



Published in final edited form as:

*J Comp Psychol.* 2011 November ; 125(4): 420–430. doi:10.1037/a0024396.

## Acoustic and Perceptual Categories of Vocal Elements in the Warble Song of Budgerigars (*Melopsittacus undulatus*)

Hsiao-Wei Tu, Edward W. Smith, and Robert J. Dooling

Department of Psychology, University of Maryland, College Park

### Abstract

The warble songs of budgerigars (*Melopsittacus undulatus*) are composed of a number of complex, variable acoustic elements that are sung by male birds in intimate courtship contexts for periods lasting up to several minutes. If these variable acoustic elements can be assigned to distinct acoustic-perceptual categories, it provides the opportunity to explore whether birds are perceptually sensitive to the proportion or sequential combination of warble elements belonging to different categories. By the inspection of spectrograms and by listening to recordings, humans assigned the acoustic elements in budgerigar warble from several birds to eight broad, overlapping categories. A neural-network program was developed and trained on these warble elements to simulate human categorization. The classification reliability between human raters and between human raters and the neural network classifier was better than 80% both within and across birds. Using operant conditioning and a psychophysical task, budgerigars were tested on large sets of these elements from different acoustic categories and different individuals. The birds consistently showed high discriminability for pairs of warble elements drawn from between acoustic categories and low discriminability for pairs drawn from within acoustic categories. With warble elements reliably assigned to different acoustic categories by humans and birds, it affords the opportunity to ask questions about the ordering of elements in natural warble streams and the perceptual significance of this ordering.

### Keywords

budgerigar; categorization; perceptual categories; warble

---

Bird vocalizations, especially the learned songs of oscines, have been intensively studied since the 1950s (see review in Marler, 2004) such that songbirds, for better or worse, represent the best, if not the only, animal models for understanding the neurobiological processes underlying human speech and language learning (Brainard & Doupe, 2002; Doupe & Kuhl, 1999; Goldstein, King, & West, 2003; Marler, 1970; Marler & Peters, 1981; Todt, 2004; Wilbrecht & Nottebohm, 2003). However, songbird songs are generally short (a few seconds) and stereotyped. Though there are exceptions, songbirds usually sing one or more “song types” (particular patterns of elements), sometimes repeatedly, in a song bout

---

Correspondence concerning this article should be addressed to Hsiao-Wei Tu, 954 Gatewood Road, Atlanta, GA 30329.

hsiaowei.tu@emory.edu.

Author Note: Hsiao-Wei Tu is now at Department of Psychology and Yerkes National Primate Research Center, Emory University.

(Catchpole & Slater, 2008). In most cases, the sequential order of song elements in a song is stereotyped and predictable so that little opportunity exists for information to be encoded by varying the order or combinations of song elements. Consider, as an extreme example, male zebra finches which sing one highly stereotyped, learned song throughout life (Zann, 1996). Bengalese finches sing a limited repertoire of stereotyped elements with a limited finite state syntax (e.g., Okanoya, 2004) and starlings show iterated motifs that appear to form basic perceptual units of a song (Gentner & Hulse, 1998, 2000). These songbird examples are in stark contrast to the syntactical rules for combining words that allow a vast range of expressions possible in human language (Kirby, 2002; Marler, 2000). Therefore, as broad models of human communication systems, vocal learning aside, the short, stereotyped nature of most bird song is one major limitation of songbird vocalizations as models of human communication.

By contrast, the warble song of a non-territorial non-songbird species, the budgerigar (*Melopsittacus undulatus*), offers some new avenues of exploration of a complex acoustic communication system. Budgerigars are small parakeets native to Australia. They usually live in large groups (from hundreds of individuals to over 25 thousand) and highly depend on vocal communication to coordinate social and reproductive behaviors in the flock (Brockway, 1964a, 1964b; Farabaugh & Dooling, 1996; Trillmich, 1976; Wyndham, 1980). The warble song of male budgerigars is a complex, melodic, multi-syllabic vocalization that can last as long as several minutes at an average rate of two to three syllables per second (Brockway, 1964b; Farabaugh, Brown, & Dooling, 1992). A representative stream of warble song from a male budgerigar lasting 20 seconds is shown in Figure 1. Even a cursory examination of this song shows that similar acoustic elements recur throughout a long song bout. But these elements are sufficiently variable within and across birds that they are not easily assigned to exclusive categories (Farabaugh et al., 1992). Finally, unlike the song of songbirds, warble song is typically delivered in close quarters during intimate interactions between individuals, especially mates, and accompanied with other intimate behaviors (Brockway, 1964b, 1965, 1969). All of these features of budgerigar acoustic communication offer new potential parallels with human speech.

Despite the importance of warble to budgerigars, not much is known about how birds learn different warble elements, how they perceive them, or whether there is any significance to the ordering of these elements. Farabaugh and her colleagues (1992) categorized budgerigar warble elements by having human subjects sort spectrograms of warble syllables. In this analysis, approximately 300 syllables per bird were recorded from 6 normal males, 2 isolated males, and 1 female and pooled together for a total of 2800 syllables, which were then sorted into 42 groups. Among them, 15 were basic “elemental” units that could not be easily subdivided while another 27 groups were compound units where two or more elemental units were combined. Generally, these 15 basic groups could be further described as consisting of narrowband (including contact call-like elements), nonharmonic broadband (including alarm call-like elements), and harmonic broadband (Farabaugh et al., 1992) sounds. This study was the first to establish that budgerigar warble elements can be acoustically categorized, but it also showed that the acoustic categories are complex and overlapping. This leaves open the question as to whether budgerigars would perceive warble

elements as belonging to different acoustic categories. The extent to which the acoustic categories of warble elements defined by human observers are useful in exploring vocal communication in budgerigars is largely dependent on whether the birds treat these acoustic categories warble elements as perceptual categories.

Defining distinct acoustic categories is typically an important first step in analyzing an animal's vocal repertoire (Deecke & Janik, 2006). If there are no demonstrable acoustic categories, then every sound may be perceived as unique (Roitblat & von Fersen, 1992), and there would be no opportunity to convey information beyond the single element as, for instance, in the proportion or sequence of elements. Thus, the first goal of the present study was to confirm, using a much larger set of warble elements, the Farabaugh et al. (1992) finding that budgerigar warble elements can be assigned to distinct acoustic categories with high agreement among human raters. To do this, we developed a neural network-based classification program that reliably and efficiently assigns large numbers of warble elements into human-defined acoustic categories. Once warble elements were assigned to acoustic categories, the second goal of the present experiments was to determine whether the birds perceive these warble elements as if they belong to the acoustic categories as defined by human raters and the neural network.

## General Method

### Vocal Recording

Warble songs were obtained from four adult male budgerigars. Three birds (Buzz, Ricky, and Puffy) were initially housed together with about 40 other budgerigars in a large aviary at the University of Maryland. Warble for the fourth was obtained from archival recordings from Farabaugh et al. (1992). The three male budgerigars were observed warbling to a specific female on several occasions, indicating pair bonding. Approximately four weeks prior to the start of recording, these three pairs were removed from the large flock and housed together to promote pair bonding and increase male warbling. Animals had *ad libitum* access to both food and water at all times.

One male budgerigar was recorded at a time. The selected pair was separated and placed in a small animal acoustic isolation chamber (Industrial Acoustic Company model AC-1, New York, NY) over night before recording. In the next morning, the doors of the chambers were opened and a recording session was begun. To stimulate warbling by the male, a low-amplitude recording of all the birds from the budgerigar flock room was played softly in the background and the subject's mate was placed in close proximity (Brockway, 1964b). While females do not warble, to avoid any contamination from any vocalizations of the female, a single directional microphone (Audio-Technica Pro 35ax clip-on instrument microphone, Audio-Technica, Inc., Stow, OH) was aimed at the male's cage such that the female was always behind the microphone. All vocalizations were stored as a single channel of a PCM WAV file at a sampling rate of 48 kHz on a Marantz PMD670 digital recorder (Marantz America, Inc., Mahwan, NJ). An aggregation of more than one hour of warble was collected over approximately four hours of recording for each bird.

