

# Formant-frequency discrimination of synthesized vowels in budgerigars (*Melopsittacus undulatus*) and humans

Kenneth S. Henry<sup>a)</sup>

Department of Otolaryngology, University of Rochester, Rochester, New York 14642, USA

Kassidy N. Amburgey

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14642, USA

Kristina S. Abrams

Department of Neuroscience, University of Rochester, Rochester, New York 14642, USA

Fabio Idrobo

Universidad de Los Andes, Bogotá, Colombia

Laurel H. Carney

Department of Biomedical Engineering, University of Rochester, Rochester, New York 14642, USA

(Received 23 May 2017; revised 29 August 2017; accepted 28 September 2017; published online 13 October 2017)

Vowels are complex sounds with four to five spectral peaks known as formants. The frequencies of the two lowest formants,  $F1$  and  $F2$ , are sufficient for vowel discrimination. Behavioral studies show that many birds and mammals can discriminate vowels. However, few studies have quantified thresholds for formant-frequency discrimination. The present study examined formant-frequency discrimination in budgerigars (*Melopsittacus undulatus*) and humans using stimuli with one or two formants and a constant fundamental frequency of 200 Hz. Stimuli had spectral envelopes similar to natural speech and were presented with random level variation. Thresholds were estimated for frequency discrimination of  $F1$ ,  $F2$ , and simultaneous  $F1$  and  $F2$  changes. The same two-down, one-up tracking procedure and single-interval, two-alternative task were used for both species. Formant-frequency discrimination thresholds were as sensitive in budgerigars as in humans and followed the same patterns across all conditions. Thresholds expressed as percent frequency difference were higher for  $F1$  than for  $F2$ , and were unchanged between stimuli with one or two formants. Thresholds for simultaneous  $F1$  and  $F2$  changes indicated that discrimination was based on combined information from both formant regions. Results were consistent with previous human studies and show that budgerigars provide an exceptionally sensitive animal model of vowel feature discrimination. © 2017 Acoustical Society of America. <https://doi.org/10.1121/1.5006912>

[MLD]

Pages: 2073–2083

## I. INTRODUCTION

Discrimination of vowels plays a crucial role in perception of speech, especially under challenging listening conditions with noise and reverberation (Kewley-Port *et al.*, 2007; Ladefoged and Maddieson, 1996). Natural voiced vowels are harmonic sounds with frequency components located at integer multiples of the fundamental frequency ( $F0$ : 100–200 Hz; also known as voice pitch; Fig. 1). The vowel spectrum is filtered by the vocal tract. During speech production, changes in the position of the tongue and lips alter the resonance of the vocal tract (Fant, 1960) and hence the amplitude of each frequency component. Natural vowels are produced with resonant peaks at four to five frequencies, known as formant frequencies. The two lowest formant frequencies,  $F1$  and  $F2$ , are sufficient for vowel identification (Hillenbrand *et al.*, 1995).

Nonhuman animal models of vowel discrimination have substantial value, considering the fundamental role of speech

perception in everyday life. Previous behavioral studies show that a variety of birds and mammals are capable of discriminating vowels [e.g., chinchilla (Burdick and Miller, 1975), baboon (Hienz and Brady, 1988), ferret (Bizley *et al.*, 2013), rat (Eriksson and Villa, 2006), blackbird and pigeon (Hienz *et al.*, 1981), budgerigar (Dooling and Brown, 1990); reviewed by Kriengwatana *et al.* (2015)]. However, few studies have determined vowel discrimination limits, that is, the minimum difference in formant frequency necessary for discrimination. In mammals, Japanese macaques can detect a 2.5% frequency change in  $F1$  and a 1.6% change in  $F2$  of synthesized vowels (Sommers *et al.*, 1992). Cats perform similarly, with  $F2$  thresholds of 2.3% (Hienz *et al.*, 1996). Lower formant-frequency discrimination thresholds have been reported for humans. Kewley-Port and colleagues found human discrimination thresholds of 2.1%–2.9% for  $F1$  (~500 Hz) and 1.1%–1.3% for  $F2$  (Kewley-Port *et al.*, 1996; Kewley-Port and Watson, 1994). Lyzenga and Horst (1997, 1998) found thresholds below 1% for some  $F1$  and  $F2$  conditions. Different results between studies should be interpreted with caution, however, because formant-frequency discrimination thresholds can vary considerably with small differences

<sup>a)</sup>Electronic mail: kenneth\_henry@urmc.rochester.edu

in methodology [e.g., choice of  $F_0$  (Kewley-Port *et al.*, 1996); positioning of formant peaks relative to harmonic frequencies (Kewley-Port and Watson, 1994; Kewley-Port and Zheng, 1998; Lyzenga and Horst, 1997, 1998)]. Therefore, the degree to which formant-frequency discrimination abilities in these animal models match or deviate from human performance limits is uncertain.

Birds provide an interesting system for further studies of formant-frequency discrimination based on the importance of vocal communication behavior, vocal learning, and even the capacity to mimic human speech in some species. To our knowledge, only one study in the budgerigar has examined formant-frequency discrimination thresholds in birds (Henry *et al.*, 2017). The budgerigar is a small parrot species (*Psittacine*) with the capacity to learn and imitate new vocalizations, including speech sounds, throughout life. The budgerigar study quantified  $F_2$  discrimination thresholds for harmonic stimuli with  $F_0$  of 200 Hz and an unnatural triangular spectral envelope. Budgerigar thresholds overlapped extensively with human thresholds measured using the same stimuli and behavioral paradigm. Thresholds ranged from 0.17%–0.55% in budgerigars and from 0.16%–0.40% in humans when the formant frequency of the standard stimulus was centered between two harmonics. Thresholds in both species increased when the standard formant frequency aligned with a harmonic, concomitant with reduced temporal envelope cues, and increased similarly with the addition of background noise.

The present study examined formant-frequency discrimination thresholds in budgerigars and humans using synthesized vowel stimuli with a spectral envelope similar to natural speech (Fig. 1). Stimuli were modeled after those used in several previous humans studies (Lyzenga and Horst, 1997, 1998). Thresholds in each species were measured using identical stimuli and the same single-interval, two-alternative, non-forced discrimination task. Experiment I measured discrimination thresholds for stimuli with a single formant located at  $F_1$  or  $F_2$ . Experiment II measured discrimination thresholds for two-formant stimuli with one changing formant frequency ( $F_1$  or  $F_2$ ) and one stationary formant. Experiment III measured discrimination thresholds for two-formant stimuli in which both  $F_1$  and  $F_2$  changed in the same direction.  $F_0$  was held constant at 200 Hz in all experiments. The results show that for synthesized vowels with a natural spectral envelope, formant-frequency discrimination thresholds in budgerigars are as sensitive as human thresholds.

## II. EXPERIMENT I

### A. Methods

#### 1. Subjects

Behavioral experiments were conducted in four budgerigars (*Melopsittacus undulatus*; two female; two male) and four human subjects (two female; two male). Budgerigars were 2–3 yrs of age at the time of testing and weighed 40–50 g. Human subjects ranged in age from 19–38 yrs and had pure-tone thresholds less than 20 dB hearing level at

