Perception of warble song in budgerigars (*Melopsittacus undulatus*): evidence for special processing

Hsiao-Wei Tu and
Department of Psychology, University of Maryland, College Park, MD 20742, USA

Robert J. Dooling
Department of Psychology, University of Maryland, College Park, MD 20742, USA

Abstract

The long, rambling warble song of male budgerigars is composed of a large number of acoustically complex elements uttered in streams lasting minutes a time and accompanied by various courtship behaviors. Warble song has no obvious sequential structure or patterned repetition of elements, raising questions as to which aspects of it are perceptually salient, whether budgerigars can detect changes in natural warble streams, and to what extent these capabilities are species-specific. Using operant conditioning and a psychophysical paradigm, we examined the sensitivity of budgerigars, canaries, and zebra finches to changes in long (>6 min) natural warble sequences of a male budgerigar. All three species could detect a single insertion of pure tones, zebra finch song syllables, budgerigar contact calls, or warble elements from another budgerigar's warble. In each case, budgerigars were more sensitive to these changes than were canaries or finches. When warble elements from the ongoing warble stream were used as targets and inserted, out of order, into the natural warble stream so that the only cue available was the violation of the natural ordering of warble elements, only budgerigars performed above chance. When the experiment was repeated with all the ongoing warble stream elements presented in random order, the performance of budgerigars fell to chance. These results show species-specific advantages in budgerigars for detecting acoustic changes in natural warble sequences and indicate at least a limited sensitivity to sequential rules governing the structure of their species-specific warble songs.

Keywords

Ordering; Warble; Sequence; Perception; Sensitivity

Introduction

Bird vocalizations, especially the learned songs of oscines, have led to many insights over the years into the neurobiology of vocal learning and vocal communication (e.g., Catchpole and Slater 2008; Marler 2004), including parallels with the learning of speech in humans (e.g., Marler 1970; Doupe and Kuhl 1999; Wilbrecht and Nottebohm 2003; Goldstein et al. 2003). While acoustically complex and comparable to human speech in many aspects, the sequential arrangement of song elements in most songbird songs is stereotyped and predictable so that little variation exists in the order or combinations of elements within a song, standing in marked contrast to the elaborate and unique syntactical rules for

Present Address: H.-W. Tu, Department of Psychology, Virginia Polytechnic Institute and State University, 230 Williams Hall, Blacksburg, VA 24060, USA

combining words that allow the unlimited expressions possible in human language (Kirby 2002; Marler 2000; Hauser et al. 2002). For example, there is a limited finite-state syntax in songs of Bengalese finches (Okanoya 2004) and nightingales (Todt and Hultsch 1998). And while perceptual tests have shown that some songbird species can accurately recognize acoustic patterns defined by an infinite recursive grammar (Gentner et al. 2006), alternative mechanisms other than the abstract syntactic structures of the stimuli may be used to explain these results (Corballis 2007; van Heijningen et al. 2009). Recently, more in-depth investigations into the linguistics of birdsong in these species continue to call into question the uniqueness of human syntactical abilities (Berwick et al. 2011; Abe and Watanabe 2011).

The present study is on a non-oscine parrot species, the budgerigar (*Melopsittacus undulatus*). Compared to the songbird species typically used in syntactical investigations, this species has a much larger and more complicated vocal repertoires characterized by extreme variations in syllable structure, sequential structure, and duration (Tu et al. 2011b; Farabaugh et al. 1992). These very features tend to make the vocalizations of this species difficult to study, but they also offer several new parallels with the way humans communicate with speech as described below.

The budgerigar is a non-territorial, group-living parrot native to Australia but a popular domesticated species because of its complex behaviors and vocal mimicry (Brockway 1969; Gramza 1970). The long, rambling warble song of adult budgerigars is composed of multiple acoustic elements delivered at the rate of 3–4 elements per second in no discernible order (Fig. 1a; Tu et al. 2011b; Farabaugh et al. 1992; Brockway 1969). Warble is almost exclusively produced by males when courting females in close quarters, accompanied by various body, head, and eye movements and other intimate behaviors. It also reinforces the pair bond between mates and promotes ovarian development and egg laying (Brockway 1964b, 1965, 1969; Farabaugh et al. 1992). Recent work has shown that the complex acoustic elements making up the warble of male budgerigars can be reliably and objectively assigned to at least broad acoustic categories and, more importantly, are perceived in categories by the birds (Tu et al. 2011b). Because the length and complexity of this song makes it experimentally intractable, there has to date been virtually no information on how budgerigars perceive complexity in warble songs.

Instead of focusing on the bird's ability to learn new warble sequences or discriminate among shortened, artificial warble sequences, we chose to tackle the problem of how birds listen to complex, natural warble sequences when they are presented for periods of time that approximate what is observed in breeding pairs under normal conditions. We used operant conditioning and a psychophysical paradigm to determine the extent to which birds could detect whether an out-of-context or an out-of-order sound had been inserted into natural warble stream. A human speech equivalent of this paradigm would be to ask subjects to listen to a 6-min sample of human speech from a single talker and to detect whether a sound or word was inserted out of context or out of order into the speech stream. Canaries and finches were also tested to address the issue of species specificity of the results with budgerigars.

