Relative salience of envelope and fine structure cues in zebra finch song

Beth A. Vernaleo[a] and Robert J. Dooling
Neuroscience and Cognitive Science Program, University of Maryland, College Park, Maryland 20742

Zebra finches produce a learned song that is rich in harmonic structure and highly stereotyped. More is generally known about how birds learn and produce this song than how they perceive it. Here, zebra finches were trained with operant techniques to discriminate changes in natural and synthetic song motifs. Results show that zebra finches are quite insensitive to changes to the overall envelope of the motif since they were unable to discriminate more than a doubling in inter-syllable interval durations. By contrast, they were quite sensitive to changes in individual syllables. A series of tests with synthetic song syllables, including some made of frozen noise and Schroeder harmonic complexes, showed that birds used a suite of acoustic cues in normal listening but they could also distinguish among syllables simply on the basis of the temporal fine structure in the waveform. Thus, while syllable perception is maintained by multiple redundant cues, temporal fine structure features alone are sufficient for syllable discrimination and may be more important for communication than previously thought. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3560121]

I. INTRODUCTION

Zebra finches are closed-ended learners that have a single sensitive period for song learning, after which new song cannot be learned. The result of this sensitive period is a single, complex, highly stereotyped song containing both spectral and temporal modulations within each syllable (Eales, 1985). Previous research by Braaten *et al.* (2006) suggests that birds attend to local syllable structure rather than global syllable ordering when categorizing songs. Furthermore, removal of one or more syllables from the song motif (a temporal envelope change affecting the whole motif) does not affect birds' ability to discriminate between two conspecific songs (Cynx, 1993). Together, these results suggest that syllable fine structure plays an important role in song recognition and discrimination.

While spectral structure in zebra finch song has been studied extensively, less is known about the temporal structure features of song, especially temporal fine structure (i.e., phase and harmonic structure). Much of the early work studying auditory perception in zebra finches and other birds has tested singular synthetic stimuli and has focused on discrimination of amplitude, or temporal envelope changes (Dooling and Haskell, 1978; Dooling, 1982; Maier and Klump, 1990). Results from these studies showed no difference in temporal resolution compared with humans.

However, more recent work has shown zebra finches to be extremely sensitive to temporal fine structure changes in synthetic harmonic stimuli that were designed to resemble natural zebra finch calls (Lohr and Dooling, 1998; Dooling and Lohr, 2006; Lohr *et al.*, 2006; Dooling *et al.*, 2002). These birds have been shown to discriminate single period reversals within harmonic stimuli (Lohr *et al.*, 2006), phase changes within Schroeder harmonic waveforms (Dooling *et al.*, 2002), and mistuned harmonics within complex stimuli (Lohr and Dooling, 1998). All of these changes affect temporal fine structure, while keeping amplitude envelope and spectrum intact. With natural stimuli, one study also showed that zebra finches can discriminate between song syllables presented in isolation using only timbre cues (Cynx *et al.*, 1990).

Temporal fine structure is becoming of increasing interest in human speech studies, specifically speech intelligibility under various listening conditions (Sheft *et al.*, 2008; Hopkins and Moore, 2009; Gilbert and Lorenzi, 2010). It is important for both voice quality and the timbre of human speech and may be more communicatively relevant than previously thought in perception and recognition (Rosen, 1992). Thus, zebra finches may provide an interesting animal model with which to test some of these issues. The present study asks whether temporal fine structure discrimination by zebra finches extends to perception of natural song, and how salient temporal fine structure features are within song compared to other song features.

II. GENERAL METHODS

A. Subjects

Four male and three female zebra finches (*Taeniopygia guttata*) were used. There were no gender differences in any of the experiments, so the results from all seven birds are reported together. The same seven birds were used for all three experiments presented. Birds were housed at the University of Maryland in an avian vivarium and kept on the photoperiod corresponding to the current season. Birds were maintained at 85%–90% of their free feeding weights and were given free access to water. All procedures were in

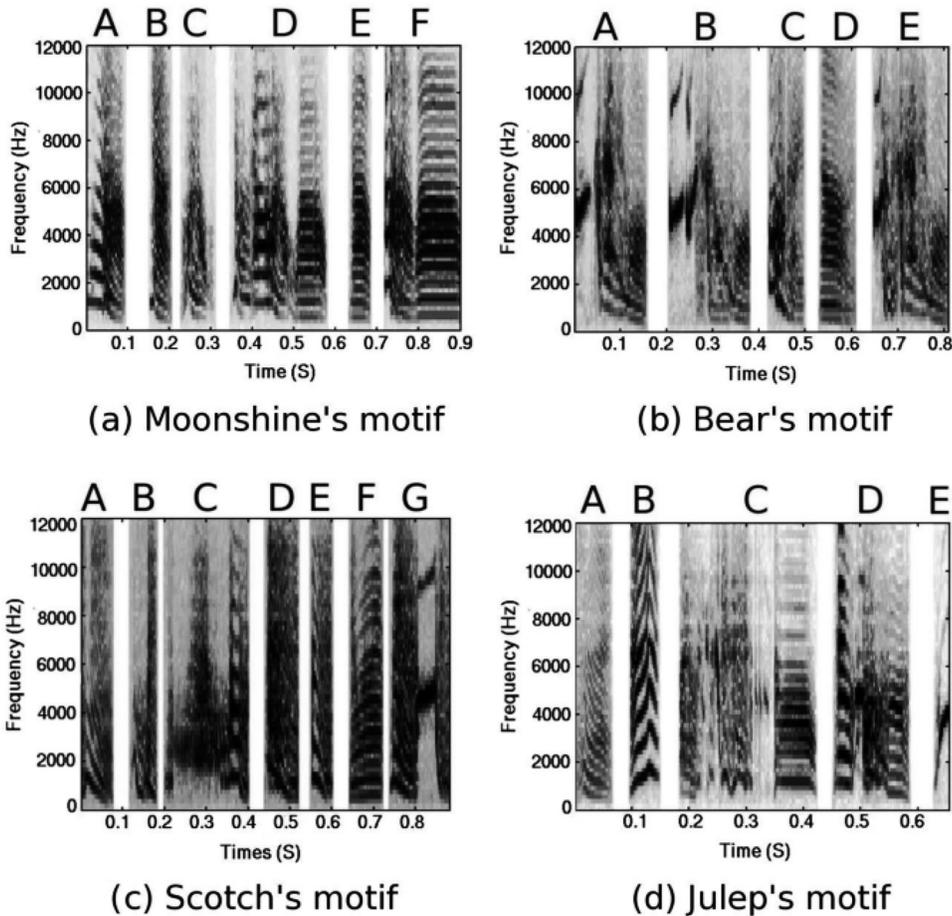[a]Author to whom correspondence should be addressed. Electronic mail: beth@netpurgatory.com

FIG. 1. Spectrograms of the four motifs tested. The *x*-axis represents time, the *y*-axis represents frequency, and amplitude is represented by the darkness of the grayscale.

accordance with the University of Maryland Institutional Animal Care and Use Committee (IACUC).