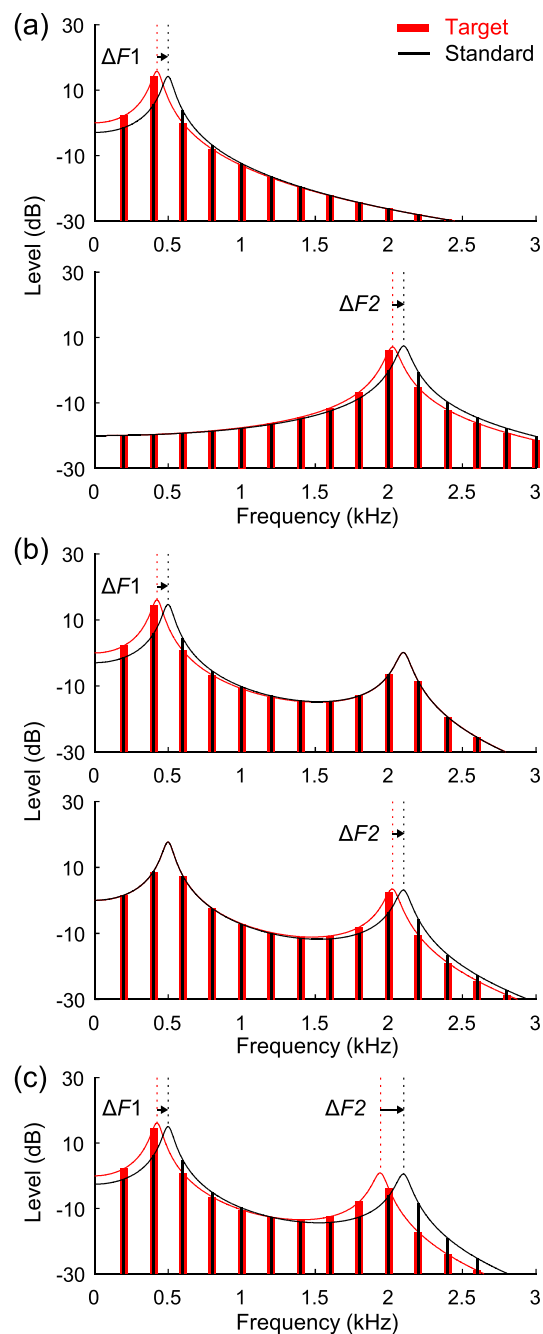


FIG. 1. (Color online) Spectra of synthetic vowel stimuli used to study formant-frequency discrimination. (a) Single-formant stimuli from experiment I used for estimation of  $F_1$  (top) at  $F_2$  (bottom) discrimination thresholds. Solid vertical lines are harmonic frequency components of the standard stimulus (black; thin lines) and a target stimulus with lower formant frequency (red; thick lines). Smooth curves are transfer functions of the Klatt synthesizer, which define the spectral envelope of each stimulus. Vertical dotted lines indicate the formant frequency of each stimulus. The formant frequency of the target stimulus was increased during tracking sessions (rightward pointing arrows) until the test stimulus could no longer be reliably discriminated from the standard. Note that a change in formant frequency alters the relative amplitude of individual harmonic components rather than their frequency. Harmonic frequencies remain unchanged because fundamental frequency ( $F_0$ ; voice pitch) was held constant (200 Hz) in all experiments. (b) Two-formant stimuli from experiment II used for estimation of  $F_1$  (top) and  $F_2$  (bottom) discrimination thresholds in the presence of a competing stationary formant. (c) Two-formant stimuli from experiment III used to study discrimination thresholds for simultaneously changing formants. The test stimulus shown is from the  $F_1$ -weighted condition, for which the percent frequency change in  $F_1$  (15%) is 2 times the percent frequency change in  $F_2$  (7.5%).

octave-spaced frequencies from 0.25–8 kHz. Experimental procedures in budgerigars were approved by the University Committee on Animal Resources at the University of Rochester. Procedures in humans were approved by the Research Subjects Review Board of the University of Rochester.

## 2. Experimental apparatus

Behavioral experiments in budgerigars were conducted in a single-walled acoustic isolation chamber (0.3 m<sup>3</sup> inside volume) lined with 6.7 cm of sound-absorbing foam insulation. The chamber contained an overhead loudspeaker (Polk Audio MC60; Polk Audio, Baltimore, MD), light emitting diode house light, and a video camera for monitoring the animal's behavior. Birds were perched in a wire-mesh cage during testing, with access to three horizontally arranged response switches and the feeding trough of a seed dispenser (ENV-203 Mini; Med Associates, St. Albans, VT). Response switches were located 19 cm above the floor of the chamber and 20 cm from the overhead loudspeaker. The house light and seed dispenser were controlled by a PC, data acquisition board (PCI-6151; National Instruments Corporation, Austin, TX), microcontroller (Arduino Leonardo; Turin, Italy), and laboratory-built hardware, which also generated acoustic stimuli (50-kHz sampling frequency) and processed input from the response switches. Behavioral test programs were written in MATLAB (The MathWorks, Natick, MA). Stimuli were convolved with a pre-emphasis filter prior to digital-to-analog conversion, power amplification (Crown D-75 A; Elkhart, IN), and presentation through the overhead loudspeaker. The pre-emphasis filter compensated for the frequency response of the acoustic system, which was determined from the output of a calibrated microphone (Brüel and Kjær type 4134; Brüel and Kjær, Marlborough, MA) placed at the approximate location of the bird's head inside the test chamber. Tones were presented during calibration at 249 log-spaced frequencies from 0.05–15.1 kHz.

Behavioral experiments in humans were conducted with subjects seated in a double-walled walk-in acoustic isolation chamber. Sessions were performed using a touchscreen PC and test program written in MATLAB. The test program interface consisted of three horizontally arranged response pushbuttons and a feedback window for displaying the result of each trial (correct or incorrect response). Stimuli were presented diotically through calibrated audiometric headphones (TDH-30; Telephonics, Farmingdale, NY).

## 3. Stimuli

Stimuli were produced with a single formant frequency at  $F1$  (425–500 Hz) or  $F2$  (2025–2100 Hz) using one resonator of a Klatt synthesizer [Klatt and Klatt, 1990; Fig. 1(a)]. Resonator bandwidth was 70 Hz for the  $F1$  stimulus and 90 Hz for  $F2$ . The resonator was stimulated with an impulse train and low-pass filtered at 5 kHz (5000-point finite impulse response) to produce single-formant stimuli with constant  $F0$  of 200 Hz. Note that changes in formant

frequency altered the relative amplitude of harmonic components rather than their frequency. Stimuli were 250 ms in duration with 25-ms cosine-squared onset and offset ramps. Preliminary investigations demonstrated absolute level cues associated with variation in formant frequency, especially when the formant was positioned near a harmonic frequency component rather than between two harmonics. Pilot behavioral experiments showed no substantial effect on thresholds, but we nonetheless focused on between-harmonic stimuli because level cues for these signals were less than 2 dB over the range of formant-frequency values presented during behavioral tests. To further reduce the utility of level cues, the presentation level of stimuli was randomly varied (“roved”) over a 16 dB range (Green *et al.*, 1983). Presentation levels were uniformly distributed with a step size of 1 dB and median value of 80.5 dB sound pressure level (SPL) for  $F1$  stimuli and 68 dB for  $F2$  stimuli. These values correspond to the typical level difference between  $F1$  and  $F2$  in multi-formant stimuli. Identical stimuli were presented to budgerigars and humans.

## 4. Procedure

Behavioral sessions were conducted using a single-interval, two-alternative, non-forced choice task. In budgerigars, birds started a trial by pecking the center observing-response switch, which initiated presentation of a single standard stimulus or a single target stimulus. Each block of ten trials consisted of five standard trials and five target trials in random sequence. The bird was then given a 3-s reaction window to make a reporting response by pecking one of the two switches located to the left and right. The correct response to a target trial (hit) was a peck on the left-hand switch, and the correct response to a standard trial (correct rejection) was a peck on the right-hand switch. All responses resulted in immediate termination of the stimulus if still playing. Correct responses were reinforced with 1 or 2 millet seeds, dependent on response bias (see below). Incorrect responses (misses and false alarms) produced a 5-s timeout during which the house light was turned off. Responses during the timeout period extended the timeout (the 5-s timeout timer was reset). A shorter 2-s timeout was imposed in rare instances when the bird did not make a reporting response within 3 s, or responded by pecking the center switch. Response bias toward the right or left switch was calculated as  $-0.5$  times the sum of the Z-score of the hit rate and the Z-score of the false alarm rate. Bias was calculated for each block of 50 trials, and controlled by adjusting the percentage of trials on each side for which two-seed reinforcement was delivered. Sessions with overall absolute values of bias greater than or equal to 0.3 were excluded from subsequent analyses (9.5% of sessions).

Initial training to perform the behavioral task described above was conducted in three stages. Birds were first trained to peck the left or right switch at any time to receive a seed reward; the center switch was covered for this training step. Pecking behavior was encouraged by attaching seeds to the switches with double-sided tape and loading the tape with seed as needed. In the second stage, acoustic stimuli (50%

target, 50% standard) were presented with 5-s duration and a variable period of silence between stimuli (3–5 s). A reaction window was also specified, extending from stimulus onset to 0.1 s after stimulus offset. Pecks on the left and right switches during the stimulus resulted in immediate termination of sound. Correct responses (on the left switch for standard stimuli, right switch for target stimuli) during the reaction window were followed with a seed reward, while incorrect responses were not. Stimulus duration was reduced across sessions, as birds learned to respond within the reaction window, until the stimulus duration was 0.5 s. The reaction window for 0.5-s stimuli was 1 s in duration. In the final stage, birds were trained to trigger each single-interval trial by pecking the center switch. This was accomplished by initially loading the center switch with seed as in training step one. The duration of stimuli and the reaction window were as in regular testing (0.25 and 3 s, respectively). Short timeouts were also introduced for incorrect responses. Timeout duration was increased across sessions to a final value of 5 s, after which birds commenced regular testing.

