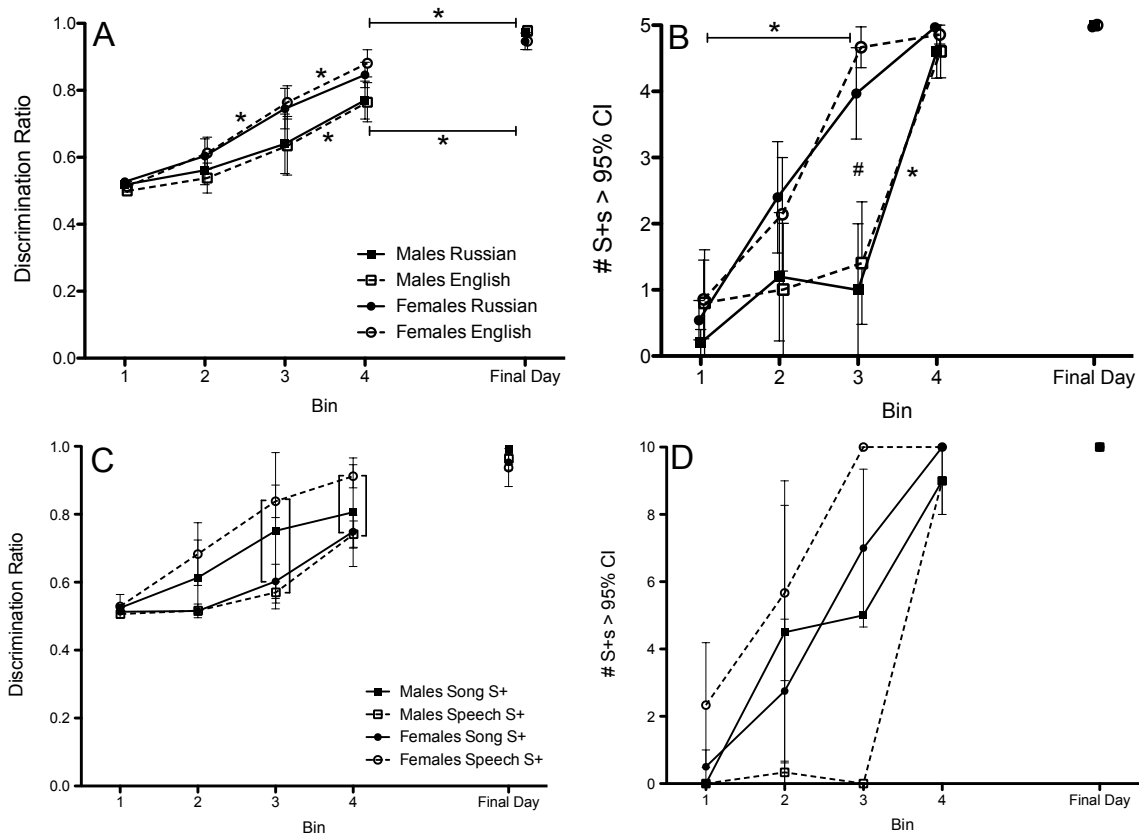
In the case of each measure, we conducted mixed analyses of variance (ANOVAs). All statistical analyses were conducted in IBM SPSS Statistics 21 (IBM).

## Results

### Acquisition: Discrimination Ratio

Figure 2A shows the discrimination ratios over the first four bins (2000 trials) and on the final day of training for male and female zebra finches. A mixed sex  $\times$  language  $\times$  bin ANOVA on discrimination ratios (top panel) revealed no main effect of sex or language, all  $F_s < 1$ , but a main effect of bin,  $F(4, 40) = 69.97$ ,  $p < 0.001$ . Post hoc tests showed that overall, discrimination ratios increased significantly from bin 2 to bin 3, from bin 3 to bin 4, and from bin 4 to the final day of training. We also found a sex  $\times$  language interaction,

$F(1, 10) = 6.09, p = 0.03$ , however the sex  $\times$  bin,  $F(4, 40) = 1.95, p = 0.12$ , language  $\times$  bin,  $F(4, 40) = 1.78, p = 0.15$ , and the sex  $\times$  language  $\times$  bin,  $F(4, 40) = 1.15, p = 0.35$ , interactions were not significant.



**Figure 2. Discrimination ratios and number of stimuli discriminated for English and Russian stimuli.** Panels A and B show the acquisition data plotted for male and female zebra finches separated by language; with discrimination ratio on the left (a) and the number of S+ stimuli learned on the right (b). Panels C and D show the acquisition data plotted for male and female zebra finches separated by whether song or speech was the rewarded stimulus, with discrimination ratio on the left (c) and the number of S+ stimuli on the right (d). Error bars represent standard errors of the means. Asterisks represent differences at  $p < 0.05$  between connected points, # represents a difference between groups.