Warble from the fourth bird, Yuri, was obtained from archival recordings from the Farabaugh et al. (1992) study as a check on whether the present methods of classification gave similar results to the previous method from our laboratory based on the sorting of sound spectrograms. Recordings were digitized at a 48 kHz sampling rate and stored on a computer together with the new warble files.

### Segmentation of Warble

A MATLAB program was created to segment long streams of warble recording into individual elements. This program advanced through each warble recording, computing root-mean-square amplitude values using a 0.83 ms window. From the resulting amplitude envelope of the whole warble song, an intensity threshold was determined based on the overall amplitude of the recording. If signal amplitude exceeded this threshold for longer than 1 ms, bordered by inter-segment intervals greater than 25 ms, the signal was considered as a warble element. Inter-segment intervals shorter than 25 ms were considered amplitude modulation within a single element. These parameters were chosen for the segmentation program so as to give results that most closely and consistently matched those of human observers.

### Experiment 1: Acoustic Categorization

Once individual elements were extracted from warble streams, we used a combination of human classification and neural network classification to assign warble elements to categories. First, three human observers categorized a subset of all the warble elements we recorded. Then a neural network was trained using the results of human categorization to adjust the weights that make additive connections among the neurons in the model. This is not a new process since neural network models have recently been applied to the classification of animal vocalizations with some success (e.g., Dawson, Charrier, & Sturdy, 2006; Ranjard & Ross, 2008).

### Method

**Vocal stimuli**—The vocalizations of three birds, Ricky, Puffy, and Yuri were used to develop the neural-network classification procedure described below. Ricky and Puffy were from an existing colony in the vivarium. Yuri, on the other hand, was a budgerigar recorded by Farabaugh et al. (1992). Data from this bird was also used in human classification and to develop the network so that it could handle variation across birds. Once trained, the neural network classification procedure was tested for validity against the warble elements from the fourth bird, Buzz.

**Preparation of the training set**—Three human raters, experienced with classification of budgerigar vocalizations, were asked to categorize a random subset of 860 warble segments from three budgerigars (283 segments from Puffy; 291 segments from Ricky; 286 segments from Yuri). Raters both listened to the vocalization and viewed the spectrogram in completing the task. A MATLAB program was developed to aid in the classification of warble segments. In short, users listened to the vocalizations as needed while examining spectrograms of the sound on the computer screen in order to assign them to an open-ended

number of groups. Each segment could only be assigned to one group and raters were encouraged to ‘clump’ as much as possible (i.e., to reduce the number of groups). Users were also required to give a description, based on sound and the spectrogram, of each group after classification. After humans categorized these 860 elements, the groups were saved and used to train the neural network to operate as an automatic classification machine. Finally, a subset of 5 warble sessions was randomly selected and used to calculate an inter-rater reliability score.

**An artificial neural network-based voting pool**—To assist humans in classifying warble elements so that large numbers of warble elements could be processed, a classifier was developed using twenty measures extracted from each element. Given that some of these measures are correlated with one another, a voting pool of three- and four-layer feed forward neural networks was used (Battiti & Colla, 1994). This approach provided twenty-six artificial neural networks (ANNs). Each network had twenty inputs – one for each measure, and one output – for the class to which this network assigns a segment. In order to make each network structurally different, the number of hidden layers varied from one to two, and the number of hidden neurons in each hidden layer varied from eight to twenty.

Eight hundred and sixty segments previously classified by three experienced humans were parsed with the Matlab function “dividevec” and used to train the ANNs by the MATLAB training function “train.” After training, new warble elements other than these 860 segments in the training set were input for classification. When classifying one segment, all twenty measures extracted from that segment were presented to each ANN, and each ANN independently made a classification decision. A majority-rule vote across all independent classifications from the pool of 26 ANNs was then taken to make the final decision about category membership for that segment. Consistency between ANNs in the pool was observed to vary, which is in general agreement with the recommendation of Battiti & Colla (1994).

**Input measures for the ANN-based voting pool classifier**—Twenty measures roughly followed the descriptions of those in Avisoft-SASLab Pro Version 4.40 (Berlin, Germany) were input to the classifier to cover a wide range of measures typically used in bird song analysis.

- ***Spectral Roughness*** measures the degree of spectral variation over time in a segment. It counts the number of times the instantaneous frequency varies significantly from a running average frequency. Spectral Roughness is unitless.
- ***Tonality*** provides an indication of the extent to which a segment is a pure tone at each moment in time. Tonality is computed as the ratio of strong spectral lines' strength to total acoustic power in a segment. The units of Tonality are dBs.
- ***Duration*** is the temporal extent of a segment, in milliseconds.
- ***Harmonic Strength*** is similar to Tonality, but Harmonic Strength considers the total strength of all spectral lines, rather than just strong spectral lines.

- **1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> Frequency Quartiles** are based on the power spectrum of a segment. The 1<sup>st</sup> Frequency Quartile indicates the lowest frequency that bounds one fourth of the total acoustic power in the segment; the second quartile bounds one half of the acoustic power, and the third quartile bounds three-fourths of the power. The units for Frequency Quartiles are Hertz.
- **Skewness of Power:** For a segment, the instantaneous power vector is computed for each sample point by squaring the amplitude of each sample point. The Matlab skewness function is then applied to the instantaneous power vector. Skewness of Power is a unitless number.
- **Zero-Crossing Frequency** describes the frequency content of a segment. It is based on the number of times the segment waveform crosses zero on the voltage axis. The units of Zero-Crossing Frequency are Hertz.
- **Average Peak Spacing** provides information about how far apart local frequency maxima in a segment are. The units of Average Peak Spacing are Hertz.
- **Amplitude Modulation** provides a notion of the short-term temporal variation in a segment's amplitude. The segment is divided into twenty equal-duration subsegments, and absolute differences in the powers of consecutive segments are averaged. The units of Amplitude Modulation measure are power per second.
- **Number of Harmonic Lines** is an attempt to describe the tonal complexity of a segment. The number of spectral lines is counted. This measure is unitless.
- **Frequency of Maximum Amplitude** is based on the power spectrum of the segment. The frequency of the strongest single spectral bin is reported. There could well be cases in which most of a segment's power is not at a frequency close to this value. The units of Frequency of Maximum Amplitude are Hertz.
- **80% Bandwidth** is based on the power spectrum of an entire segment. To compute it, the minimum number of power spectral bins which together contain 80% of the total segment power is counted. Those bins need not be contiguous. The units are Hertz.
- **Entropy** is based directly on the digital samples of the segment. First, the ratio of the geometric mean of the absolute values to the arithmetic mean of the absolute values is computed. Entropy is the base e logarithm of this ratio, and it is unitless.
- **Time to Peak Amplitude** is the amount of time between the onset of a segment and the maximum instantaneous power of the segment. The units are milliseconds.
- **Kurtosis of Power:** For a segment, the instantaneous power vector is computed for each sample point by squaring the amplitude of each sample point. The Matlab kurtosis function is then applied to the instantaneous power vector. Kurtosis of Power is a unitless number.
- **Frequency Modulation** indicates how much a signal varies in frequency over time. It is computed by dividing a segment into twenty equal-duration subsegments, and



identifying the peak-amplitude frequency of each subsegment. The frequency changes between consecutive segments are summed. Units are Hertz per second.

- **Standard Deviation of Power** is based on the distribution of the power in a segment. It is standard deviation of the power in each sample in the segment. The units are watts.
- **Average Power Per Sample** is the mean of the power of each sample in the segment. The units are watts.