General methods

Subjects

Four budgerigars, two zebra finches *(Poephila guttata castanotis)*, and two canaries (*Serinus canarius*) were used in these experiments. The birds in these experiments were either bred in the laboratory from a large communal population where outside birds were constantly introduced to avoid inbreeding or purchased from a local breeder. Both the canaries and

finches failed to learn some of the tasks even after extensive training, and therefore data for these two species are reported for only some experiments.

The budgerigars were bred in the laboratory from a large communal population where outside birds were constantly introduced to avoid inbreeding. They were housed individually in small cages and kept on a constant 12/12-h light–dark cycle. Since food was used as reinforcement, all birds were maintained at approximately 85–90 % of their free-feeding weight with ad libitum access to water at all times. In addition to the birds used in perceptual testing, several other male budgerigars were used to obtain the vocalizations used in perceptual tests. The budgerigars whose vocalizations were recorded to provide stimulus material and for further analyses were all housed together in a large cage in a separate vivarium and also kept on a constant light–dark cycle. They had ad libitum access to food and water all the time.

Vocal recording

Warble songs were recorded from four male budgerigars individually with their paired mate in close proximity so that the two could interact visually and acoustically. Before a recording session, the pair was separated and placed in a small animal acoustic isolation chamber (Model IAC-1, Industrial Acoustics Company, Bronx, NY, USA) for at least 1 h. The birds were visually reunited during the recording session when the doors of the chambers were opened and playback of a low-amplitude recording of birds from the budgerigar flock room began to facilitate warble production.

Three males were recorded using a single directional microphone (Model PRO35A, Audio-Technica USA, Inc., Stow, OH, USA) aimed at the male's cage. This ensured that the male's warble recordings were not contaminated by calls from the neighboring female. All vocalizations were recorded 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., Mahwah, NJ, USA). One male used in the study of Farabaugh et al. (1992) was also included here. This bird was recorded by a Realistic Electret dynamic microphone and a Panasonic Omnivision VHS hifi-stereo video cassette recorder, model P-4960 (Farabaugh et al. 1992). All vocalizations were digitized at a 48 kHz sampling rate and stored on a computer together with the other warble recordings. An aggregation of more than 1 h of warble was collected over approximately 4 h of recording. Budgerigar contact calls and zebra finch songs were also recorded in a similar way.

Warble recordings were band-pass filtered between 300 Hz and 12.0 kHz using Adobe Audition 2.0. Then, each warble recording was segmented into acoustic elements by a custom-written MATLAB program described in Tu et al. (2011b). Briefly, this program computes root-mean-square amplitude values using a 0.83 ms window through each warble sequence and determines an intensity threshold based on the noise level of each recording session. A warble element was defined when the root-mean-square amplitude envelope continuously exceeded the amplitude threshold for longer than 1 ms. Additionally, quiet intervals between elements that were shorter than 25 ms were ignored, and the two syllables surrounding the interval were combined and counted as one. Finally, individual WAV files of each warble element were saved along with a log file, indicating the sequential order of each segment in the original warble recording.

Apparatus

Birds were trained and tested in a small wire cage ($23 \times 25 \times 16$ cm^3) mounted in a sound-attenuated chamber (Model IAC-3, Industrial Acoustics Company, Bronx, NY, USA). Inside the test cage, a perch was mounted on the floor in front of an opening through which food

was accessible when a hopper was raised by activation of a solenoid. A control panel containing two microswitch response keys within easy reach of the bird was mounted in front of the perch. The observation key (left) and the report key (right) were approximately 5 cm apart, and each key had an 8-mm light emitting diode (LED) attached.

All test sessions were conducted using custom-designed MATLAB software driving a PC microcomputer controlling Tucker-Davis Technologies (TDT, Gainesville, FL, USA) System III modules. Acoustic stimuli were stored digitally and output via the D/A channel of a TDT signal processor (Model RX6, TDT) at a sampling rate of 24.4 kHz. The output of the D/A converter was fed to a digital attenuator (Model PA5, TDT), which imposed a 3 dB rove on each syllable. This signal was then passed to an analog summer (Model SM5, TDT), amplified (Model D-75, Crown Audio, Inc., Elkhart, IN, USA), and sent to a loudspeaker (KEF Model 80V, GP Acoustics, Inc., Marl-boro, NJ, USA) in the sound-attenuated test chamber where warble was produced at a mean level of 70 dB SPL at the bird's head.

Stimulus calibration was performed 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 birds' head during testing. For calibration, the amplitude of each stimulus was first normalized to a constant root-mean-square value. Then, a 2,500 Hz, 150 ms pure tone with the same root-mean-square amplitude was created to represent the sound pressure level of the entire stimulus set. The sound pressure level of that tone was taken as the sound pressure level of the complex stimulus set with the same root-mean-square amplitude.