B. Apparatus

All psychoacoustic experiments took place in a wire cage anchored inside of a sound-attenuated chamber (Industrial Acoustics Company, Bronx, NY). Birds sat on a perch and had access to food through an opening in the floor of the cage. Millet was delivered through a food hopper, which was brought up to the food opening through activation of a solenoid. The two response keys were mounted to the wall of the cage, directly in front of the perch. The keys consisted of 8-mm light emitting diodes (LEDs) separated by 5 cm, each attached to a microswitch. The left LED served as the observation key, and the right LED served as the report key. All experiments were monitored from the outside via an overhead video camera.

Stimuli were generated in MATLAB (Mathworks, Natick, MA) or Adobe Audition (Adobe, San Jose, CA) as wav files (48 000 Hz sampling rate) and stored on an Intel Core 2 Duo computer (Mid Atlantic Data Systems, Gaithersburg, MD), which controlled all experiments. The computer operated a Tucker Davis Technologies System 3 module (TDT, Gainsville, FL), in which sounds were sent to a digital–to–analog converter (TDT RX-6), then output to a programmable attenuator (TDT PA-5), and a signal mixer (SM-5). Then sounds were played out of the speaker in the psychoacoustics chamber at a sampling rate of 24 414 Hz. Sounds were presented via a speaker (KEF Model 80C, Kent, UK) mounted to the

chamber ceiling, angled 45° downward toward the cage. The speaker was approximately 40 cm away from the bird's head.

C. Stimuli

Songs consist of a specific ordering of syllables, termed a motif, that are repeated several times throughout the song bout. Motifs can range from three to eight syllables long (Sossinka and Bohner, 1980), and each syllable is acoustically distinct, containing multiple cues which result in a unique sound. For all experiments using natural birdsong, vocalizations were recorded from male birds in a foam-lined acoustic chamber, using a Marantz portable solid state recorder (Model PMD670) at a sampling rate of 48 000 Hz. Individual song motifs were extracted from song recordings using Adobe Audition, and high-pass filtered with a cutoff frequency of 350 Hz using Raven Pro 1.3 (Cornell Lab of Ornithology, Cornell, NY). Inter-syllable intervals were band-reject filtered at all frequencies to produce pure silence between syllables. A 5-ms cosine rise and fall time was imposed on all motifs to prevent clipping artifacts. This did not change the overall envelope of the beginning and ending syllables in the motif. One isolated motif from each of the four male birds tested was used as a stimulus. Thus, each male bird was tested on his own song, and three conspecific songs. Since all birds tested were housed in the same room, the test motifs were familiar stimuli. The four test motifs are shown in Fig. 1.

All synthetic motifs were generated in MATLAB and were modeled after these motifs to maintain the overall temporal envelope (i.e., syllable and interval durations were the same as in the natural motifs, keeping the overall motif rhythm intact).

D. Testing procedure

Zebra finches were trained to peck the observation key during a repetition of a background sound, until they heard a different sound alternated with the background, the target sound. When they heard this alternation between the background and target, birds were required to peck the report key to receive a food reward (i.e., 2 s of access to the food hopper). After the first peck of the observation key, a random interval between 2 and 6 s occurred before presentation of the target. Once the target was presented, if the bird pecked the report key within 2.5 s, this was recorded as a *hit*, and the bird received a food reward. If birds did not peck within this time, this was recorded as a *miss* and was not rewarded.

Birds ran 100 trials within a session, which consisted of ten 10-trial blocks. For each block, seven targets were presented. The remaining trials were sham trials in which the background sound was presented as the target. If birds pecked the response key during the response interval of sham trials, this was recorded as a *false alarm* and was punished by a 5-s blackout period in which all of the lights in the chamber were extinguished. If birds correctly withheld pecking during a sham trial, this was recorded as a *correct rejection* but was not rewarded. Birds were run until they could complete 300 trials with a false alarm rate below 20%, and the last 200 trials were used for analysis.

E. Analysis

For each set of 100 trials, the percent correct hit rate was calculated for each target. For each trial, the response latency was recorded. For response latencies, missed targets in which birds failed to peck the report key were assigned a latency of 2500 ms, which corresponded to the maximum response interval. Latencies began from the start of the change presented in the target motif.

III. EXPERIMENT 1: PERCEPTION OF TEMPORAL ENVELOPE AND FINE STRUCTURE CUES IN NATURAL SONG

Zebra finch song contains both temporal envelope (i.e., overall motif) and fine structure (i.e., within-syllable) cues. Temporal envelope cues determine the overall rhythm and timing of song and include syllable sequencing as well as the durations of syllables and intervals. Thus, the temporal envelope refers to amplitude fluctuations over the entire motif, rather than within individual syllables (i.e., syllable envelope), which will be examined in detail in experiment 2. Fine structure cues, on the other hand, are small changes that take place within individual syllables, and encompass spectral fine structure, temporal fine structure, and syllable envelope cues. Temporal fine structure specifically refers to temporal changes that occur within a single period, such as

phase and harmonic structure. While both temporal envelope and fine structure cues are well represented in zebra finch song, the relative salience of these two cues is an open question.

The following experiment proceeded in two phases. The first phase focused on the relative salience of temporal envelope versus fine structure cues in song. The second phase examined perception of syllable fine structure. In this second phase, single syllable reversals at all locations in the motif, and from all four motifs were tested. Differences in syllable reversal discriminability should give insight into what acoustic cues within the syllables birds use to make these discriminations.

A. Stimuli

All birds were tested on changes made to the four song motifs discussed in the general methods. In generating the original motif, the inter-syllable intervals were band-reject filtered at all frequencies to eliminate noise between the syllables.

Temporal envelope changes consisted of doubling the duration of single inter-syllable intervals in the motif. This was done by adding amplitude values of zero that were the same duration as the original interval. Fine structure changes consisted of reversing single syllables in time, while keeping the order of syllables in the motif intact. In reversing syllables, the overall spectral content remains the same, whereas fine structure (i.e., small scale timing) is changed. These two types of targets are shown in Fig. 2.

B. Procedure

All subjects were tested on the same four stimulus sets. To prevent practice effects, the order in which stimulus sets were tested was randomized. Each block of ten trials contained seven targets and three sham trials, presented in random order. For the first phase of experiment 1, the seven targets consisted of three interval doublings and four syllable reversals. Because the motifs differed in the number of syllables and intervals, the first, middle, and last positions within the motif were tested. For syllable reversals, two syllables in the middle position were tested, as well as the first and last syllables. For interval doublings, one middle interval was tested, as well as the first and last intervals.

For the second phase of experiment 1, targets consisted of single syllable reversals (example shown in Fig. 2), at each location within the song motif. For the two motifs that contained only five syllables, two additional targets consisted of a reversal of a portion of a complex syllable (termed a sub-syllable, see syllable G in Fig. 2 for an example). The sub-syllables tested were from syllable A in Bear's song, and syllable C in Julep's song. This was done to determine the resolution with which birds can discriminate changes to song syllables. For one motif that contained six syllables, the remaining target consisted of a motif in which all syllables were reversed but remained in sequential order (i.e., ABCDEF).