**Procedure**—After the neural network was trained and incorporated into the classification program, 500 segments from each of the four budgerigars (Ricky, Puffy, Yuri and Buzz) were chosen randomly and categorized both by this automatic classifier and by human experimenters. The resulting human-program reliability ensured whether the classifier was in agreement with human sorters.

Next, the relative merit of each measure was evaluated by the extent of the change in the grouping of the training set before and after each of the measures was excluded in the neural network. If a measure was crucial to “correctly” group the elements according to human criteria, eliminating it in the program would result in large differences in the grouping of elements (i.e., more elements would be put into the “wrong” groups). In the end, all segments from these four budgerigars were categorized, and the proportion of elements in each warble element category was compared among individuals by a Chi-Square test.

## Results

Table 1 shows the relative merit of each acoustic measure used in the automatic categorization procedure. Quality parameters such as spectral roughness and tonality were relatively more important than other acoustic features, while amplitude parameters like standard deviation of power and average power per sample were less influential in categorization.

With an inter-rater category reliability of 89.3%, three human raters categorized 860 warble segments into 7 elemental groups and two “special” groups – one contained segments that have a contact call-like element immediately followed by a broadband sound (Group H) and the other was cage noise produced by the bird moving during recording. The raters' descriptions of the 7 elemental groups were as follows: A) alarm call-like elements which were loud, broadband non-harmonic sounds, approximately 100 ms in duration; B) contact call-like elements which were frequency-modulated tonal sounds, approximately 100-300 ms; C) long harmonic calls which were harmonic sounds longer than 100 ms; D) short harmonic calls which were any harmonic sound that is shorter than 100 ms; E) “noisy” calls which were any broadband sound that sounds noisy (not harmonic) and were shorter than 70 ms; F) clicks which were extremely short broadband calls that sound like clicks; and G) pure tone-like elements which were elements that show little or no frequency modulation. The representative sound spectrograms of each group are shown in Figure 2. Moreover, the average human-program reliability in sorting 2000 segments (500 segments from each bird) was 83.2% (82.2% for Puffy, 88.2% for Ricky, 81.6% for Yuri, and 80.8% for Buzz).

All warble segments recorded from the four birds were categorized by the classification program. Eliminating those sounds identified as cage noise left 7357 segments in Puffy's warble streams, 5633 segments in Ricky warble, 7204 segments in Yuri warble, and 6027 segments in Buzz warble. Overall, contact call-like elements (Group B) were the most common segments, comprising 33% of warble; while pure tone-like elements (Group G) were the least common segments, comprising only 4% of warble.

Figure 3 shows that the distribution of elements in the eight categories varied considerably across individuals ( $\chi^2_{(21)} = 1379.89, p < 0.001$ ). Most of the variations between individuals existed in the noisy group (Group E). Yuri, the bird whose warble streams were used in the study of Farabaugh et al. (1992), differed from other three birds only in the number of clicks (Group F) and pure tone-like segments (Group G) produced. Approximately 15% of Buzz, Ricky, and Puffy's warble were clicks, but only about 10% of Yuri warble were clicks. Yuri had less than 1% of pure tone-like segments, but Buzz, Ricky, and Puffy had 7%, 4%, and 7% respectively.

## Discussion

The first goal of this experiment was to define acoustic categories for budgerigar warble elements that were consistent across raters. With the aid of a neural network classifier, we could simulate human categorization decisions and process a much larger number of warble elements than used in the Farabaugh et al. (1992) study. Our human raters agreed on eight acoustic categories (seven elemental categories and one compound category). This is fewer than the 15 elemental groups found in the Farabaugh et al. (1992) study, but reasonable in view of the fact that nine groups in this previous work contained no more than 2% of all the warble samples. The major differences between the two studies are the stringent interval criterion in the present study and the instructions given to the raters to 'lump' rather than 'split' in categorizing warble elements. This instruction reduced the number of categories with only a few elements in them. Most importantly, the resulting categories were 'universal' in that they could be reliably applied to warble from different budgerigars. Of the 27 compound classes from earlier work (Farabaugh et al., 1992), only one compound group (a contact call-like element immediately followed by a broadband sound) was found in the present analysis. This was due to greater precision in applying a criterion for the inter-element-interval during segmentation, which was set to minimize the number of compound segments and to produce as many single elements as possible.

When all warble streams were analyzed, the results showed that the relative proportion of elements in each category varied across the four birds analyzed (Figure 3). The category with the largest variation in the present analysis was the noisy group. This is perfectly consistent with the earlier analysis in that this group is a combination of four nonharmonic broadband subgroups found in the previous study by Farabaugh and her colleagues (1992). In both the present analysis and the Farabaugh et al. (1992) study, Group B (contact call-like elements) was the largest group.

The relative merit of each acoustic measure used in the neural network categorization showed that quality parameters, especially spectral roughness and tonality, were weighted more heavily than other acoustic measures when human experimenters (and later the neural



network) sorted elements into categories. This, in fact, was generally similar to what Farabaugh et al. (1992) used to manually sort warble spectrograms. Here, the neural network technique as applied in this study also ended up relying on similar parameters in simulating the raters' decision when categorizing warble elements.

Overall, the average human-program reliability was 83.2%, which was acceptable given the large, overlapping variations in budgerigar warble elements. It is worth noting that the human-program correspondence for warble elements from the bird, Buzz, who was not included in the training set, was 80.8% showing that the automatic categorization procedure is applicable across birds without additional modification or training.

Budgerigars that live together show some sharing of their warble classes (Farabaugh et al., 1992). The similarities between the warble songs of our three budgerigars (Buzz, Ricky, and Puffy) who lived together in a large flock of 100 birds for a month, and Yuri (recorded almost two decades earlier) strongly support the fact that there are common elements and element classes across budgerigars. On the other hand, there are also differences. The relatively small number of pure tone-like elements in Yuri's warble, compared to the other three birds, suggests that this element may be a newly shared element within the current flock (Farabaugh et al., 1992).

## Experiment 2: Evidence for Perceptual Categories

The experiment above showed that budgerigar warble elements from multiple birds can be sorted into eight acoustic categories with a high degree of reliability between human raters, and this classification behavior can be simulated with a neural network categorization program. However, the existence of human-defined acoustic categories does not reveal whether these categories have any perceptual relevance for the bird. In other words, human groupings done by spectrogram analysis, even when supplemented with listening, do not tell us whether birds perceive the acoustic variations of these warble elements in a similar way.

The purpose of this experiment is to examine the perception of these acoustic categories by budgerigars to see whether the acoustic differences between these seven acoustic categories (the compound group was omitted from this experiment). Such findings would be evidence of perceptual categories for these sounds (Brown, Dooling, & O'Grady, 1988; Dooling, Park, Brown, Okanoya, & Soli, 1987). The experimental strategy employed here is to use a random selection of multiple warble elements (i.e. different birds, different recordings) from the same acoustic category as a repeating background, and then to use a random selection of multiple warble elements (i.e. different birds, different recordings) from all of seven acoustic categories, including the background category, as targets in a discrimination paradigm. If birds are more successful in discriminating other-category targets from the repeating background, and much less successful in discriminating same-category targets from the background, we can safely assume they are responding primarily to those relevant acoustic variations that reliably differentiated the background category from all other warble element categories and ignoring irrelevant acoustic variation within categories. Note that this is evidence of perception of acoustic categories as occurs when humans and animals categorize

natural and varying speech sounds such as vowels in the face of talker variation (e.g., Burdick & Miller, 1975; Kuhl & Miller, 1975; Ohms, Gill, van Heijningen, Beckers, & ten Cate, 2010). Our findings are not a demonstration of the strict form of categorical perception which requires both labeling and discrimination functions (e.g., Liberman & Mattingly, 1985). If birds are shown to attend to the acoustic variation across warble categories and ignore the acoustic variation within acoustic categories, it strongly suggests that they treat elements within the same acoustic category as the same.