Procedure

The procedure for training and testing birds has been described previously (Tu et al. 2011a). Basically, birds were trained to detect an inserted target sound in a repeating background consisting of natural or artificial warble sequences (Fig. 2). Each background sequence was longer than 6 min and recurred throughout a session of approximately 20 min. While the background sequence was played continuously, a peck on the observation key began a random interval of 2–6 s during which additional pecks to the observation key were ineffective. This random interval prevented the bird from being able to learn when a target stimulus would be inserted into the background warble. The next peck on the observation key after the random interval expired resulted in the presentation of a target stimulus following the last warble element in the repeating background. If the bird detected the inserted target and pecked the report key within 2 s, the food hopper was activated for 1.5 s giving the bird momentary access to food. This response was recorded as a "hit." Failure to peck the report key within 2 s of the presentation of a target was recorded as a "miss." On the other hand, withholding a peck to the report key during sham trials (where no sound was inserted into the warble stream) was recorded as a "correct rejection." Pecks to the report key during sham trials were recorded as "false alarms" and punished with a blackout period (2–10 s) during which all of the room lights were turned off and no sound was played. Report key pecks during any other time of the warble stream were also recorded and punished with blackouts. Following a blackout, testing resumed with the same trial (or next trial if it was a false alarm). Each test session consisted of approximately 90–120 trials, among which 20–30 % were sham trials. Each test session typically lasted about 20 min. Birds were tested twice a day, 5 days a week.

Data analysis

Birds' behavior (hit/miss/correct rejection/false alarm) in each trial was recorded and later pooled together to calculate averaged hit rate and false alarm rate. To minimize any species differences in response bias, these two numbers were then used to derive d',

$$d' = z(\text{hit rate}) - z(\text{false alarm rate})$$

To avoid infinite values, 100 % correct and 0 % false alarm rates were converted to $1 - 1/(2N)$ and $1/(2N)$, respectively, where N is the number of trials on which the percentage was based (Macmillan and Creelman 2005). An arbitrary d' of 1.0 was taken as a reference well above chance to evaluate the performance. It is a conventional level commonly seen in psychophysics and signal detection theory (e.g., Kastak and Schusterman 1999; Bregman et al. 2012; Demirel et al. 2009). Comparing d' between two conditions was made in the conventional way, in which the standard error (square root of the variance, Macmillan and Creelman 2005; Gourevitch and Galanter 1967) of d' was calculated and used to construct a 95 % confidence interval around the d' value of each condition. If the two 95 % confidence intervals overlapped, there was no significant difference in the sensitivity of these two conditions. If the two intervals did not overlap, the sensitivity in these two conditions differed significantly (Macmillan and Creelman 2005).

Experiment 1: Detection of sounds inserted in natural warble sequences

In this experiment, five different types of sounds were inserted into continuous natural warble sequences and the experiment proceeded in four phases. In the first two phases, non-budgerigar sounds were inserted into warble. In phase 1, 13 pure tones (non-biological sounds) spanning the frequency range and duration of natural warble elements were used as targets. In the second phase, zebra finch song syllables (non-budgerigar vocalizations) were used as targets. In both cases, these stimuli were quite distinct from budgerigar warble elements even to the human ear and were thus quite easy for the budgerigars to detect. In a third phase, budgerigar contact calls were inserted into warble sequences. Contact calls are short, highly frequency modulated and produced as single utterances by both males and females (Brockway 1964a). These are the most common sounds produced by budgerigars but they do not normally occur in warble (Tu et al. 2011b). Finally, in phase 4, a specific type of warble element (warble calls) was used as targets. These warble calls resemble contact calls produced as single utterances and they are also the most frequent elements in a warble bout (Tu et al. 2011b; Farabaugh et al. 1992). Some of these warble calls were taken from warble produced by male budgerigars other than the bird that produced the background warble stream used in the test session. In addition, the target set in phase 4 also included warble calls from the same individual as the one who produced the background warble stream. In this latter case, the only cue left for the bird to detect insertions would be any underlying sequential "rule" that governs the arrangement of warble elements.

Methods

Training—Training stimuli were restricted. Birds were trained with only one background sequence and target sounds consisting of only a limited set of pure tones, zebra finch syllables, and budgerigar contact calls. The main purpose of training was to ensure that the birds were accustomed to detecting oddballs in extraordinarily long streams of stimuli, instead of comparing one single background sound and one single target sound as they were trained before. The stimuli used to train birds were not used in final data collection sessions to avoid the possibility that subjects memorized the training stimuli and thus confounded the results.

During training sessions, background sounds were introduced at lower amplitude (40 dB SPL) at first and then gradually increased to the same level as the insertion (70 dB SPL). The number of training sessions required to bring birds to a level of 80 % correct varied between species but generally required several days to several weeks.

Background sequences—Background warble streams at least 6 min in length were recorded on three separate occasions from each of the four male budgerigars, making a total of 12 background warble sets. Each background contained more than 900 elements and was played back in its natural sequence with a constant 150 ms of silence between elements (Fig. 1b). In other words, the order of the elements was perfectly preserved, but the "tempo" of the background is somewhat less variable than that found in natural warble. This difference is slight and difficult to discern by human listeners. The background warble stream in each session was randomly chosen from 12 background warble sets recorded from 4 budgerigars. No two consecutive running sessions used a background from the same bird.