The same behavioral paradigm was used in humans, but observing and reporting responses were made by pressing the pushbuttons of the touchscreen computer. Furthermore, reinforcement consisted of a cartoon image for correct responses and a black rectangle for incorrect responses, and bias was controlled by reporting the bias value to the subject after each session and reminding subjects to reduce bias by using both reporting-response options in cases when they were unsure of the correct answer. Biased sessions (absolute value of bias greater than or equal to 0.3; 7.8% of sessions) were excluded from further analyses.

Behavioral test sessions in budgerigars were conducted 7 days per week. Sessions were 20–30 min long and conducted once in the morning and once in the afternoon. During initial non-tracking sessions, birds discriminated between the standard stimulus, with formant frequency of 500 Hz for  $F1$  or 2100 Hz for  $F2$ , and a fixed target stimulus with lower formant frequency by 50 Hz. Non-tracking sessions were repeated with the same stimulus condition until discrimination performance exceeded 90%. Birds typically required 10–30 non-tracking sessions to attain this performance level, for a new stimulus condition. Thereafter, four to five behavioral tracks were conducted per day in each bird to allow repeated estimation of formant-frequency discrimination thresholds (i.e., one threshold per track, each track consisting of  $\sim 120$  trials). Within each track, the formant-frequency of the target stimulus was systematically varied from the starting point 50 Hz below the standard formant frequency using a two-down one-up adaptive staircase method (Levitt, 1970). The absolute frequency difference between target and standard was decreased following each pair of consecutive hits for the same target value, and conversely, increased following each miss [Fig. 2(a)]. An initial step size of 8 Hz was reduced to 4 Hz after 2 reversals in the step direction (up vs down) of the behavioral tracking session. The step size was further reduced to 2 Hz after 4 track reversals and to 1 Hz after 6 track reversals. Each track continued until (1) at least 15 reversals occurred, (2) the standard deviation (SD) of the last eight

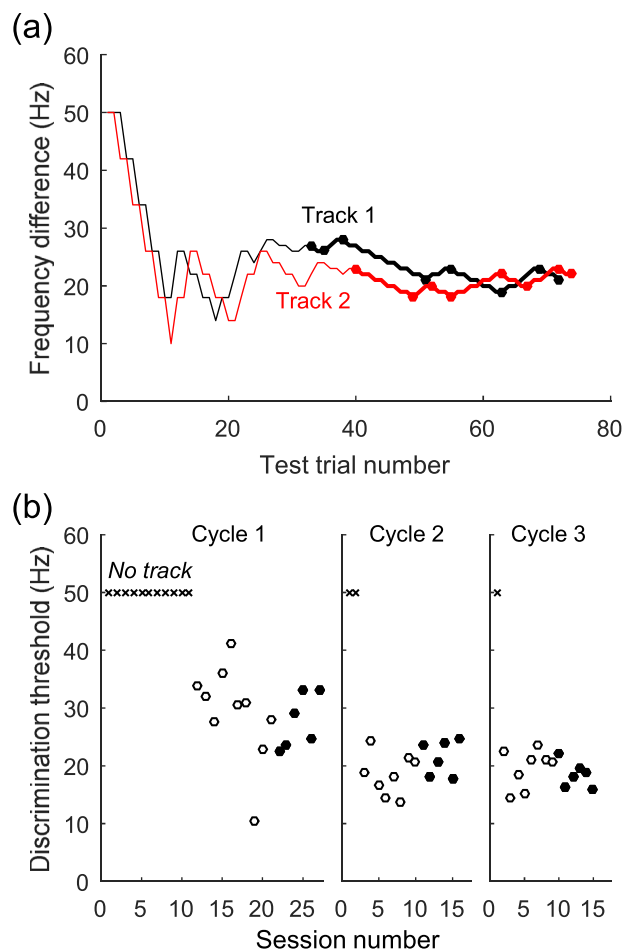


FIG. 2. (Color online) (a) Representative behavioral results from two tracking sessions in budgerigar B47. The formant frequency of the test stimulus was varied according to a two-down, one-up, adaptive tracking method. Each trace plots the absolute value of the formant-frequency difference between the target stimulus and the standard stimulus versus target trial number. Discrimination threshold was calculated as the mean  $y$ -value of the last eight reversal points (circles). Thresholds for the sessions shown, for  $F1$  discrimination with  $F2$  present (experiment II), were  $23.50 \pm 3.21$  Hz (black) and  $20.8 \pm 2.05$  Hz (red; means  $\pm$ SD). (b) Representative behavioral results showing threshold variation across test sessions. Thresholds are for  $F1$  discrimination with  $F2$  present, in B47.  $\times$  markers near the top of the plot indicate non-tracking sessions. Sessions for this condition were conducted in three cycles, between which other stimulus conditions were tested (36 sessions between test cycles one and two; 120 sessions between test cycles two and three). A new vertical axis is drawn at the start of each new test cycle. For each cycle, threshold was calculated as the mean of the last six session thresholds (filled symbols). The third testing cycle was conducted due to a significant decrease in threshold between the first ( $28.67 \pm 4.05$  Hz) and second cycles ( $21.46 \pm 3.04$ ;  $t_{10} = -3.486$ ,  $P = 0.006$ ). The threshold for this condition ( $20.00 \pm 2.99$  Hz) was calculated as the mean of the last two test-cycle thresholds, which were not significantly different from each other ( $t_{10} = 1.872$ ,  $P = 0.091$ ).

formant-frequency reversals was less than 5 Hz, and (3) the absolute difference between the mean of the last four formant-frequency reversals and the mean of the previous four formant-frequency reversals was less than 5 Hz (Costa and Cancado, 2012). The threshold for each track was estimated as the difference between the mean of the last eight formant-frequency reversals and the standard formant frequency. Thresholds were excluded from further analyses if any of the last eight formant-frequency reversals was equal to the standard formant frequency.



Thresholds were estimated a minimum of 13 times until two stability criteria were met: (1) the standard deviation of the last six threshold estimates was less than 10 Hz, and (2) the absolute difference between the mean of the last three thresholds and the previous three thresholds was less than 5 Hz [Fig. 2(b)]. The last six thresholds were then used to calculate a mean threshold for the stimulus condition. The two stimulus conditions of experiment I were conducted in a random sequence in each bird. Thereafter, experiments II and III were conducted, followed by repetition of the two conditions from experiment I. This testing cycle was repeated until no significant change was observed in the threshold for the condition (two-sample *T*-test;  $\alpha=0.05$ ). Thresholds were usually stable after two testing cycles, but occasionally required three cycles, or in one instance a fourth. The reported threshold for each condition, including those from experiments II and III, is the mean threshold of the last two testing cycles.

Behavioral test sessions in humans were conducted similarly, except that subjects did not perform non-tracking sessions prior to tracking sessions, and the starting formant frequency of the test stimulus was 75 Hz below the standard formant frequency. Furthermore, the minimum number of test sessions for each stimulus condition was 8, versus 13 in budgerigars, and each condition was tested once unless stability criteria could not be met within 10–12 sessions. In these unusual cases, the subject moved on to experiments II and III before returning to complete additional test sessions. The reported threshold for each condition is the mean of the last six thresholds.

### 5. Statistical analyses

Thresholds were log transformed and analyzed in R (version 3.4.1) using linear mixed-effects model analyses (Bates *et al.*, 2015). The models included species as a between-subject effect and formant frequency, number of formants, and formant weighting scheme as within-subject effects. Subject intercepts were modeled as a random effect. Interactions were included between fixed effects and dropped when not significant ( $p > 0.05$ ) in order of decreasing  $p$  value. Degrees of freedom for *F* tests and pairwise comparisons of least-squares means were calculated based on the Satterthwaite approximation. Visual inspection of model results showed that residuals were normally distributed.