## Method

**Subjects**—Four budgerigars (two males and two females), two zebra finches, and two canaries were trained to perform the repeating background discrimination task. They were housed individually in small cages and kept on a constant 12-12 light-dark cycle at the University of Maryland. Since food was used as reinforcement, they were maintained at approximately 85-90% of their free-feeding weight with *ad libitum* access to water at all times. None of the subjects had been previously housed with those budgerigars whose warble elements were used as stimuli in this experiment (Ricky, Puffy, and Yuri).

**Vocal stimuli**—All stimuli were randomly chosen from the warble segments of Puffy, Ricky, and Yuri that were analyzed in the previous experiment. The loudness of individual stimuli was calibrated with a Larson-Davis sound level meter (Model 825, Provo, UT) with a 20-foot extension cable attached to a ½ inch microphone. The microphone was positioned in the place normally occupied by the bird's head during testing.

Each test session consisted of 90 trials (70 target trials, and 20 sham trials where no target was presented). Each target stimulus was presented against a large repeating background set of other warble elements. To ensure an adequate representation of within category variation, the 70 target trials included 10 elements randomly selected from each of the 7 element categories. The 10 elements from each group were randomly drawn from a larger set of 30 elements (i.e., 10 elements from each bird). In all, the birds were run for 3 sessions for each background type so that all 30 target elements were used (10 elements per session). In other words, the total target set consisted of a total of 210 elements (7 categories × 10 elements × 3 birds).

The background set consisted of 150 elements randomly selected from one element category. The elements were evenly drawn from the 3 individuals (i.e., 50 elements from each bird). Since there were 7 categories, 7 different background sets were constructed (one for each element category). In total, each subject bird was tested for 21 test sessions (3 target sets × 7 background sets).

All told, in this experiment, there were a total of 28 possible pairs of group comparisons (7 within-category pairs and 21 between-category pairs). Since all possible combinations of background and target categories were tested and the birds received equal numbers of background and target categories, it follows that they were not trained to respond to particular category differences over others.

**Apparatus**—Birds were trained and tested in a small wire cage ( $23 \times 25 \times 16 \text{ cm}^3$ ) mounted in a sound-attenuated chamber (Industrial Acoustics Company, Bronx, NY, IAC-3). The test cage contained a perch on the floor, a food hopper in front of the perch, and a control panel with two microswitch response keys mounted vertically in front of the perch. The keys were approximately 5 cm apart and each key had an 8 mm light emitting diode (LED) attached.

The experiment was controlled by a PC microcomputer operating Tucker-Davis Technologies (TDT, Gainesville, FL) System III modules. Stimuli were stored digitally and output via a 2-channel signal processor (TDT, Model RX6) at a sampling rate of 24.4 kHz. Each signal was then output at a mean level of about 70 dB SPL with a 3 dB rove from a separate channel of the D/A converter to a separate digital attenuator (TDT, Model PA5), combined in an analog summer (TDT, SM5) and then amplified (Model D-75, Crown Audio, Inc., Elkhart, IN) to a loudspeaker (KEF Model 80V, GP Acoustics, Inc., Marlboro, NJ) in the sound-attenuated chamber. All test sessions were automated by a custom-made MATLAB program.

**Procedure**—Birds were trained to peck one key (observation key) repeatedly during a continuous presentation of the background set (multiple elements from the same category) and to peck the other key (report key) when they detect a token from the target set (multiple elements from all categories) presented alternately with the background tokens. The interval between onsets of two consecutive stimuli was 1 second. Since the duration of each stimulus was different, the inter-stimulus interval varied accordingly. To human listeners, this rate sounded slightly slower than the delivery rate in a nature warble song, but here we were more concerned about the perception of each sound element rather than the tempo of a long sequence.

If the bird detected the target and pecked the report key within 3 seconds (two background-target alternations), the food hopper was raised by activation of a solenoid for 1.5 seconds and the bird received access to food. This was recorded as a “hit” and the bird's response latency was recorded. If the bird failed to peck the report key, this was recorded as a “miss” and a response latency of 3 seconds was recorded. If the bird did not peck the report key during sham trials where no target was presented, it was recorded as a “correct rejection.” If the bird pecked to the report key when there was no target presented, it was recorded as a “false alarm” and punished with variable time-out period (2 to 10 seconds). The same trial (or next trial if it is a false alarm) resumed after the blackout period.

Generally, each session lasted for approximately 20 minutes. Sessions with false alarm rates higher than 20% were not used for analysis. Birds were tested twice a day, 5 days a week. For each bird, different experimental conditions were tested in a random order to avoid biases and practice effects.

**Data Analysis**—For each possible pair, birds' response latency and its behavior (hit/miss/correct rejection/false alarm) were recorded. For each budgerigar run on each type of background, average response latencies were calculated for all trials involving between-category discriminations and for those involving within-category discriminations. Misses

count as 3000 ms, the maximum response interval. The resulting data were compared by paired  $t$  test with the assumption that longer latencies reflect more similar perception of the two stimuli, and shorter latencies indicate greater differences between them (Dooling, Brown, Park, Okanoya, & Soli, 1987; Dooling & Okanoya, 1995; Dooling et al., 1987; Okanoya & Dooling, 1988).

On the other hand, for each type of background, all trials involving between category discriminations were pooled together for a hit rate and a false alarm rate, and so were those involving within category discriminations. The numbers were then used to derive a pooled  $d'$  for each of these two conditions by the formula:

$$d' = z(\text{Hit rate}) - z(\text{False alarm rate})$$

To avoid infinite values, 100% correct and 0% false alarm rates were converted to  $1/(2N)$  and  $1 - 1/(2N)$ , respectively, where  $N$  is the number of trials which the percentage was based on (Macmillan & Creelman, 2005).

The  $d'$  is a standard measure of a subject's sensitivity in discrimination experiments (Jesteadt, 2005; Macmillan & Creelman, 2005). When the number of trials is small, data can be combined across multiple stimuli, sessions, or subjects to derive a pooled  $d'$ . Pooled  $d'$  values are generally lower and less variable than average  $d'$  values (Macmillan & Creelman, 2005).

To evaluate differences in  $d'$  between two conditions, the standard error of each  $d'$  was calculated and used to construct a 95% confidence interval (CI) around each  $d'$  value. When the two 95% CIs overlapped, we conclude that there was no significant difference in the subject's sensitivity in these two discrimination tasks. On the other hand, no overlapping indicated a significant difference in discrimination sensitivity between the two stimulus sets (Macmillan & Creelman, 2005).

## Results

In fewer than 10% of the test sessions did birds have a false alarm rate higher than 20% and these sessions were discarded. Average response latencies for the four budgerigars are shown in Figure 5A. Budgerigars responded significantly faster (easier discrimination) on between category discriminations than on within category discriminations ( $t = 40.26$ ,  $p < 0.001$ ). These results are also reflected in the  $d'$  analysis. Figure 5B shows that for each element category, budgerigars were generally more sensitive to between category acoustic differences than within category acoustic differences.

Figure 6 shows that all three species were all significantly better at discriminating two elements drawn from different acoustic categories (budgerigar: 95% CI,  $d' = 3.49-3.76$ ; canary: 95% CI,  $d' = 1.99-2.28$ ; zebra finch: 95% CI,  $d' = 2.34-2.60$ ) than they were at discriminating two elements drawn from the same acoustic category (budgerigar: 95% CI,  $d' = 0.71-1.01$ ; canary: 95% CI,  $d' = 0.11-0.52$ ; zebra finch: 95% CI,  $d' = 0.08-0.44$ ).