Target sounds—The total number of target sounds used in a test session varied across target type. Thus, the number of sham trials was also varied across sessions so that sham trials constituted approximately 20 % of a test session. The total number of trials per session varied accordingly. The target stimuli used in the four phases of Experiment 1 were as follows:

- *Pure tones* (*102 trials × 4 sessions*)

  One session included a seven-sound-duration gradient (25, 50, 100, 150, 300, 400, and 670 ms) where the frequency was fixed at 1,500 Hz, and a seven-sound-frequency gradient (500, 100, 1,500, 2,000, 2,500, 3,000, and 4,000 Hz) where the duration was fixed at 150 ms, making a total of 13 stimuli (one stimulus, a 1,500-Hz tone of 150 ms, was the same in these two gradients). These durations and frequencies covered the range of warble element characteristics. Every stimulus repeated 6 times (a total of 78 trials), plus 24 sham trials in each session.

- *Zebra finch song syllables* (*111 trials × 4 sessions*)

  Songs of six zebra finches were obtained and segmented into syllables. Each bird contributed the syllables of one song to one session, where all syllables repeated twice, making a total of 84 targets. Additionally, there were 27 sham trials per session.

- *Contact calls* (*100 trials × 8 sessions*)

  One session contained 20 sham trials and 80 signal trials (i.e., sham trials occurred on 20 % of the trials in a session). Each signal trial used one unique contact call. In other words, 80 contact calls recorded from four birds (20 from each) were used in one session. Because each calls only appeared once throughout the experiment, it was impossible for the birds to memorize the target calls and peck the key whenever a familiar call was played.

- *Warble calls* (*96 trials × 8 sessions*)

  Similarly, one session contained 24 sham trials and 72 warble calls, 18 from each of the four birds that contributed the background warble. In other words, 54 calls were recorded from a bird other than that in the background and 18 calls were directly extracted from the background elements.

Results

The results for all three species for all five types of target insertions are shown in Fig. 3. All three species showed a high level of performance when detecting pure tones and zebra finch song syllables inserted in natural warble sequences. Budgerigars were clearly more sensitive than zebra finches and canaries. There was no significant difference in the detectability of pure tones and that of zebra finch syllables for any of the species.

When detecting contact calls embedded in warble sequences, the d' for all three species remained above 1. Budgerigars were very sensitive to these contact call targets ($d' = 3.86$), but somewhat less sensitive than when detecting zebra finch syllables ($d' = 4.24$). Performance for zebra finches and canaries, on the other hand, dropped significantly when detecting contact calls ($d' = 2.56$ in zebra finches; $d' = 1.69$ in canaries) compared to zebra finch syllables ($d' = 3.58$ in zebra finches; $d' = 2.85$ in canaries).

When the inserted sounds were warble calls, all three species performed significantly better when the targets were warble calls recorded from different birds than the one produced the background warble stream (e.g., Bird A's calls tested against Bird B's warble background) than when the same bird provided both the warble call target and the background (e.g., Bird A's calls tested against Bird A's warble background) (Fig. 3). Warble call insertions from the same individual who produced the background warble sequence were the most difficult to detect and the performance of all birds decreased significantly (d' dropped from 3.86 to 1.11 in budgerigars, from 2.56 to 0.39 in zebra finches, from 1.69 to 0.35 in canaries). Even though budgerigars performed worse when the target warble calls were from the same individual as the background warble stream, they still remained above our reference of $d' = 1.0$ and were significantly better than the performance of zebra finches and canaries, which fell to chance.

Discussion

These results show that budgerigars, zebra finches, and canaries can be trained in a psychophysical task to detect 'oddball' sounds inserted into long, natural warble sequences. Results also show that these birds are capable of some form of voice recognition in these complex streams of vocalizations since the warble calls were fairly easily detected in the warble stream of an unrelated individual for all three species. When warble call targets and the background warble stream were from the same individual, budgerigars showed evidence of being able to perceive changes in the warble streams. This was not true of canaries or zebra finches tested in exactly the same way. The next experiment explores the budgerigars' sensitivity to the ordering of elements in natural warble sequences in greater detail.

Experiment 2: Detection of warble calls in altered warble sequences

Since finches and canaries were unable to detect warble calls inserted into natural warble streams using ordering cues embedded in natural warble sequences, only budgerigars were tested in this experiment. In this experiment, the entire background warble sequences, as opposed to the targets, were manipulated in order to examine the cues budgerigars were using to detect insertions.

Methods

Background warble sequences in Experiment 1 were manipulated in three different ways (see below) in this experiment. Each bird experienced four sessions of each manipulation, and the backgrounds of these four sessions were randomly selected from each of the four recorded individuals.

Background sequences

- Natural sequences

  The background elements were played continuously in their natural sequence as in Experiment 1 in this phase.

- Randomized sequences

Here, the same background warble elements were used, but the elements were played in a random order rather than their natural sequence.

- Natural sequences of reversed warble elements

  In this phase, the same background warble was used and each element was played in the correct order. However, each element was temporally reversed. Thus, the duration and overall spectrotemporal features of each warble element remained the same, but the fine structure of each of the elements making up the sequence was changed.