### B. Results and discussion

Thresholds for discrimination of a formant-frequency change were examined in four budgerigars using single-formant synthesized vowels and operant conditioning procedures [Fig. 3(a); “single formant” stimulus conditions]. Thresholds were measured for an *F1* condition, with standard frequency of 500 Hz (left panel), and an *F2* condition with standard frequency of 2100 Hz (right). Formant-frequency discrimination thresholds decreased from 3.48% ( $\pm 0.81$ ) for *F1* to 1.12% ( $\pm 0.22$ ) for *F2* [means ( $\pm$ SD)]. Discrimination thresholds were lower for *F2* than for *F1* in all test subjects.

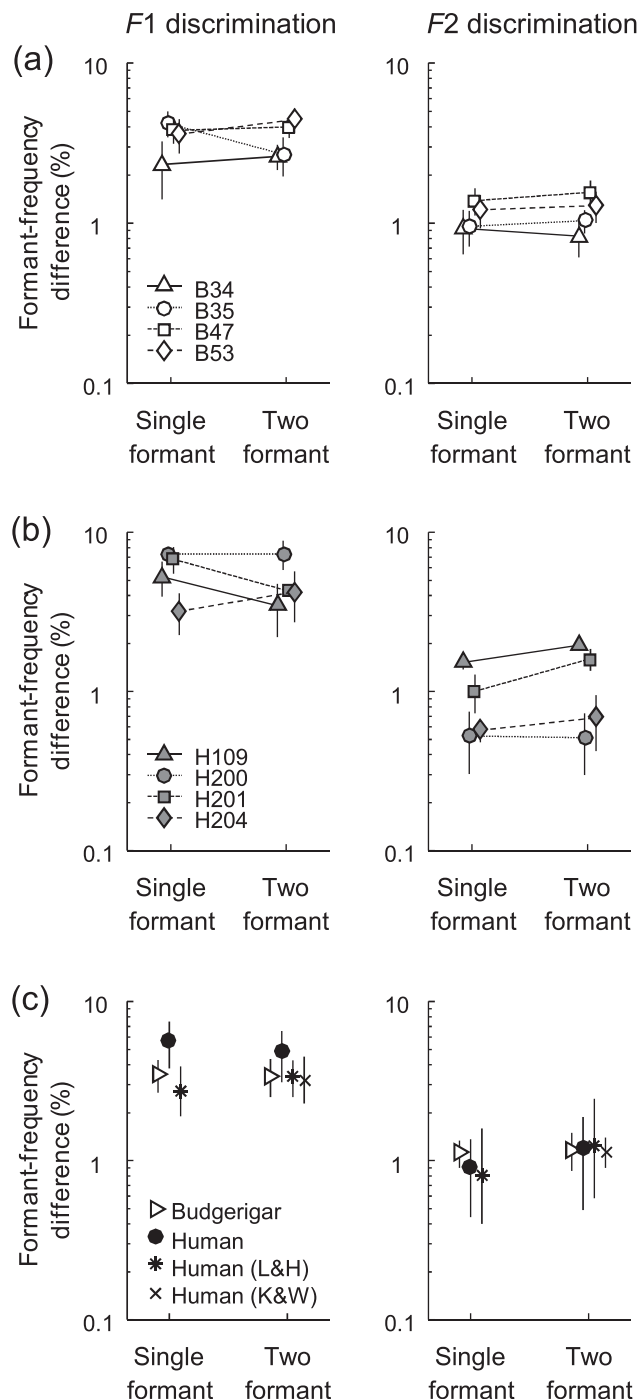


FIG. 3. (a) Thresholds of individual budgerigars ( $n=4$ ) for discrimination of a frequency change in *F1* (left panel) or *F2* (right panel). Each panel shows thresholds obtained with a single-formant stimulus (left), with one formant peak, and a two-formant stimulus (right) for which one formant-frequency changed while the other remained constant. Thresholds of the same test subject are plotted with the same symbol (see legend) and connected with a line segment to indicate the change between conditions. Vertical bars indicate the standard deviation of each threshold estimate. (b) Thresholds of individual human subjects ( $n=4$ ) for discrimination of a frequency change in *F1* or *F2*, presented as in (a). Thresholds were obtained using the same stimuli and behavioral paradigm as in budgerigars. (c) Mean thresholds of budgerigars and humans in the present study, from individual data in (a) and (b), compared to human thresholds from previous studies (L&H: Lyzenga and Horst 1997, 1998; K&W: Kewley-Port and Watson 1994). Error bars for data from the present study and from Lyzenga and Horst (1997, 1998) indicate the standard deviation across subjects. Error bars for Kewley-Port and Watson (1994) indicate the range of median thresholds observed across several similar test conditions (see text).

Formant-frequency discrimination thresholds were measured in four human subjects using identical stimuli and the same behavioral paradigm [Fig. 3(b)]. Human thresholds decreased from 5.63% ( $\pm 1.84$ ) for  $F1$  to 0.90% ( $\pm 0.46$ ) for  $F2$ , and overlapped extensively with budgerigar thresholds for both stimulus conditions [Fig. 3(c)]. Human thresholds were lower for  $F2$  than for  $F1$  as in budgerigars, and showed considerable variation across subjects as in previous psychophysical studies (e.g., Lyzenga and Horst, 1997; see below). A mixed-model analysis of thresholds in both species showed a main effect of formant frequency ( $F1$  vs  $F2$ ;  $F_{1,6} = 90.89$ ,  $p = 0.0001$ ) but not species ( $F_{1,6} = 0.18$ ,  $p = 0.69$ ). The species by formant-frequency interaction approached significance ( $F_{1,6} = 5.69$ ,  $p = 0.054$ ), due to a slightly greater threshold difference between formant frequencies in humans than in budgerigars.

Previously, Lyzenga and Horst (1997) measured formant-frequency discrimination thresholds in humans using similar Klatt-synthesized stimuli with a single formant [Fig. 3(c); stars]. As in the present study, stimuli were presented with  $F0$  of 200 Hz, roving level, and standard frequencies of 500 and 2100 Hz for  $F1$  and  $F2$ , respectively. Notable differences of the previous work from the present study include narrower  $F1$  bandwidth (50 Hz), broader stimulus bandwidth extending to at least 6 kHz, and the use of a three-interval, three-alternative, forced-choice discrimination task. Thresholds ranged from 1.9%–3.9% for  $F1$  discrimination (mean = 2.7%) and from 0.4%–1.6% for  $F2$  discrimination (mean = 0.8%). These values are in close agreement with budgerigar and human thresholds found here. Thus, we conclude that formant-frequency discrimination thresholds for single-formant stimuli are as sensitive in budgerigars as in humans.

### III. EXPERIMENT II

Natural vowels contain multiple formant peaks that decrease in amplitude with increasing frequency due to low-pass filtering by the vocal tract. As such, the possibility exists for one formant, especially  $F1$  due to its high amplitude, to interfere with (mask) frequency discrimination of other formants. The few previous studies of formant masking have produced different results in different species. In humans,  $F2$  discrimination thresholds can increase by a factor of 2 or more in the presence of a second, stationary formant at  $F1$ , whereas  $F1$  thresholds are relatively unaffected by  $F2$  presence (Lyzenga and Horst, 1998). In contrast,  $F2$  discrimination in Japanese macaques appears unaffected by  $F1$  presence (Sommers *et al.*, 1992). The extent to which these conflicting findings reflect species differences versus differences in stimulus design or the choice of behavioral paradigm is unclear. The goal of experiment II was to determine if the presence of an additional stationary formant influences formant-frequency discrimination thresholds in budgerigars. We determined the effects of (1)  $F1$  presence on  $F2$  discrimination and (2)  $F2$  presence on  $F1$  discrimination, in budgerigars and in humans.

### A. Methods

Behavioral experiments were conducted in the same budgerigar test subjects and human listeners from experiment I to allow direct comparison of behavioral data. The apparatus and behavioral test procedure were unchanged from experiment I. Stimuli were produced with formant frequencies at  $F1$  (425–500 Hz) and  $F2$  (2025–2100 Hz) using two cascaded resonators of a Klatt synthesizer [Fig. 1(b)]. These formant frequency ranges correspond roughly to those observed for the English vowel / $\epsilon$ /. For assessment of  $F1$  discrimination thresholds, with standard frequency of 500 Hz, the frequency of  $F2$  was held constant at 2100 Hz. For assessment of  $F2$  discrimination thresholds, with standard frequency of 2100 Hz, the frequency of  $F1$  was held constant at 500 Hz. Thus, the frequency of the stationary formant was between harmonics in both cases. Thresholds were evaluated for a decrease in formant frequency relative to the standard stimulus. Median stimulus level was 80 dB SPL for both conditions. Other stimulus details, including the use of roving presentation level, were the same as in experiment I.