To further analyze the discrimination ratios over acquisition, we looked at males and females separately using a repeated measures language  $\times$  bin ANOVA for each. For males, there was a main effect of language,  $F(1, 4) = 10.33, p = 0.03$ , where overall, Russian had a higher discrimination ratio than English. There was a main effect of bin,  $F(4, 16) = 25.86, p < 0.001$ , where post-hoc tests indicated DRs increased from bin 3 to 4 and from 4 to the final day. For females, there was a different result: there was no main effect of language,  $F(1, 6) = 1.09, p = 0.34$ , but a main effect of bin,  $F(4, 24) = 49.17, p < 0.001$ , where post-hoc tests indicated DRs increased from bin 2 to 3, 3 to 4, and 4 to the final day. There was no language  $\times$  bin interaction for either sex.

### Acquisition: Number of S+s learned

Figure 2B also shows the number of S+s successfully discriminated over the first four bins and on the final day of training for males and females. A mixed sex  $\times$  language  $\times$  bin ANOVA on the number of CIs successfully discriminated revealed a main effect of sex,  $F(1, 9) = 5.33, p = 0.046$ , no main effect of language,  $F(1, 9) = 1.18, p = 0.31$ , and a main effect of bin,  $F(4, 36) = 30.10, p < 0.001$ . There was also a sex  $\times$  bin interaction,  $F(4, 36) = 4.53, p = 0.005$ , but no other interaction was significant. To further analyze the sex  $\times$  bin interaction, we performed language  $\times$  bin ANOVA separately for each sex. For both sexes, there was no main effect of language,  $F < 1$ , or a language by bin interaction, males,  $F < 1$  and females,  $F(4, 20) = 1.51, p = 0.24$ . For males, post-hoc tests on the significant main effect of bin,  $F(4, 16) = 19.73, p < 0.001$ , revealed the number of S+s discriminated did not change significantly from bin 1 to bin 3, increased from bin 3 to bin 4, and then did not change significantly from bin 4 to the final bin. For females, post-hoc tests on the significant main effect of bin,  $F(4, 20) = 14.38, p < 0.001$ , revealed the number of S+s discriminated increased significantly from bin 1 to bin 3, and then did not change to bin 4 or to the final day. In addition, comparisons of the number of S+s learned between males and females, on each bin revealed females had learned significantly more Russian S+s,  $t(9) = -3.57, p = 0.006$ , and English S+s,  $t(10) = -2.56, p = 0.03$ , than males at bin 3, but there were no sex differences between males and females at any other bin.

### Song vs. Speech as S+

Figure 2C shows discrimination ratios over acquisition, this time plotted separately for males and females depending on which stimulus (song or speech) was the S+. We did not expect differences between S+ groups, however, a mixed sex  $\times$  S+ group  $\times$  bin ANOVA revealed a sex  $\times$  S+ group  $\times$  bin interaction,  $F(4, 32) = 6.697, p < 0.001$ , and a main effect of bin,  $F(4, 32) = 97.51, p < 0.001$ . There was no main effect of sex or S+ group, nor a sex  $\times$  bin or S+ group  $\times$  bin interaction,  $F_s < 1$ . To further analyze the three-way interaction, we first examined sexes separately with S+ groups  $\times$  bin ANOVAs. For males, there was a main effect of bin,  $F(4, 12) = 33.73, p < 0.001$ , but no main effect of S+ group,  $F(1, 3) = 2.79, p = 0.19$ , nor an interaction,  $F(4, 12) = 1.60, p = 0.24$ , meaning that there was no difference in acquisition for males that had song or speech as the S+. In contrast, for females there was no main effect of S+ group,  $F(1, 5) = 5.45, p = 0.07$ , a main effect of bin,  $F(4, 20) = 68.48, p < 0.001$ , and an S+ group  $\times$  bin interaction,  $F(4, 20) = 6.79, p < 0.001$ . T-tests showed that females with speech as S+ had higher discrimination ratios at bin 3,  $t(5) = -3.35, p = 0.02$ , and bin 4,  $t(5) = -2.62, p = 0.047$ , than females with song as S+, but there was no difference at bins 1, 2, and the final bin. In addition, when we compared males and females within S+ groups using sex  $\times$  bin ANOVAs, there were no main effects of sex, song S+:  $F(1, 4) = 2.84, p = 0.17$ , speech S+:  $F(1, 4) = 5.26, p = 0.08$ , and main effects of bin, song S+:  $F(4, 16) = 33.82, p < 0.001$ , speech S+:  $F(4, 16) = 89.18, p < 0.001$ . There was no sex  $\times$  bin interaction,  $F(4, 16) = 1.22, p = 0.34$ , for the song S+ group, meaning males and females learned the song S+ at the same rate, but there was a significant difference for the speech S+,  $F(4, 16) = 10.66, p < 0.001$ . T-tests showed that for the speech S+ discrimination, females had higher discrimination ratios than males at bin 3,  $t(4) = -4.70, p = 0.009$ , and bin 4,  $t(4) = -3.29, p = 0.03$ , but no other bin, including the final day. A sex  $\times$  S+ group  $\times$  bin ANOVA on the number S+s discriminated successfully revealed only a main effect of bin,  $F(4, 32) = 24.85, p < 0.001$ ; see Figure 2D).