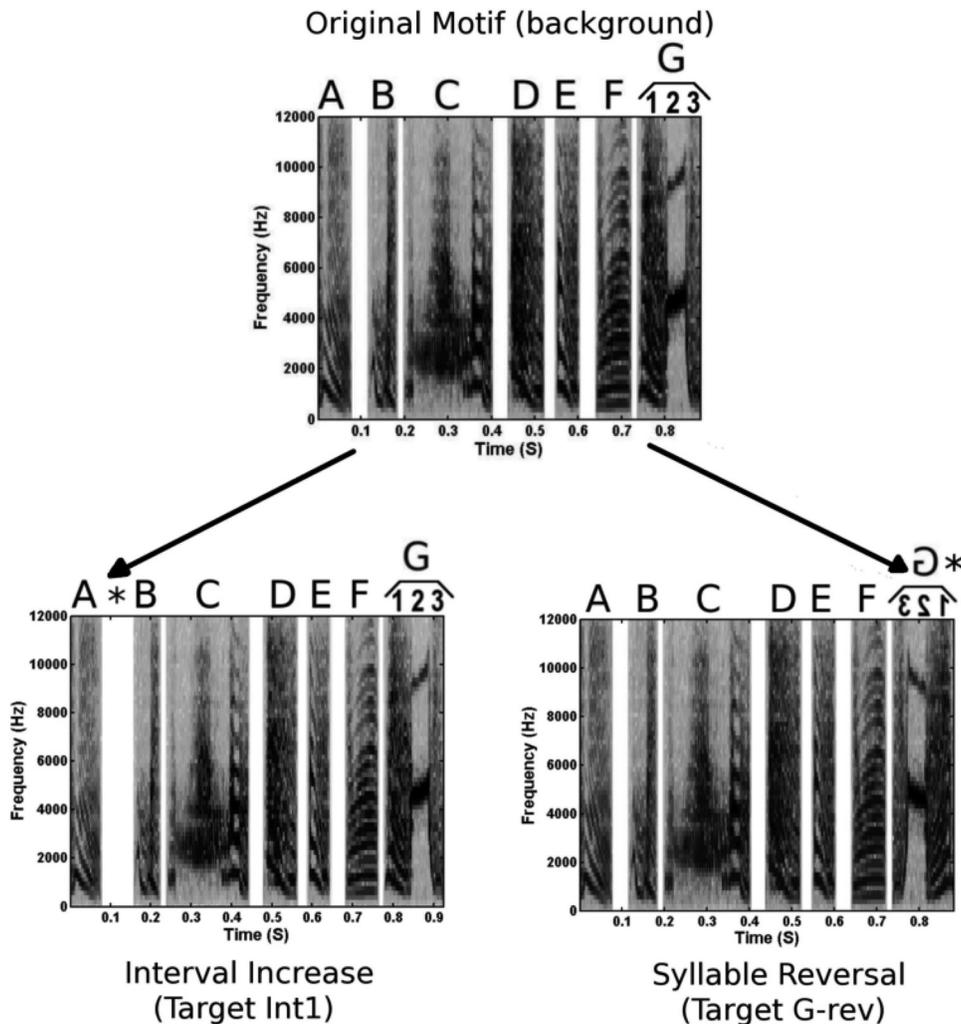Song motifs were presented at 70 dB SPL, at a rate of once per 1500 ms. Birds were run on 300 trials for each

FIG. 2. Example of a background motif (top), and the two types of targets presented: single interval increase (bottom left) and single syllable reversal (bottom right). Syllable G is broken down into its three component sub-syllables.

stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate did not exceed 20%.

C. Results and discussion

Birds discriminated single interval doublings on average less than 5% of the time, and single syllable reversals on average greater than 90% of the time (with the exception of one syllable). In other words, all birds tested were much better at discriminating single syllable reversals than they were at discriminating single interval doublings, regardless of the song motif. A two way analysis of variance (ANOVA) with target type (interval doubling, syllable reversal) and position within the motif (first, middle, last) as factors showed a significant main effect of target type $[F(1,183) = 2972.874, p < 0.001]$ such that single syllable reversals had significantly higher hit rates $(M = 93.94, \sigma = 13.89)$ than single interval doublings $(M = 3.15, \sigma = 5.03)$. The main effect of position within the motif was not significant $[F(2,183) = 0.805, p = 0.449]$. There was no significant interaction effect $[F(2,183) = 0.967, p = 0.382]$. This shows that changes to syllables in general were more salient than changes to intervals, and that this was

independent of position within the motif. These results are illustrated in Fig. 3.

Previous research has shown that zebra finches produce intervals with greater variability than they produce syllables, and this may explain a perceptual insensitivity to changes in interval duration. Glaze and Troyer (2006) examined the durations of intervals and syllables in many renditions of a set of songs and found that the coefficient of variation is about 1.5 times greater for intervals than it is for syllables. In addition, tempo changes in song affect the durations of intervals more than syllables. In other words, when songs are sped up or slowed down, the intervals tend to stretch and compress whereas syllable durations are more stable. It is possible that because intervals are normally sung with some amount of variability, changes to interval duration are not particularly salient to birds. Changes to the fine structure of syllables, however, are quite salient. The next phase of this experiment examined whether all syllable reversals are equally discriminable, and if syllable type, syllable duration, or syllable location within the motif was a determinant of discriminability.

When tested on single syllable reversals at all locations within the song motif, birds discriminated nearly all syllable reversals within the four song motifs greater than 80% of the

(a) Moonshine's motif

(b) Bear's motif

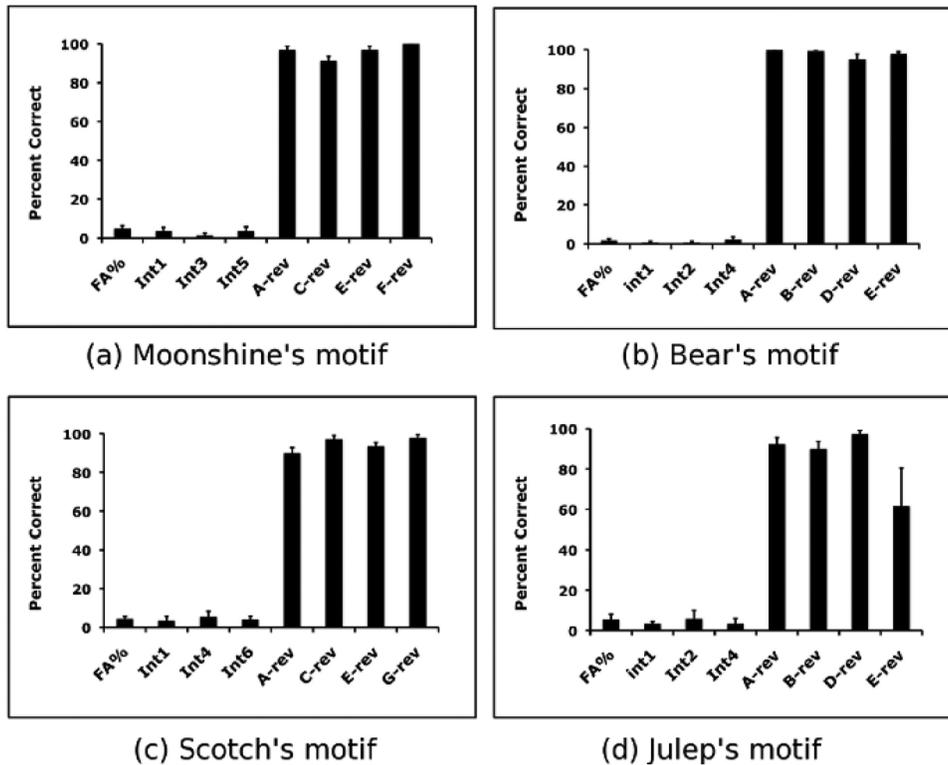(c) Scotch's motif

(d) Julep's motif

FIG. 3. Percent correct discrimination performance on single interval doublings and single syllable reversals in natural song, tested in the same session. Results from all seven birds were averaged together. One female, whose false alarm exceeded 20% for Julep's motif was not included in the analysis for that motif. Error bars show standard error of the mean.

time. For all birds, hit rates were well above the false alarm rate, indicating that this task is relatively easy and birds are performing above chance, and in fact near perfect on many syllables. Although performance on most syllables was quite good, there were still differences in discrimination hit rates for different syllables. To explain these differences in performance, we examined three qualities of syllables: duration, position, and syllable type.