Moreover, budgerigars performed significantly better than zebra finches and canaries at both within-category and between-category discriminations.

## Discussion

The birds in these experiments were all tested in a discrimination paradigm. We infer the existence of perceptual categories from these discrimination data when birds have significantly greater difficulty discriminating variations among stimuli within the same category than they do discriminating between stimuli that span two categories (Goldstone, 1994; Horn & Falls, 1996). The present results show that acoustic categories of budgerigar warble elements also represent perceptual categories for the budgerigars. In other words, these data show that birds focus on the essential acoustic differences between categories and ignore the irrelevant acoustic differences that occur across multiple renditions of the same warble element or across elements produced by different individuals.

These acoustic categories are obviously quite robust since both zebra finches and canaries also show perceptual behavior that correlates with these human-defined acoustic categories of warble. This should not be too surprising as there are many examples of mammals and birds perceiving the categories of human speech sounds as humans do (e.g., Burdick & Miller, 1975; Dooling & Brown, 1992). Nevertheless, there are interesting species differences in the present results as well. Budgerigars performed at a higher level on both between and within category discriminations than either zebra finches or canaries. These results are consistent with earlier studies on discrimination of conspecific and heterospecific contact calls in these three species (Dooling & Brown, 1992; Okanoya & Dooling, 1991). In these earlier studies, all three species showed clear perceptual categories corresponding to the calls of each species, but the birds were also more efficient (i.e., shorter response latencies) at discriminating among contact calls of their own species compared with the contact calls of the other species.

## General Discussion

The warble song of male budgerigars is unique for its length, the complexity and overlapping nature of its acoustic elements, its multi-syllabic structure and temporal organization, and the fact that it is delivered in intimate social contexts. Similar acoustic elements recur throughout warble, raising the possibility that budgerigars might perceive these elements as belonging to one group or category. The present experiments used a combination of signal analysis, neural networks, and psychophysics to establish that budgerigars perceive their extremely long and variable warble songs as discrete components which can be assigned to at least seven basic acoustic-perceptual categories.

Categorization is strictly defined as the process in which ideas or objects are sorted according to the perceived similarity (Horn & Falls, 1996; Pothos & Chater, 2002), and it is a critical aspect of communication where senders and receivers share the same code to exchange information (Horn & Falls, 1996; Seyfarth & Cheney, 2003; Smith, 1977). Sometimes, as is obviously the case of the acoustic signals of speech and many animal vocalizations, these categories are broad and overlapping in acoustic space. In human speech, there is considerable variability in different utterances of the same vowel due to

different speakers, contexts, rates of speaking, etc., but the perceptual identification of these tokens is highly accurate despite this considerable variation (Hillenbrand, Getty, Clark, & Wheeler, 1995; Pickett, 1999). The acoustic-perceptual categories of human speech are quite robust since a variety of vertebrates have been shown to be sensitive to some of these categories in the face of within-category variation in talker, intonation, and gender (e.g., Burdick & Miller, 1975; Dooling & Brown, 1990; Kluender, Diehl, & Killeen, 1987). Similarly, the seven acoustic categories of budgerigar warble elements are robust in that humans, canaries, and finches, also behave as if they perceive these acoustic categories roughly as budgerigars do.

Budgerigar warble elements are acoustically variable and overlapping in acoustic space presenting some of the challenges often seen in studies of the acoustic-perceptual categories of natural human speech (e.g., Lisker & Abramson, 1964; Peterson & Barney, 1952). Having demonstrated the existence of acoustic categories for warble elements, the perceptual problem to be addressed is whether birds perceive these as belonging in groups or categories in spite of extensive within-category acoustic variations. Here, using psychophysical tests with multiple elements from different acoustic categories produced on different occasions by different birds, we show that budgerigars behave as if they have perceptual categories corresponding to the acoustic of their own vocalization.

The stimulus sets and the task adopted in these experiments tested natural, multidimensional variation with the warble element categories and, in this sense, is analogous to the perception of different vowels and consonants spoken by different speakers and in different contexts. As with human listening to speech, we would assume that budgerigars can perceive these categories of warble elements at a rate that is at least as fast as the normal delivery rate of warble, though this has yet to be tested.

The fact that these seven acoustic-perceptual categories obtain across individuals is a promising first step and may serve as a platform to begin to investigate the importance of element frequency, combinations, and ordering in long warble streams. Even the most cursory review of warble shows that these elements are arranged in a variety of ways though all elements do not all occur with the same frequency (Farabaugh et al., 1992; Tu & Dooling, 2010). To extend the potential parallels with human speech, the uniqueness of human language clearly lies in the syntactical and grammatical capabilities that allow us to string together a limited number of words using rule-based orders to create a vast number of sentences for intra-species communication. The case for any kind of a similar capability remains to be made for budgerigars.

Finally, male warble song, delivered in close quarters, is a critical component of a very elaborate set of courtship behaviors in budgerigars (Brockway, 1964b, 1969). It would not be hard to imagine that sexual selection drives this process and there is an underlying structure to the sequences of male warble in response to female preference for more complex songs such as those of mockingbirds, blackbirds (see review in Searcy & Yasukawa, 1996), or Bengalese finches (Okanoya, 2004). With the current findings that elements can be reliably sorted into acoustic groups, and budgerigars respond as if they perceived these elements in categories, it is now possible to examine whether the sequences



of elements in natural warble have perceptual significance for the bird. There is a number of obvious potential information-bearing aspects of warble song including the relative proportion of different elements, grouping of elements, or sequences of elements. It would be interesting to know whether a bird produces the same proportion of elements on different occasions, whether different birds always show roughly the same proportion of elements, and whether female behavior, either on a short term or long term basis, alters the frequency or pattern of warble elements produced by males.

## Acknowledgments

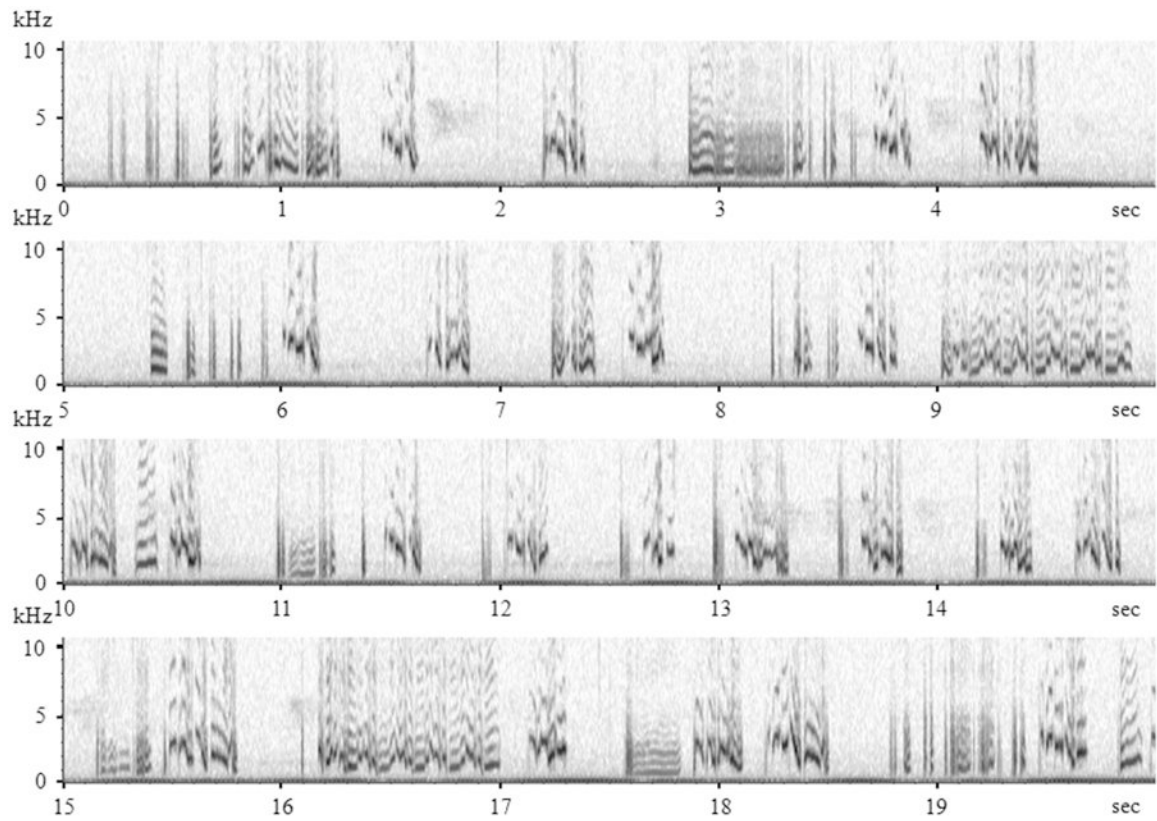
This Work was supported by NIH/NIDCD R01-DC 000198 to RJD. We wish to thank Daniel Lindsey and Christina Takara for help in collecting data; Marjorie Leek, Gerald Wilkinson, and Juan Uriagereka for comments on an earlier draft.