## Generalization Test

We used the responses from the first 500 trials of the generalization probe and calculated percentage of response to training and probe stimuli over these trials (Figure 3). One bird only responded on 5% of trials in the first 500, so we did not include its data in the analyses. We performed repeated measures stimulus type (training vs. probe)  $\times$  reward (S+ vs. S-) ANOVAs separately for each language. For both languages, there was a main effect of stimulus type, English:  $F(1, 10) = 81.69, p < 0.001$ , Russian:  $F(1, 10) = 69.47, p < 0.001$ , and reward, English:  $F(1, 10) = 296.20, p < 0.001$ , Russian:  $F(1, 10) = 195.48, p < 0.001$ , and significant stimulus type  $\times$  reward interactions, English:  $F(1, 10) = 64.09, p < 0.001$ ; Russian:  $F(1, 10) = 193.89, p < 0.001$ . To further examine the interactions, we then performed repeated measures t-tests within each language. For the English stimuli, percentage response to training S+ was significantly higher than to training S-,  $t(10) = 45.18, p < 0.001$ , and to probe S+,  $t(10) = 8.75, p < 0.001$ . Response to probe S+ was higher than to probe S-,  $t(10) = 2.93, p = 0.02$ , and training and probe S-s were not different,  $t(10) = -1.22, p = 0.25$ . For Russian stimuli, percentage response to training S+ was significantly higher than to training S-,  $t(10) = 31.34, p < 0.001$ , and to probe S+,  $t(10) = 18.68, p < 0.001$ . Probe S+ was not significantly higher than probe S-,  $t(10) = 2.26, p = 0.052$ , and training and probe S-s were not different,  $t(10) = -1.29, p = 0.23$ .