There was a significant effect of duration on syllable reversal discrimination. Discrimination performance stabilized for syllables more than 100 ms in duration. Thus, for syllables that were longer than 100 ms, reversals were discriminated nearly 100% of the time. However, for syllables shorter than 100 ms in duration, performance was positively correlated with syllable duration [Pearson, $r(12) = 0.696$, $p = 0.006$; Fig. 4]. There was no effect of syllable position on reversal discrimination performance for the five positions tested: first, second, middle, penultimate, and last [one-way ANOVA, $F(4,156) = 1.890$, $p = 0.115$]. There was, however, a significant effect of syllable type on performance [one-way ANOVA, $F(4,156) = 10.590$, $p < 0.001$]. Syllables were categorized into five types: noisy (Moonshine's syllable C), stack harmonics (e.g., Bear's syllable D), sweeps (e.g., Scotch's syllable E), high (e.g., Julep's syllable E), and combo. Combo syllables are syllables that contain two or more types, with less than 5 ms of silence separating them. All syllables over 100 ms were combo syllables, and all syllables less than 100 ms fell into the remaining four categories. Pairwise comparisons using the Tukey test showed that hit rates were significantly higher for the combo group compared with the remaining four groups ($p < 0.05$).

This syllable type effect on performance could explain the saturation in discrimination performance for syllables greater than 100 ms in duration. All of the syllables longer than 100 ms in duration were combo syllables. They contained at least two distinct types of sub-syllables that become flipped in order when they are reversed in time, making reversals easy to discriminate. This is analogous to words that contain multiple syllables. For instance, the word "BIRDSONG" becomes "ƃuosᗡᴚ�Iꓭ". Reversing these words in time not only reverses the fine structure of each individual syllable (bird and song) but also reverses the order in which these syllables occur. This is an additional cue that can be used in reversal discrimination. Syllables that were
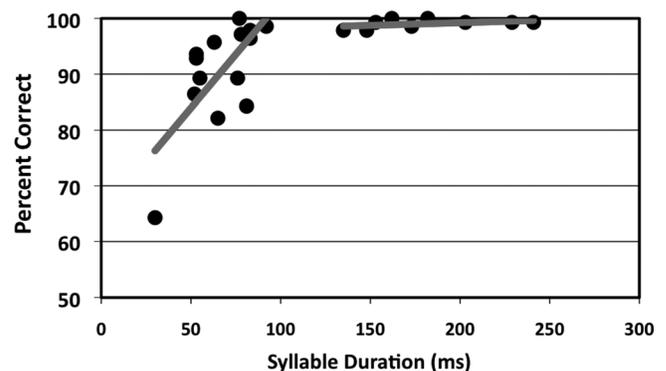

FIG. 4. Discrimination performance on single syllable reversals as a function of syllable duration. Regression lines show that for syllables greater than 100 ms, performance was nearly perfect ($y = 0.0097x + 97.291$). However, for syllables less than 100 ms in duration, performance was positively correlated with syllable duration ($y = 0.3831x + 64.812$).

shorter than 100 ms were equally distributed among the stack, sweep, noisy, and high groups, indicating that the duration effect seen for these short syllables was independent of syllable type.

There is earlier evidence that male zebra finches are most sensitive to their own songs, as they have specific neurons in the avian forebrain that respond best to the bird's own song, and very little to conspecific songs (Solis and Doupe, 1997; Theunissen and Doupe, 1998). Since hit rates for most syllable reversals were very similar and nearly 100%, response latencies were also analyzed to determine if a Bird's own song (BOS) effect existed. There was no BOS effect for male zebra finch performance on single syllable reversals. If birds were attending to song identity, then we would expect each bird's response latencies to be shortest for his own song, and longer for the conspecific songs. Instead, for each bird, the response latencies were shortest for the same motif. This result shows that birds attend to the acoustic properties of the song, rather than song identity. However, the exact features that birds listen to remains unclear. To test the salience of spectral structure, the next experiment tested reversal discrimination in a synthetic song in which spectral cues were highly reduced.

IV. EXPERIMENT 2: SYLLABLE PERCEPTION IN THE ABSENCE OF SPECTRAL STRUCTURE

Song normally contains complex, time-varying spectral structure, which is unique for each individual song syllable in the motif. This experiment examined the role of spectral structure in syllable discrimination by testing the ability of birds to discriminate syllable reversals in a song that had highly reduced spectral cues. The song envelope was filled with random Gaussian noise in order to reduce spectral cues while keeping the overall timing and rhythm of the song motif intact. In this way we were able to determine the role that spectral structure plays in syllable reversal discriminability. Although the spectrum remains stable over time, Gaussian noise does contain fine structure cues. The role of fine structure in syllable perception was tested by comparing performance on two types of synthetic songs: songs that contained the same piece of noise within each syllable and songs that contained a unique piece of noise within each syllable.

A. Stimuli

Noise motifs were created by filling the amplitude envelopes of the four motifs shown in Fig. 1 with random Gaussian noise. For each bird's motif, the syllables were isolated using Adobe Audition, and each syllable envelope was extracted using a Hilbert transform. Each syllable envelope was either filled with the same piece of random Gaussian noise (termed same-seed noise), or a unique piece of noise (termed random noise). Noise bursts were generated in MATLAB using the randn function. Same-seed noise was generated by specifying the seed. A separate seed was used for each motif. Each burst was then multiplied by its corresponding syllable envelope to generate a noise syllable similar to the natural syllable. These noise syllables were then concatenated with amplitude values of zero that were the same duration as

the natural song intervals, resulting in a noise song that had the same syllable envelopes, syllable durations, and interval durations as natural song.

