UCSF

UC San Francisco Previously Published Works

Title

Neural Encoding and Integration of Learned Probabilistic Sequences in Avian Sensory-Motor Circuitry

Permalink

https://escholarship.org/uc/item/9vp3w0g9

Journal

Journal of Neuroscience, 33(45)

ISSN

0270-6474

Authors

Bouchard, Kristofer E
Brainard, Michael S

Publication Date

2013-11-06

DOI

10.1523/jneurosci.2181-13.2013

Peer reviewed

Systems/Circuits

Neural Encoding and Integration of Learned Probabilistic Sequences in Avian Sensory-Motor Circuitry

Kristofer E. Bouchard[1,2,3,4] **and Michael S. Brainard**[3,4,5]

[1]Department of Neurosurgery, University of California, [2]Center for Neural Engineering and Prosthesis, University of California-San Francisco and University of California-Berkeley, and [3]Department of Physiology and [4]Center for Integrative Neuroscience, University of California, San Francisco, California 94158 and, [5]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815

Many complex behaviors, such as human speech and birdsong, reflect a set of categorical actions that can be flexibly organized into variable sequences. However, little is known about how the brain encodes the probabilities of such sequences. Behavioral sequences are typically characterized by the probability of transitioning from a given action to any subsequent action (which we term "divergence probability"). In contrast, we hypothesized that neural circuits might encode the probability of transitioning to a given action from any preceding action (which we term "convergence probability"). The convergence probability of repeatedly experienced sequences could naturally become encoded by Hebbian plasticity operating on the patterns of neural activity associated with those sequences. To determine whether convergence probability is encoded in the nervous system, we investigated how auditory-motor neurons in vocal premotor nucleus HVC of songbirds encode different probabilistic characterizations of produced syllable sequences. We recorded responses to auditory playback of pseudorandomly sequenced syllables from the bird's repertoire, and found that variations in responses to a given syllable could be explained by a positive linear dependence on the convergence probability of preceding sequences. Furthermore, convergence probability accounted for more response variation than other probabilistic characterizations, including divergence probability. Finally, we found that responses integrated over >7–10 syllables (~700–1000 ms) with the sign, gain, and temporal extent of integration depending on convergence probability. Our results demonstrate that convergence probability is encoded in sensory-motor circuitry of the song-system, and suggest that encoding of convergence probability is a general feature of sensory-motor circuits.

Introduction

Songbirds provide an excellent model for investigating neural encoding of sequence statistics. The learned song of each bird is composed of a unique set of categorical acoustic elements, termed syllables, organized into sequences (Fig. 1 *A*, *B*). For the Bengalese finch these sequences are highly variable, though the statistical structure of this variability is stable over time (Okanoya and Yamaguchi, 1997; Woolley and Rubel, 1997; Hampton et al., 2009; Warren et al., 2012). This variability in sequencing provides an opportunity to investigate if and how the probabilistic sequencing of events is encoded in sensory-motor circuitry.

We focus on nucleus HVC of the songbird forebrain, which is analogous to vocal premotor areas of human cortex (Doupe and Kuhl, 1999). HVC provides descending motor control for the temporal structure of learned song (Fig. 1*E*) and participates in syllable sequencing (Nottebohm et al., 1976; McCasland, 1987;

Vu et al., 1994; Yu and Margoliash, 1996; Hahnloser et al., 2002; Fee et al., 2004; Ashmore et al., 2005; Sakata and Brainard, 2008; Long et al., 2010). In HVC, as in many sensory-motor circuits, including the human speech system, the same neurons that are responsible for the control of behavior are also responsive to the sensory stimuli associated with the behavior (i.e., the sound of one's own voice) (Rizzolatti and Craighero, 2004; Prather et al., 2008; Sakata and Brainard, 2008; Edwards et al., 2010). For example, HVC neurons respond more strongly to playback of the bird's own song (BOS) than to manipulated versions of BOS, such as songs with reverse syllable ordering, suggesting that auditory responses of HVC neurons can be used as a read-out of how sequences are encoded in the song system (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995; Nishikawa et al., 2008).

Previous work has primarily characterized the statistics of behavioral sequences as the probability of transitioning from a given event to subsequent events, here termed divergence probability. Figure 1*C* displays a divergence diagram for the songs of an individual Bengalese finch. In this diagram, the thickness of each edge reflects the divergence probability of transitioning from a given syllable (x) to any other syllable (y) [defined for pairwise transitions as follows: $D(xy) = P(y|x)$]. Because syllable sequences are produced according to divergence probability, it has been hypothesized that HVC encodes divergence probability (Nakamura and Okanoya, 2004; Nishikawa et al., 2008; Jin, 2009). In contrast, for neural networks that are sequentially activated by sensory or motor patterns, postsynaptic Hebbian learning rules could result in the encoding of an alternative

Figure 1. Bengalese finch song, probabilistic characterization of sequence statistics, and sensory-motor organization of the song system. **A**, **B**, Example of Bengalese finch (Bf) song. **A**, Oscillogram (sound amplitude vs time) of a single Bf song segment, below which are the labels arbitrarily assigned to syllables. **B**, Spectrogram (power at frequency vs time) of the same song. **C**, Divergence diagram of syllable sequencing compiled from a large corpus of songs from this bird. Each node corresponds to a unique syllable from the bird's repertoire. Edge thickness corresponds to the observed divergence probability for transitions between syllables, and dots denote transition destination. The divergence probability for transition "xy," denoted $D(xy)$, reflects the probability that an "x" will be followed by a "y" [i.e., $P(y|x)$]. **D**, Convergence diagram compiled from the same corpus of songs as in **C**. The convergence probability for transition "xy," denoted $C(xy)$, reflects the probability that a "y" will be preceded by an "x" [i.e., $P(x|y)$]. Convergence and divergence probabilities can differ for the same transition. For example, in this case the divergence probability for transitions from "a" to "b" was approximately one-half [$D(ab) = P(b|a) = 0.55$], whereas the convergence probability for transitions to "b" from "a" was 1 [$C(ab) = P(a|b) = 1.0$]. **E**, Diagram of the avian song system. Acute recordings were made in nucleus HVC (used as proper name). There is a premotor latency of 30–50 ms (ΔT_{Motor}) between activity in HVC and subsequent vocalization. In addition, there is a latency of 15–20 ms ($\Delta T_{Auditory}$) for auditory activity to reach HVC. This makes for a total auditory-motor latency between premotor activity and resulting auditory feedback of 45–70 ms.

representation of sequence statistics, which we term convergence probability (Bouchard et al., 2012). Figure 1D shows a convergence diagram, in which each edge reflects the probability of transitioning to a given syllable from any other syllable [$C(xy) = P(x|y)$ where C indicates convergence probability]. Note that for variable sequences, divergence and convergence probabilities can differ (Fig. 1C, D). In this study, we take advantage of the distinction between different probabilistic characterizations of variably sequenced syllables to examine whether and how convergence probability is encoded in sensory-motor networks.

Materials and Methods

For each of the seven birds in this study, we extensively characterized the statistics of syllable sequencing during singing to generate individually cus-

tomized stimuli that would allow us to investigate how responses of HVC neurons depend on sequence probability. For each bird, we analyzed a sufficient number of syllable transitions (up to several thousand) to achieve a reliable characterization of the underlying probabilistic structure of syllable sequences. To investigate the effects of syllable sequence on the auditory responses of HVC neurons, we conducted acute electrophysiological recordings from individual neurons (single unit) and small clusters of neurons (multiunit) in sedated, adult male Bengalese finches. Based on the observed response characteristics and previous *in vivo* and *in vitro* studies (Hahnloser et al., 2002; Mooney and Prather, 2005), it is likely that the neurons we recorded are primarily HVC inhibitory interneurons. Previous studies have shown that this population receives convergent input from neurons that transmit descending motor control signals (Mooney and Prather, 2005; Rosen and Mooney, 2006; Bauer et al., 2008; RA projection neurons, Fig. 1E). Therefore, HVC interneurons are an appropriate target for monitoring the

population of RA projection neurons, and thus the integrated auditory response properties of neurons providing descending motor commands for song.

Animals. Seven adult male Bengalese finches (age >110 d) were used in this study. During the experiments, birds were housed individually in sound-attenuating chambers (Acoustic Systems), and food and water were provided *ad libitum*. Light:dark (14:10) photocycles were maintained during development and throughout all experiments. Birds were raised with a single tutor. All procedures were performed in accordance with established animal care protocols approved by the University of California–San Francisco Institutional Animal Care and Use Committee.

Electrophysiology. Several days before the first experiment was conducted on a bird, a small preparatory surgical procedure was performed. Briefly, birds were anesthetized with Equithesin, a patch of scalp was excised, and the top layer of skull was removed over the caudal part of the midsagittal sinus where it branches, as well as 0.2 mm rostral, 1.9 mm lateral of the branch point. A metallic stereotaxic pin was secured to the skull with epoxy and a grounding wire was inserted under the dura. For neural recordings, birds were placed in a large sound-attenuating chamber (Acoustic Systems) and stereotaxically fixed via the previously implanted pin. For the first penetration into a hemisphere, the second layer of skull was removed, the dura was retracted, and polydimethylsiloxane (Sigma) was applied to the craniotomy. Using the caudal branch point of the midsagittal sinus as a landmark, extracellular electrodes were inserted at stereotaxic coordinates for HVC (see coordinates above). At the end of a recording session, craniotomies were cleaned and bone wax was used to cover the brain. For subsequent penetrations into a previously exposed hemisphere, bone wax was removed and dura was retracted.

During electrophysiological recordings, birds were sedated by titrating various concentrations of isoflurane in O_2 using a non-rebreathing anesthesia machine (VetEquip). Typical isoflurane concentrations were between 0.0125% and 0.7% and were adjusted to maintain a constant level of sedation. Sedation was defined as a state in which the bird's eyes were closed and neural activity exhibited high rates of spontaneous bursts and maintained responsiveness to auditory stimuli. Within this range, both bursting baseline activity and highly selective BOS responses were similar to those previously found in the sleeping bird or in urethane-anesthetized birds. In one experiment, urethane was used instead of isoflurane. The results of this experiment were qualitatively similar to those of non-urethane experiments and were pooled with the rest of the data. Recordings were performed in sedated, head-fixed birds to achieve stable neuronal recordings in a controlled environment for the long durations required of the experimental paradigm. This preparation allowed for multiple experiments to be conducted in the same bird over the course of several weeks. Throughout the experiment, the state of the bird was gauged by visually monitoring the eyes and respiration rate using an IR camera. Body temperature was thermostatically regulated (39°C) through a heating blanket and custom equipment. Sites within HVC were at least 100 μm apart and were identified based on stereotaxic coordinates, baseline neural activity, and auditory response properties. Experiments were controlled and neural data were collected using in-house software (Multikrank; written by B. D. Wright and D. Schleef, University of California, Berkeley, CA). Electrophysiological data were amplified with an AM Systems amplifier (1000×), filtered (300-10,000 Hz), and digitized at 32,000 Hz.

Playback of auditory stimuli. Stimuli were band-pass filtered between 300 and 8000 Hz and normalized such that BOS playback through a speaker placed 90 cm from the head had an average sound pressure level of 80 dB at the head (A scale). Each stimulus was preceded and followed by 0.5–1 s of silence and a cosine-modulated ramp was used to transition from silence to sounds. The power spectrum varied <5 dB across 300–8000 Hz for white-noise stimuli. All stimuli were presented pseudorandomly.

Song analysis. All behavioral analyses and stimulus creation were done using custom code written in MATLAB software (Mathworks). Several days before a pinning surgery was performed, an adult male Bengalese finch was placed in a sound-attenuating chamber (Acoustic Systems) to collect audio recordings. An automated triggering procedure was used to record and digitize (44,100 Hz) several hours of the bird's singing. These recordings were then scanned to ensure that >50 bouts were obtained.

Bouts were defined as continuous periods of singing separated by at least 2 s of silence.

Bengalese finch songs typically consist of between 5 and 12 distinct acoustic events termed syllables organized into probabilistic sequences. Each bout of singing consists of several renditions of sequences, with each sequence containing between 1 and ~40 examples of a particular syllable. The syllables from 15 to 50 bouts were hand labeled and then analyzed as described below.

Analysis of song syntax. In this study, we examine how auditory responses to a target syllable, that we designate s_0, vary as a function of preceding sequences of length L (S^L). For example, in the sequence "*abc*," $L = 2$, $S^2 = $ "*ab*" and $s_0 = $ "*c*." We refer to sequences of length L preceding a particular target syllable s_0 as $S^L s_0$. To characterize the statistical structure of syllable sequences, several related measures of sequence production were calculated, each with distinct meanings:

The marginal probability of a unique sequence of length L characterizes the overall frequency with which a particular sequence of a given length (occurring N_{SL} times) was produced relative to the number of all sequences of that length (N_L):

$$P(S^L) = \frac{N_{SL}}{N_L} \tag{1}$$

Note that when $L = 1$, this is equivalent to $P(s)$, which is the overall frequency of syllable s relative to all other syllables.

The convergence probability characterizes the probability that a particular sequence (S^L) occurred before a given target syllable (s_0; see also Fig. 1):

$$C(S^L s_0) = P(S^L | s_0) \tag{2}$$

In terms of sequential ordering relative to the target syllable (s_0), in the special case where $L = 1$, $C = P(s_{-1} | s_0)$. For first-order Markov sequences, $P(S^L | s_0)$ can be decomposed into the product of pairwise probabilities:

$$C(S^L s_0) = P(S^L | s_0) = \Pi_{i=0}^{(L-1)} P(s_{i-1} | s_i) \tag{3}$$

However, given the non-Markovian nature of Bengalese finch syllable sequences, this decomposition is only an approximation. Therefore, our core analyses uses the empirically measured values of $C(S^L s_0)$ instead of the Markov approximation.

The divergence probability characterizes the probability that a particular target syllable (s_0) followed a given sequence (S^L) (Fig. 1):

$$D(S^L s_0) = P(s_0 | S^L) \tag{4}$$

in terms of sequential ordering relative to the target syllable (s_0) in the special case where $L = 1$, divergence probability $D = P(s_0 | s_{-1})$.

Marginal probability, convergence probability, and divergence probability are related through Bayes theorem:

$$D(S^L s_0) P(S^L) = P(S^L, s_0) = C(S^L s_0) P(s_0) \tag{5}$$

where $P(S^L, s_0)$ here denotes the probability of the combined sequence S^L followed by s_0 [$P(S^L s_0)$], which for our purposes is simply the probability of the sequence $S^L s_0$ occurring.

We estimated the sampling error in our calculation of probabilities by randomly resampling (with replacement) 50% of the labeled songs 1000 times. For each of the 1000 resamplings, we calculated convergence probabilities on each half of the dataset. We used the mean absolute difference between convergence probabilities calculated on each half as our measure of sampling error. Across sequence lengths, $L = [1:9]$, the mean estimation error was <2%.

Construction of synthetic BOS stimuli. As noted above, Bengalese finch song consists of many acoustically distinct syllables separated by periods of silence called intersyllable gaps. Over several bouts of singing, a bird will produce hundreds of examples of each of these syllables and each example, although identifiable as a particular syllable, will have subtle acoustic differences. The same holds for the intersyllable gaps. Our goal was to examine how auditory responses to individual syllables are modulated by sequence variability. To isolate sequence variability from these

other sources of variability found in song, we created synthetic stimuli (sBOS and pseudorandom, see below). These stimuli consisted of one representative exemplar of each syllable selected from among those naturally produced by the bird; in addition, these stimuli were constructed using the median of all intersyllable gaps for each intersyllable gap.

To arrive at an objective procedure for selecting the exemplar syllable, first, for each syllable, we calculated a feature vector that incorporates parameters based upon the syllable's amplitude, spectral, and temporal profile (Sakata and Brainard, 2008). Next, these feature vectors were grouped according to syllable type. Therefore, each distinct syllable type is represented by hundreds of points in a feature cloud corresponding to the individual examples of that syllable. Finally, we chose a representative exemplar from near the center of mass of each cloud. To evaluate how far each example of a syllable was from the center of mass, for all syllables (x_j) belonging to each unique syllable type (s), the Mahalanobis distance (MD) between an instance of the syllable and the group mean (μ_s) was calculated as follows:

$$MD\left(x_j, \mu_s, \sum_s\right) = \sqrt{(x_j - \mu_s) \sum_s^{-1} (x_j - \mu_s)'} \quad (6)$$

where Σ_s is the within-group covariance matrix.

Creation of pseudorandom stimuli. To probe how sequences are encoded in the population of HVC neurons, a stimulus set that consisted of 10 strings of 1000 pseudorandomly ordered syllables was constructed. We used the same exemplar syllables and median intersyllable gap as were used for synthetic BOS stimuli (see above). These 10 strings consisted of two distinct subsets of sequences randomly strung together: those that were naturally produced by the bird and those that contained no naturally occurring syllable transitions. To generate natural strings of length L, we randomly selected from among all unique sequences of syllables of length L produced by the bird. All natural sequences of a particular length had an equal probability of being chosen. To generate non-natural sequences of length $L = [1:10]$, we first randomly generated sequences of syllables in which each syllable that the bird produced had an equal probability of occurring in any of the L positions. Sequences with $C \leq 0.01$ were considered non-natural. These sequences were then screened and only those sequences with no natural transitions were selected. Natural and non-natural sequences were included with equal probability and with equal probability of being of any length ($L = [1: 10]$). This procedure was repeated until 10 strings of 1000 syllables were created. The equal balancing of natural and non-natural sequences was used to achieve good sampling of both types of sequences because of the relatively small number of naturally occurring sequences compared with all N^L combinations of N syllables organized into length L sequences. During playback, these 10 strings were presented randomly interleaved, separated by 2 s of silence. Each of the 10 strings was presented between 10 and 20 times per site.

In pseudorandom stimuli, by design, every sequence produced by the bird (of length 1–10) has an equal probability of occurring. The probability of a given sequence being inserted in the pseudorandom stimulus is $1/M$, where M is the total number of unique sequences of a given length for that bird. Note that the number of possible sequences rapidly increases with sequence length: for N unique elements, the number of possible sequences of length L is N^L. Therefore, the number of occurrences of a given sequence will tend to decrease with increasing sequence length. However, for naturally occurring sequences, this effect is independent of the probability of the sequence because all sequences of a given length have equal probability of being included in the stimulus during construction. Further, because Bengalese finch syllable sequences occupy a relatively small subset of the space of all possible sequences of a given length, the effect of increasing sequence length on frequency of occurrence is much less than the theoretical value of $1/N^L$.

This stimulus set is analogous to a "colored" white-noise stimulus like that which might be used to map out the receptive field of primary sensory neurons using the reverse correlation technique. Reverse correlation works by correlating the occurrence of spikes with values of the preceding stimulus. The general technique of reverse correlation calculates functions that describe how the system transforms, or filters, the input (stimulus) into the output (spikes). By pseudorandomly arranging

syllables into long strings, we have in effect created colored noise in sequence space. Tracking the neural response to single syllables preceded by different sequences allows us to determine how the evoked response depends on the length and probability of the preceding sequence, which is analogous to reverse correlation.

Spike sorting and calculation of instantaneous firing rates. Single units were identified events exceeding 6 SDs from the mean and/or were spike sorted using in-house software based on a Bayesian inference algorithm, as described previously (Sakata and Brainard, 2008). Multiunit neural data were thresholded to detect spikes >3 SDs away from the mean. Both single and multiunit spike times were binned into 5 ms compartments and then smoothed using a truncated Gaussian kernel with a SD of 2.5 ms and total width of 5 ms.

Characterization of neural responses to individual syllables. To characterize the responses to individual target syllables (s_0), we defined a response window that started 15 ms after the onset of the syllable and extended 15 ms after the offset of that same syllable. The 15 ms offset reflects HVC response latency and was derived by looking for the peak in the cross-correlation of the smoothed neural response with a binary representation of song. A binary representation of song was one in which the amplitude waveform has been replaced with 1 s throughout an acoustic event and 0 s elsewhere. This response latency is similar to that found in awake Bengalese finches (Sakata and Brainard, 2008).

We created stimuli consisting of long strings of pseudorandomly sequenced syllables that contain many instances of a particular sequence $S^L s_0$. An entire pseudorandom stimulus (10,000 syllables) was repeated between 10 and 20 times at each recording site. Within the pseudorandom stimuli, we define a sequence $S^L s_0$ as above (see Analysis of song syntax) with the additional constraint that the transition to the first syllable in the sequence (S^L) constitute a non-natural transition. The time varying spike rate in response to a particular instance of $S^L s_0$ found in the pseudorandom stimuli, $R(S^L s_0, t)$, was defined as the average spike rate during $S^L s_0$, where the average was taken across repetitions of the stimuli. These responses were then averaged across all instances of $S^L s_0$ found in the stimulus, yielding $\bar{R}(S^L s_0, t)$. The mean response rate to the last syllable in a sequence (s_0), $\bar{R}(s_0|S^L)$, was calculated by averaging each $R(S^L s_0, t)$ across the target syllable's response window and then averaging across all instances of $S^L s_0$. In the text, when the response to a particular $S^L s_0$ is displayed, we plot $\bar{R}(S^L s_0, t)$ or $\bar{R}(s_0|S^L)$ and the SE measure is calculated across all instances of $S^L s_0$ found in the stimulus.

Syllables that naturally occur as repeats (e.g., "$dddd$") exhibited spike rate adaptation in response to repeated transitions. We therefore included in our analysis only the response to the first repeat (e.g., "d") in such sequences. This excluded 2/40 syllables from our analysis, yielding the 38 syllables that were the focus of our main analysis.

Modulation of responses to individual syllables by probability in pseudorandom stimuli. In our analysis of how responses to individual syllables were modulated by preceding sequences, each of the 38 unique syllables in the pseudorandom stimuli contributes one set of data points. First, at each of the 18 recording sites, the response to each s_0 when preceded by a particular sequence [$\bar{R}(s_0|S^L)$] was extracted for each length as described above. Because results across different sites within a bird were significantly correlated (average between site correlation coefficient, $r = 0.42$; Sutter and Margoliash, 1994), these responses were then averaged across sites within a bird. We note that, given the strong correlation between multiunit sites within a bird, the averaging of responses across multiunit sites is the statistically conservative approach because it effectively reduces the sample size used for statistical testing. All results hold whether this averaging is performed or not. When we report the slopes of response modulation by probability, $\Delta R/\Delta P$, we use the slopes from linear regressions (β_1; see below, Univariate regression analysis of the un-normalized responses [$\bar{R}(s_0|S^L)$] at each site versus the (un-normalized) convergence or divergence probabilities.

To combine data across syllables and sites, the extracted responses [$\bar{R}(s_0|S^L)$] were first normalized across sequences of a given length for each site so that the maximum value was 1 and then averaged across sites, yielding $\bar{R}_N(s_0|S^L)$. This quantifies the average amount by which the response to s_0 was modulated by S^L relative to all other sequences of length L. We used these normalized response modulations to summarize the

effect of convergence and divergence probability on auditory responses across syllables and sequence lengths. Because both C and D will vary across the different syllables and will generally change systematically with increasing sequence length, to combine data from different syllables and lengths, C and D were normalized to the maximum for each syllable at a given sequence length. Responses to sequences with identical probability were averaged together.

Calculation of integration functions. To characterize how responses to individual syllables were modulated by the length and probability of preceding sequences, we binned convergence probability into three increasing, nonoverlapping categories [low: $C = 0$; medium: $0 < C < \max(C)$; high: $\max(C)$]. For selection of the high-probability sequences, for each syllable, the most probable sequence of 5 preceding syllables ($S^5 s_0$) was found. Sequences of length $L = [1:4]$ were chosen as the appropriate length substring of $S^5 s_0$ and sequences of length $L = [6:9]$ were selected as the most probable sequence of the corresponding length that contained the previous length sequence. For example, if the target syllable was "g" and the most probable length $L = 5$ preceding sequence was $S^5 = $ "bcdef," the length $L = 4$ sequence would be $S^4 = $ "cdef," and the length $L = 6$ sequence would be $S^6 = $ "abcdef," provided that "a" was the most likely syllable to precede "bcdef." Therefore, for the most probable sequences, sequences of length L contained the length L-1 sequence. Therefore, a change in the response to the last syllable when increasing the number of preceding syllables results strictly from the inclusion of the additional syllable. Medium-probability sequences were the naturally produced sequences that were not included in the high probability set. Low-probability sequences were composed of increasingly long sequences of syllables containing no natural transitions. Due to the stochastic nature of the stimuli, there were occasionally cases in which appropriate sequences for a particular length and probability category for a given syllable were not present. Therefore, to be included in these analyses, we required that for a given probability category, each syllable have an appropriate sequence for at least 7 lengths.

For calculation of average integration functions, at each site we first calculated $\bar{R}(s_0|S^L)$ as above (see above, Characterization of neural responses to individual syllables). $\bar{R}(s_0|S^L)$ was then normalized by the maximum response across string lengths and probability category. These normalized integration functions were then averaged across sites. To average data from different syllables and compare responses across probability categories, we subtracted the average response to s_0 independent of preceding sequence identity (i.e., the response at $L = 0$) independently for each probability category, where the average is taken over syllables. This ensures that the average response modulation starts at zero and allows negative values. We then divisively normalized so that the average maximum value across probability categories and lengths was 100. This yields $\bar{R}_N(s_0|S^L, C, L)$, which quantifies the average amount by which the response to s_0 was modulated by the length and convergence probability of the preceding sequence. Therefore, the average integration functions (Fig. 6D) describe how, on average, the magnitude of the neural response to a syllable was modulated by the length and probability of the preceding sequence, expressed as a percentage of the average maximum difference from the response at $L = 0$.

Statistics. Data are reported as mean ± SE unless otherwise indicated. Data in Figure 2B and Figure 3C were analyzed using the Wilcoxon's signed-rank test. Data in Figure 4F were analyzed using the one-sample, two-way Student's t test against the null hypothesis of 0 mean. Data in Figure 5B were analyzed using Wilcoxon's signed-rank test. The first sequence length to reach asymptote for high probability integration functions (Fig. 6D) was determined using a serial Z test between the data at each sequence length and the estimated asymptotic value and SD of the estimate. Results were considered significant if the probability of type I errors was $< \alpha = 0.05$. Bonferroni corrections were used to adjust α levels when multiple comparisons were performed.

Throughout the results, we averaged neural data across sites within a bird to avoid pseudoreplication due to strong correlations between sites (mean between site correlation coefficient, $r = 0.45$).

Univariate regression analysis. Linear regressions of response to syllable s_0 as a function of convergence probability were performed with the model:

$$\bar{R}(s_0|S^L, C(S^L s_0)) = \beta_0 + \beta_1 C(S^L s_0) \qquad (7)$$

Analogous models were used for regressions of responses as a function of divergence probability, as well as normalized responses versus normalized probabilities (see above, Modulation of responses to individual syllables by probability in pseudorandom stimuli).

Regressions of response modulation as a function of sequence length for the probability categories low, medium, and high (see above, Calculation of integration function) were performed with the linear model:

$$\bar{R}_N(s_0|S^L, L, C) = \beta_0 + \beta_1 L, \ C \ \epsilon \ [Low, Medium, High] \qquad (8)$$

and the sigmoidal model:

$$\bar{R}_N(s_0|S^L, L, C) = \frac{\beta_0}{1 + e^{-\beta_1(\beta_2 - L)}}, \ C \ \epsilon \ [Low, Medium, High] \qquad (9)$$

The sigmoidal model was fit by an iterative minimization of the RMS error using the *nlinfit* function in MATLAB.

We used the Akaike's Information Criterion (AIC) to determine the goodness of fit for linear versus sigmoid models of the high and medium probability integration functions as follows:

$$AIC = 2p + 2 \log(\sigma_r^2) \qquad (10)$$

where p is the number of parameters in a given model (linear: $p = 2$, sigmoidal: $p = 3$) and σ_r^2 is the estimated residual variance. By this metric, the high-probability sequences were more parsimoniously fit with a sigmoid than a line, whereas the converse was true for medium probability sequences.

General linear model analysis. Syllable transitions in Bengalese finch song can be non-Markovian such that probabilities for some transitions depend not only on the current syllable, but also on the preceding syllable transitions (data not shown, see also Nakamura and Okanoya, 2004; Jin and Kozhevnikov, 2011; Katahira et al., 2011; Warren et al., 2012). Because of this complexity, we analyzed in two complementary ways the statistics of syllable sequencing and how they relate to auditory responses of HVC neurons. In the first approach (as described above), which contributes to the primary analysis presented here, we measured the relative probabilities with which entire sequences of syllables occurred. This approach captures non-Markovian contributions to sequence statistics, but does not assess directly how responses depend on the probabilities of transitions at specific locations within the sequences. In the second approach, which contributes to our analysis using general linear models (GLMs), we measured the relative probabilities with which pairwise transitions between syllables occurred. This approach assumes that transitions are Markovian and thus only approximates the actual statistics of sequencing, but it enables an analysis that directly assesses contributions of transitions at specific locations within sequences.

One consequence of the non-Markovian nature of Bengalese finch song is that as the length of the preceding sequence increases, the divergence and convergence probabilities between two particular syllables will change. Increasing sequence length causes divergence probabilities to approach 1, reflecting the intuition that, as the length of the conditioning sequence increases, the probability of the upcoming syllable becomes more certain; conversely, convergence probability shrinks to a finite number, reflecting the intuition that, given a single syllable, the probability of having transitioned from a particular sequence becomes lower as the length of the sequence increases. Therefore, the distributions of C and D will be different for long sequences, reflecting the different kinds of information contained in these characterizations. This raises the possibility that differences observed using the entire conditioning sequence may be the result of numerical differences in the underlying distributions at large sequence lengths. We therefore used a GLM as a second approach to
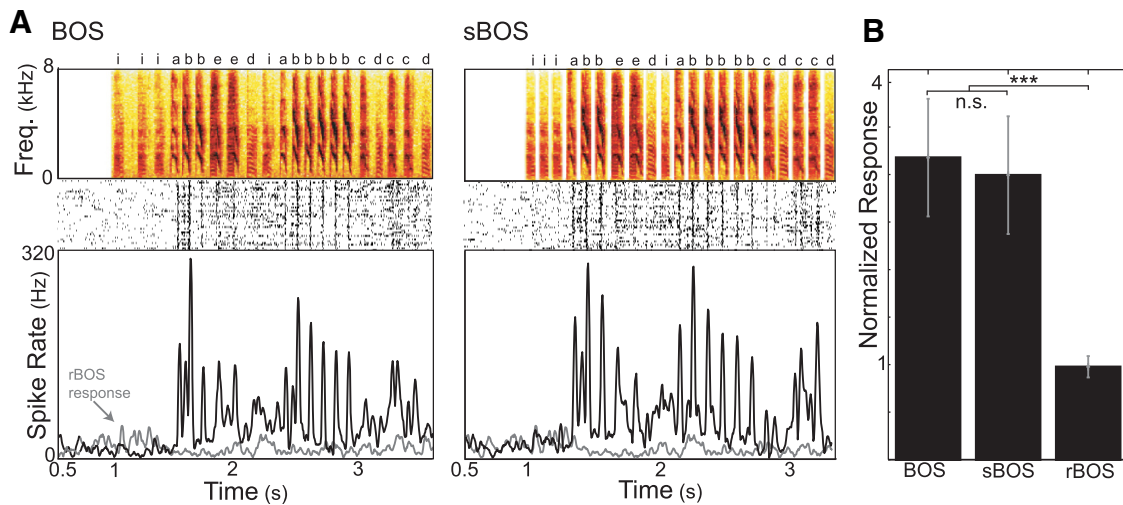
Figure 2. Synthetic version of BOS, which isolates sequence variability, is an effective stimulus. **A**, Example spectrograms of BOS and sBOS. Below each spectrogram is the spike raster and associated instantaneous spike rate from one HVC site recorded over 50 trials; black lines indicate responses to BOS or sBOS, gray lines indicate responses to a control stimulus, rBOS. **B**, Average auditory responses (divisively normalized to baseline firing rate, mean ± SE) to BOS, sBOS, and rBOS. (nonsignificant: $p > 0.5$, Wilcoxon signed-rank test; ***$p \leq 10^{-10}$, with Bonferroni correction; data from 47 sites in $n = 7$ birds).

investigate how the statistics of syllable sequencing relate to the auditory responses of HVC neurons.

The GLM used to investigate how responses to syllables were modulated by different probabilistic characterizations of the preceding sequence (Fig. 5B) was as follows:

$$R(s_0) = \alpha + \sum_{i=0}^{-(L-1)} \beta_i F(s_i, s_{i-1}), \; L \in [1 \quad 9] \quad (11)$$

where $R(s_0)$ is the response to the target syllable and $F(s_i, s_{i-1})$ denotes the various naturally occurring local probabilities of s_i and s_{i-1}. Here, $F(s_i, s_{i-1})$ was either: convergence probability $[P(s_{i-1}|s_i)]$, divergence probability $[P(s_i|s_{i-1})]$, the marginal probability of individual syllables $[P(s_{i-1})]$, or the logarithm of convergence probability $[\log_{10}P(s_{i-1}|s_i)]$, For example, if we are examining the response to $s_0 = "d"$ as a function of the convergence probability of the length 3 sequence ($S^3 = "abc"$) we have:

$R(d) = \alpha + \beta_0 P(c|d) + \beta_{-1} P(b|c) + \beta_{-2} P(a|b)$

To maintain a paired comparison of the data going into the different probability functions F, we restricted our analysis to naturally occurring sequences: $P(s_{i-1}|s_i) > 0$ or, equivalently, $P(s_i|s_{i-1}) > 0$. We note that this imposition slightly alters the interpretation of the marginal probability of the preceding syllable $P(s_{i-1})$ because it is now restricted to $P(s_{i-1}) \mid P(s_{i-1}|s_i) > 0$, $i \in [0 \; L]$. We believe that this is a more appropriate quantity to examine because HVC auditory responses are greater for natural versus unnatural sequences (Margoliash and Fortune, 1992; Lewicki and Arthur, 1996). The model was fit simultaneously for all syllables using all natural sequences of a particular length preceding a particular target syllable s_0 found in our dataset.

The GLM used to investigate how responses to syllables were modulated by the convergence probability at a particular transition in the preceding sequence in Figure 7A was as follows:

$$R(s_0) = \alpha + \sum_{i=0}^{-8} \beta_i P(s_{i-1}|s_i), \quad (12)$$

where $R(s_0)$ is the response to a particular instance of s_0 in the stimulus (averaged across sites after normalization across lengths) and the subscript denotes location in the sequence, with 0 being the target syllable and indices increasing with location in preceding sequence. The model was fit for each syllable individually using all sequences of length 9 preceding the target syllable found in the pseudorandom dataset (i.e., no averaging across multiple instances of the same sequence). The weight assigned to each transition location (β_i) quantifies the relative influence of the convergence probability at that location in the sequence on the auditory response to the terminal syllable [$R(s_0)$].