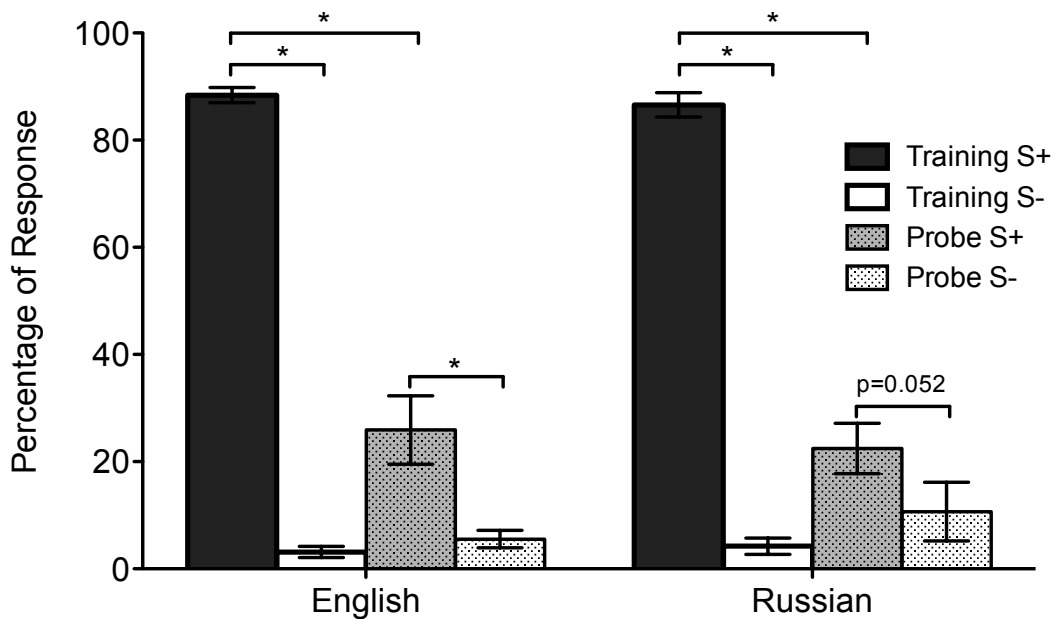


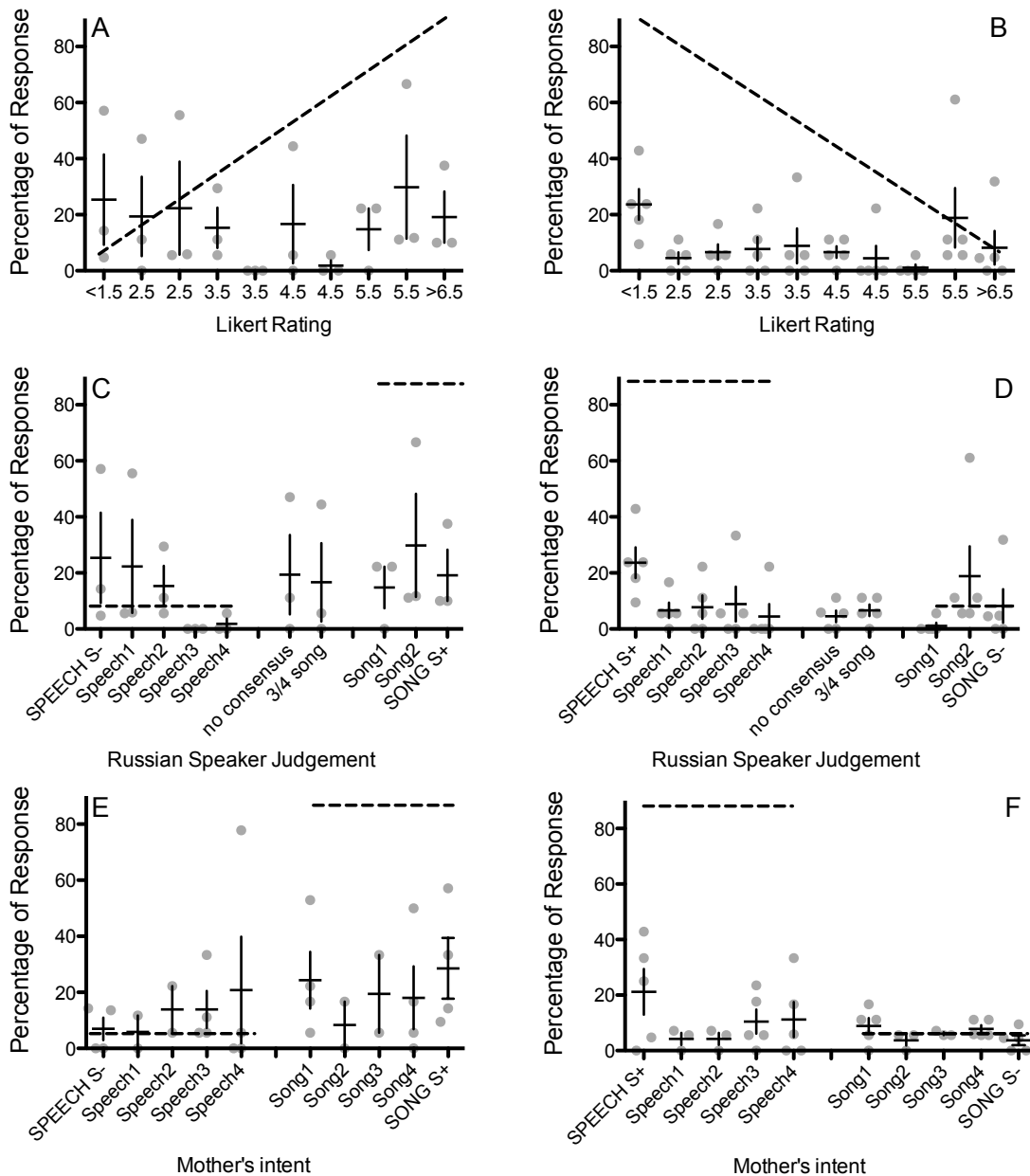
Figure 3. Percentage of response to training stimuli and generalization stimuli separated by language. Darker bars show responses to training stimuli (S+ and S-) and lighter bars show responses to generalization probe stimuli the same categories as the training stimuli but not previously heard.

## Ambiguous Probe Tests

We presented four stimuli of each language in each of two probe tests. However, we excluded data from a set of stimuli from a probe test if a bird did not respond to probe stimuli at all or responded to one probe stimulus at a low rate ( $< 5.5\%$ ). This meant we excluded data for two birds from the first probe test, and for six birds from the second probe test (four birds had song as the S+ and two birds had speech as S+). There was a large amount of individual variability in response to the stimuli, so we plotted individual data in Figure 4 and evaluated responses qualitatively.