Targets consisted of single noise syllable reversals, at all locations with the noise motif. These noise syllable reversals correspond to the natural syllable reversals tested in the previous experiment, with the exception that sub-syllable reversals were not tested in noise motifs. For motifs which contained less than seven syllables, the remaining targets were motifs in which all syllables were reversed but remained in sequential order (i.e., FEDCBA). An example of a noise motif is illustrated in Fig. 5. For both same-seed and random noise motifs, the syllable envelopes remained the same and it is only the fine structure of the noise that differed. In the random noise case, the noise was unique for each syllable. In the same-seed case, the noise was shared between syllables. Whether or not performance differs for the two types of noise will provide insight into the degree to which birds listen to the fine structure of the noise, versus the amplitude envelope of the syllables. Our hypothesis was that if birds listen to the syllable envelope when making these discriminations, then they should perform similarly for both types of noise motifs. However, if they listen to the fine structure in the noise more than the syllable envelope, then there should be a difference in performance. We expected birds to perform worse on the motifs containing random noise. This is because of the increased complexity of the
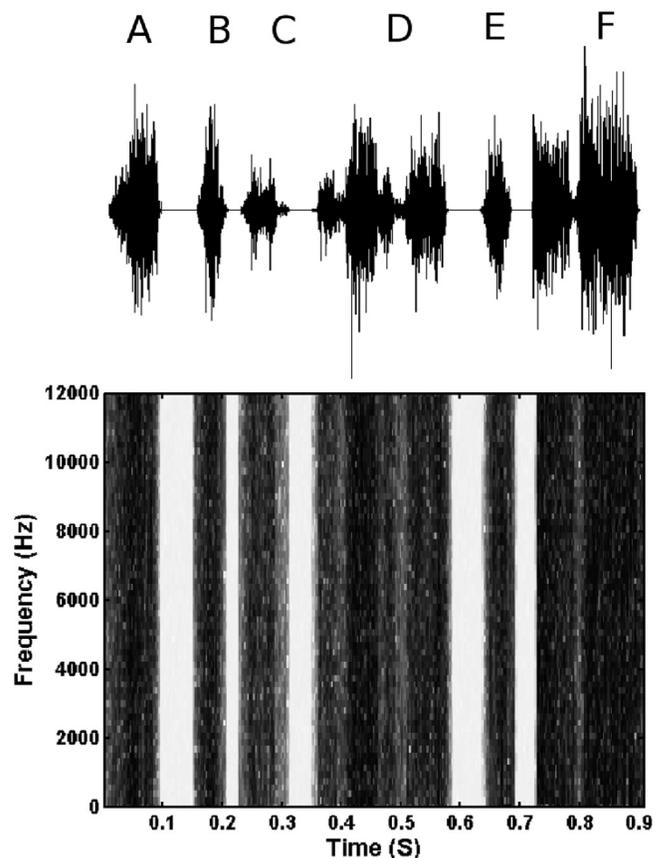

FIG. 5. Amplitude envelope (top) and spectrogram (bottom) of a noise motif, created by filling the song envelope of Moonshine's motif [Fig. 1(a)] with random Gaussian noise.

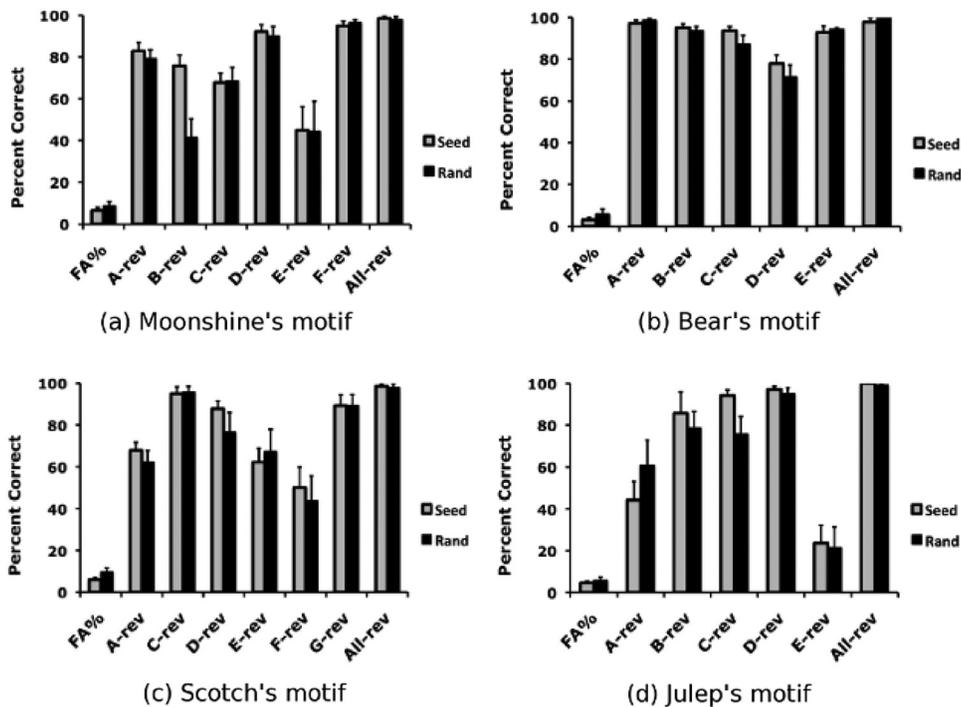B. A. Vernaleo and R. J. Dooling: Salience of cues in zebra finch song

FIG. 6. Percent correct discrimination performance on single syllable reversals in same-seed and random noise motifs. Results from all seven birds were averaged together. Error bars show standard error of the mean.

motif, as it does not have any shared noise among adjacent syllables.

B. Procedure

Birds were tested on the four noise song motifs, with each stimulus set containing single noise syllable reversals at all locations within the motif. The order in which the stimulus sets were tested was randomized.

Motifs were presented at 70 dB SPL, at a rate of once per 1500 ms. Each block of ten trials contained seven targets and three sham trials, presented in random order. Birds were run on 300 trials for each stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate did not exceed 20%.

C. Results and discussion

Even in the absence of spectral structure, birds were able to discriminate syllable reversals 70% of the time or greater, for the majority of the noise syllables tested. These results are shown in Fig. 6. Overall discrimination performance on single syllable reversals in natural motifs ($M = 94.42$, $\sigma = 13.87$) was significantly higher compared to discrimination of reversals in same-seed noise motifs ($M = 77.82$, $\sigma = 24.88$, $t = 7.226$, $p < 0.001$) and random noise motifs ($M = 74.12$, $\sigma = 27.87$, $t = 8.088$, $p < 0.001$). The decline in average performance was negatively correlated with syllable duration [Pearson, seed: $r(20) = -0.627$, $p = 0.002$; rand: $r(20) = -0.640$, $p = 0.001$]. Thus, the removal of spectral structure affected performance on shorter syllables, but not longer ones. As with experiment 1, there was a syllable duration effect for both types of noise motifs [Pearson, seed: $r(11) = 0.582$, $p = 0.037$; rand: $r(11) = 0.645$, $p = 0.017$] such that reversal discrimination of

syllables shorter than 100 ms was positively correlated with syllable duration.

There was no difference in performance between same-seed and random noise motifs. A two way ANOVA with noise type (same-seed, random) and syllable identity (Moonshine's syllable A, Bear's syllable B, etc.) as factors showed a main effect of syllable identity [$F(21,264) = 19.263$, $p < 0.001$], likely attributed to differences in noise syllable duration, as seen above. However, there was no significant main effect of noise type on performance [$F(1,264) = 3.407$, $p = 0.066$]. There was also no significant interaction effect [$F(21,264) = 1.016$, $p = 0.443$]. This shows that changing the noise within the song envelope had little effect upon performance. This suggests that the shared noise in the syllables of the same-seed motif does not provide an advantage in the listening task. Shared noise for each syllable means that birds could complete that task either by comparing each syllable in the background with the same syllable in the target or by comparing adjacent syllables within a motif. Since changing the noise for each syllable did not affect performance, birds were not likely comparing the noise in adjacent syllables. Performance was the same regardless of the noise that was inside of the envelope. Thus, syllable envelope is likely the main cue birds attend to when listening to noise songs.