Target sounds—Each test session consisted of 100 trials, 75 were test trials in which a single warble call target was inserted, and the rest 25 were sham trials with no target inserted. Of the 75 target warble calls used in a test session, 50 were taken from the warble sequences of other birds and 25 were taken from the warble sequence of the same bird that produced the background warble stream. From Experiment 1, we knew that it was relatively difficult to rely purely on sequential cues to detect a warble element of the same bird in the background sequence. Mixing target sounds from the background bird and another bird in one session controlled the level of difficulty and maintained subjects' motivation, reducing the risk that our subjects might become frustrated and stop running. However, the main variable of interest is the performance on warble calls from the same individual that produced the warble background stream since performance on these stimuli can only be explained by the birds' sensitivity to the ordering of the sequences in a natural warble stream.

Results

Figure 4 shows that the d' value for budgerigars detecting warble call targets inserted in naturally ordered warble streams was 0.98, similar to that resulted in Experiment 1 ($d' = 1.11$). However, when the background was randomized (i.e., a new and unnatural sequence was presented to the birds), d' declined significantly to chance ($d' = 0.18$). When reversed elements were used in the background warble streams, but the warble stream sequences were played in their natural order, budgerigars again performed at a very high level ($d' = 2.62$), similar to their performance in detecting pure tones inserted in warble.

Discussion

The results of this experiment confirm that budgerigars are able to use an ongoing sequential cue to detect an out-of-order target warble element. All targets were only presented once during a test session, eliminating the possibility that memorizing each target sound would increase performance. Moreover, four different warble streams were used in the background across the four test sessions. Although it was difficult to gauge how familiar each subject was with each background warble sequence, we judged it unlikely that they could remember the entire sequences from only a few exposures. The fact that budgerigars' performance fell to chance when the background warble elements were randomized in the control condition is also consistent with a sensitivity to subtle changes in the rhythm or prosody of the background warble sequence.

Finally, temporally reversed warble elements preserved the overall acoustic complexity of budgerigar warble stream (i.e., overall spectral content, intensity, and duration), but distorted the temporal fine structure of these complex sounds. This would be the bird equivalent of humans listening to a long speech stream where the words are time-reversed (Galbraith et al. 2004; Binder et al. 2000; Dehaene-Lambertz et al. 2002). In such a listening task, it is quite easy for humans to detect the occurrence of a word played in the "forward" direction amid other words all played in reverse. The same is obviously true of budgerigars. The birds had

no trouble in detecting a 'forward' warble call target against a background of reversed warble elements. It is worth noting that human listeners are utterly incapable of detecting forward warble call insertions in reversed warble background streams. Indeed, human listeners also cannot distinguish forward versus reversed warble streams.

General discussion

The warble song of male budgerigars is highly effective in coordinating reproductive efforts (Brockway 1969) but, until now, little was known about how budgerigars listen to this song or which features they attend to. Using operant conditioning and a psychophysical paradigm, budgerigars, zebra finches, and canaries were all tested on long streams of natural warble at durations typically produced by male birds. All three species showed excellent sensitivity for detecting alien sound insertions such as tones, zebra finch song syllables, or budgerigar contact calls that do not naturally occur in warble. But in every case, budgerigars outperformed both canaries and finches, revealing a clear species-specific advantage in these tasks. Budgerigars were also much more proficient than canaries or finches at detecting the insertion of warble call elements from different budgerigars into another budgerigar's natural warble stream. Most interestingly, budgerigars can detect inserted warble call elements taken from that background warble stream that were simply inserted out of order. Canaries and finches could not perform above chance on this task.

The length and complexity of budgerigar warble afforded the opportunity to ask unique questions about how birds listen to this long vocalization and perceive alterations in its acoustic structure. One possibility is that birds listen to warble song "synthetically" and gauge the suitability of the male from the overall complexity of warble song, or its volume, or tempo, or the proportion of elements occurring over long periods of time. Another possibility is that birds listen more "analytically" in an element-by-element mode and pay exquisite attention to the acoustic details of individual elements. Still another possibility is that the effectiveness of warble is carried in the ordering of some or all of the elements in warble. A human analog of this experimental design would be as if listeners were asked to listen to a 6-min speech stream from a single talker and to detect whether tones, nonsense words, words spoken by different talkers, or words from the speech stream were simply inserted out of order. This is an easy task for human listeners unless they are listening to foreign language speech with which they are unfamiliar. Then, the complexity of the speech stream becomes somewhat overwhelming to the listener, making it difficult to detect all but the most non-speech-like insertions.

Here, results show that all three species were remarkably proficient at detecting all but one of the various sound insertions in warble. Canaries and finches listening to budgerigar warble could still detect warble call insertions from other budgerigars, analogous to humans identifying different talkers even though the speech stream is in a foreign language. This suggests that these birds were listening analytically to the warble streams and were quite sensitive to acoustic features of individual elements in spite of the complex background. Moreover, the high performance also suggests an ability to distinguish individual bird voices of another species.