## References

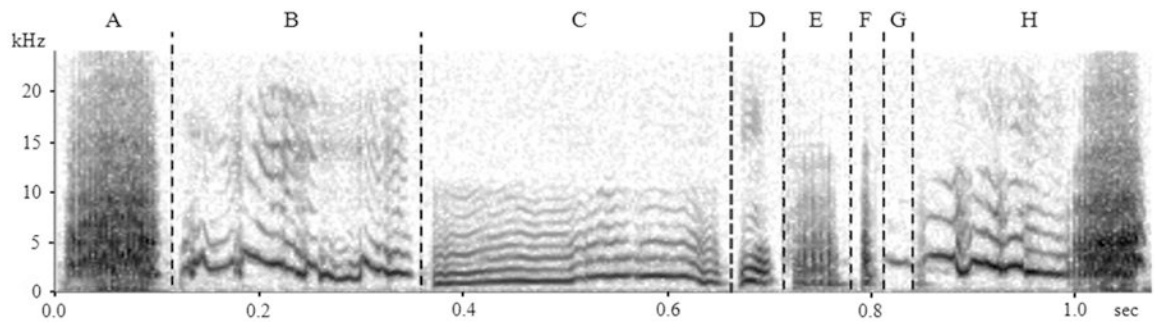
- Battiti R, Colla AM. Democracy in neural nets: Voting schemes for classification. *Neural Networks*. 1994; 7:691–707.
- Brainard MS, Doupe AJ. What songbirds teach us about learning. *Nature*. 2002; 417:351–358. [PubMed: 12015616]
- Brockway BF. Ethological studies of the budgerigar: non-reproductive behavior. *Behaviour*. 1964a; 22:193–222.
- Brockway BF. Ethological studies of the budgerigar: reproductive behavior. *Behaviour*. 1964b; 23:294–324.
- Brockway BF. Stimulation of ovarian development and egg laying by male courtship vocalization in budgerigars (*Melopsittacus undulatus*). *Animal Behaviour*. 1965; 13:575–578. [PubMed: 5894177]
- Brockway, BF. Roles of budgerigar vocalization in the integration of breeding behavior. In: Hinde, RA., editor. *Bird Vocalizations*. London: Cambridge University Press; 1969. p. 131-158.
- Brown SD, Dooling RJ, O'Grady K. Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): III. Contact calls. *Journal of Comparative Psychology*. 1988; 102:236–247. [PubMed: 3180731]
- Burdick CK, Miller JD. Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*. 1975; 58:415–427. [PubMed: 1184835]
- Catchpole, CK.; Slater, PJB. *Bird Song: Biological Themes and Variations*. 2. New York, NY: Cambridge University Press; 2008.
- Dawson MRW, Charrier I, Sturdy CB. Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types. *Journal of the Acoustical Society of America*. 2006; 119:3161–3172. [PubMed: 16708971]
- Deecke VB, Janik VM. Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls. *Journal of the Acoustical Society of America*. 2006; 119:645–653. [PubMed: 16454318]
- Dooling RJ, Brown SD. Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels. *Perception & Psychophysics*. 1990; 47:568–574. [PubMed: 2367177]
- Dooling RJ, Brown SD. Auditory perception of conspecific and heterospecific vocalizations in birds: Evidence for special processes. *Journal of Comparative Psychology*. 1992; 106:20–28. [PubMed: 1555398]
- Dooling RJ, Brown SD, Park TJ, Okanoya K, Soli SD. Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): I. Pure tones. *Journal of Comparative Psychology*. 1987; 101:139–149. [PubMed: 3608422]
- Dooling, RJ.; Okanoya, K. Psychophysical methods for assessing perceptual categories. In: Klump, GM.; Dooling, RJ.; Fay, R.R.; Stebbins, W.C., editors. *Methods in Comparative Psychoacoustics*. Basel, Switzerland: Birkhäuser Verlag; 1995. p. 307-318.

- Dooling RJ, Park TJ, Brown SD, Okanoya K, Soli SD. Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals. *Journal of Comparative Psychology*. 1987; 101:367–381. [PubMed: 3691060]
- Doupe AJ, Kuhl PK. Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*. 1999; 22:567–631.
- Farabaugh SM, Brown ED, Dooling RJ. Analysis of warble song of the budgerigar *Melopsittacus undulatus*. *Bioacoustics*. 1992; 4:111–130.
- Farabaugh, SM.; Dooling, RJ. Acoustic communication in parrots: Laboratory and field studies of budgerigars, *Melopsittacus undulatus*. In: Kroodsma, DE.; Miller, EH., editors. *Ecology and Evolution of Acoustic Communication in Birds*. Ithaca, NY: Cornell University Press; 1996. p. 97-120.
- Gentner TQ, Hulse SH. Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour*. 1998; 56:579–594. [PubMed: 9784206]
- Gentner TQ, Hulse SH. Perceptual classification based on the component structure of song in European starlings. *Journal of the Acoustical Society of America*. 2000; 107:3369–3381. [PubMed: 10875382]
- Goldstein MH, King AP, West MJ. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*. 2003; 100:8030–8035.
- Goldstone R. Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*. 1994; 123:178–200. [PubMed: 8014612]
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*. 1995; 97:3099–3111. [PubMed: 7759650]
- Horn, AG.; Falls, JB. Categorization and the design of signals: The case of song repertoires. In: Kroodsma, DE.; Miller, EH., editors. *Ecology and Evolution of Acoustic Communication in Birds*. Ithaca, NY: Cornell University Press; 1996. p. 121-135.
- Jesteadt W. The variance of  $d'$  estimates obtained in yes-no and two-interval forced choice procedures. *Perception & Psychophysics*. 2005; 67:72–80. [PubMed: 15912873]
- Kirby, S. Learning, bottlenecks and the evolution of recursive syntax. In: Briscoe, T., editor. *Linguistic Evolution through Language Acquisition*. Cambridge, UK: Cambridge University Press; 2002. p. 173-204.
- Kluender KR, Diehl RL, Killeen PR. Japanese quail can learn phonetic categories. *Science*. 1987; 237:1195–1197. [PubMed: 3629235]
- Kuhl PK, Miller JD. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*. 1975; 190:69–72. [PubMed: 1166301]
- Lieberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition*. 1985; 21:1–36. [PubMed: 4075760]
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*. 1964; 20:384–422.
- Macmillan, NA.; Creelman, CD. *Detection Theory: A User's Guide*. 2nd. Mahwah, NJ: Lawrence Erlbaum Associates; 2005.
- Marler P. Birdsong and speech development: Could there be parallels? *American Scientist*. 1970; 58:669–673. [PubMed: 5480089]
- Marler, P. Origins of music and speech: Insights from animals. In: Wallin, NL.; Merker, B.; Brown, S., editors. *The Origins of Music*. Cambridge, MA: Massachusetts Institute of Technology; 2000. p. 31-48.
- Marler, P. Science and birdsong: the good old days. In: Marler, P.; Slabbekoorn, H., editors. *Nature's Music: The Science of Birdsong*. San Diego, CA: Elsevier Academic Press; 2004. p. 1-38.
- Marler, P.; Peters, S. Birdsong and speech: Evidence for special processing. In: Eimas, PD.; Miller, JL., editors. *Perspectives on the Study of Speech*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1981. p. 75-112.
- Ohms VR, Gill A, van Heijningen CAA, Beckers GJL, ten Cate C. Zebra finches exhibit speaker independent phonetic perception of human speech. *Proceedings of the Royal Society B*. 2010; 277:1003–1009. [PubMed: 19955157]