We first plotted responses to Russian ambiguous stimuli according to the Likert scale ratings by native English speakers (Figure 4), separately for birds with song as S+ (Panel A) and speech as S+ (Panel B). We also plotted for comparison responses to generalization probes and expected percentages of response if the birds responded as if on a continuum from lowest (similar to S-) to highest (similar to S+). The fit of the bird data to this model of expected responses is poor. We then plotted the same responses to the same Russian ambiguous stimuli according to how native Russian speakers categorized the Russian ambiguous stimuli (Figure 4, Panels C and D). Four native Russian speakers were asked to listen to each of the probe stimuli presented to the birds and decide whether the person they heard was singing or speaking, or if they could not tell. Stimuli where there was 100% agreement among the raters was called “song” ( $n = 2$ ) or “speech” ( $n = 4$ ); one stimulus was categorized as song by three out of the four raters (3/4 song) and for the last stimulus (“no consensus”) two raters indicated the mother was singing, one indicated the mother was speaking, and one could not determine what the mother was doing. Again, the birds’ responses did not respond as if they were consistently categorizing stimuli similarly to the way the native Russian speakers did. Although some birds responded to some stimuli as if they were S+ or S-, there was no consistency within a particular bird or to a particular stimulus. For example, in Figure 4C, Speech 1 should be treated like an S-, but one bird responds as if it is an S+, and Song 2 should be treated like an S+, and two birds treat it like an S-. Another example in Figure 4C and D: the Song S+ birds respond to Speech 3 and 4 as if they were S-, as expected, however, only one Speech S+ bird responded as if they were S+.

We did not have Likert ratings for English stimuli (because the stimuli were confounded by the language content—English-speaking raters had significant difficulty separating the melody from the words and so the ratings obtained were not reliable or valid). We instead plotted responses to English probe stimuli according to what the mother was asked to produce in the original recording (Figure 4 Panels E and F). All of the intended songs were responded to as if they were S- by all of the birds that were trained with song as S- (Panel F). However, not all the intended speech stimuli were responded to as if they were S+ (see Speech 1, 2 in Panel F). Responses by Song S+ birds were more variable.



**Figure 4. Average (lines) and individual (dots) responses to the Russian and English probe stimuli plotted separately for birds that had Song stimuli as S+ (left column) and birds that had Speech stimuli as S+ (right column).** In Panels A and B, responses to Russian ambiguous stimuli are plotted according to the Likert scale ratings of each individual stimulus; responses to generalization probe stimuli (1.5 = speech, 6.5 = song) are also plotted. The dashed line starts and ends at the average percentage of response to the song and speech training stimuli on the last day of training. In Panels C and D, responses to Russian ambiguous stimuli are plotted according to how native Russian speakers categorized the ambiguous stimuli with response to stimuli in the generalization probe shown for comparison. Dashed lines show average percentage of response to Russian song and speech on the last day of training. In Panels E and F, responses to English ambiguous stimuli are plotted according to what mothers were asked to produce (song or speech) with response to generalization probe song and speech shown for comparison. Dashed lines show average percentage of response to English song and speech on the last day of training.

## Discussion

In this study we asked whether, like infants, zebra finches could discriminate between infant-directed song and infant-directed speech in both English and Russian. Overall, zebra finches learned to discriminate all five English and Russian ID song stimuli from all five English and Russian ID speech stimuli equally well, and generalized this learning to one novel stimulus of each type. We did find a minor sex difference: after 1500 trials females had learned between four and five of each of the S+ types while males had only learned one. However, 500 trials later males had also learned all five stimuli. Unexpectedly, ease of learning differed depending on whether song or speech was the rewarded stimulus: males and females learned at the same rate when song was the S+, but females learned faster than males when speech was the S+.

### Discrimination by Songbirds vs. Infants

Both songbirds and human infant listeners are able to discriminate between ID song and ID speech stimuli in both English and Russian. In a standard head-turn preference paradigm (Tsang et al., 2016), infants looked longer to ID song than to ID speech, indicating that infants can discriminate between the two categories of stimuli, and furthermore that infants showed an attentional preference to listen to ID song over ID speech. The present study shows that the two stimulus categories are clearly discriminable to songbirds as well—indicating that non-language specific acoustic cues separate the two stimulus categories (see also Tsang et al. for detailed acoustic analyses). Using songbirds also helps us to dissociate the extent to which language development in humans relates to speech-song processing. Overall, the infants looked longer to Russian stimuli than English stimuli, suggesting that exposure to and familiarity with native language impacts the perception of speech and song. In addition, innate processes of language acquisition that occur during the second half of the first year of life further differentiate native vs. nonnative language stimuli for human infants (Werker & Tees, 1984). Therefore, both familiarity with English and the specific age of the infants could have influenced the infants' preferences. The data in this experiment provided complementary information to the infant results: we could measure how long it took for birds to learn the category differences, we could compare acquisition between languages, and we could ensure there was no confound of language as the birds were not learning to speak English or Russian.