The ability of budgerigars to detect insertions in natural warble streams where the only cue is a violation of the sequential ordering of elements points to a potential rule that governs the sequential organization of warble elements in natural warble song and is perceptually salient to budgerigars but unavailable to the other two species. The complexity of warble and the high degree of proficiency shown by budgerigars suggest two conclusions. One is that budgerigars are listening to warble streams in an analytical fashion with a considerable ability to focus on each individual elements and their relative sequential position in the

entire sequence. Human listeners find this either challenging or impossible except for cases where the inserted elements are pure tones. One drawback of the present test paradigm is that there is no control over where in the warble stream a trial occurs. It is possible, therefore, that warble call insertions following or preceding particular elements violate a local rule but do not violate any rules when they follow or precede other elements. Nevertheless, this still requires that budgerigars have some sensitivity to ordering of elements in their warble in order to succeed at this task.

The other conclusion is that this sensitivity to warble element sequences is specific to budgerigars and beyond the capability of canaries and finches that were tested under identical conditions. However, it is still unclear whether this ability is specific to warble. Future studies might be interesting to test the sensitivity of budgerigars to sequence changes in heterospecific songs, tonal melodies, and human speech using the same experimental paradigm in the present study.

Human language has long been thought as unique and special for the production of an infinite range of expressions by recombining a finite set of lexica (Kirby 2002). Whether animals have this linguistic capacity is still a matter of considerable debate by cognitive psychologists and linguists (e.g., Gentner et al. 2006; Hauser et al. 2002; Corballis 2007). Most comparative investigations into animals aimed at these issues have involved intensive training and testing using sequences that are either relatively short and simple compared to human language or artificially elongated beyond the animal's normal repertoire. In the present study, the length and complexity of budgerigar warble offers a more natural system for the comparative study of these aspects of human speech. The perceptual salience of the ordering of warble elements demonstrated in the present study suggests an intriguing animal model for examining cognitive and perceptual differences observed in humans when listening to speech in one's native language versus foreign language speech. The present animal model may open the door to some deeper comparisons between animal vocalization and human speech.
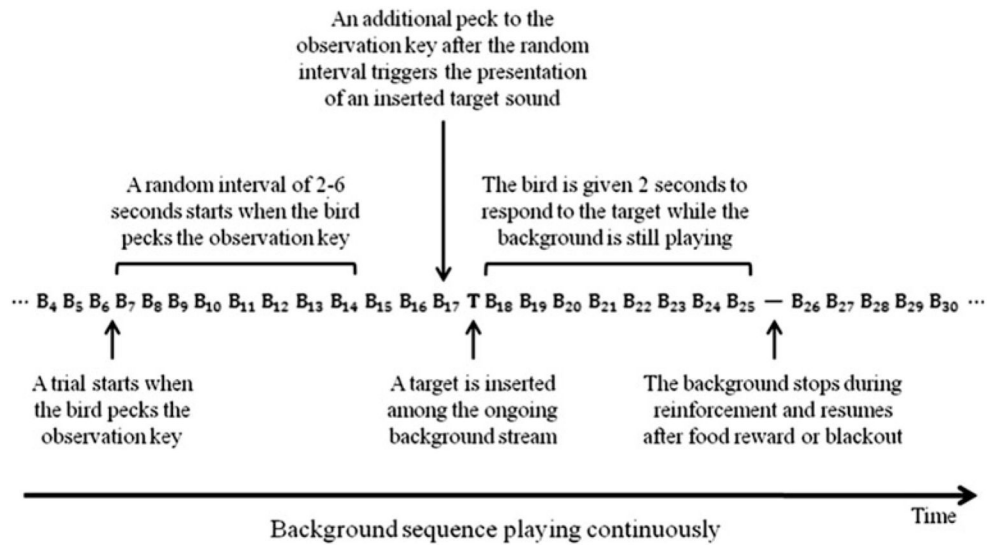
Acknowledgments

References

Abe K, Watanabe D. Songbirds possess the spontaneous ability to discriminate syntactic rules. Nat Neurosci. 2011; 14:1067–1074. [PubMed: 21706017]

Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ. Songs to syntax: the linguistics of birdsong. Trends Cogn Sci. 2011; 15:113–121. [PubMed: 21296608]

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 2000; 10:512–528. [PubMed: 10847601]

Bregman MR, Patel AD, Gentner TQ. Stimulus-dependent flexibility in non-human auditory pitch processing. Cognition. 2012; 122:51–60. [PubMed: 21911217]

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 (*Melo-psittacus undulatus*). Anim Behav. 1965; 13:575–578. [PubMed: 5894177]

Brockway, BF. Roles of budgerigar vocalization in the integration of breeding behavior. In: Hinde, RA., editor. Bird vocalizations. Cambridge University Press; London: 1969. p. 131-158.

Catchpole, CK.; Slater, PJB. Bird song: biological themes and variations. 2. Cambridge University Press; New York: 2008.

Corballis MC. Recursion, language, and starlings. Cogn Sci. 2007; 31:697–704. [PubMed: 21635313]

Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. Functional neuroimaging of speech perception in infants. Science. 2002; 298:2013–2015. [PubMed: 12471265]

Demirel S, Fortune B, Fan J, Levine RA, Torres R, Nguyen H, Mansberger SL, Gardiner SK, Cioffi GA, Johnson CA. Predicting progressive glaucomatous optic neuropathy using baseline standard automated perimetry data. Invest Ophthalmol Vis Sci. 2009; 50:674–680. [PubMed: 18936149]

Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. Ann Rev Neurosci. 1999; 22:567–631. [PubMed: 10202549]

Farabaugh SM, Brown ED, Dooling RJ. Analysis of warble song of the budgerigar *Melopsittacus undulatus*. Bioacoustics. 1992; 4:111–130.