To test this hypothesis, the rise and fall rates of syllable envelopes were calculated to determine if performance was correlated with forward/reverse symmetry of syllable envelopes. The rise/fall rates of the first and last 10 ms were calculated. This gives a measure of the attack and decay of the signal. The absolute value of the difference was taken as a measure of envelope asymmetry. Performance was significantly correlated with rise/fall asymmetry of the first and last 10 ms of the syllable for random noise motifs [Pearson,

$r(20) = 0.459$, $p = 0.032$], but not for natural [$r(20) = 0.224$, $p = 0.317$] or same-seed noise motifs [$r(20) = 0.398$, $p = 0.067$]. This suggests that when the motif became more complex by the addition of unique noise for each syllable, birds relied on syllable envelope cues to make forward/reverse discriminations.

While envelope cues provide a large amount of song information, recent evidence of fine temporal resolution in zebra finch auditory perception suggests that temporal fine structure in song may provide additional communication information. The final experiment in this paper examined the role of temporal fine structure in syllable perception, by removing spectral and envelope cues from song while keeping fine structure cues intact.

V. EXPERIMENT 3: PERCEPTION OF TEMPORAL FINE STRUCTURE WITHIN A SONG-LIKE ENVIRONMENT

Research over the last decade has shown that zebra finches are quite proficient in detecting temporal fine structure changes within synthetic stimuli, specifically changes to phase and harmonic structure (Lohr and Dooling, 1998; Lohr et al., 2006; Dooling et al., 2002). While these experiments test the perceptual limits of fine temporal processing in zebra finches, they do not ask whether birds use the same abilities when listening in a more natural setting to behaviorally relevant stimuli. All previous experiments testing zebra finches' ability to discriminate between phase changes have been in the context of short, single sounds (about the duration of a single syllable).

The goal of this experiment was to test whether zebra finches are also able to discriminate changes to only temporal fine structure, when the stimuli had the same overall timing cues as song. This was done by testing reversals in motifs made of Schroeder harmonic waveforms. Schroeder waveforms are harmonic complexes in which the long-term amplitude envelope and spectrum remain constant over time, but phase changes occur across frequencies. The result is a stimulus that only changes in phase information when reversed in time.

A. Stimuli

Schroeder waveforms were generated in MATLAB. The waveforms had a fundamental frequency of 640 Hz, which is within the normal range of fundamental frequencies for both zebra finch contact calls (Simpson and Vicario, 1990), and zebra finch song syllables (Williams et al., 1989; Williams, 2001). The waveforms consisted of ten harmonic components, had a frequency range of 640–6400 Hz, and had a 10 ms ramp at each end. The starting phases for each harmonic were determined by a modified version of the algorithm developed by Schroeder (1970),

$$\theta_n = C\pi n(n-1)/N, \quad (1)$$

where C is a scalar, n is the nth harmonic component, and N is the total number of harmonics in the waveform. This results in complexes in which the phase is either monotonically increasing or monotonically decreasing across frequency, and results

in either upward or downward frequency sweeps within each period of the complex. The scalar term determines the speed and direction of frequency sweeps, and in this experiment we used $+1$ and -1. This scalar represents the slowest speed for frequency sweeps, resulting in a flat temporal envelope. A scalar of $+1$ means that the phase is monotonically increasing (positive phase Schroeder), and a scalar of -1 means that the phase is monotonically decreasing (negative phase Schroeder).

For each song, Schroeder waveforms were generated to be the same duration as natural song syllables and were concatenated with amplitude values of zero that were the same duration as natural song intervals. The resulting stimulus was a string of Schroeder waveforms with a fundamental frequency of 640 Hz, which had the same rhythm as that of natural song. For the background Schroeder motif, only positive waveforms were concatenated. For targets, a single Schroeder waveform was reversed in time so that there was one negative waveform, and the rest were positive waveforms. The task was to discriminate single Schroeder syllable reversals within a Schroeder motif. For the motifs that contained less than seven syllables, the remaining targets were motifs in which all Schroeder waveforms were reversed, but remained in sequential order (as in the previous experiment). An example of a positive and negative Schroeder waveform, and of a Schroeder motif is shown in Fig. 7.

B. Procedure

Birds were tested on the four Schroeder motifs, with each stimulus set containing single Schroeder reversals at all locations within the motif. The order in which the stimulus sets were tested was randomized.

Motifs were presented at 60 dB SPL, at a rate of once per 1500 ms. Because Schroeder motifs lacked an amplitude envelope, motifs appeared much louder compared to the previous stimuli. Thus, a presentation level of 60 dB was chosen for a more comfortable listening level for the birds. Each block of ten trials contained seven targets and three sham trials, presented in random order. Birds were run on 300 trials for each stimulus set, and the last 200 continuous valid trials were used for analysis. Valid trials were trials in which the false alarm rate did not exceed 20%.

C. Results and discussion

Birds were able to discriminate single Schroeder syllable reversals at a high level of performance, even though the only acoustic cue present in reversals was phase information. Although overall performance on Schroeder syllable reversals ($M = 83.64$, $\sigma = 15.28$) was significantly lower compared to reversals of natural song syllables ($M = 94.42$, $\sigma = 13.87$, $t = 6.484$, $p < 0.001$), it was significantly higher compared to either type of noise motif (Schroeder v. seed: $t = 2.470$, $p = 0.014$; Schroeder v. rand: $t = 3.715$, $p < 0.001$). Thus, performance on Schroeder waveform reversals was the closest approximation to performance on natural syllable reversals, even in the absence of spectral or syllable envelope cues. Similar to the previous two experiments, there
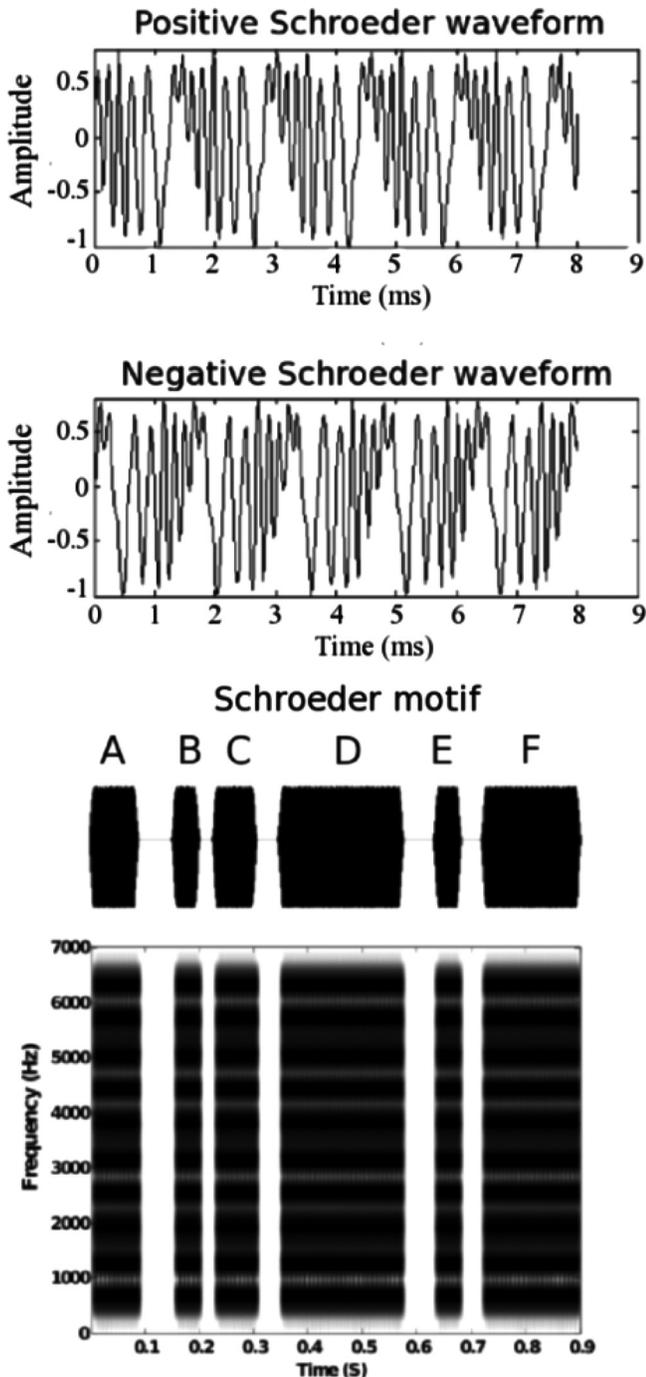
FIG. 7. Examples of positive and negative phase Schroeder waveforms (top), and an amplitude envelope and spectrogram of a Schroeder motif (bottom) modeled after Moonshine's natural song motif.