In the infant study, one group heard Russian stimuli and another heard English stimuli, but with songbirds we could present stimuli in both languages to all subjects, allowing for a direct comparison of responses to each language. Overall, songbirds did not show a bias to one language over another: generally birds learned English and Russian at the same rate (although female birds learned slightly faster than males), suggesting that there was nothing inherent in any one language that would make the speech-song discrimination and categorization easier. Thus any differences in how the infant language groups responded is not likely due to the discriminability of the acoustic features, something we would not know without the bird data.

What we cannot determine, however, is which specific cues the birds (or infants) were using to do the discrimination task; we would have to present stimuli with only one particular feature available and determine if the discrimination was still possible. Despite pitch contour being a highly salient cue for humans, including infants (e.g., Plantinga & Trainor, 2009; Trainor, Wu & Tsang, 2004) and despite the fact that birds can use pitch contour for discrimination (Spierings & ten Cate, 2014) it is unlikely the birds (and infants) were using pitch contour to solve this discrimination task. In this study, pitch contour was an unreliable cue as the contours

were matched and balanced across the individual stimuli in each category. However, we could ask whether the birds were using the same cues as humans by comparing their responses to ambiguous stimuli.

### **Responses to Probe Stimuli: Songbirds vs. Adult Humans**

A limitation of the head-turn preference paradigm used in Tsang et al. (2016) with infants was that no new category exemplars could be introduced beyond the initial category sets. However, with the songbirds, we were able to do so following the training phase. We presented birds with various stimuli not heard during training to test how they responded to them in comparison to the training stimuli, and also in comparison to how humans categorized (or attempted to categorize) these ambiguous stimuli. If birds and humans responded similarly to particular stimuli, it might be reasonable to conclude that they were using the same acoustic cues to perform the task.

For the ambiguous Russian stimuli, we first plotted bird responses according to the Likert Scale ratings done by English listeners. If birds and humans were attending to the same cues, we might expect responding in birds to increase along a continuum (as on the Likert scale) between the average response rates for the non-rewarded and rewarded training stimuli. However, the average responses of the birds did not line up with this pattern, nor did the responding of any individual bird. We then re-plotted the probe test responses according to how native Russian speakers categorized the stimuli. In this case, we expected more categorical responses by the birds (as depicted by the dashed lines in Panels C and D of Figure 4) to the stimuli the native Russian speakers agreed were song and speech, but again, the birds did not match the humans. It is possible birds responded categorically to specific stimuli, but with no general pattern it is impossible to tell if seemingly appropriate responses were spurious. However, if birds responded as Russian speakers did, it would be unlikely that they were using the same cues as the human listeners: birds would not be able to use non-acoustic cues, such as specific words or phrases, that would give the human listeners an advantage at categorizing the stimuli (e.g., Xu, Gandour, & Francis, 2006). We therefore need more specific testing to determine what birds, and non-Russian speaking humans, are using to categorize these stimuli.

In contrast to Russian probe stimuli, responses to English probe stimuli seemed more consistent within and across birds. Responses plotted according to whether English mothers were asked to sing or speak to their infant when the stimuli were recorded seemed to be the best fit for this data (Figure 4 Panels E and F). For example, the most consistent were the responses to ID song by birds that had song as an S- (Panel F). It is possible then that the acoustic features of the stimuli recorded when the mothers sang or spoke to their infants aligned with the ones the birds were using to discriminate the song from speech stimuli. When these English stimuli were recorded, mothers were instructed specifically to sing or speak to their infants, perhaps causing the English mothers to exaggerate the song or speech like qualities of their vocalizations. By contrast, Russian mothers were freely communicating with their infants, and no specific direction was given as to what to produce. Perhaps this small methodological difference made the acoustic features, and thus categories, more distinct in the English than Russian stimuli, and explains why we get a better fit with English than Russian data.

## Sex Differences in Songbirds

In the present study, we used both male and female adult zebra finches and found that there were slight differences in performance between the sexes. This has been observed previously in other songbirds including zebra finches (see Kriengwatana, Spierings & ten Cate, 2016 for discussion). It is possible that females are slightly better at auditory discrimination because quality of male song is a critical component of mate selection. Although young males must listen to adult male song during development to memorize a model to copy during vocal learning, females must continue to listen to and assess male quality throughout adulthood (Zann, 1996). Perhaps then females are more able to discern particular features of acoustic stimuli, translating to better discriminative abilities. This sex difference may be unique to songbirds—while some studies have shown that female infants have a slight advantage for processing speech- and language-based stimuli (e.g., Molfese & Molfese, 1979), to our knowledge, there is no evidence to suggest that speech or song perception would be significantly different between human males and females. Establishing true sex differences in human discrimination often requires large samples and/or many trials, something that is difficult to do in infant behavioural perception tasks. Tsang et al. found no differences in looking times to speech and song stimuli between male and female infants (unreported data). The extent to which the learning advantage demonstrated by female zebra finches in the speech/song discrimination can be generalized to human listeners remains for future research studies to examine.