### B. Results and discussion

Thresholds were measured in four budgerigars for discrimination of a frequency change in either  $F1$  or  $F2$  using synthesized vowels with two formants. Thresholds decreased from 3.43% ( $\pm 0.92$ ) for  $F1$  to 1.18% ( $\pm 0.32$ ) for  $F2$  [means ( $\pm$ SD); Fig. 3(a); “two formant” stimulus conditions]. These thresholds were essentially unchanged from those obtained using a single-formant stimulus in experiment I. Relative to single-formant thresholds,  $F1$  thresholds for two-formant stimuli were similar in two test subjects (B34 and B47), slightly elevated in a third (B54), and reduced in the fourth subject (B35).  $F2$  thresholds appeared unchanged in all birds.

For the same two-formant stimuli, thresholds of human listeners ( $n = 4$ ) decreased from 4.81% ( $\pm 1.71$ ) for  $F1$  to 1.19% ( $\pm 0.70$ ) for  $F2$  [Fig. 3(b)]. Human thresholds overlapped considerably with budgerigar thresholds for the same stimulus conditions [Fig. 3(c)]. Furthermore as in budgerigars, the thresholds of each human subject showed no consistent change between measurements with a single-formant stimulus or a two-formant stimulus. Compared to single-formant thresholds from experiment I,  $F1$  thresholds for two-formant stimuli were similar or slightly lower, while  $F2$  thresholds were similar or slightly higher.

A mixed-model analysis of thresholds from both species in experiments I and II showed a significant effect of formant frequency ( $F1$  vs  $F2$ ;  $F_{1,21} = 150.71$ ,  $p < 0.0001$ ) but not species ( $F_{1,6} = 0.28$ ,  $p = 0.62$ ) or the number of formants present (one or two;  $F_{1,21} = 0.04$ ,  $p = 0.84$ ). A significant interaction was observed between formant frequency and species ( $F_{1,21} = 6.87$ ,  $p = 0.016$ ), but not between formant frequency and the number of formants present ( $F_{1,19} = 0.83$ ,  $p = 0.37$ ) or between the number of formants present and species ( $F_{1,19} = 0.01$ ,  $p = 0.92$ ). These results show that in both budgerigars and humans, formant-frequency discrimination thresholds are lower for  $F2$  than  $F1$  and not strongly influenced by the presence of an additional formant.

Thresholds are similar between species, but may differ more between  $F1$  and  $F2$  in humans than in budgerigars.

The  $F1$  and  $F2$  discrimination thresholds found in the present study are consistent with the results of previous human studies that used similar multi-formant stimuli. Across several stimulus conditions with  $F0$  of 200 Hz and standard formant frequencies similar to those investigated here ( $F1$ : 450–550 Hz;  $F2$ : 1950–2300 Hz), Kewley-Port and Watson (1994) found human thresholds ranging from 2.3%–4.5% for  $F1$  discrimination and from 0.9%–1.4% for  $F2$  discrimination [Fig. 3(c); “×” markers]. Similarly for  $F0$  of 200 Hz and the same standard formant frequencies used here ( $F1$ : 500 Hz;  $F2$ : 2100 Hz), Lyzenga and Horst (1998) found human thresholds of 2.5%–4.3% for  $F1$  (mean = 3.4%) and 0.6%–2.5% for  $F2$  [mean = 1.2%; Fig. 3(c); stars].

Our finding that  $F2$  discrimination is unaffected by the addition of a stationary  $F1$  formant agrees with results from a previous study in Japanese macaques (Sommers *et al.*, 1992). However, Lyzenga and Horst (1998) observed a two-to-three fold increase in human  $F2$  thresholds with the addition of  $F1$ . These large increases were observed when  $F1$  was aligned with a harmonic rather than between harmonics. For more closely matched conditions to the present study, with  $F1$  between harmonics, the observed increase in  $F2$  threshold was smaller [i.e., from 0.8% to 1.2%; Fig. 3(c); stars, right panel] and consistent with the present results.

#### IV. EXPERIMENT III

Natural vowel discrimination requires behavioral sensitivity to multiple formant-frequency differences rather than a single formant-frequency difference in isolation. Several previous studies suggest that human listeners combine information between  $F1$  and  $F2$  regions to discriminate stimuli with multiple changing formants (Hawks, 1994; Lyzenga and Horst, 1998; Mermelstein, 1978). The objective of experiment III was to compare thresholds for simultaneous changes in formant frequency between budgerigars and humans.  $F1$  and  $F2$  changed together according to two relationships. Relative to the standard stimulus [ $F1 = 500$  Hz;  $F2 = 2100$  Hz; Fig. 1(c)], the percent change in  $F1$  was either two times the percent change in  $F2$  (the “ $F1$ -weighted” condition) or equal to the percent change in  $F2$  (the “equal-weight condition”).

##### A. Methods

Behavioral experiments were conducted in the same budgerigars and human subjects from experiments I and II to allow direct comparison of results. The experimental apparatus and test procedures were the same as in experiments I and II. Stimuli were generated as in experiment II using two cascaded resonators of a Klatt synthesizer, and with roving presentation level. Formant frequencies of target stimuli were both lower than those of the standard stimulus.

##### B. Results and discussion

The threshold for a combined change in formant frequency can be expressed as the percent frequency difference