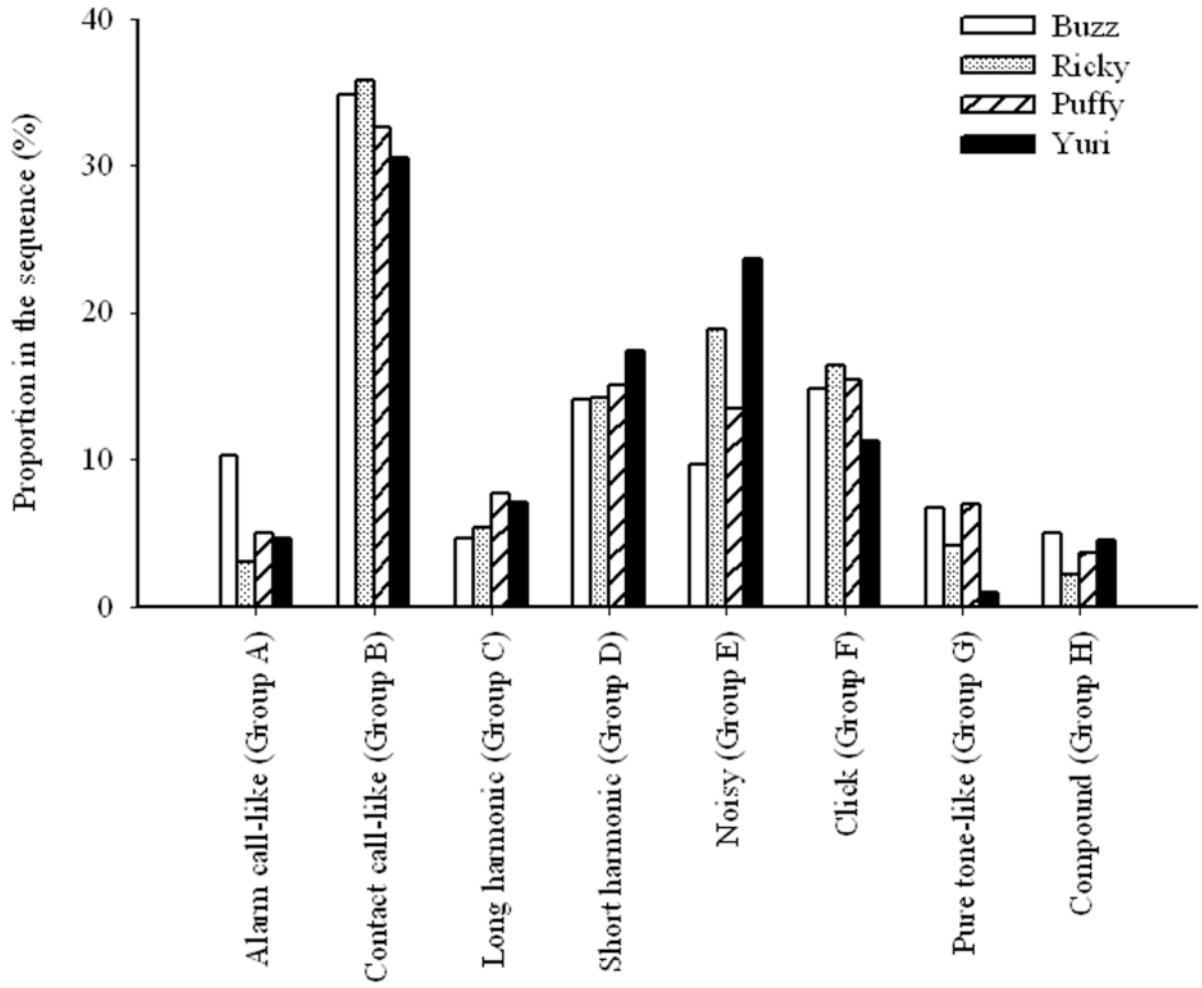
- Okanoya K. The Bengalese finch: A window on the behavioral neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences*. 2004; 1016:724–735. [PubMed: 15313802]
- Okanoya K, Dooling RJ. Obtaining acoustic similarity measures from animals: A method for species comparisons. *Journal of the Acoustical Society of America*. 1988; 83:1690–1693. [PubMed: 3372876]
- Okanoya K, Dooling RJ. Perception of distance calls by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*): Assessing species-specific advantages. *Journal of Comparative Psychology*. 1991; 105:60–72. [PubMed: 2032457]
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Pickett, JM. Prosodic and tonal features. In: Pickett, JM., editor. *The Acoustics of Speech Communication*. Needham Heights, MA: Allyn & Bacon; 1999. p. 75-98.
- Pothos EM, Chater N. A simplicity principle in unsupervised human categorization. *Cognitive Science*. 2002; 26:303–343.
- Ranjard L, Ross HA. Unsupervised bird song syllable classification using evolving neural networks. *Journal of the Acoustical Society of America*. 2008; 123:4358–4368. [PubMed: 18537387]
- Roitblat HL, von Fersen L. Comparative cognition: Representations and processes in learning and memory. *Annual Review of Psychology*. 1992; 43:671–710.
- Searcy, WA.; Yasukawa, K. Song and female choice. In: Kroodsma, DE.; Miller, EH., editors. *Ecology and Evolution of Acoustic Communication in Birds*. Ithaca, NY: Cornell University Press; 1996. p. 454-473.
- Seyfarth RM, Cheney DL. Signalers and receivers in animal communication. *Annual Review of Psychology*. 2003; 54:145–173.
- Smith, WJ. *The behavior of communicating: an ethological approach*. Cambridge, MA: Harvard University Press; 1977.
- Todt D. From birdsong to speech: a plea for comparative approaches. *Annals of the Brazilian Academy of Sciences*. 2004; 76:201–208.
- Trillmich F. Spatial proximity and mate specific behaviour in a flock of budgerigars (*Melopsittacus undulatus*; Aves, Psittacidae). *Zeitschrift für Tierpsychologie*. 1976; 41:307–331.
- Tu HW, Dooling RJ. Perception of alterations in natural budgerigar (*Melopsittacus undulatus*) warble: Implications of animal “syntactical capability”. Paper presented at the 159th Meeting of the Acoustical Society of America. 2010
- Wilbrecht L, Nottebohm F. Vocal learning in birds and humans. *Mental Retardation and Development Disabilities Reserach Reviews*. 2003; 9:135–148.
- Wyndham E. Diurnal cycle, behavior, and social organization in the budgerigar (*Melopsittacus undulatus*). *Emu*. 1980; 80:25–33.
- Zann, RA. *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. New York, NY: Oxford University Press; 1996.