Unexpectedly, the results also show that there are sex differences depending on whether song or speech was the S+. Female birds were better than males at the discrimination when speech was the S+, while there was no difference between the sexes when song was the S+. There is some evidence that suggests that the exaggerated acoustics of infant-directed speech may facilitate perception of speech sounds and that infant-directed stimuli are generally more attentionally engaging (see Soderstrom, 2007 for a review). To our knowledge, there are no studies that have specifically examined language facilitation effects of infant-directed singing on speech perception during infancy, thus it is difficult to know whether the same result would be expected for infant listeners. Although we know that zebra finches can discriminate among speech sounds (i.e., phonemes; see ten Cate, 2014 for review) it remains unclear why females should show facilitation effects for human song over human speech. Females do show slight advantage over males in auditory discrimination of their own vocalizations (see discussion above), so if human song is more like birdsong than is human speech, this advantage may have translated to the sex difference observed here. However, there have been no direct comparisons of the acoustics and perception of human song compared to birdsong; future studies should more closely examine whether facilitation effects for infant-directed speech related to language acquisition are also true in the case of infant-directed song.

## Limitations and Summary

The results of this study and Tsang et al. (2016) provide further evidence that, despite the fact that all of the stimuli used in the study share a perceived “musicality” by human adult listeners, infant-directed speech and song comprise two different acoustic categories that are clearly discriminable by both songbirds and infants. However, there remain some limitations to consider. The stimuli used in the discrimination task for birds were altered from those presented to the infants to adapt for species differences. Although we do not believe this substantially changed the stimuli, a future study could use these altered stimuli in a discrimination (rather than preference) paradigm with humans to first confirm they are not significantly different from the originals used with infants, but also to determine if they are as discriminable for humans as for birds. We could

also begin to examine which specific acoustic features the birds are using for this task, and whether it is similar to those used by humans, by designing stimuli specifically to answer this question.

Responses to probe stimuli by birds were highly variable, and some birds didn't respond to probe stimuli at all, likely due to neophobia; this has been observed in similar studies with finches (e.g., van Heijningen, Chen, van Laatum, van der Hulst, & ten Cate, 2013). A change to procedure, presenting stimuli to be used for probe tests before discrimination training begins may help this (e.g., Hahn et al., 2015). We also did not explicitly demonstrate that birds were using the categories of song and speech by including a group that was forced to memorize the stimuli (e.g., Bloomfield et al., 2008; Hahn et al., 2016). Despite these limitations, using songbirds as objective listeners allowed us to detail the rate of learning in each of the categories, with learners that were not learning the languages used. Comparative studies such as these continue to be a productive approach to answering complex research questions.

### Acknowledgments

The authors wish to thank Dr. Judy Plantinga and Dr. Sandra Trehub for providing the English stimuli used in this study, and Sean Roach and Sean Aitken for help with the birds. All research using animals done according to CCAC standards and was approved by the local animal ethics committee at Dalhousie University (Protocol number 14-067).

### References

- Amy, M., Monbureau, M., Durand, C., Gomez, D., They, M., & Leboucher, G. (2008). Female canary mate preferences: differential use of information from two types of male-male interaction. *Animal Behaviour*, *76*, 971–982.
- Bloomfield, L. L., Farrell, T. M., & Sturdy, C. B. (2008). Categorization and discrimination of "chick-a-dee" calls by wild-caught and hand-reared chickadees. *Behavioural Processes*, *77*, 166–176.
- Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [computer program] Version 5.3.50, retrieved 21 May 2013 from <http://www.praat.org>.
- Doupe, A., & Kuhl, P. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Reviews of Neuroscience*, *22*, 567–631.
- Falk, S. (2011a). Temporal variability and stability in infant-directed sung speech: Evidence for language-specific patterns. *Language and Speech*, *54*, 167–180.
- Falk, S. (2011b). Melodic vs. intonational coding of communicative functions - A comparison of tonal contours in infant-directed song and speech. *Psychomusicology*, *21(1&2)*, 53–68.
- Fernald, A. (1989). Intonation and communicative content in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.
- Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, *10*, 279–293.
- Gentner, T. Q., Hulse, S. H., Bentley, G. E., & Ball, G. F. (2000). Individual vocal recognition and the effect of partial lesions to HVC on discrimination, learning, and categorization of conspecific song in adult songbirds. *Journal of Neurobiology*, *42*, 117–133.
- Hahn, A. H., Hoang, J., McMillan, N., Campbell, K., Congdon, J., & Sturdy, C. B. (2015). Biological salience influences performance and acoustic mechanisms for the discrimination of male and female songs. *Animal Behaviour*, *104*, 213–228.
- Hahn, A. H., Hoeschele, M., Guillette, L. M., Hoang, J., McMillan, N., Congdon, J. V., ... Sturdy C. B. (2016). Black-capped chickadees categorize songs based on features that vary geographically. *Animal Behaviour*, *112*, 93–104.
- Jusczyk, P. W., Frederici, A. D., Wessels, J. M., Svenkreund, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, *32*, 402–420.

- Kemler-Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*, 111–116.
- Kriengwatana, B., Spierings, M. J., & ten Cate, C. (2016). Auditory discrimination learning in zebra finches: effects of sex, early life conditions, and stimulus characteristics. *Animal Behaviour*, *116*, 99–112.
- Molfese, D. L., & Molfese, V. J. (1979). Hemisphere and stimulus differences as reflected in cortical responses of newborn infants to speech stimuli. *Developmental Psychology*, *15*, 505–511.
- Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development*, *27*, 455–464.
- Phillmore, L. S., Sturdy, C. B., & Weisman, R. G. (2003). Does reduced social contact affect discrimination of distance cues and individual vocalizations? *Animal Behaviour*, *65*, 911–922.
- Phillmore, L. S., Sturdy, C. B., Ramsay, S. M., & Weisman, R. G. (1998). Discrimination of auditory distance cues by black-capped chickadees (*Poecile atricapillus*) and zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, *112*, 282–291.
- Plantinga, J., & Trainor, L. J. (2005). Memory for melody: infants use a relative pitch code. *Cognition*, *98*, 1–11.
- Searcy, W. A., & Beecher, M. D. (2009). Song as an aggressive signal in songbirds. *Animal Behaviour*, *78*, 1281–1292.
- Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*, 501–532.
- Spierings, M. J., & ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society B*, *281*, 2014-480.
- Spierings, M. J., & ten Cate, C. (2016). Zebra finches as a model species to understand the roots of rhythm. *Frontiers in Neuroscience*, *10*, 345. doi:10.3389/fnins.2016.00345
- Sturdy, C. B., & Weisman, R. G. (2006). Rationale and methodology for testing auditory cognition in songbirds. *Behavioural Processes*, *72*, 265–272.
- ten Cate, C. (2014). On the phonetic and syntactic processing abilities of birds: From songs to speech and artificial grammars. *Current Opinion in Neurobiology*, *28*, 157–164.
- ten Cate, C., Spierings, M., Hubert, J., & Honing, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Frontiers in Psychology*, *7*, 730, doi:10.3389/fpsyg.2016.00730
- Trainor, L. J. (1996). Infant preferences for infant-directed versus non infant-directed playsongs and lullabies. *Infant Behavior & Development*, *19*, 83–92.
- Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior & Development*, *20*, 383–396.
- Trainor, L. J., Wu, L., & Tsang, C.D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*, 289–296.
- Trehub, S., Trainor, L. J., & Unyk, A. M. (1993). Music and speech processing in the first year of life. *Advances in Child Development and Behavior*, *24*, 1–35.
- Tsang, C. D., Falk, S., & Hessel, A. (2016). Infants prefer infant-directed song over speech. *Child Development*. doi:10.1111/cdev.12647
- van Heijneingen, C.A.A., 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? *Animal Cognition*, *16*, 165–175.
- Weisman, R. G., Njegovan, M., Sturdy, C. B., Phillmore, L. S., Coyle, J., & Mewhort, D. J. (1998). Frequency-range discriminations: Special and general abilities in zebra finches (*Taeniopygia guttata*) and Humans (*Homo sapiens*). *Journal of Comparative Psychology*, *112*, 244–258.
- Weisman, R. G., Brownlie, L., Olthof, A., Njegovan, M., Sturdy, C., Mewhort, D. (1999). Timing and classifying brief acoustic stimuli by songbirds and humans. *Journal of Experimental Psychology – Animal Behavior Processes*, *25*, 139–152.
- Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*, 1063–1074.
- Zann, R. (1996). *The zebra finch: a synthesis of field and laboratory studies*. Oxford, UK: Oxford University Press.

**Financial conflict of interest:** This research was supported by a sub-grant to CDT and LSP as part of a SSHRC Major Collaborative Research Initiative grant to Annabel Cohen (University of Prince Edward Island, Charlottetown, PEI, Canada) and an NSERC Discovery Grant to LSP.

**Conflict of interest:** No stated conflicts.

*Submitted: October 14<sup>th</sup>, 2016*

*Resubmitted: April 19<sup>th</sup>, 2017*

*Accepted: April 20<sup>th</sup>, 2017*