was a duration effect such that Schroeder waveform duration and performance was positively correlated for waveforms shorter than 100 ms [Pearson, $r(11) = 0.809$, $p < 0.001$]. A comparison of performance on natural syllables and Schroeder harmonic syllables is shown in Fig. 8.

The results in this experiment confirm previous results regarding temporal fine structure discrimination in birds. Here, we showed that birds are capable of temporal fine structure discrimination in a song-like acoustic environment. Schroeder harmonics in this experiment had a fundamental

frequency of 640 Hz, similar to that of zebra finch vocalizations. Previous work has indicated that zebra finches may have greater temporal precision than other birds and humans, as they are able to discriminate between positive and negative phase Schroeder waveforms up to 1000 Hz. Performance in budgerigars and canaries drops for frequencies higher than 700 Hz, and performance in humans drops for frequencies higher than 300 Hz (Dooling *et al.*, 2002; Lohr *et al.*, 2006; Lauer *et al.*, 2006). Since frequency sweeps occur once per period, Schroeder's harmonics with higher fundamental frequencies have shorter periods and the frequency glide occurs over a shorter time. Zebra finches are able to discriminate changes that occur over 1–2 ms, but other birds and humans need longer periods. Our results demonstrate that zebra finches use this temporal acuity when listening to biologically relevant, song-like stimuli, and that temporal fine structure is an important acoustic cue in song.

VI. GENERAL DISCUSSION

The song of the zebra finch, though simple in its syllable sequencing, is quite complex in its acoustic structure. Song motifs contain between three and eight syllables, each of which has a unique pattern of spectral, temporal, and amplitude envelope cues. Motifs are sung in a repetitive manner, with little variation from rendition to rendition. Syllables and the intervals separating them have consistent durations that provide an overall rhythm to the song. In addition, syllables contain amplitude envelope, spectral, and temporal fine structure cues. Through the use of auditory discrimination experiments, we have been able to determine the relative salience of time scales, and acoustic cues within zebra finch song motifs. The results presented in this study add to the way in which we think about zebra finch song, as a mode of animal communication, and as a model for human speech development.

A great deal has been learned about the neural mechanisms of song learning and production through electrophysiological recordings of the neurons in the anterior forebrain and motor production pathways, both in awake behaving zebra finches, and those asleep (Theunissen and Doupe, 1998; Solis and Doupe, 1997; Doupe and Solis, 1997; Nick and Konishi, 2005). However, only a few studies have been done to test predictions made from electrophysiology in actual behavior, specifically song perception (Lohr *et al.*, 2006). One prediction we tested was preference for the BOS, based on neural responses. Results found in experiment 2 show that during syllable discrimination tasks, zebra finches attend to the acoustic structure of the songs, and not the actual identity of the songs. All four males showed a similar pattern in which the average response latencies were the shortest for the same song. Thus, it appears that the BOS does not provide an advantage in perceiving changes to song. We did however find agreement with electrophysiological results in experiment 1, which tested the relative salience of temporal envelope and fine structure changes to the song motif. Neurons in HVC (proper name) are still sensitive to the BOS even when intervals between syllables are greatly increased.
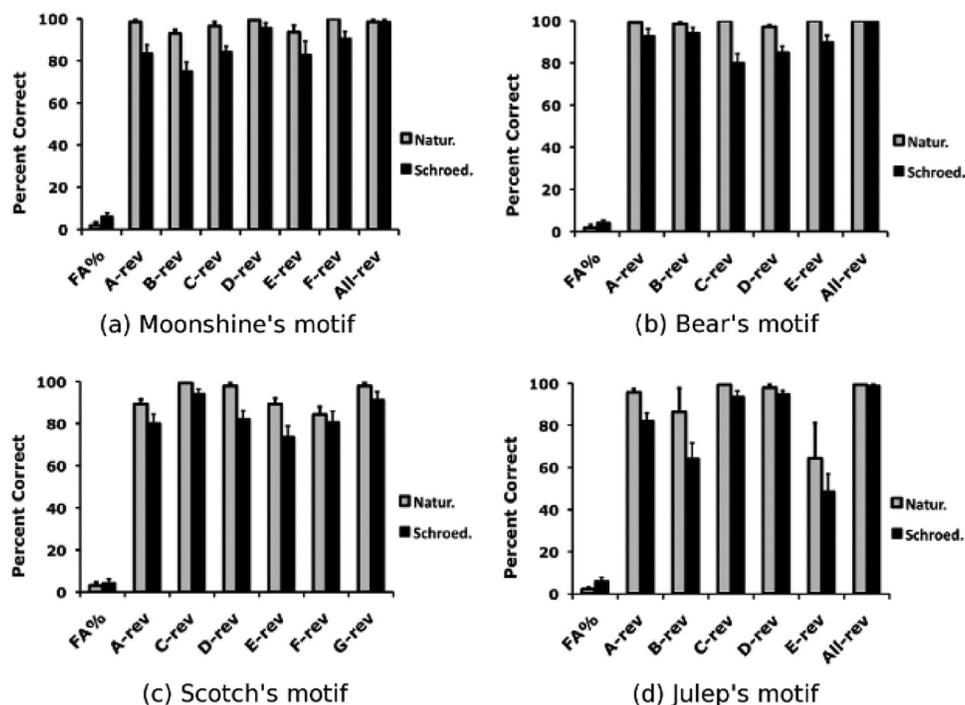
FIG. 8. Percent correct discrimination performance on single syllable reversals in natural and Schroeder motifs. Results from all seven birds were averaged together. Error bars show standard error of the mean.

(a) Moonshine's motif

(b) Bear's motif

(c) Scotch's motif

(d) Julep's motif

Margoliash and Fortune (1992) found that for HVC neurons that respond best to several syllables, neuronal responses remained strong even when an interval between the syllables was increased by over 200 ms. Behaviorally, we found that zebra finches were unable to discriminate between the original motif, and a motif in which a single interval had been doubled in duration. In this case, the sensitivities of neurons in the anterior forebrain matched perceptual sensitivities.