Galbraith GC, Amaya EM, de Rivera JMD, Donan NM, Duong MT, Hsu JN, Tran K, Tsang LP. Brain stem evoked response to forward and reversed speech in humans. NeuroReport. 2004; 15:2057–2060. [PubMed: 15486481]

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. Recursive syntactic pattern learning by songbirds. Nature. 2006; 440:1204–1207. [PubMed: 16641998]

Goldstein MH, King AP, West MJ. Social interaction shapes babbling: testing parallels between birdsong and speech. Proc Nat Acad Sci. 2003; 100:8030–8035. [PubMed: 12808137]

Gourevitch V, Galanter E. A significance test for one-parameter isosensitivity functions. Psychometrika. 1967; 32:25–33. [PubMed: 5232570]

Gramza AF. Vocal mimicry in captive budgerigars (*Melo-psittacus undulatus*). Z Tierpsych. 1970; 27:971–998.

Hauser MD, Chomsky N, Fitch WT. The faculty of language: what is it, who has it, and how did it evolve? Science. 2002; 298:1569–1579. [PubMed: 12446899]

Kastak, D.; Schusterman, RJ. Loss of hearing sensitivity with depth in a free diving California sea lion (*Zalophus californi-anus*). Paper presented at the 13th biennial conference on the biology of marine mammals; Maiu, HI. 1999.

Kirby, S. Learning, bottlenecks and the evolution of recursive syntax. In: Briscoe, T., editor. Linguistic evolution through language acquisition. Cambridge University Press; Cambridge: 2002. p. 173-204.

Macmillan, NA.; Creelman, CD. Detection theory: a user's guide. 2. Lawrence Erlbaum Associates; Mahwah: 2005.

Marler P. Birdsong and speech development: could there be parallels? Am Sci. 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. Massachusetts Institute of Technology; Cambridge: 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. Elsevier Academic Press; San Diego: 2004. p. 1-38.

Okanoya K. The bengalese finch: a window on the behavioral neurobiology of birdsong syntax. Ann NY Acad Sci. 2004; 1016:724–735. [PubMed: 15313802]

Todt D, Hultsch H. How songbirds deal with large amount of serial information: retrieval rules suggest a hierarchical song memory. Biol Cybern. 1998; 79:487–500.

Tu H-W, Osmanski MS, Dooling RJ. Learned vocalizations in budgerigars (*Melopsittacus undulatus*): the relationship between contact calls and warble song. J Acoust Soc Am. 2011a; 129:2289–2297. [PubMed: 21476684]

Tu H-W, Smith EW, Dooling RJ. Acoustic and perceptual categories of vocal elements in the warble song of budgerigars (*Melopsittacus undulatus*). J Comp Psychol. 2011b; 125:420–430. [PubMed: 22142040]

van Heijningen CAA, de Visser J, Zuidema W, ten Cate C. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. Proc Natl Acad Sci. 2009; 106:20538–20543. [PubMed: 19918074]

Wilbrecht L, Nottebohm F. Vocal learning in birds and humans. Ment Retard Dev Disabil Res Rev. 2003; 9:135–148. [PubMed: 12953292]

(a) Natural warble sequence

**(b)** Background warble sequence used in Experiment 1
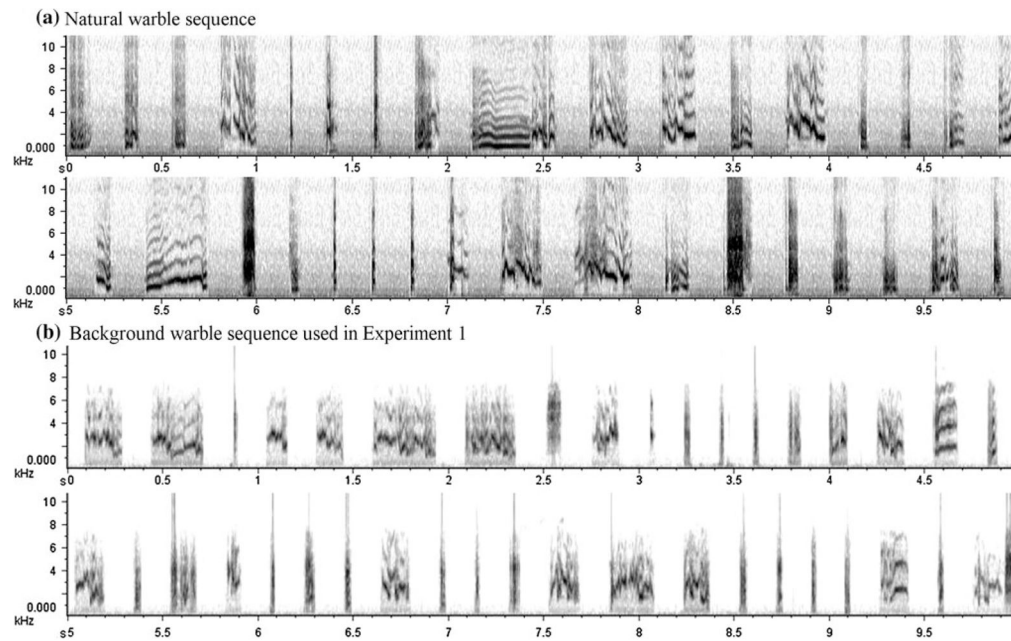
Fig. 1.
a An example spectrogram of approximately 10 s of natural warble sequence. **b** An example spectrogram of approximately 10 s of a background warble stream used in Experiment 1. All elements were played back in natural sequence with a constant 150-ms interval. (Time in second on the *x*-axis and frequency in kHz on the *y*-axis.)

An additional peck to the
observation key after the random
interval triggers the presentation
of an inserted target sound

A random interval of 2-6
seconds starts when the bird
pecks the observation key

The bird is given 2 seconds to
respond to the target while the
background is still playing

··· B_4 B_5 B_6 B_7 B_8 B_9 B_{10} B_{11} B_{12} B_{13} B_{14} B_{15} B_{16} B_{17} T B_{18} B_{19} B_{20} B_{21} B_{22} B_{23} B_{24} B_{25} — B_{26} B_{27} B_{28} B_{29} B_{30} ···

A trial starts when
the bird pecks the
observation key

A target is inserted
among the ongoing
background stream

The background stops during
reinforcement and resumes
after food reward or blackout

Background sequence playing continuously

Time

Fig. 2.
An illustration of the presentation of a background sequence (*B*) and an inserted target
sound (*T*) during one trial. Each letter represents one warble element, and the number in
subscript indicates the order of each element in the sequence. In this example, the
background warble stream was played in its natural sequence, and a target sound was
inserted between the 17th and the 18th element in the sequence. The background sequence
was played continuously except during reinforcement (food reward on "hit" or blackout on
"false alarm")
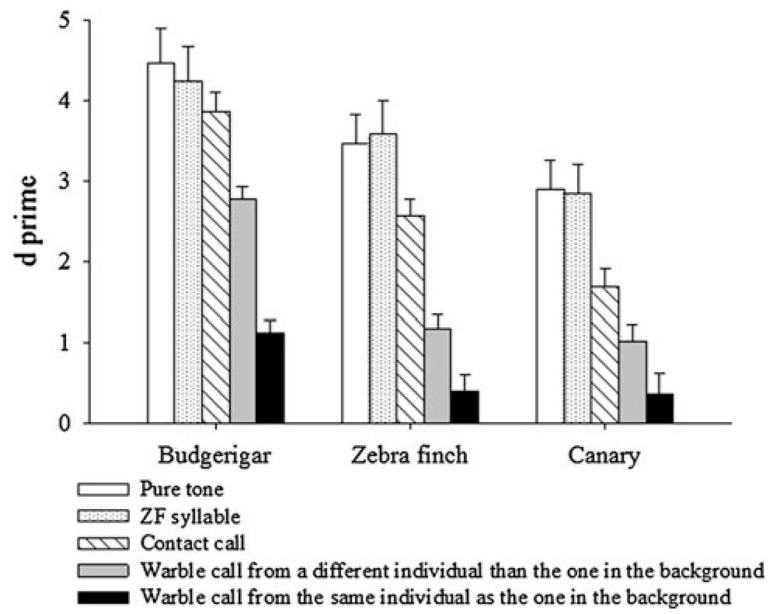
Fig. 3.
The performance of all three species detecting insertions of pure tones, zebra finch syllables, contact calls, and warble calls is shown as d' values. The 95 % confidence intervals are shown as *error bars*
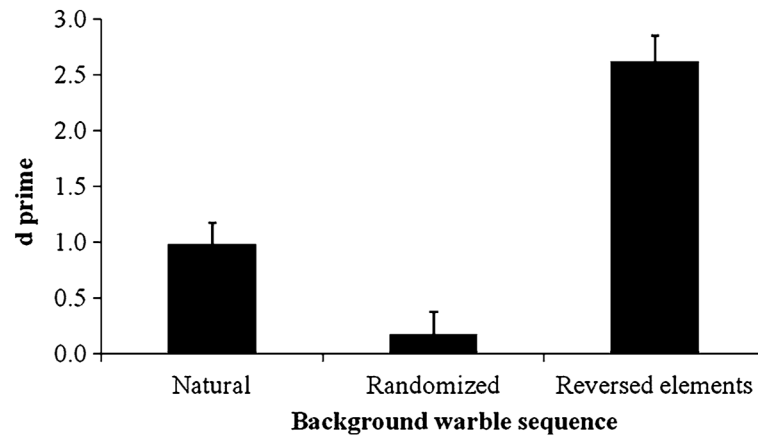
Fig. 4.
Comparison of budgerigars' ability to detect insertions of warble calls in different types of background warble sequences. The 95 % confidence intervals are shown as *error bars*