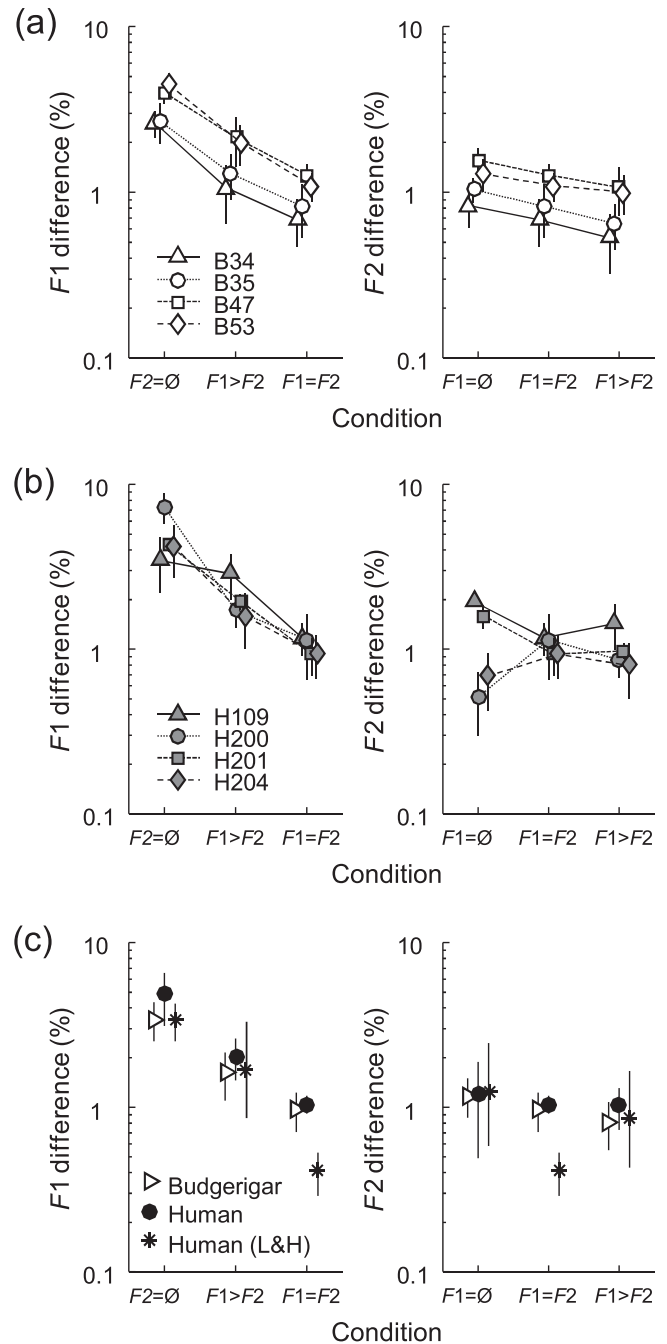


FIG. 4. (a) Thresholds of individual budgerigars ( $n = 4$ ) for discrimination of simultaneous changes in  $F1$  and  $F2$ . The percent change in  $F1$  was either equal to the percent change in  $F2$  ( $F1 = F2$  condition) or two times the percent change in  $F2$  ( $F1 > F2$  condition). Thresholds with respect to  $F1$  (left panel) are plotted together with isolated  $F1$  thresholds ( $F2 = \emptyset$  condition; left) and arranged left to right in order of increasing  $F2$  change. Thresholds with respect to  $F2$  are plotted with isolated  $F2$  thresholds ( $F1 = \emptyset$  condition; left) and arranged left to right in order of increasing  $F1$  change. The meaning of the symbols is the same as in Fig. 3(a). (b) Thresholds of individual human subjects ( $n = 4$ ) for the same stimuli with simultaneously changing formant frequencies, presented as in (a). (c) Mean thresholds of budgerigars and humans in the present study, from individual data in (a) and (b), compared to human thresholds from a previous study (L&H: Lyzenga and Horst 1998). Error bars in all cases indicate the standard deviation across subjects.

from the standard stimulus with respect to either  $F1$  or  $F2$ . For the  $F1$ -weighted condition of the present study, the threshold expressed with respect to  $F1$  is 2 times the threshold with respect to  $F2$ . For the equal-weight condition, the



threshold is the same whether expressed in terms of  $F1$  or  $F2$ . When both (1) the threshold with respect to  $F1$  is lower than the threshold for an isolated  $F1$  change and (2) the threshold with respect to  $F2$  is lower than the threshold for an isolated  $F2$  change, the pattern suggests discrimination based on combined information between formant regions.

Thresholds for simultaneous changes in formant frequency were studied in four budgerigars. For the  $F1$ -weighted condition, the threshold was 1.62% ( $\pm 0.52$ ) with respect to  $F1$  and 0.81% ( $\pm 0.26$ ) with respect to  $F2$  [means ( $\pm$ SD); Fig. 4(a); “ $F1 > F2$ ” condition]. For comparison, thresholds from experiment II for an isolated formant-frequency change were 3.43% ( $\pm 0.92$ ) for  $F1$  and 1.18% ( $\pm 0.32$ ) for  $F2$  [Fig. 4(a); “ $F1 = \emptyset$ ” and “ $F2 = \emptyset$ ” conditions, respectively]. In every test subject, the threshold with respect to  $F1$  was lower than the isolated  $F1$  threshold and the threshold with respect to  $F2$  was lower than the isolated  $F2$  threshold. These results show that budgerigars combined information between  $F1$  and  $F2$  to discriminate stimuli with simultaneously changing formant frequencies. A similar pattern was observed for the equal-weight condition, where the threshold expressed with respect to either formant was 0.97% [ $\pm 0.26$ ; Fig. 4(a); “ $F1 = F2$ ” condition]. In all test subjects, this threshold was lower than both the isolated  $F1$  and isolated  $F2$  threshold.

Discrimination thresholds for the same stimuli in human subjects ( $n = 4$ ), whether expressed with respect to  $F1$  or  $F2$  [Fig. 4(b)], overlapped extensively with budgerigar thresholds and followed the same general pattern across conditions [Fig. 4(c)]. Thresholds for the  $F1$ -weighted condition were 2.03% ( $\pm 0.58$ ) with respect to  $F1$  and 1.02% ( $\pm 0.29$ ) with respect to  $F2$ . For comparison, thresholds for an isolated formant-frequency change were 4.81% ( $\pm 1.71$ ) for  $F1$  and 1.19% ( $\pm 0.70$ ) for  $F2$ , from experiment II. The threshold with respect to  $F1$  was lower than the isolated  $F1$  threshold in every subject, while the threshold with respect to  $F2$  was lower in two subjects (H109 and H201) but not the other two. The pattern suggests that some human subjects combine information between formant regions to perform the discrimination task, while others rely more heavily on  $F2$  (i.e., receive less benefit from  $F1$  variation). The same pattern was observed in humans for the equal-weight condition as for the  $F1$ -weighted condition [Figs. 4(b) and 4(c)]. The threshold with respect to either formant was 1.04% ( $\pm 0.13$ ). This threshold was lower than the isolated  $F1$  threshold in every subject and lower than the isolated  $F2$  threshold in two subjects (the same as for the  $F1$ -weighted condition) but not the other two.

A mixed-model analysis of budgerigar results showed a significant effect of weighting condition on thresholds with respect to  $F1$  ( $F_{2,6} = 293.56$ ,  $p < 0.0001$ ) and thresholds with respect to  $F2$  ( $F_{2,6} = 59.05$ ,  $p = 0.0001$ ). For both weighting schemes, the threshold with respect to  $F1$  was lower than the isolated  $F1$  threshold (equal weight:  $T_6 = -24.06$ ,  $p < 0.0001$ ;  $F1$  weighted:  $T_6 = -14.49$ ,  $p < 0.0001$ ; pairwise comparisons of least-square means) and the threshold with respect to  $F2$  was lower than the isolated  $F2$  threshold (equal weight:  $T_6 = -5.60$ ,  $p = 0.001$ ;  $F1$  weighted:  $T_6 = -10.87$ ,  $p < 0.0001$ ). In contrast, an analysis combining results

between budgerigars and humans showed an effect of weighting condition for thresholds with respect to  $F1$  ( $F_{2,14} = 106.88$ ,  $p < 0.0001$ ) but not thresholds with respect to  $F2$  ( $F_{2,14} = 1.52$ ,  $p = 0.25$ ). Effects of species and the species by weighting condition interaction were not significant ( $F1$  thresholds; species:  $F_{1,6} = 1.95$ ,  $p = 0.21$ ; species  $\times$  weighting condition:  $F_{2,12} = 0.68$ ,  $p = 0.53$ ;  $F2$  thresholds; species:  $F_{1,6} = 0.12$ ,  $p = 0.74$ ; species  $\times$  weighting condition:  $F_{2,12} = 1.10$ ,  $p = 0.36$ ). The difference between the full model and the model including budgerigar results only was caused by divergent results in listeners H200 and H204, both of whom had unusually low isolated  $F2$  thresholds.

Taken together, these results show that budgerigars and some human listeners combine information between formant regions to discriminate stimuli with simultaneously changing formant frequencies. Several previous human studies reached the same conclusion. For formant-frequency changes similar to the equal-weight condition studied here, Lyzenga and Horst (1998) found a discrimination threshold of 0.4%, for  $F1$  or  $F2$  [Fig. 4(c); stars]. This threshold was considerably lower than both the 3.4% threshold for  $F1$  and the 1.2% threshold for  $F2$  observed for two-formant stimuli with one changing formant. Similar conclusions were drawn by earlier studies (Hawks, 1994; Mermelstein, 1978), suggesting that human listeners combine  $F1$  and  $F2$  information to detect simultaneous formant changes.

## V. GENERAL DISCUSSION

The present study compared formant-frequency discrimination thresholds between budgerigars and humans using identical synthesized-vowel stimuli and the same behavioral paradigm. Discrimination thresholds were quantified for stimuli with a single formant (experiment I) and for two-formant stimuli where either one formant frequency changed while the other remained constant (experiment II) or both formant frequencies changed in the same direction (experiment III). Behavioral thresholds of budgerigars were as sensitive as human thresholds for all stimulus conditions and followed the same pattern across conditions.

These results show that budgerigars discriminate formant-frequency differences of synthesized vowels with performance limits that closely match those of human listeners. These new results are consistent with those of a previous budgerigar study that used less natural single-formant stimuli synthesized with a triangular spectral envelope (Henry *et al.*, 2017). The stimuli of the previous study exhibited large differences in the modulation depth of temporal envelope fluctuations depending on whether the formant peak aligned with a harmonic or fell between two harmonics. Behavioral thresholds were lowest, and highly similar between species, when the formant frequency of the standard stimulus was centered between two harmonics [budgerigar: 0.36% ( $\pm 0.18$ ); human: 0.30% ( $\pm 0.09$ ); means  $\pm$ SD; Henry *et al.*, 2017]. Thresholds increased in background noise and when the formant frequency of the standard aligned with a harmonic. Threshold elevation in these cases coincided with reduced temporal envelope cues, and with diminished neural coding of envelope fluctuations in the auditory midbrain



(inferior colliculus, IC) of budgerigars. Average-rate-based neural coding of envelope fluctuations could account for behavioral thresholds in quiet but not in noise, even after optimal information pooling across units with a decoder analysis. In contrast, IC envelope synchrony was sufficient to account for behavioral performance under all test conditions. These results demonstrate the importance of envelope synchrony at the midbrain processing level. Finally, budgerigar IC responses closely resembled predicted responses of a mammalian IC model, consistent with conserved neural coding mechanisms for simple harmonic sounds in the midbrain of birds and mammals.

Earlier behavioral studies showed that budgerigars group natural and synthesized vowels into the same phonetic categories identified by human listeners. Perceptual grouping in budgerigars was assessed using behavioral response latency, which increases with greater stimulus dissimilarity in perceptual space (Dooling *et al.*, 1987). In one study of natural vowel discrimination, budgerigars were found to perceive acoustic differences among vowels as more salient than differences between talkers, even when  $F_0$  differences were large (Dooling and Brown, 1990). Thus, as in other birds and mammals (reviewed by Kriengwatana *et al.*, 2015), budgerigars can accurately identify different vowels despite lack of acoustic invariance in  $F_0$ . A follow-up study examined discrimination of synthesized stimuli along the /ra-/la/ continuum. The rising frequency transition in  $F_3$  at the onset of the stimulus ranged from 0 Hz for /la/ to 1.2 kHz for /ra/, with seven intermediate  $F_3$ -transition stimuli. Budgerigars showed enhanced discrimination near the /ra-/la/ human perceptual boundary compared to stimuli falling within either phonetic category (Dooling *et al.*, 1995). Differences in discrimination performance were abolished for sinewave versions of the same stimuli, consistent with human results (Best *et al.*, 1989). These studies highlight similarities between budgerigars and humans in perceptual grouping of vowel-like sounds that probably help support the budgerigar's well-known ability to mimic human speech.

Previous studies in mammalian species suggest that formant-frequency discrimination thresholds are slightly less sensitive than human thresholds for matched stimuli with low  $F_0$ . Japanese macaques can detect an isolated frequency change of 2.5% for  $F_1$  and 1.6% for  $F_2$  (Sommers *et al.*, 1992), while cats can detect a 2.3%  $F_2$  change (Hienz *et al.*, 1996). Human formant-frequency discrimination thresholds were not studied using equivalent stimuli and procedures in these studies, but for similar low- $F_0$  stimuli, Kewley-Port and colleagues (1996) found thresholds of 2.1% for  $F_1$  (500 Hz standard) and 1.2% for  $F_2$ . Similarly, for their "between-harmonic" stimulus conditions, Lyzenga and Horst (1998) found thresholds of 0.8% for  $F_1$  and 1.1% for  $F_2$ . In another study that did use equivalent stimuli and test procedures in macaques and humans (Sinnott and Kreiter, 1991), thresholds for discrimination along three vowel continua were found to be 2 to 3 times higher in macaques than in human subjects. Vowel stimuli were synthesized with three formants along the /I-i/, /æ-ε/, and /a-Λ/ continua, each of which exhibits opposite changes in  $F_1$  and  $F_2$  (i.e.,  $F_1$  decreases whereas  $F_2$  increases).

The observation that formant-frequency discrimination in the macaque and cat "approaches" human behavioral performance contrasts markedly with findings for pure-tone stimuli in these species. Whereas humans can detect a 0.16% frequency change in a 2-kHz tone (Wier *et al.*, 1977), this threshold is 4.3% in macaques (Sommers *et al.*, 1992) and 5.9% in cats (Hienz *et al.*, 1993; note that Heinz and colleagues could not replicate the lower thresholds reported by Elliott *et al.*, 1960). Thus, while humans are more sensitive to frequency differences in tones than in formants, the macaque and cat show the opposite pattern, with greater formant-based sensitivity. In budgerigars, thresholds for pure-tone frequency discrimination range from 0.62%–0.75% for frequencies between 1 and 4 kHz (Dooling and Saunders, 1975). These pure-tone discrimination thresholds are lower than those found here for formant-frequency discrimination, suggesting that the budgerigar may conform more to the human pattern. Note however, that the previous study used different behavioral test procedures based on shock-avoidance conditioning (Dooling and Saunders, 1975). Therefore, the exact relationship between frequency discrimination of pure tones versus formants remains uncertain in the budgerigar.

Formant-frequency discrimination of the Klatt-synthesized vowels could be accomplished through sensitivity to the amplitude differences between harmonic frequency components near the formant peak. This amplitude difference is low when the formant falls between two harmonics, as was the case for the standard stimuli used here (see Fig. 1), and increases as the formant comes into alignment with a single harmonic (Lyzenga and Horst, 1997). Hence, in the present study, budgerigar and human test subjects may have relied on an increase in the amplitude difference between harmonic components near the formant peak to detect downward shifts in formant frequency. Note that the same amplitude-difference cue would not reliably support formant-frequency discrimination for more natural stimuli with sufficient variation in  $F_0$ . Therefore, the thresholds reported here, as in previous studies of synthetic vowel discrimination (Hienz *et al.*, 1996; Kewley-Port *et al.*, 1996; Kewley-Port and Watson, 1994; Lyzenga and Horst, 1997, 1998; Sinnott and Kreiter, 1991; Sommers *et al.*, 1992), should be viewed as the performance limit of the system observed under conditions of minimal stimulus uncertainty and optimal listening. Further studies are needed to determine the mechanisms that support discrimination and grouping of naturally spoken vowels.

Differences in cochlear anatomy between birds and mammals raise the question of whether the budgerigar discriminates complex sounds based on coding mechanisms also found in mammals. The cochlear duct is shorter in birds than in mammals (2.5 mm in the budgerigar; Manley *et al.*, 1993), uncoiled in structure, and relatively broad across the sensory epithelium with 50 or more hair cells spanning its width at some apical locations. While clearly lacking in high-frequency sensitivity (>5–6 kHz), auditory-nerve (AN) studies in the pigeon and starling show that response patterns are surprisingly similar to those in typical mammals (Manley *et al.*, 1985; Sachs *et al.*, 1974). AN fibers in these avian species exhibit irregular spontaneous discharge activity as in

mammals. Spontaneous discharge rates are unimodally distributed (versus bimodal in mammals) and range from 5–200 spikes/s across fibers (median: 50–90 spikes/s). Avian AN fibers phase lock to pure-tone frequencies up to 4–5 kHz and exhibit dynamic ranges (the sound level difference between threshold and response saturation) between 10 and 50 dB (mean: 25 dB; similar to cat data). Maximum discharge rates in birds are higher than those observed in mammals. Tuning curves are V-shaped and have similar sharpness to mammalian tuning curves but lack the low-frequency tail region typically observed in high-frequency mammalian fibers. Similar coding principles observed in the periphery extend to at the least the midbrain processing level, where recent studies highlight important parallels between avian and mammalian coding schemes (Woolley and Portfors, 2013). In the budgerigar IC, most recording sites exhibit V-shaped tuning in response to pure tones, with tuning sharpness similar to cat IC units (Henry *et al.*, 2016, 2017). Notably, many budgerigar IC sites show bandpass modulation tuning, where stimuli with modulation rates near the best modulation frequency of the cell evoke greater discharge activity than higher and lower modulation rates. Bandpass modulation tuning is also common in the IC of mammals (Joris *et al.*, 2004; Krishna and Semple, 2000; Langner and Schreiner, 1988; Nelson and Carney, 2007), and may contribute to the modulation filter bank of psycho-physical masking models (Dau *et al.*, 1997; Ewert *et al.*, 2002; Kay, 1982; Kay and Matthews, 1972).