**Figure 1.**  
A spectrogram of a 20-second selection of warble recorded from a male budgerigar;  
frequency in kHz on the y-axis and time in second on the x-axis.



**Figure 2.**  
Sonograms of examples of the seven elemental groups (A–G) and the compound group (H) which is a contact call-like element immediately followed by a broadband sound.



**Figure 3.**  
A comparison of the distribution of the elements in different categories among individuals.

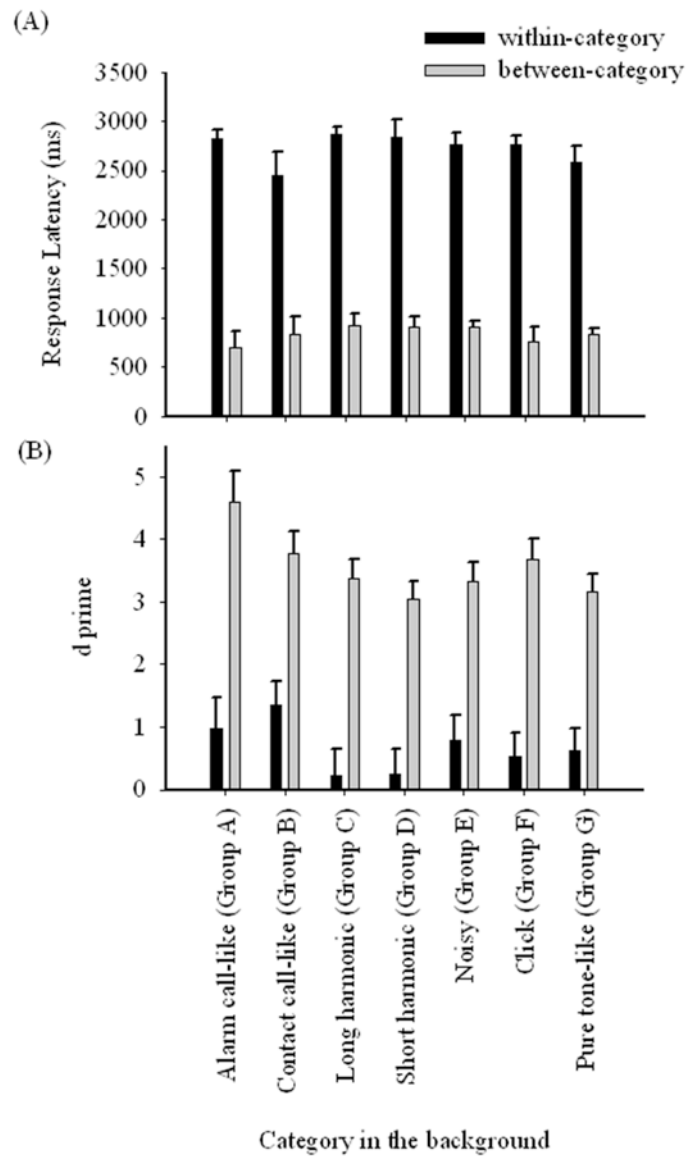


## DISTRIBUTION OF TARGET AND BACKGROUND SOUNDS

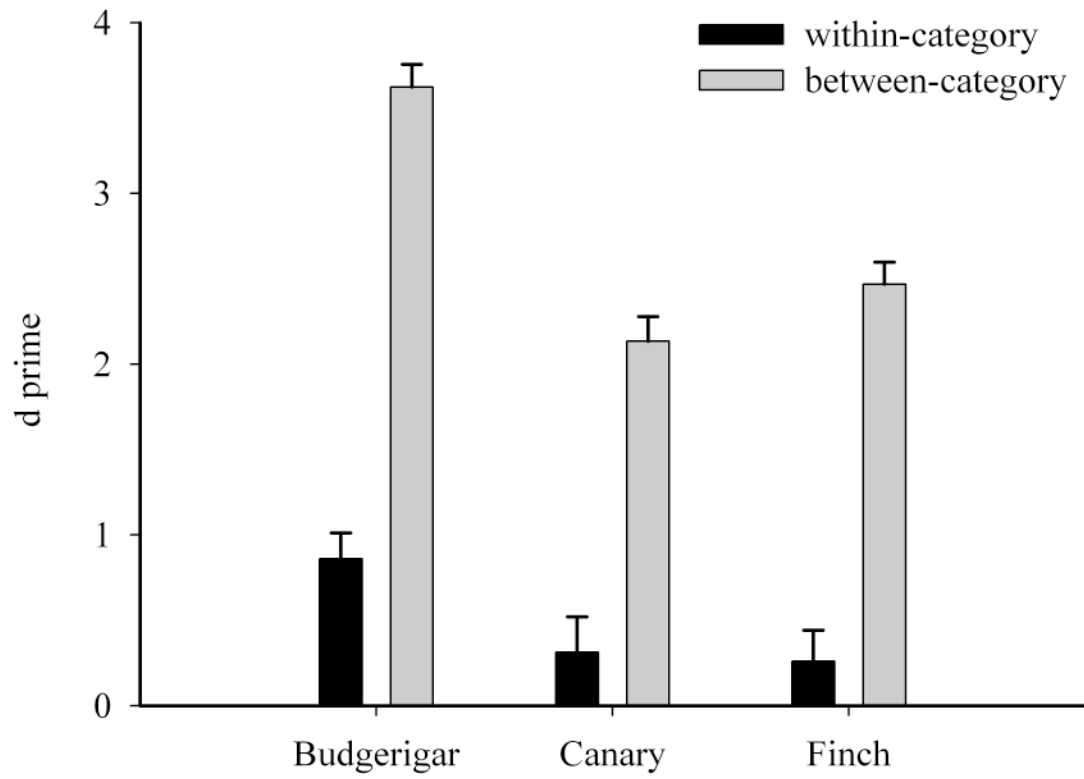


**Figure 4.**

An illustration of the random presentation of background and target sounds during a sequence of two trials. Bird identity is indicated in superscript (R = Ricky; P = Puffy; Y = Yuri). Token number is indicated in subscript (randomly selected from each bird). There were a total of 150 tokens in each background set, including 50 tokens from each bird. There were a total of 70 tokens in each target set, including 10 tokens from each acoustic category that were randomly selected from 3 birds. This schema shows two trials in a session where elements from Group A were in the background set. There were 7 acoustic categories, making up 7 background sets. Each background set was repeatedly used for 3 sessions for 3 unique target sets. In other words, each subject was run 21 sessions to complete this experiment.



**Figure 5.** Comparison of between-category and within-category discriminations by (A) response latency, error bars indicate standard deviation among four individuals; (B) d prime, error bars indicate 95% confidence interval.



**Figure 6.** Comparisons of  $d$  primes of between-category and within-category discriminations. Error bars indicate 95% confidence interval.

**Table 1**  
**The Relative Merit of Twenty Acoustic Measures in Categorizing Warble Segments**

Measures	Parameters	Relative merit (%)
Spectral Roughness	Quality	83
Tonality	Quality	61
Harmonic Strength	Frequency	48
1 <sup>st</sup> Frequency Quartile	Frequency	47
Duration	Temporal	28
Skewness of Power	Amplitude	24
Zero-Crossing Frequency	Frequency	22
3 <sup>rd</sup> Frequency Quartile	Frequency	18
2 <sup>nd</sup> Frequency Quartile	Frequency	13
Average Peak Spacing	Frequency	11
Amplitude Modulation	Amplitude	9
Number of Harmonic Lines	Frequency	7
Frequency of Max Amplitude	Frequency	6
80% Bandwidth	Frequency	5
Entropy	Quality	5
Time to Peak Amplitude	Temporal	3
Kurtosis of Power	Amplitude	3
Frequency Modulation	Frequency	2
Standard Deviation of Power	Amplitude	0
Average Power Per Sample	Amplitude	0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript