Title

Zebra finches are sensitive to combinations of temporally distributed features in a model of word recognition.

Journal

The Journal of the Acoustical Society of America, 144(2)

ISSN

0001-4966

Authors

Knowles, Jeffrey M
Doupe, Allison J
Brainard, Michael S

Publication Date

2018-08-01

DOI

10.1121/1.5050910

Peer reviewed

Zebra finches are sensitive to combinations of temporally distributed features in a model of word recognition[a)]

Jeffrey M. Knowles,[1,b)] Allison J. Doupe,[1,c)] and Michael S. Brainard[2,d)]
[1]*Center for Integrative Neuroscience, University of California, San Francisco, 675 Nelson Rising Lane, San Francisco, California 94158, USA*
[2]*Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, California 94158, USA*

(Received 1 June 2018; accepted 21 July 2018; published online 21 August 2018)

Discrimination between spoken words composed of overlapping elements, such as "captain" and "captive," relies on sensitivity to unique combinations of prefix and suffix elements that span a "uniqueness point" where the word candidates diverge. To model such combinatorial processing, adult female zebra finches were trained to discriminate between target and distractor syllable sequences that shared overlapping "contextual" prefixes and differed only in their "informative" suffixes. The transition from contextual to informative syllables thus created a uniqueness point analogous to that present between overlapping word candidates, where targets and distractors diverged. It was found that target recognition depended not only on informative syllables, but also on contextual syllables that were shared with distractors. Moreover, the influence of each syllable depended on proximity to the uniqueness point. Birds were then trained birds with targets and distractors that shared both prefix and suffix sequences and could only be discriminated by recognizing unique combinations of those sequences. Birds learned to robustly discriminate target and distractor combinations and maintained significant discrimination when the local transitions from prefix to suffix were disrupted. These findings indicate that birds, like humans, combine information across temporally distributed features, spanning contextual and informative elements, in recognizing and discriminating word-like stimuli. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5050910

[AMS] Pages: 872–884

I. INTRODUCTION

Human speech consists of a relatively small set of phonemes combined in sequences to form an enormous corpus of meaningful utterances (Hauser *et al.*, 2002). Though speech comprehension requires aggregating information across a range of time scales, word recognition is one example where sequence combinations are essential. To recognize a spoken word, human listeners must map an incoming sequence of acoustic elements onto internal representations of known word candidates and decide which candidate is consistent with the auditory sequence (Marslen-Wilson, 1987). The neural processes underlying auditory word recognition are poorly understood (DeWitt and Rauschecker, 2012), in part because of the difficulty in carrying out mechanistic studies in human subjects, and the lack of animal models for lexical processing. Here, we set out to model multi-syllabic word recognition in a non-human animal by training zebra finches (ZFs) to discriminate naturalistic, word-like, song motifs and testing whether and how birds' behavioral responses exhibit dependencies on stimulus features that parallel aspects of human word processing.

Birdsongs are rare examples of non-human animal vocalizations that contain complex spectrotemporal sequences similar to multi-syllabic human words (Doupe and Kuhl, 1999). The ZF song occurs in "motifs" that, like words, consist of discrete identifiable "syllables" that are organized in stereotyped sequences lasting hundreds of milliseconds (Brainard and Doupe, 2002; Doupe and Kuhl, 1999; Greenberg *et al.*, 2003). Songs are ethologically relevant in that birds naturally recognize individual identity based on the acoustic structure of the song (Woolley and Doupe, 2008; Reibel, 2009). Moreover, birds readily can be trained in operant paradigms to recognize and differentiate song stimuli (Cynx, 1993; Scharff *et al.*, 1998; Gentner, 2008; Nagel *et al.*, 2010). Previous studies demonstrate that ZFs are highly sensitive to the local acoustic structure of syllables (Braaten *et al.*, 2006; Vernaleo and Dooling, 2011; Dooling and Prior, 2017), but ZFs and other songbirds can also learn to classify songs according to the sequencing of both syllables and motifs (Chen and ten Cate, 2015; Chen *et al.*, 2015; van Heijningen *et al.*, 2013; Comins and Gentner, 2013; Gentner *et al.*, 2006; ten Cate, 2018). These indications that ZFs are sensitive to both local acoustic features and more global sequential organization of stimuli suggest that they could serve as a good model for studying aspects of word recognition.

We focus specifically on processes engaged in recognizing a given "target" word among a cohort of alternatives that

[a)]Portions of this work were presented in "Selective responses in the zebra finch auditory cortex reflect the time course of salient information" at the Society of Neuroscience Annual Meeting, Washington DC, November 2014 and "Sequence context influences behavioral recognition of and neural responses to familiar courtship song in zebra finches" at the Association for Research in Otolaryngology Annual Meeting, Baltimore, MD, February 2017.
[b)]Deceased June 7, 2018.
[c)]Deceased October 24, 2014.
[d)]Electronic mail: msb@phy.ucsf.edu

share overlapping features. In the case of multi-syllabic word recognition, humans attend to local features (phonemes and syllables), but also rely on combinations of these elements in temporal sequences ranging from hundreds of milliseconds to seconds (Marslen-Wilson, 1987). While some words are distinguished by elements that are unique to a target word, most words are combinations of elements that overlap with other possible candidates (Marslen-Wilson, 1987). For example, the initial utterance "cap" (/kaep/) could indicate any of a set of possible words such as "capital," "captive," "caption," or "capricious." Only when the listener subsequently hears the utterance "tain" (/tIn/), for instance, can the sequence be identified as "captain," and the other possible candidates rejected (Zwitserlood and Schriefers, 1995). The first phoneme at which a phonetic sequence is consistent with only a single candidate, in this case "tain", has been termed the uniqueness point (UP; Marslen-Wilson and Tyler, 1980). The prefix of a multi-syllabic word (preceding the UP) can be defined as temporal "context" since it is shared among multiple candidates. But attending to this context is also required to prevent confusion with other words that share the same suffix; for example, the context "cap" (/kaep/), enables distinction between "captain" (/kaep/-/tIn/) and a cohort of other words that share the suffix /tin/, such as "cretin" (/krE/-/tIn/). Thus, in these examples, sequential combinations of phonemes preceding and following the UP are critical to word recognition because each potential target shares overlapping segments with other possible candidates.

These studies of human word recognition indicate the importance of combinatorial processing of sound segments. Listeners must attend to initial acoustic elements preceding the UP that may be shared among a set of possible word candidates, but recognition of a specific target word requires combining information from this contextual prefix with acoustic elements following the UP. While processing of such temporally extended sequence combinations is required to identify specific words, humans carry this out in a rapid and automatic fashion that does not require conscious awareness of the analysis and integration of information preceding and following the UP (Marslen-Wilson, 1987). Nevertheless, several lines of experiments indicate that the UP plays a privileged role in mapping phoneme sequences onto internal representations of specific words (Marslen-Wilson and Tyler, 1980); in particular, both direct (Marslen-Wilson, 1984; Marslen-Wilson and Tyler, 1980) and indirect (Marslen-Wilson and Zwitserlood, 1989; Radeau *et al.*, 1989) measures show that word recognition is temporally locked to the initial acoustic features following the UP. Moreover, though evidence for the prominence of the UP in the processing of fluent speech remains mixed (Balling, 2010), recent studies suggest that word-specific neural activity is aligned to the UP (Kocagoncu *et al.*, 2017; Zhuang *et al.*, 2014).

Given these observations for human word recognition, we were interested in testing whether birds challenged with an analogous recognition task would similarly combine information about local acoustic features and their more global sequential organization in order to differentiate word-like targets from a cohort of distractors composed of overlapping elements. We additionally wondered whether bird target recognition, like human word recognition, would exhibit a pronounced sensitivity to UPs in the stimuli that corresponded to the earliest acoustic features that enabled discrimination of targets from distractors.

To examine whether ZF motif discrimination exhibits these features of human word recognition, we trained birds in an operant paradigm to respond to a particular target motif and suppress responses to a cohort of distractor motifs in a go/no-go task. We then systematically probed birds' utilization of individual syllables and syllable combinations in responding to the targets. In experiment 1, we tested whether birds based their responses solely on syllables that were explicitly "informative" about the target compared to the distractors, or whether their responses also depended on combinations of experimentally defined contextual and informative syllables. In experiment 2, we explicitly required birds to discriminate between targets and distractors based on syllable combinations, and examined how discrimination depends on particular syllables and their sequential organization.

II. METHODS

A. Experiment 1

Five groups of birds (22 total) were trained to discriminate a single target motif from a set of distractor motifs. The stimuli were all sequences of six syllables and were structured $ABCX_iY_iZ_i$, where letters indicate syllable identity. All the motifs shared the three-syllable prefix ABC and were distinguished by the three-syllable suffixes $X_iY_iZ_i$, which were unique for the target and each distractor [Figs. 1(a) and 1(b)]. The prefix ABC was "contextual" in the sense that it did not help to distinguish the target from the distractors. Four birds in each of Groups 1–3 (12 total) were trained on a round-robin in which individual syllables in the contextual (ABC), target ($X_tY_tZ_t$), and distractor ($X_{di}Y_{di}Z_{di}$) positions were rotated such that individual syllables served in context, target, and distractor positions for different groups [Table I(a)]. In Groups 1–3, the target was a fixed sequence $X_tY_tZ_t$ whereas the distractors had the six permutations each of $[X_{di},Y_{di},Z_{di}]$ (36 distractors in total). Two additional groups of 5 birds (Groups 5 and 6) were trained to discriminate one target $ABCX_tY_tZ_t$ from 30 distractors $ABCX_{di}Y_{di}Z_{di}$, where each distractor contained a unique syllable string $X_{di}Y_{di}Z_{di}$ [Table I(b)]. Once birds reached a criterion level of performance (less than 5% response to distractors and typically at least 80% response to targets—see Fig. 1(d), % response histograms at left), probes stimuli were presented in which syllables and groups of syllables of the target motif were omitted, substituted, or shifted in frequency. Probe stimuli were randomly interleaved across these different types and were presented on 10% of the trials. Responses to probe stimuli were unrewarded ("contingency-neutral"): regardless of response, birds received neither reward nor time out, and could initiate a new trial after 5 s had elapsed. Results presented in Figs. 2–4 reflect responses to stimuli following the initiation of probe trials.

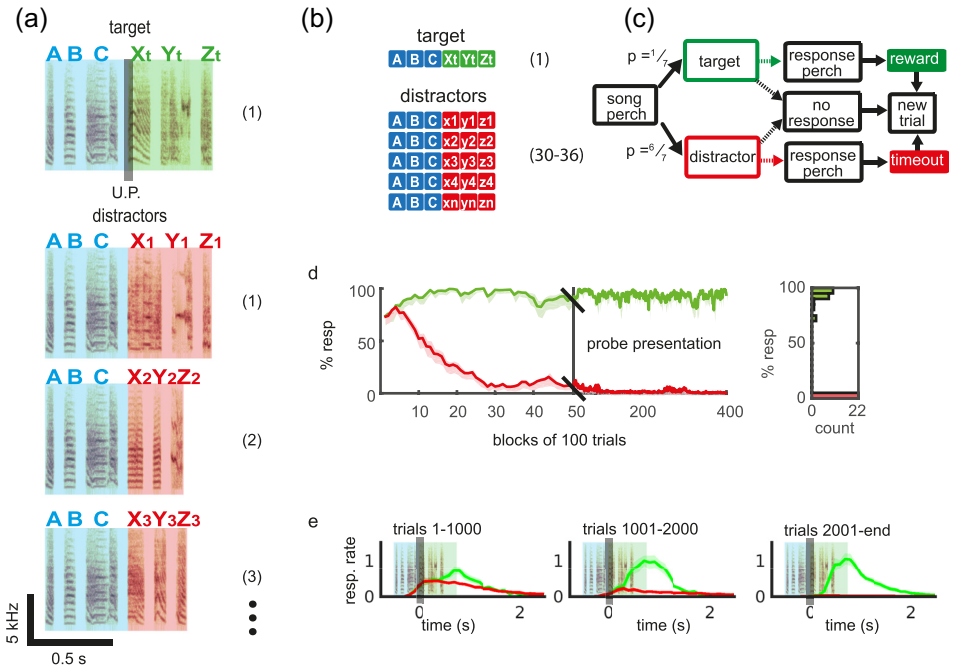

FIG. 1. (Color online) OC paradigm to examine the role of temporal context in behavioral discrimination of a target song motif. (a) Spectrograms of example target and distractor motifs for one group in experiment 1. The prefix syllables (ABC, blue shading) were constant among all motifs for a particular bird, whereas the informative syllables (XYZ, green and red shading) differed between the target (green) and each of the distractors (red). The UP occurs at the first unique syllable position (UP, gray bar). (b) Schematic representation of target and distractor stimuli in experiment 1. Birds discriminated 1 target from 30 to 36 distractors (see text). (c) State diagram of the go/no-go operant task (see text). (d) Percent responses to target and distractor stimuli during initial learning (trials 1–5000, expanded left) and probe presentation (trials 5000–40 000, compressed right) among $n = 22$ birds in experiment 1. Green curve: percent responses to target motif. Red curve: mean percent response to distractors for each bird. Shading: ±SEM. Far right: histogram of each bird's response to the target (green) and distractor (red) motifs at the start of probe presentation. (e) Response latency distributions calculated with a bin size of 100 ms and normalized by the number of trials presented (units: responses/second/trial) for the initial 1000 trials (left), second 1000 trials (middle), and all remaining trials (right), each averaged across 22 birds. Green curves: response rate to the target motif. Red curves: response rate to the distractor motifs. Response times have been aligned to the UP for each stimulus group ($t = 0$). Shading represents ±one standard error of the mean (SEM) among birds.

B. Experiment 2

Five birds were trained to discriminate two target motifs (ABCDEF, uvwxyz) from two distractor motifs (uvwDEF, ABCxyz). In this task, none of the syllables were individually informative about the identity of targets vs distractors, and birds were required to combine features from the suffix and the prefix [Fig. 5(a)]. The stimuli were balanced across experiments such that the distractors for two of the birds were the targets for the other three birds. Once birds reached asymptotic performance [Fig. 5(b)], probe stimuli were presented in which syllables in the target and distractor motifs were substituted or shuffled in position.

C. Animals

Female ZFs were obtained from an in house breeding colony, where they were raised along with male siblings by their parents before graduating to single-sex group cages. Beginning at 100 days old, females were isolated and placed in operant conditioning (OC) boxes for behavioral training.

TABLE I. Syllables used to generate stimuli for experiment 1. (a) Thirty-six syllables were used to generate target and distractor stimuli for groups 1–3. For each group, the prefix and target suffix consisted of three syllables in fixed string as indicated (fixed string denoted by 1–2–3). For each group, the distractor suffixes were the six permutations of the six sets of syllables (set of syllables denoted by [1,2,3]), creating 36 distractors for each group. (b) Ninety-six syllables were used to generate the target and distractor stimuli for groups 4–5. For each group the target prefix and suffix consisted of the fixed strings indicated. For each group, the distractor suffixes were the 30 fixed strings indicated for each group.

	Prefix	Target suffix	Distractor suffixes	Novel prefix
a: Stimuli for experiment 1 Groups 1–3				
Group 1	1–2–3	10–11–12	[4,5,6], [13,14,15], [19,20,21], [22,23,24], [25,26–27], [28,29,30]	7–8–9
Group 2	4–5–6	13–14–15	[7,8,9], [16,17,18], [19,20,21], [22,23,24], [25,26,27], [28,29,30]	1–2–3
Group 3	7–8–9	16–17–18	[1,2,3], [10,11,12], [25,26,27], [28,29,30], [31,32,33], [34,35,36]	4–5–6
b: Stimuli for experiment 1 Groups 4–5				
Group 4	1–2–3	4–5–6	7–8–9, 10–11–12, 13–14–15, 16–17–18, 19–20–21,…, 94–95–96	97–98–99
Group 5	1–2–3	10–11–12	4–5–6, 7–8–9, 13–14–15, 16–17–18, 19–20–21,…, 94–95–96	97–98–99

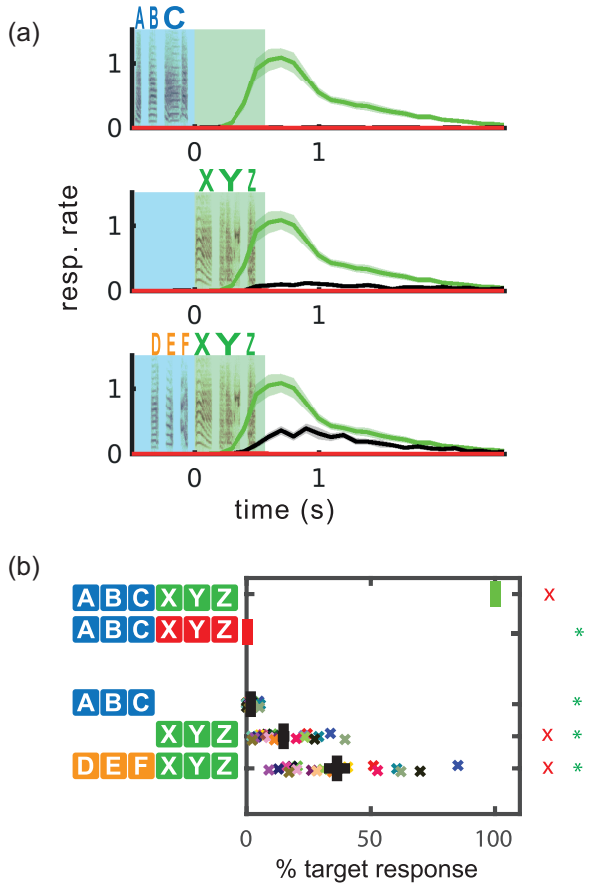

FIG. 2. (Color online) Manipulations of both contextual and informative syllables affect behavioral responses. (a) Probe stimuli in which the informative syllables were omitted (top), the contextual syllables were omitted (middle), and the contextual syllables were substituted (bottom). Response rate for these probe stimuli is shown in black. Response rate for target stimuli (green) and distractor stimuli (red) are superimposed on each plot for comparison. For each stimulus, the response latency distribution was calculated with a bin size of 100 ms and normalized by the number of trials (responses/trial/s). Behavioral responses have been aligned to the UP (time 0) for each group, and are plotted aligned to the spectrograms for group 1. The contextual and informative syllables have been shaded blue and green, respectively. Error shadings reflect mean response ±SEM across 22 birds. (b) Percent response of each bird to each stimulus divided by the response to the target. Row 1: Target (green); row 2: distractors (red); row 3: omission of informative syllables. Row 4: omission of the contextual syllables. Row 5: substitution of the contextual syllables. Colored x's indicate individual bird means. Heavy black lines indicate mean ±SEM across birds. For each stimulus, green *'s indicate significant difference from target response and red x's indicate significant difference from distractor responses (sign-rank; $p < 0.01$).

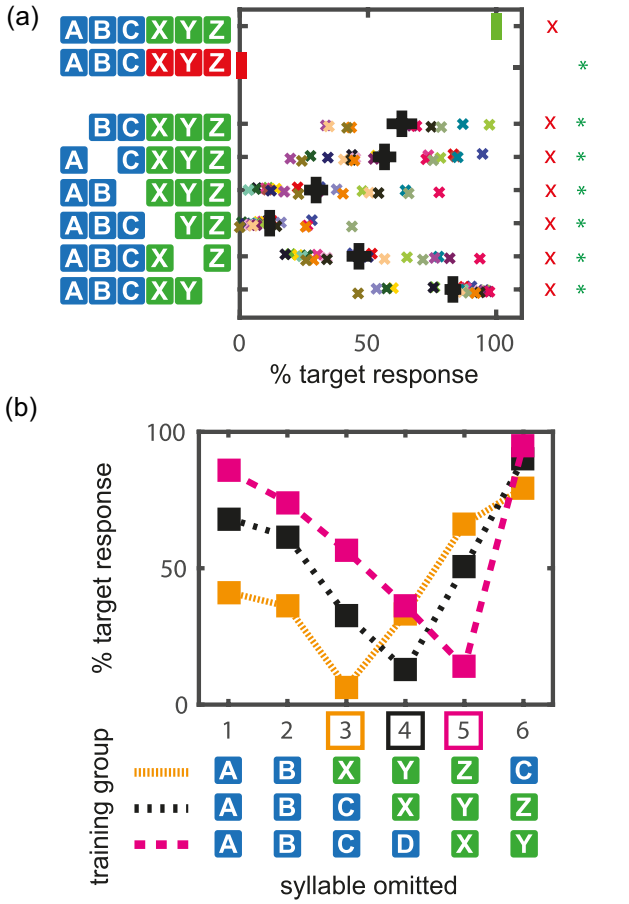

FIG. 3. (Color online) Syllable position relative to the UP determines influence on responses. (a) Responses to probes in which each syllable in each position was replaced with silence ($n = 22$ except for omission of syllable A, $n = 12$). Individual and group data are normalized as in Fig. 2. Green *'s indicate significant difference from target response; red x's indicate significant difference from distractor responses (sign-rank; $p < 0.01$). (b) Summary of responses to syllable omission probes across groups of birds trained on stimuli with the UP in position 2 (orange; $ABX_iY_iZ_iC$; $n = 2$), position 3 (black; $ABCX_iY_iZ_i$; $n = 22$), and position 4 (pink; $ABCDX_iY_i$; $n = 2$). Across experiments and UP positions, birds showed a consistent weighting of each syllable based on its position relative to the UP (onset of syllable X in each case, indicated by a colored square around the UP syllable position for that group). See text for statistical analyses. See supplementary material for data from individual birds trained with the UP in positions 2 and 4.[1]

All procedures conformed with the Institutional Animal Care and Use Committee at UC San Francisco.

D. Operant system

Birds were placed in custom-built operant training boxes containing a speaker, a song perch (SP), a response perch (RP), light emitting diode lighting, and an automatic feeder. The training apparatus was similar to that used in previous studies (Gentner and Margoliash, 2003; Nagel *et al.*, 2010). The behavior task, including stimulus delivery and each of the response contingencies illustrated in Fig. 1, was controlled by a custom system designed and implemented in house. Birds were housed in the behavior cages and performed approximately 500 to 2000 trials during each 12 h light cycle. Birds performed the task to obtain food, but their weight and health was monitored throughout training to make sure they were well fed. Water was provided *ad libitum*.

E. Pretraining

Birds were initially trained on a series of pretraining tasks lasting 3–5 days using practice stimuli. Birds first learned to hop on the SP to trigger song playback then on the RP to trigger the automatic feeder, then hop on the SP and RP in sequence to access food. Once they were consistently performing this task, birds began training on discrimination tasks (Nagel *et al.*, 2010).

F. Target discrimination task

In discrimination mode, birds were required to discriminate between target and distractor motifs [Fig. 1(c)]. Birds

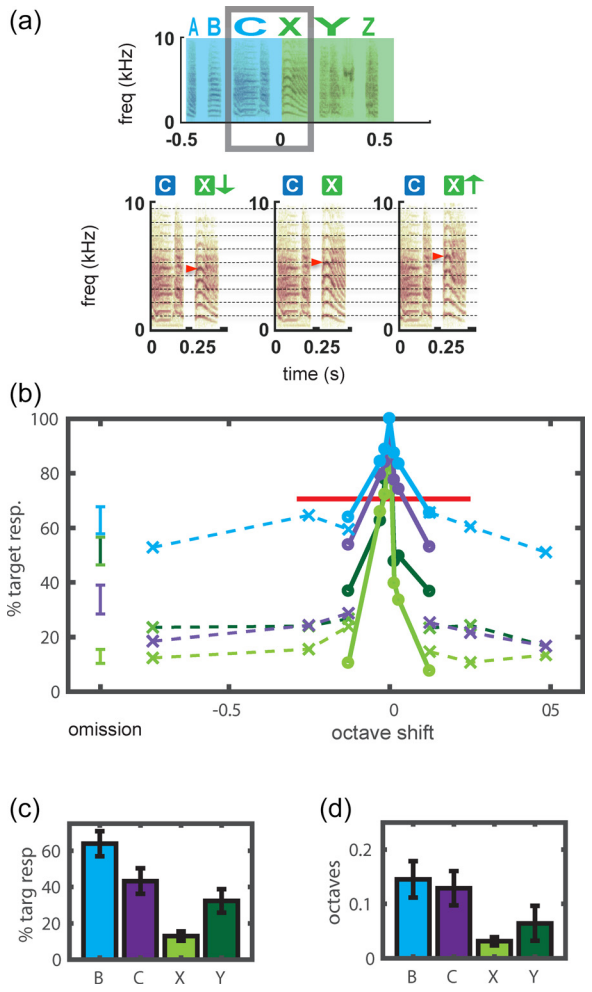

FIG. 4. (Color online) Responses to frequency-shifted syllables demonstrate tuning to individual syllables in context. (a) Illustration of probe stimuli with frequency shifted renditions of individual syllables. In this example, the syllable X has been shifted down (left) and up (right) by 0.125 octaves. Red arrows point to corresponding harmonics of each syllable to illustrate the magnitude of the shift. (b) Mean percent target response among birds to probes in which each of syllables B, C, X, and Y were omitted (left) or replaced with frequency shifted renditions (right). One group of birds ($n = 6$) was presented with grossly shifted renditions ranging from ± 0.125 to ± 0.5 octaves (dotted lines and x's). Another group of birds ($n = 10$) were presented with narrowly shifted renditions (solid lines and circles). Blue: syllable B; Purple: C; light green: X; dark green: Y. (c) Responses for all birds ($n = 16$) for ± 0.125 octave shifts of each syllable. Responses to stimuli with shifts of X were less than responses to all other shifted stimuli, and responses to stimuli with shifts of Y and C were less than responses to stimuli with shifts of B ($p < 0.01$ in all cases). (d) Bandwidth of tuning curves for each syllable at 75% of target response (red line). Bandwidth for shifts of X and Y were significantly less than bandwidths for shifts of C and B ($p < 0.01$ and $p < 0.05$, respectively).

initiated a trial by hopping on the SP, which triggered playback of a stimulus. Birds could respond anytime up to 5 s after trial initiation by hopping on the RP. On target trials, hopping on the RP triggered the feeder to deliver a seed reward for 5 s. On distractor trials hopping on the RP triggered a timeout period of 30–90 s in which the box light was extinguished and all perches were inactivated.

G. Stimuli

Motifs were generated from a database of syllables extracted from ZF songs. A 10 ms exponential ramp was

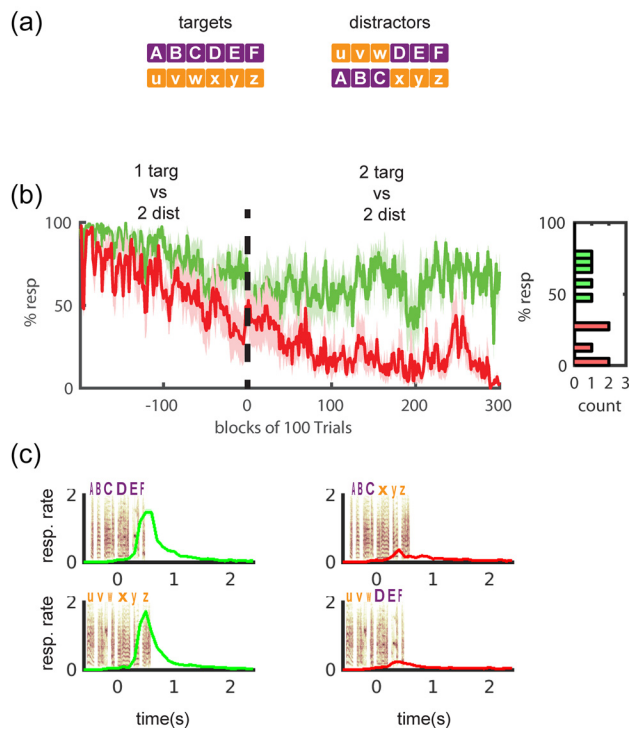

FIG. 5. (Color online) Behavioral paradigm to test sensitivity to sequence combinations. (a) Stimuli for experiment 2. Birds were trained to discriminate targets ABCDEF and uvwxyz from distractors uvwDEF and abcXYZ (letters denote distinct syllables). (b) Left: percent response to target and distractor combinations over the course of learning in experiment 2. Birds were first trained to discriminate each target combination one at a time from the distractor combinations, before discriminating both targets from both distractors. Right: histogram of each bird's percent response to targets and distractors at asymptotic performance. (c) Response rate over time for an example bird in experiment 2. Plots represent mean and 95% confidence intervals for the mean.

applied and syllables were adjusted to a constant root-mean-square. Motifs were presented at 70 dB sound pressure level. Motifs consisted of six syllables with durations ranging from 50 to 200 ms, separated by 50 ms gaps.

H. Analysis

Behavioral responses were recorded along with all other events by the behavior controller software and saved to Json text files that were loaded into MATLAB. In all figures, error bars for group data represent standard error of the mean across birds. All analyses and statistical calculations were performed using custom software written in MATLAB. For plots showing response latencies, latency distributions were calculated with a bin size of 100 ms and normalized by the number of trials presented (units: responses/second/trial), such that the area under each curve corresponds to the percent response rate for the corresponding stimulus.

III. RESULTS

A. Experiment 1

We first trained ZFs in an operant procedure that required the recognition of a single target motif from a cohort of partially overlapping distractors. In experiment 1, ZFs learned to respond to a target of the form $ABCX_tY_tZ_t$

and suppress responses to 30–36 distractors of the form $ABCX_{di}Y_{di}Z_{di}$ [Figs. 1(a)–1(c); see Sec. II]. The letters here denote syllables, which were fixed (ABC) or variable ($X_iY_iZ_i$) across stimuli. The structure of these stimuli made syllables in the suffix positions $X_iY_iZ_i$ explicitly informative because they indicated the identity of the stimulus as a target vs distractor, and the prefix positions ABC contextual since they did not. Thus, this paradigm models the discrimination among a set of words that differ in their final segments but share a common initial segment (such as captain and captive).

Birds learned to respond reliably to the targets and suppress responses to the distractors within ~3000 trials [Fig. 1(d)]. Across 22 birds trained in experiment 1 we computed the percentage of trials on which birds responded correctly to the target [Fig. 1(d), green] and incorrectly to the distractors [Fig. 1(d), red]. Initially, birds responded with high probability to both targets and distractors. Correspondingly, birds received a food reward on a high proportion of target trials, but experienced time out periods on a majority of distractor trials. Over the course of training, birds maintained high levels of responding to targets, but gradually learned to suppress responses to distractors, such that good asymptotic performance was achieved by about 3000 trials. Once birds reached a criterion level of performance (distractor response <5%), we presented probe stimuli in which syllables in the target motif were omitted, replaced with novel syllables, or spectrally shifted, in order to assess sensitivity of target responses to different stimulus features.

1. Response latencies reflect learned sensitivity to the UP

Similar to the case of multi-syllabic words, the ordering of syllables $ABCX_iY_iZ_i$ created a clear point in the sequence when birds begin to receive information about motif identity (the transition from C to X_i). We define this position as the UP [Fig. 1(a)]. The UP in this task is analogous to UPs in spoken word recognition (Marslen-Wilson, 1987); in each case the UP is the position in an auditory sequence at which acoustic information uniquely differentiates a word or motif from other candidates in the lexicon. Because the UP is hypothesized to play a prominent role in word discrimination, we were interested in the extent to which the presence of the UP in our stimuli similarly contributed differentially to shaping target responses.

We first assessed how the timing of the UP related to the latency of behavioral responses. Birds were allowed to respond to target and distractor stimuli at any time up to 5 s from trial initiation. The response time might reflect both the progression of learning and the features of the stimulus to which birds learn to attend in deciding to respond. In particular, to successfully discriminate targets from distractors, birds must learn to wait at least until the UP (syllable X), when acoustic differences between the target and distractors first occur. However, since birds were not required to respond quickly, it might be advantageous to listen to the entire stimulus before responding.

At the beginning of the discrimination task [Fig. 1(e), trials 1–1000], responses sometimes occurred during the playback of contextual syllables (ABC), prior to the UP. Since there is no information present in the contextual syllables that differentiates targets from distractors, it is not surprising that this period includes a high proportion of incorrect responses to distractors [Fig. 1(d)]. As birds learned to respond to the target and suppress responses to distractors, the latency distribution shifted such that birds only responded after the UP [green lines, Fig. 1(e), trials 1001–2000 and 2001–end]. This demonstrates that birds learned to withhold responses until after the UP, as required to successfully discriminate targets from distractors. However, birds did not typically wait until the last informative syllable (Z) before responding. Rather, on many trials, responses were initiated shortly after the UP and before the end of the stimulus. This suggests that birds weigh heavily the first informative syllable following the UP relative to other informative syllables in generating target responses. This possibility was further assessed (see below) by measuring responses to probe stimuli in which both contextual and informative syllables in each position were systematically manipulated.

2. Both contextual syllables and informative syllables contributed to target responses

We designed the stimuli in experiment 1 so that birds had to utilize the informative syllables ($X_iY_iZ_i$) to discriminate targets from distractors. After birds reached asymptotic performance, we confirmed that target responses depended on these informative syllables by measuring responses to probe stimuli that contained the contextual prefix, ABC, but omitted the informative suffix, $X_tY_tZ_t$. The rate of responding to these stimuli was very low, and not significantly different from the rate of responding to distractor stimuli [Figs. 2(a) and 2(b), "ABC___"].

The contextual syllables (ABC) provided no information that differentiates targets from distractors, so that birds did not need to attend to these syllables in order to differentiate targets from distractors. However, we hypothesized that if birds learned to recognize targets and distractors as word-like sequences, then target responses might depend on the appropriate combination of both contextual prefixes and informative suffixes. Indeed, we found that omission of the contextual syllables, ABC, nearly eliminated target responses; the response to $X_tY_tZ_t$ when presented in isolation was only 10% of target response [$r(X_tY_tZ_t) = 10\%$; Figs. 2(a) and 2(b), "___XYZ"].

It could be that birds required sound in the positions normally occupied by contextual syllables but were not sensitive to the specific acoustic properties of the contextual syllables. To test this, we substituted novel syllables DEF in the positions normally occupied by ABC for each group of birds. The presence of novel syllables in the contextual positions rescued responses significantly compared to omitting the contextual syllables, but responses to stimuli with novel syllables in contextual positions were only 40% of those to the original target [$r(DEFX_tY_tZ_t) = 40\%$; Figs. 2(a) and 2(b),

"DEFXYZ"]. These results indicate that the contextual syllables influenced responses to the informative syllables, even though there was no information in the contextual syllables that differed between targets and distractors. They suggest that in generating target responses, birds normally listen for the presence of combinations of syllables that span both the contextual prefix and informative suffix.

3. Birds demonstrated a systematic weighting of syllables based on proximity to the UP

To measure how individual syllables influenced target responses, we presented probe stimuli in which syllables in each position were manipulated one at a time (Fig. 3). Omitting any of the informative syllables significantly reduced target responses. However, the magnitude of reduction depended on the position of the syllable that was omitted. Omitting the first informative syllable, X, had the largest effect, reducing the response rate to 12% of the target response rate, while omitting Y and Z reduced the response rate to 51%, and 90% of the target response, respectively. We similarly found that there was a graded effect of omitting each of the contextual syllables, with the effect of omission depending on proximity to the UP (Fig. 3). Omitting the last contextual syllable, C, had the largest effect, reducing the response rate to 32% of the target response rate, while omitting B and A reduced the response to 62% and 70% of the target response, respectively.

One might expect that informative syllables would have a larger influence on target responses than contextual syllables. Strikingly, however, the designation of contextual vs informative did little to explain a syllable's impact on a bird's responses. A one-way analysis of variance (ANOVA) showed no effect of this variable on behavioral responses ($F = 0.03$; $DF = 1$; $p = 0.8$). Instead, we found that the influence of individual syllables depended on their position in the sequence rather than their direct relevance to differentiating targets from distractors. A one-way ANOVA showed a strong effect of sequence position ($F = 40.15$; $DF = 5$; $p < 0.0001$). *Post hoc* comparisons showed that birds responded significantly less to probe stimuli with omissions (Fig. 3) and substitutions (data not shown) of each syllable than to the full target motif. Birds' responses were most disrupted by manipulating the syllable X_t followed in order by C,Y_t,B,A,Z_t (*Post hoc* rank sum tests show that responses for manipulations of X were significantly less than for manipulations of C ($p < 0.001$), responses for manipulations of C were less than for manipulations of Y, B, and A ($p < 0.01$ in all cases), and responses for manipulations of Y, B, and A were significantly less than for manipulations of Z ($p < 0.01$ in all cases).

These data indicate that a syllable's position in the sequence determined its influence on target responses, but they leave two possible explanations of how position influences syllable weighting. First, the global position of a syllable in the motif (first, second, third, etc.) could be essential, such that syllables at a certain point after the motif onset or before the motif offset tend to be the most salient. Second, temporal proximity to important sounds could be essential, with syllables surrounding the UP having the greatest influence. To disambiguate the effect of global position from that of position relative to the UP (which were equivalent for birds trained to discriminate among stimuli of the form $ABCX_iY_iZ_i$), we trained two additional groups of birds to discriminate targets of the form $ABX_iY_iZ_iC$ and $ABCDX_iY_i$ from distractors of the same form [Fig. 3(b)]. For these groups, the position of the UP was shifted to different absolute positions in the motif, so among all birds position 3, 4, or 5 contained the first informative syllable.

These additional experiments demonstrated that birds' sensitivity to each syllable was determined by its position relative to the UP. Responses from birds from all three groups were run through a multi-factor ANOVA to compare the effect of absolute motif position (1,2,3,4,5,6) vs the position relative to the UP (−4,−3,−2,−1,0,1,2,3), as well as informative vs contextual category. We found that relative position explained the variance in birds' responses much better than other factors. In a three-way ANOVA, relative position had a significant effect ($F = 4.05$; $DF = 7$; $p < 0.001$), absolute position had no significant effect ($F = 1.06$; $DF = 5$; $p = 0.2$), and informative vs contextual category had no significant effect ($F = <0.11$; $DF - 1$; $p = 0.7$;). Together, the results from experiment 1 demonstrate that birds are sensitive to disruptions of all syllables of the target stimulus, but that sensitivity to disruptions depends systematically on a syllable's proximity to the UP.

4. Responses to the full motif were stronger than the sum of responses to extracted components

Given that omitting C or X_t strongly decreased responses, we wondered whether presenting these syllables on their own could drive responses, or whether the most important syllables only drove responses in the normal context of the motif. To address this question, we measured how responses to the full motif differed from linear sums of the responses to various components. In general, responses to the motifs were supra-linear sums of responses to subsets of syllables. For example, we found that responses to the syllable combination CX_t presented alone, r(CX_t), plus the response to the rest of the target motif, r(AB_ _Y_tZ_t) was only 27% ± 25% of the response to the entire target motif. Similarly, r(ABC) +r($X_tY_tZ_t$) was only 10% ± 15% of the response to the full target. Although omitting the contextual and target syllables had similar effects on target responses, the two classes of syllables had different independent effects on birds' responses. This contrast is demonstrated by r(ABC), which was zero or negligible for all birds tested, and r($X_tY_tZ_t$), which was significantly greater than zero for all birds tested (Fig. 2).

These results demonstrate that the motif structure is key to driving responses in this paradigm: the target syllables most necessary to driving responses (C and X_t) were not themselves sufficient to drive responses if they were presented out of their normal sequence positions. In this sense, the finding that r(CX_t) ≪ r($ABCX_tY_tZ_t$) indicates that the surrounding "motif context" AB_ _Y_tZ_t, was also necessary to elicit strong target responses.

5. Birds demonstrated a graded sensitivity to spectral shifts depending on syllable position

The experiments described above using omissions demonstrate that the presence and identity of all the syllables in the target sequence affected birds' responses, but do not test whether birds attend to the detailed spectral properties of all the syllables. For instance, the extended gap when a syllable has been omitted might itself act as a salient cue, like a rest in music. To test the sensitivity of birds to spectral properties of each syllable in the sequence, we presented motifs containing spectrally shifted renditions of the syllables B, C, X_t, and Y_t (Fig. 4). Previous studies (Bregman *et al.*, 2012; Nagel *et al.*, 2010) used analogous shifts applied to the entire motif to examine birds' selectivity and tolerance to pitch variation in song classification. Here, this parametric manipulation allowed us to measure how effects of varying the pitch of individual syllables compared with those arising from gross manipulations such as omissions and substitutions.

We found that birds were sensitive to spectral shifts of all syllables tested [Figs. 4(b) and 4(c)]. There was a symmetric and monotonic decrease in response rate relative to target as individual syllables were shifted either upwards or downwards in frequency, and shifting of individual syllables in each position by ± 0.12 octaves resulted in responses that were comparable to those elicited by stimuli in which the same syllables were completely omitted. Consistent with the earlier observation that effects of syllable omissions depended on proximity to the UP, we additionally found that the percent reduction in response caused by a given shift in frequency was greatest for the first informative syllable (syllable X), followed by syllables Y, C, and B [Fig. 4(c)]. Similarly, we measured the bandwidth of effective syllables {width of spectral shifts that elicited at least 75% of target responses [red line Fig. 4(b)]}. The bandwidth of tuning curves was narrowest for syllable X followed by Y, C, and B [Fig. 4(d)]. These probe responses indicate that birds were sensitive to the detailed spectral content of both informative and contextual syllables surrounding the UP.

B. Experiment 2

The results from experiment 1 indicate that the sequential structure of motifs, in addition to the presence, identity, and spectral content of particular sequence elements, was important in driving target recognition. The reliance of target recognition on both contextual and informative syllables could reflect that birds are "listening for" the specific combinations of these features that are normally present in the target. However, an alternative possibility is that abnormal features in the positions of the contextual syllables disrupts or "vetoes" responses to the subsequent informative syllables. In experiment 2, we therefore were interested in whether birds could learn to differentiate targets from distractors by virtue of recognizing specific combinations of prefix and suffix elements.

To test this, we trained birds to discriminate targets of the form ABCDEF and uvwxyz from distractors ABCxyz and uvwDEF [Fig. 5(a)]. To perform this task, birds were required to condition their responses to the second three syllables of each motif (DEF or xyz) on the identity of the first three syllables in the motif [ABC or uvw; Fig. 5(a)]. This stimulus organization created a UP analogous to that in experiment 1. However, unlike experiment 1, birds had to explicitly recognize combinations of the prefixes and suffixes in order to identify the target, since each suffix could be part of a target or distractor depending on the context.

1. Birds could discriminate targets from distractors based on sequence combinations

Birds were able to learn the dual-target, dual-distractor task in experiment 2. However, in preliminary experiments, we found that birds had difficulty learning if they were presented with all four stimuli at the outset. Instead, we successfully trained birds by first presenting them with one target vs two distractors at a time (1t vs 2d), rotating between targets as birds learned to discriminate each target from the two distractors [Fig. 5(b)]. After birds reached asymptotic performance in 1t vs 2d, we transitioned to both targets vs both distractors (2t vs 2d). Birds never reached the level of performance displayed in experiment 1, but they learned to respond robustly to each target and suppress responses to each distractor [Figs. 5(b) and 5(c)]. These results demonstrate that birds can combine information from the prefix and suffix to recognize the target stimuli.

2. Target recognition did not depend exclusively on first order transitions

Experiment 2 demonstrates that ZFs are capable of discriminating based on syllable combinations. But recognition of the target and distractor combinations could depend only on the local transition from the prefix to the suffix (CD and wx for targets vs Cx and Wd for distractors). In this case, the discrimination might only involve the recognition of the transition itself (a syllable-length feature), rather than the discrimination of syllable combinations over greater temporal extents.

To test which feature combination(s) birds utilized in the task, we presented sets of unrewarded probe stimuli in which both the target and distractor motifs were altered by substituting syllables surrounding the transitions from prefix to suffix, or rearranging the order of syllables surrounding the transition from prefix to suffix. The rationale for these probe stimuli is that if birds significantly discriminate among probe stimuli in which particular features are manipulated symmetrically in both the targets and the distractors, then those features are not essential to birds' discrimination among the original training stimuli.

We found that birds continued to respond more strongly to target probes than to distractor probes across stimuli in which single novel syllables were substituted into each position [Fig. 6(a)]. This indicates that birds were not exclusively reliant on any specific pairwise transitions between adjacent syllables in recognizing the target. However, for substitutions into position 4, the discrimination between target and distractor probes was no longer significant, suggesting that, as in experiment 1, the first informative syllable following the UP was especially important to target recognition. For

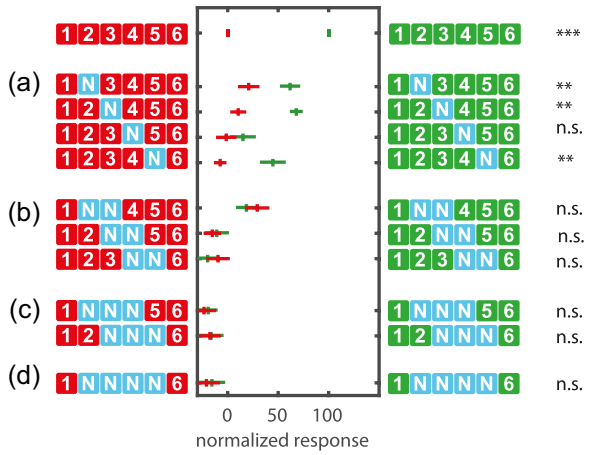

FIG. 6. (Color online) Responses to substitutions of syllables in the target and distractor motifs. Each row depicts a particular manipulation applied to the target (green) and distractor (red) motifs. Responses are normalized by subtracting each bird's mean response to the distractor stimuli then dividing by each bird's mean response to target stimuli. Top: normalized response to targets (100%) and distractors (0%). (a) Probes in which one syllable (in positions 2, 3, 4, and 5) was replaced with novel syllables (N). (b) Probes in which two syllables (in positions 2–3, 3–4, and 5–6) were replaced with novel syllables (N). (c) Probes in which three syllables (positions 2–3–4 and 3–4–5) were replaced with novel syllables (N). (d) Probes in which four syllables (positions 2–3–4–5) were replaced with novel syllables (N). For each manipulation and each bird, the average normalized response to the manipulated targets and manipulated distractors was calculated, and the mean across birds ±SEM is plotted. *'s indicate significant difference between the responses to target and distractor probes across $n = 5$ birds (paired t-test; ***$p < 0.0001$; **$p < 0.01$ *$p < 0.05$).

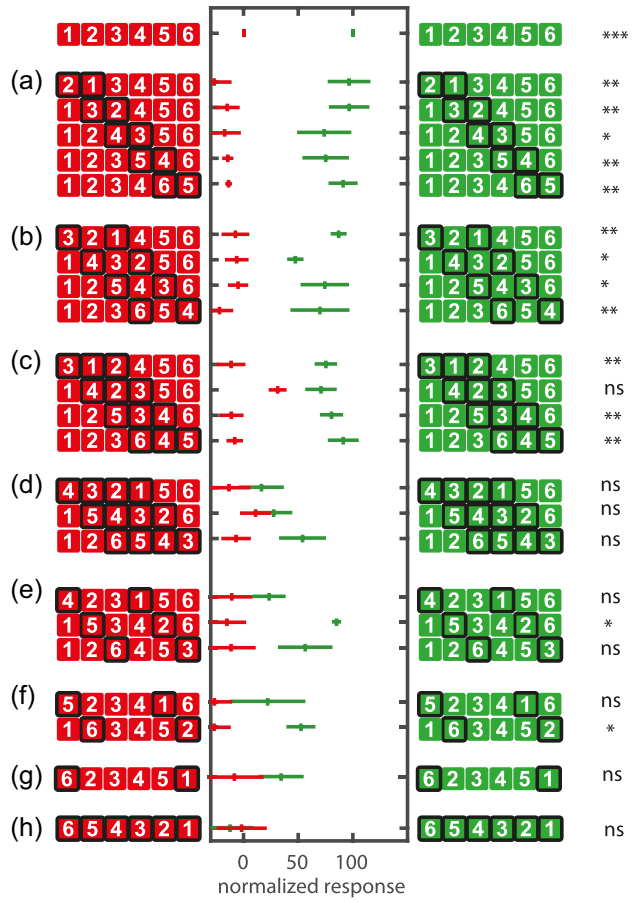

FIG. 7. (Color online) Responses to probes in which subsequences were reversed or shuffled. Each row depicts a particular manipulation applied to the target (green) and distractor (red) motifs. Responses are normalized by subtracting each bird's mean response to the distractor stimuli then dividing by each bird's mean response to target stimuli. Top: normalized response to targets (100%) and distractors (0%). (a) Probes in which two neighboring syllables were reversed (changes are highlighted by black outlines; positions 1–2, 2–3, 3–4, 4–5, and 5–6). (b) Probes in which triplets of syllables were reversed (positions 1–2–3, 2–3–4, 3–4–5, and 4–5–6). (c) Probes in which triplets of syllables were shuffled by one (positions 1–2–3, 2–3–4, 3–4–5, and 4–5–6). (d) Probes in which four syllable subsequences were shuffled (positions 1–2–3–4, 2–3–4–5, and 3–4–5–6). (e) Probes in which pairs of syllables separated by 3 were reversed (positions 1–4, 2–5, and 3–6). (f) Probes in which pairs of syllables separated by 4 were reversed (positions 1–5 and 2–6). (g) Probes in which the first and last syllable were reversed in position (positions 1–6). (h) Probes in which all six syllables, (positions 1–2–3–4–5–6) were reversed. For each manipulation and each bird, the average normalized response to the manipulated targets and manipulated distractors was calculated, and the mean across birds ±SEM is plotted. *'s indicate significant difference between the responses to target and distractor probes across $n = 4$ birds (paired t-test; ***$p < 0.0001$; **$p < 0.01$ *$p < 0.05$).

grosser disruptions in which two or more syllables were substituted, discrimination between target probes and distractor probes was completely eliminated [Figs. 6(b)–6(d)]. The loss of discrimination for these stimuli could reflect the removal of specific features that are normally important for differentiating targets from distractors, or alternatively could reflect the introduction of abnormal features associated with novel syllables.

In order to assess the importance of syllable ordering without the confound of introducing novel syllables, we tested responses to stimuli in which only the positions of the original syllables were rearranged. We found that birds continued to discriminate robustly between all target and distractor probes in which each neighboring pair of syllables were switched [Fig. 7(a)]. Indeed, discrimination continued to be significant for most stimuli in which adjacent triplets of syllables were rearranged [Figs. 7(b) and 7(c)] and was only grossly reduced for more global rearrangements of syllables [Figs. 7(d)–7(h)]. These data further demonstrate that birds did not rely on any single pairwise transitions in discriminating among stimuli, but instead relied on distributed combinations of features that distinguish targets from distractors.

IV. DISCUSSION

ZF motifs are similar to human words in time scale and spectrotemporal complexity (Doupe and Kuhl, 1999), providing an opportunity to examine whether ZF motif processing exhibits parallels to human word recognition. In this study, we asked whether ZFs process motifs as word-like units and to what degree birds are sensitive to word-like syllable combinations present in the stimuli. We found several striking parallels to human lexical processing associated with word identification and discrimination. First, we found that birds naturally learn to discriminate target stimuli from overlapping distractors by relying on combinations of temporally distributed features. Second, birds' recognition of target stimuli depends not only on combinations of informative stimulus features, but also includes sensitivity to the presence of contextual features that do not differ between targets and distractors. Third, birds exhibit particular sensitivity to

syllables and syllable combinations surrounding the UP, where target motifs diverge from the cohort of distractors. These parallels suggest that ZFs may provide a useful model for investigation of underlying mechanisms that contribute to perceptual processing associated with word recognition and discrimination.

A. ZFs are sensitive to combinations of contextual and informative elements

The ability to discriminate a target word from candidates with an overlapping initial sequence (captain vs captive) as well as those with an overlapping final sequence (captain vs cretin) demonstrates that word recognition depends on extended temporal combinations. While word recognition might occur quickly once sufficient evidence becomes available, correct recognition nonetheless depends on sensitivity to extended sequence features preceding and following the UP (Marslen-Wilson, 1987); captain and captive can be discriminated based on the informative suffixes, but are only meaningfully interpreted in the setting of the contextual prefixes. While there are many aspects of language acquisition and lexical processing of fluent speech that differ from our operant training paradigm, we sought specifically to model the combinatorial processing that is required for the perception and discrimination of word-like acoustic sequences.

Analogous to overlapping initial sequences in words, the prefix in experiment 1 (ABC) defines the contextual setting in which birds are confronted with informative suffixes that enable distinction between the target ($ABCX_tY_tZ_t$) and a cohort of distractors ($ABCX_{di}Y_{id}Z_{di}$). In principle, birds need not attend to the contextual syllables in the prefix to discriminate between the targets and distractors (Fig. 1). Yet omitting the contextual syllables (ABC) dramatically reduced target responses (Fig. 2). Indeed, omitting the last contextual syllable (C), reduced target responses nearly as much as omitting the first informative syllable (X) (Fig. 2). Further tests demonstrated that the presence, identity, and spectral properties of the contextual features were necessary to elicit full target responses (Figs. 2, 3, and 4). Hence, target responses depended on contextual syllables, even when they were not informative about the identity of targets vs distractors. Importantly, there was significant generalization of target responses to probe stimuli in which each individual syllable was omitted or modified, indicating that target responses did not rely exclusively on local acoustic features present in the target stimuli (including, for example, the transition from C to X). Together, these data indicate a natural tendency of ZFs, like humans, to combine information across temporally distributed features, spanning contextual and informative elements, in recognizing and discriminating word-like stimuli.

Experiment 2 demonstrates a further capacity of birds to combine information across features preceding and following the UP in recognizing targets. In this experiment, birds discriminated two targets, ABCDEF and uvwxyz from two distractors ABCxyz and uvwDEF. Because the target and distractor classes contained the same 12 syllables, birds could not simply base their discrimination on the presence or absence of any specific syllables (as was in principle possible for experiment 1). Rather, to accomplish this task, birds were required to recognize unique combinations of syllables from the prefix and suffix. Again, we found that birds maintained reduced but significant target responses to probe stimuli in which syllables in each position were disrupted, demonstrating that target recognition did not depend exclusively on local features or neighboring elements in the sequences. This natural tendency of birds to develop sensitivity to temporally distributed features in our experiment is consistent with a recent report demonstrating that birds can be trained to discriminate stimuli based on combinations of two syllables with non-adjacent dependencies (Chen and ten Cate, 2017). Overall, experiment 2 indicates that birds, like humans, can differentiate word-like targets from distractors by associating distinct responses with unique combinations of a shared set of prefix and suffix elements. Together with experiment 1, these results indicate that recognition of target stimuli does not rely exclusively on sensitivity to local acoustic features. Rather, birds develop sensitivity to temporally distributed features and can learn to discriminate between stimuli based on specific sequence combinations.

B. ZF motif responses reflect sensitivity to the UP

Based on prevalent ideas about the effect of serial position on recall and recognition, one might have expected syllables in the early or later sequence positions to have the greatest influence on recognition or reaction times (Wright *et al.*, 1985). Alternatively, syllables across all positions in the stimulus might have contributed equally to recognition; indeed, previous songbird studies that did not explicitly model discrimination of motifs with overlapping, word-like structure, found no consistent weighting of syllables as a function of position (Cynx, 1993; Vernaleo and Dooling, 2011). Instead of these possibilities, we found that for sets of stimuli with a well-defined UP, the influence of stimulus features was strongly correlated with their proximity to the UP (Fig. 3).

Prominent theories of auditory word recognition posit that the UP is a privileged position in the auditory sequence that is especially important in mapping phoneme sequences onto internal representations (Marslen-Wilson, 1987). This theory is supported by experiments indicating that humans recognize a word as early as possible in the auditory sequence; humans respond within hundreds of milliseconds after the UP when directed to detect target words (Marslen-Wilson, 1984; Marslen-Wilson and Tyler, 1980) or to make decisions that require reference to lexical knowledge of a target word (Radeau *et al.*, 1989; Marslen-Wilson and Zwitserlood, 1989).

ZF responses in our paradigm reflected sensitivity to the UP that parallels these observations for lexical processing in human speech. In experiment 1, response latencies and birds' sensitivity to local manipulations were both locked to the position of the UP. Similarly, in experiment 2, probe stimuli suggest that birds utilized the first syllables following the UP (D in ABCDEF and x in uvwxyz) in order to

discriminate the targets and distractors, though their performance of the task did not rely exclusively on the first order transitions from prefix to suffix (CD and wx vs Cx and wD). These data indicate that birds in both experiments recognized and responded to the target at a short latency following the UP, even though there was additional information about stimulus identity provided by each of the syllables following the UP.

These results suggest that the recognition of word-like stimuli by birds, as for humans, places a premium on rapidity. In speech, fast recognition of words is thought to reflect an efficient process that assesses word candidates in parallel (Marslen-Wilson, 1987), and as such could minimize the high order neural resources required to map sound sequences onto meanings (Christiansen and Chater, 2016). In our specific task, where trials were spaced a minimum of 5 s apart, there would seem to be little premium on responding immediately following the UP, rather than attending to the remaining informative syllables (which were completed within a couple of hundred milliseconds). That birds responded quickly after the UP suggests that the neural processes involved in target detection focused on the earliest available sounds that were informative about the target and distractor motifs.

C. Semi-local representation of song sequences

Previous studies of song recognition by birds have drawn a distinction between reliance on syllabic or subsyllabic features of birdsongs, termed "local features," and transitional or syntactic features that depend on the ordering of multiple syllables in the motif, which have been termed "global features" (ten Cate and Okanoya, 2012; Vernaleo and Dooling, 2011; Braaten *et al.*, 2006; Comins and Gentner, 2013; van Heijningen *et al.*, 2009; Gentner *et al.*, 2006; Okanoya *et al.*, 2000). Such studies also have indicated that some of the sensitivity to the global sequential structure of stimuli may be built up from sensitivity to "semi-local" representations of short sequences of syllables, such as pairwise transitions between syllables (Chen and ten Cate, 2015; Chen *et al.*, 2015; van Heijningen *et al.*, 2009, 2013). In our experiments, we similarly found that target recognition could depend on combinations of syllables distributed throughout the stimuli, but that birds were most influenced by syllables in close proximity to the UP.

The graded disruption of target recognition in our task arising from alteration of syllables around the UP provides a particularly salient demonstration of the presence and temporal extent of sensitivity to such semi-local features; for example, in experiment 1, target discrimination only required attending to the informative syllables (XYZ) that followed the UP, but behavioral performance reflected "unnecessary" sensitivity to the contextual syllables (ABC) that preceded the UP. This sensitivity was remarkably strong for the presence of the pairwise combination of syllables CX, with a progressive decrease in sensitivity to both contextual and informative syllables further from the UP (corresponding to a falloff over a period of many 10 s of milliseconds). The importance of semi-local features observed here parallels observations for human language processing that suggest a narrow temporal window, perhaps limited by some form of acoustic or lexical short term memory, over which features of speech can be combined (Christiansen and Chater, 2016).

Such sensitivity to semi-local syllable combinations in our experiments could potentially arise in at least two conceptually distinct ways. The enhanced sensitivity to features around the UP might reflect the temporal extent of an "attentional window" focused on the onset of informative syllables following the UP. In this case, birds might learn to recognize the target based on the co-occurrence of features (including contextual syllables) that fall within this window. Alternatively, alterations of the acoustic features preceding the UP might influence the recognition of subsequent informative syllables. Such an influence of preceding "acoustic context" on perception of successive speech sounds has been demonstrated for both humans (Holt and Lotto, 2008; Mann, 1980) and birds (Lachlan and Nowicki, 2015; Lotto *et al.*, 1997) and potentially could reflect interactions in the processing of successive sounds even at the earliest stages of the auditory system, prior to any learned representation of the target stimulus.

D. Neural mechanisms involved in sequence recognition

The behavioral sensitivity of ZFs to semi-local, word-like sequence combinations raises the possibility of examining the neural substrates that underlie the representation and perception of such stimuli. Previous neurophysiological studies have demonstrated responses to playback of the bird's own song (BOS) in ZF auditory forebrain and song pre-motor areas in male ZFs. These responses exhibit striking sensitivity to syllable combinations present in the BOS sequence (Bouchard and Brainard, 2013; Lewicki and Arthur, 1996; Margoliash and Fortune, 1992). Moreover, disruptions of these brain regions can have an influence on performance of birds in song recognition tasks (Brenowitz, 1991; Scharff *et al.*, 1998). However, because responses in these song-specialized regions are primarily selective for the BOS, it remains unclear how they might effectively contribute to perceptual processing of other song stimuli. Indeed, several lines of evidence suggest that other high order auditory regions, separate from song specific structures, might be engaged in representation of behaviorally relevant song stimuli. In particular, neurons in these auditory regions respond preferentially to conspecific songs compared to other stimuli (Calabrese and Woolley, 2015; Woolley *et al.*, 2010) but can also develop enhanced representations of song stimuli that have particular behavioral relevance to an individual, including the sequential organization of syllables (Gentner and Margoliash, 2003; Schneider and Woolley, 2013).

One barrier to understanding how the neural representation of extended sensory features in such higher auditory areas relates to stimulus recognition is that natural sequences, such as songs, contain ambiguous and potentially redundant features and feature combinations. The redundant stimulus features present a challenge for interpreting both

behavioral and neural responses to natural vocal sequences because it is difficult to tell which features in the stimuli drive learned behavioral and neural responses (Gentner and Margoliash, 2003; Kiggins *et al.*, 2012). In our experiments, limiting the features and feature combinations available to birds enforced consistent and quantifiable behavioral responses that depended on experimentally defined informative features, but were also influenced by contextual features in the preceding sequence. Of particular interest is whether training on this type of operant task, which has previously been demonstrated to lead to selective representation of salient stimuli in the avian auditory system (Gentner and Margoliash, 2003; Schneider and Woolley, 2013), might lead to a specialized representation of task relevant sequence combinations, such as syllable transitions surrounding the UP (Kocagoncu *et al.*, 2017; Zhuang *et al.*, 2014). More broadly, because birds in our target discrimination task exhibit behavioral sensitivity to stimulus features that mirror aspects of human lexical processing, they offer an attractive opportunity in future studies to examine potentially shared neural mechanisms that contribute to recognition and discrimination of word-like stimuli.

ACKNOWLEDGMENTS

We thank the members of the Brainard lab and Coleman labs for discussions and comments on the manuscript. This work was supported by the Howard Hughes Medical Institute, NIDCD 5R01DC011356, and an NSF GRFP to J.M.K.

[1]See supplementary material at https://doi.org/10.1121/1.5050910 for additional data related to Fig. 3.

Balling, L. W. (**2010**). "New directions for the uniqueness point," in *Linguistic Theory and Raw Sound*, edited by P. J. Henrichsen (Samfundslitteratur, Copenhagen), pp. 87–100.

Bouchard, K. E., and Brainard, M. S. (**2013**). "Neural encoding and integration of learned probabilistic sequences in avian sensory-motor circuitry," J. Neurosci. **33**, 17710–17723.

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.

Brainard, M. S., and Doupe, A. J. (**2002**). "What songbirds teach us about learning," Nature **417**, 351–358.

Bregman, M. R., Patel, A. D., and Gentner, T. Q. (**2012**). "Stimulus-dependent flexibility in non-human auditory pitch processing," Cognition **122**, 51–60.

Brenowitz, E. A. (**1991**). "Altered perception of species-specific song by female birds after lesions of a forebrain nucleus," Science **251**, 303–305.

Calabrese, A., and Woolley, S. M. N. (**2015**). "Coding principles of the canonical cortical microcircuit in the avian brain," Proc. Natl. Acad. Sci. U.S.A. **112**, 3517–3522.

Chen, J., and ten Cate, C. (**2015**). "Zebra finches can use positional and transitional cues to distinguish vocal element strings," Behav. Processes **117**, 29–34.

Chen, J., and ten Cate, C. (**2017**). "Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird," J. Exp. Psychol. Anim. Learn. Cogn. **43**, 295–302.

Chen, J., van Rossum, D., and ten Cate, C. (**2015**). "Artificial grammar learning in zebra finches and human adults: XYX versus XXY," Anim. Cogn. **18**, 151–164.

Christiansen, M. H., and Chater, N. (**2016**). "The now-or-never bottleneck: A fundamental constraint on language," Behav. Brain Sci. **39**, e62.

Comins, J. A., and Gentner, T. Q. (**2013**). "Perceptual categories enable pattern generalization in songbirds," Cognition **128**, 113–118.

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

DeWitt, I., and Rauschecker, J. P. (**2012**). "Phoneme and word recognition in the auditory ventral stream," Proc. Natl. Acad. Sci. U.S.A. **109**, E505–E514.

Dooling, R. J., and Prior, N. H. (**2017**). "Do we hear what birds hear in birdsong?," Anim. Behav. **124**, 283–289.

Doupe, A. J., and Kuhl, P. K. (**1999**). "BIRDSONG AND HUMAN SPEECH: Common themes and mechanisms," Annu. Rev. Neurosci. **22**, 567–631.

Gentner, T. Q. (**2008**). "Temporal scales of auditory objects underlying birdsong vocal recognition," J. Acoust. Soc. Am. **124**, 1350–1359.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (**2006**). "Recursive syntactic pattern learning by songbirds," Nature **440**, 1204–1207.

Gentner, T. Q., and Margoliash, D. (**2003**). "Neuronal populations and single cells representing learned auditory objects," Nature **424**, 669–674.

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (**2003**). "Temporal properties of spontaneous speech—A syllable-centric perspective," J. Phon. **31**, 465–485.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (**2002**). "The faculty of language: What is it, who has it, and how did it evolve?," Science **298**, 1569–1579.

Holt, L. L., and Lotto, A. J. (**2008**). "Speech perception within an auditory cognitive science framework," Curr. Dir. Psychol. Sci. **17**, 42–46.

Kiggins, J. T., Comins, J. A., and Gentner, T. Q. (**2012**). "Targets for a comparative neurobiology of language," Front. Evol. Neurosci. **4**, 6.

Kocagoncu, E., Clarke, A., Devereux, B. J., and Tyler, L. K. (**2017**). "Decoding the cortical dynamics of sound-meaning mapping," J. Neurosci. **37**, 1312–1319.

Lachlan, R. F., and Nowicki, S. (**2015**). "Context-dependent categorical perception in a songbird," Proc. Natl. Acad. Sci. U.S.A. **112**, 1892–1897.

Lewicki, M. S., and Arthur, B. J. (**1996**). "Hierarchical organization of auditory temporal context sensitivity," J. Neurosci. **16**, 6987–6998.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica)," J. Acoust. Soc. Am. **102**, 1134–1140.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.

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

Marslen-Wilson, W., and Tyler, L. K. (**1980**). "The temporal structure of spoken language understanding," Cognition **8**, 1–71.

Marslen-Wilson, W., and Zwitserlood, P. (**1989**). "Accessing spoken words: The importance of word onsets," J. Exp. Psychol. Hum. Percept. Perform. **15**, 576–585.

Marslen-Wilson, W. D. (**1984**). "Function and process in spoken word recognition: A tutorial review," in *Attention and Performance: Control of Language Processes* (Erlbaum, Hillsdale, NJ), pp. 125–150.

Marslen-Wilson, W. D. (**1987**). "Functional parallelism in spoken word-recognition," Cognition **25**, 71–102.

Nagel, K. I., McLendon, H. M., and Doupe, A. J. (**2010**). "Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task," J. Neurophysiol. **104**, 1426–1437.

Okanoya, K., Tsumaki, S., and Honda, E. (**2000**). "Perception of temporal properties in self-generated songs by Bengalese finches (Lonchura striata var. domestica)," J. Comp. Psychol. **114**, 239–245 (2000).

Radeau, M., Mousty, P., and Bertelson, P. (**1989**). "The effect of the uniqueness point in spoken-word recognition," Psychol. Res. **51**, 123–128.

Reibel, K. (**2009**). "Song and female mate choice in zebra finches - A review," Adv. Stud. Behav. **40**, 197–238.

Scharff, C., Nottebohm, F., and Cynx, J. (**1998**). "Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway," J. Neurobiol. **36**, 81–90.

Schneider, D. M., and Woolley, S. M. N. (**2013**). "Sparse and background-invariant coding of vocalizations in auditory scenes," Neuron **79**, 141–152.

ten Cate, C. (**2018**). "The comparative study of grammar learning mechanisms: Birds as models," Curr. Opin. Behav. Sci. **21**, 13–18.

ten Cate, C., and Okanoya, K. (**2012**). "Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning," Philos. Trans. R. Soc. Lond. B Biol. Sci. **367**, 1984–1994.

van Heijningen, C. A., Chen, J., van Laatum, I., van der Hulst, B., and ten Cate, C. (**2013**). "Rule learning by zebra finches in an artificial grammar learning task: Which rule?," Anim. Cogn. **16**, 165–175.
van Heijningen, C. A. A., de Visser, J., Zuidema, W., and ten Cate, C. (**2009**). "Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species," Proc. Natl. Acad. Sci. U.S.A. **106**, 20538–20543.
Vernaleo, B. A., and Dooling, R. J. (**2011**). "Relative salience of envelope and fine structure cues in zebra finch song," J. Acoust. Soc. Am. **129**, 3373–3383.
Woolley, S. C., and Doupe, A. J. (**2008**). "Social context-induced song variation affects female behavior and gene expression," PLoS Biol. **6**, e62, 525–537.
Woolley, S. M. N., Hauber, M. E., and Theunissen, F. E. (**2010**). "Developmental experience alters information coding in auditory midbrain and forebrain neurons," Dev. Neurobiol. **70**, 235–252.
Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., and Cook, R. G. (**1985**). "Memory processing of serial lists by pigeons, monkeys, and people," Science **229**, 287–289.
Zhuang, J., Tyler, L. K., Randall, B., Stamatakis, E. A., and Marslen-Wilson, W. D. (**2014**). "Optimally efficient neural systems for processing spoken language," Cereb. Cortex **24**, 908–918.
Zwitserlood, P., and Schriefers, H. (**1995**). "Effects of sensory information and processing time in spoken-word recognition," Lang. Cogn. Process. **10**, 121–136.