## VI. CONCLUSIONS

The present study quantified formant-frequency discrimination thresholds in budgerigars and humans using the same behavioral test procedures in each species. Stimuli were synthesized with one or two formants and had spectral envelopes similar to natural speech. Discrimination thresholds in budgerigars were as sensitive as human thresholds for all stimulus conditions and followed the same pattern across conditions. Thresholds expressed as a percent frequency change were lower for  $F_2$  than  $F_1$ , and showed no consistent change with the presence of an additional stationary formant. Thresholds for simultaneous  $F_1$  and  $F_2$  changes in the same direction were lower than the threshold for either formant alone, suggesting that both species combine information between formant regions to discriminate combined formant differences. These findings, along with recent evidence for parallel auditory processing schemes in birds and mammals, highlight the budgerigar as an intriguing animal model for further behavioral and physiological studies of vowel discrimination.

## ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant No. R01-DC001641 to L.H.C. and Grant No. R00-DC013792 to K.S.H. The authors thank Douglas Schwarz for technical assistance, Caleb Connelly and Ixcel Valeriano for behavioral data collection, and Joyce McDonough for comments on the manuscript.

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* **67**, 1–48.
- Best, C. T., Studdert-Kennedy, M., Manuel, S. Y., and Rubin-Spitz, J. (1989). "Discovering phonetic coherence in acoustic patterns," *Percept. Psychophys.* **45**, 237–250.
- Bizley, J. K., Walker, K. M. M., King, A. J., and Schnupp, J. W. H. (2013). "Spectral timbre perception in ferrets: Discrimination of artificial vowels under different listening conditions," *J. Acoust. Soc. Am.* **133**, 365–376.
- Burdick, C. K., and Miller, J. D. (1975). "Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/," *J. Acoust. Soc. Am.* **58**, 415–427.
- Costa, C. E., and Cancado, C. R. X. (2012). "Stability check: A program for calculating the stability of behavior," *Mex. J. Behav. Anal.* **38**, 61–71.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). "Modeling auditory processing of amplitude modulation. I: Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.* **102**, 2892–2905.
- Dooling, R. J., Best, C. T., and Brown, S. D. (1995). "Discrimination of synthetic full-formant and sinewave/ra-la/continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*)," *J. Acoust. Soc. Am.* **97**, 1839–1846.
- Dooling, R. J., and Brown, S. D. (1990). "Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels," *Percept. Psychophys.* **47**, 568–574.
- Dooling, R. J., Brown, S. D., Park, T. J., Okanoya, K., and Soli, S. D. (1987). "Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): I. Pure tones," *J. Comp. Psychol.* **101**, 139–149.
- Dooling, R. J., and Saunders, J. C. (1975). "Hearing in the parakeet (*Melopsittacus undulatus*): Absolute thresholds, critical ratios, frequency difference limens, and vocalizations," *J. Comp. Physiol. Psychol.* **88**, 1–20.
- Elliott, D. N., Stein, L., and Harrison, M. J. (1960). "Determination of absolute-intensity thresholds and frequency-difference thresholds in cats," *J. Acoust. Soc. Am.* **32**(3), 380–384.
- Eriksson, J. L., and Villa, A. E. P. (2006). "Learning of auditory equivalence classes for vowels by rats," *Behav. Processes* **73**, 348–359.
- Ewert, S. D., Verhey, J. L., and Dau, T. (2002). "Spectro-temporal processing in the envelope-frequency domain," *J. Acoust. Soc. Am.* **112**, 2921–2931.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, Hague, The Netherlands).
- Green, D. M., Kidd, G., and Picardi, M. C. (1983). "Successive versus simultaneous comparison in auditory intensity discrimination," *J. Acoust. Soc. Am.* **73**, 639–643.
- Hawks, J. W. (1994). "Difference limens for formant patterns of vowel sounds," *J. Acoust. Soc. Am.* **95**, 1074–1084.
- Henry, K. S., Abrams, K. S., Forst, J., Mender, M. J., Neilans, E. G., Idrobo, F., and Carney, L. H. (2017). "Midbrain synchrony to envelope structure supports behavioral sensitivity to single-formant vowel-like sounds in noise," *J. Assoc. Res. Otolaryngol.* **18**, 165–181.
- Henry, K. S., Neilans, E. G., Abrams, K. S., Idrobo, F., and Carney, L. H. (2016). "Neural correlates of behavioral amplitude modulation sensitivity in the budgerigar midbrain," *J. Neurophysiol.* **115**, 1905–1916.
- Hienz, R. D., Aleszczyk, C. M., and May, B. J. (1996). "Vowel discrimination in cats: Thresholds for the detection of second formant changes in the vowel /e/," *J. Acoust. Soc. Am.* **100**, 1052–1058.
- Hienz, R. D., and Brady, J. V. (1988). "The acquisition of vowel discriminations by nonhuman primates," *J. Acoust. Soc. Am.* **84**, 186–194.
- Hienz, R. D., Sachs, M. B., and Aleszczyk, C. M. (1993). "Frequency discrimination in noise: Comparison of cat performances with auditory-nerve models," *J. Acoust. Soc. Am.* **93**, 462–469.
- Hienz, R. D., Sachs, M. B., and Sinnott, J. M. (1981). "Discrimination of steady-state vowels by blackbirds and pigeons," *Acoust. Soc. Am.* **70**, 699–706.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Joris, P. X., Schreiner, C. E., and Rees, A. (2004). "Neural processing of amplitude-modulated sounds," *Physiol. Rev.* **84**, 541–577.
- Kay, R. H. (1982). "Hearing of modulation in sounds," *Physiol. Rev.* **62**, 894–975, available at <http://physrev.physiology.org/content/62/3/894>.
- Kay, R. H., and Matthews, D. R. (1972). "On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones," *J. Physiol.* **225**, 657–677.

- Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (2007). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* **122**, 2365–2375.
- Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. T. (1996). "Fundamental frequency effects on thresholds for vowel formant discrimination," *J. Acoust. Soc. Am.* **100**, 2462–2470.
- Kewley-Port, D., and Watson, C. S. (1994). "Formant-frequency discrimination for isolated English vowels," *J. Acoust. Soc. Am.* **95**, 485–496.
- Kewley-Port, D., and Zheng, Y. (1998). "Auditory models of formant frequency discrimination for isolated vowels," *J. Acoust. Soc. Am.* **103**, 1654–1666.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.
- Kriengwatana, B., Escudero, P., and ten Cate, C. (2015). "Revisiting vocal perception in non-human animals: A review of vowel discrimination, speaker voice recognition, and speaker normalization," *Front. Psychol.* **6**, 1543.
- Krishna, B. S., and Semple, M. N. (2000). "Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus," *J. Neurophysiol.* **84**, 255–273.
- Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages* (Wiley-Blackwell, Hoboken, NJ).
- Langner, G., and Schreiner, C. E. (1988). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," *J. Neurophysiol.* **60**, 1799–1822.
- Levitt, H. (1970). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lyzenga, J., and Horst, J. W. (1997). "Frequency discrimination of stylized synthetic vowels with a single formant," *J. Acoust. Soc. Am.* **102**, 1755–1767.
- Lyzenga, J., and Horst, J. W. (1998). "Frequency discrimination of stylized synthetic vowels with two formants," *J. Acoust. Soc. Am.* **104**, 2956–2966.
- Manley, G. A., Gleich, O., Leppelsack, H. J., and Oeckinghaus, H. (1985). "Activity patterns of cochlear ganglion neurones in the starling," *J. Comp. Physiol. A.* **157**, 161–181.
- Manley, G. A., Schwabedissen, G., and Gleich, O. (1993). "Morphology of the basilar papilla of the budgerigar, *Melopsittacus undulatus*," *J. Morphol.* **218**, 153–165.
- Mermelstein, P. (1978). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," *J. Acoust. Soc. Am.* **63**(2), 572–580.
- Nelson, P. C., and Carney, L. H. (2007). "Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus," *J. Neurophysiol.* **97**, 522–539.
- Sachs, M. B., Young, E. D., and Lewis, R. H. (1974). "Discharge patterns of single fibers in the pigeon auditory nerve," *Brain Res.* **70**, 431–447.
- Sinnott, J. M., and Kreiter, N. A. (1991). "Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans," *J. Acoust. Soc. Am.* **89**, 2421–2429.
- Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (1992). "Formant frequency discrimination by Japanese macaques (*Macaca fuscata*)," *J. Acoust. Soc. Am.* **91**, 3499–3510.
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 178–184.
- Woolley, S. M. N., and Portfors, C. V. (2013). "Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain," *Hear. Res.* **305**, 45–56.