Birdsong has often served as an animal model for some aspects of human vocal development. The basis for this comes from parallels with speech, such as learning during a critical period, similar stages of development, and the necessity of auditory feedback. Few studies have tackled purely perceptual similarities between birdsong and human speech. One parallel emerging from our results is the use of envelope cues in the absence of spectral structure. Zebra finches were able to discriminate between forward and reversed syllables, even when the spectral structure was replaced with random Gaussian noise. A study by Drullman (1995) tested speech intelligibility in humans after filling the speech envelope of a sentence with random noise. On average, subjects were able to repeat back 98.30% of sentences. When listening to noise speech, subjects were able to use envelope cues alone to understand sentences. From these results, and the results of our experiments, amplitude envelope cues have been shown to be important in perception of both song and speech, and specifically play a role in understanding speech.

Our results also show that zebra finches can use temporal fine structure alone for syllable discrimination. Temporal fine structure in speech is important for perceiving voice quality and timbre (Rosen, 1992). These may also be important features for song perception given that related males have shared song elements in their repertoire. However, recent work has illustrated that temporal fine structure and

temporal envelope cues in speech are not entirely independent. Even when temporal envelope cues are removed from speech, the envelope can be reconstructed from the remaining fine structure information at the auditory periphery via narrow band filtering (Ghitza, 2001). While reconstructed temporal envelope cues do not contribute significantly to phoneme identification, reconstructed envelopes applied to noise or sinusoidal signals do result in somewhat intelligible speech (Gilbert and Lorenzi, 2006; Sheft et al., 2008). Using Schroeder harmonic stimuli to test perception of fine structure cues is actually an advantage, because the stimuli were constructed without an envelope and are not subject to envelope reconstruction. Thus, temporal fine structure cues were truly isolated, and zebra finches were able to perform syllable discriminations in the absence of envelope and spectral cues.

In conclusion, fine structure cues within-syllables tend to be more salient than overall temporal envelope cues within the motif, as zebra finches detect changes to syllables much more easily than changes to inter-syllable intervals. Similar to humans, zebra finches are sensitive to amplitude envelope features of syllables, particularly when spectral cues are not available. However, temporal fine structure information alone is sufficient for zebra finches to discriminate syllable-like stimuli in a song motif. Performance on Schroeder motifs most closely approximated performance on natural song motifs, which suggests that temporal fine structure may contain more information for song perception than previously thought. Thus, the correct perception/identification of syllables in song can be maintained by different cues, but perception is most robust when all cues are present in song. This is reminiscent of human speech, in which multiple redundant cues are used for recognition of phonemes.

Braaten, R. F., Petzoldt, M., and Colbath, A. (**2006**). "Song perception during the sensitive period of song learning in zebra finches (*Taeniopygia guttata*)," J. Comp. Psychol. **120**, 79–88.

Cynx, J. (**1993**). "Conspecific song perception in zebra finches (*Taeniopygia guttata*)," J. Comp. Psychol. **107**, 395–402.

Cynx, J., Williams, H., and Nottebohm, F. (**1990**). "Timbre discrimination in zebra finch (*Taeniopygia guttata*) song syllables," J. Comp. Psychol. **104**, 303–308.

Dooling, R. (**1982**). "Auditory perception in birds," in *Acoustic Communication in Birds, Vol. I: Production, Perception, and Design Features of Sounds*, edited by D. Kroodsma (Academic Press), pp. 95–130.

Dooling, R., and Haskell, R. (**1978**). "Auditory duration discrimination in the parakeet (*Meopsittacus undulatus*)," J. Acoust. Soc. Am. **63**, 1640–1642.

Dooling, R., Leek, M., Gleich, O., and Dent, M. (**2002**). "Auditory temporal resolution in birds: Discrimination of harmonic complexes," J. Acoust. Soc. Am. **112**, 748–759.

Dooling, R., and Lohr, B. (**2006**). "Auditory temporal resolution in the zebra finch (*Taeniopygia guttata*): A model of enhanced temporal acuity," Ornithol. Sci. **5**, 15–22.

Doupe, A., and Solis, M. (**1997**). "Song- and order-selective neurons develop in the songbird anterior forebrain during vocal learning," J. Neurobiol. **33**, 694–709.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–591.

Eales, L. (**1985**). "Song learning in zebra finches: Some effects of song model availability on what is learnt and when," Anim. Behav. **33**, 1293–1300.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listers to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438–2444.

Gilbert, G., and Lorenzi, C. (**2010**). "Role of spectral and temporal cues in restoring missing speech information," J. Acoust. Soc. Am. **128**, 294–299.

Glaze, C., and Troyer, T. (**2006**). "Temporal structure in zebra finch song: Implications for motor coding," J. Neurosci. **26**, 991–1005.

Hopkins, K., and Moore, B. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," J. Acoust. Soc. Am. **125**, 442–446.

Lauer, A., Dooling, R., Leek, M., and Lentz, J. (**2006**). "Phase effects in masking by harmonic complexes in birds," J. Acoust. Soc. Am. **119**, 1251–1259.

Lohr, B., and Dooling, R. (**1998**). "Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undula-tus*))," J. Comp. Psychol. **112**, 36–47.

Lohr, B., Dooling, R., and Bartone, S. (**2006**). "The discrimination of temporal fine structure in call-like harmonic sounds by birds," J. Comp. Psychol. **120**, 239–251.

Maier, E., and Klump, G. (**1990**). "Auditory duration discrimination in the European starling (*Sturnus vulgaris*)," J. Acoust. Soc. Am. **88**, 616–621.

Margoliash, D., and Fortune, E. (**1992**). "Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc," J. Neurosci. **12**, 4309–4326.

Nick, T., and Konishi, M. (**2005**). "Neural song preference during vocal learning in the zebra finch depends on age and state," J. Neurobiol. **62**, 231–242.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London B **336**, 367–373.

Schroeder, M. (**1970**). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.

Sheft, S., Ardoint, M., and Lorenzi, C. (**2008**). "Speech identification based on temporal fine structure cues," J. Acoust. Soc. Am. **124**, 562–575.

Simpson, H., and Vicario, D. (**1990**). "Brain pathways for learned and unlearned vocalizations differ in zebra finches," J. Neurosci. **10**, 1541–1556.

Solis, M., and Doupe, A. (**1997**). "Anterior forebrain neurons develop selectivity by an intermediate stage of birdsong learning," J. Neurosci. **17**, 6447–6462.

Sossinka, R., and Bohner, J. (**1980**). "Song types in the zebra finch (*Poephila guttata cas-tanottis*)," Z. Tierpsychol. **17**, 123–132.

Theunissen, F., and Doupe, A. (**1998**). "Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches," J. Neurosci. **18**, 3786–3802.

Williams, H. (**2001**). "Choreography of song, dance and beak movements in the zebra finch (*Taeniopygia guttata*)," J. Exp. Biol. **204**, 3497–3506.

Williams, H., Cynx, J., and Nottebohm, F. (**1989**). "Timbre control in zebra finch (*Taeniopygia guttata*) song syllables," J. Comp. Psychol. **103**, 366–380.