Analysis of effects of marginal probabilities on auditory responses. We used partial correlation analysis to examine the degree to which single-syllable and joint-sequence probability were correlated with the auditory responses across different syllables. Partial correlation analysis quantifies the degree of association between two variables (here, measured neural responses and probability) while controlling the effect of a set of other variables (here, the acoustic structure of the syllables). We quantified the acoustic structure of the syllables using the same 16-dimensional feature vector used for selection of representative syllables. Therefore, we can think of a 17-dimensional vector describing the acoustic structure of the syllables (16-dimensional) and the probability of the syllable occurring (1-dimensional). For each of these features, we decorrelated each one of the features with every other feature and then decorrelated the auditory responses with all of the acoustic features. Correlating the decorrelated residuals between acoustic features and probability with the decorrelated residuals between acoustic features and auditory responses results in the partial correlation coefficient between probability and auditory responses. In the current context, the partial correlation coefficient quantifies the unique contribution of the single-syllable marginal and joint-sequence probabilities to auditory responses. The results of this analysis examined the slopes and combined regressions for single-syllable marginal and joint-sequence probability. Across analysis metrics (distribution of slopes and combined regression analysis), we detected no significant modulation of auditory responses due uniquely to either of these probabilities (slopes are not different from 0, $p > 0.4$, t test and combined regressions are not significant).

Results

Synthetic stimuli that isolate sequence variability are effective at driving HVC auditory responses

To isolate the influence of sequence variability on HVC responses, we created synthetic stimuli (sBOS) in which every syllable in the bird's repertoire was replaced by a prototypical syllable of the same type and each intersyllable gap was replaced by the median of all gaps (see Materials and Methods). HVC neurons responded equally well to sBOS stimuli and the natural BOS stimuli from which they were derived. Figure 2A illustrates an example of responses to these two stimuli at one recording site in HVC. As previously reported, neurons in Bengalese finch HVC responded vigorously to BOS and responses were locked to syllables (Nakamura and Okanoya, 2004; Sakata and Brainard, 2008). In this case, BOS and sBOS evoked similar auditory responses that were greater than responses to a temporally reversed version of BOS (rBOS). Similarly, across 47 multiunit sites in 7 birds,

responses to BOS and sBOS were significantly greater than responses to rBOS (Fig. 2B; $p < 10^{-10}$, Wilcoxon signed-rank test, $n = 7$ birds, with Bonferroni correction, m = 3 comparisons) and were not significantly different from each other ($p > 0.5$). These data demonstrate that synthetic versions of BOS elicit auditory responses comparable to natural versions of BOS. Therefore, we can construct effective stimuli that allow us to manipulate sequences independent of other song parameters.

Responses to individual syllables depend on the convergence probability of preceding sequences

We next systematically examined how responses to individual "target" syllables from a bird's song varied as a function of the sequences that preceded them. To do so, we created stimuli consisting of 10,000 pseudorandomly ordered syllables constructed from the same syllables and median intervals used in the sBOS stimuli (Fig. 3A, see Materials and Methods). We found that auditory responses to the last syllable following the same long sequences (of length 7) in both sBOS and pseudorandom stimuli (Fig. 3B,C) were highly correlated ($r = 0.93$, $p \leq 10^{-8}$, $n = 11$ sites) and not significantly different from each other ($p > 0.5$, Wilcoxon signed-rank test, $n = 11$ sites). Therefore, these pseudorandom stimuli effectively drive HVC auditory responses and enable us to systematically probe how the responses to individual target syllables vary as a function of both the length and probability of the sequences that precede them.

We used these pseudorandom stimuli to test the hypothesis that HVC auditory responses to a target syllable are modulated by the convergence probability of preceding sequences. For each target syllable and preceding sequence length (for lengths $L = 1–9$), we identified each unique sequence within the pseudorandom stimuli that terminated in the target syllable. We averaged the responses to each unique sequence and calculated the convergence probability (from the corpus of naturally produced songs) for the entire preceding sequence of length L (S^L) given the terminal target syllable s_0 [$C(S^L s_0) = P(S^L|s_0)$; e.g., "kabc," $L = 3$, $S^L = $ "kab," $s_0 = $ "c," $C = P(kab|c) = 0.63$]. We then linearly regressed the responses to a target syllable against the convergence probability of the preceding sequence to determine whether the responses to syllables varied systematically with convergence probability. Figure 4A–C illustrates an example of how the response to the target syllable "c" (shaded region) depended on the convergence probability of preceding sequences of length 3. Within the bird's repertoire of songs, the sequence "abk" (left) was never produced before "c," whereas sequences "fjb," "cab," and "kab" were produced 2%, 25%, and 63% of the time given terminal syllable "c." In this case, responses to the target syllable monotonically increased with increasing convergence probability (Fig. 4B,C; slope = 196 Hz/unit change in C). Figure 4D, E illustrates a second example from a different bird of how responses to a target syllable varied as a function of the convergence probability of preceding sequences (in this case of length 4). As in the previous example, responses to the target syllable ("b") monotonically increased with increasing convergence probability (slope = 289 Hz/unit change in C).

Across all experiments with pseudorandom stimuli, we consistently observed an increase in responses to target syllables as a function of convergence probability. Our dataset contained a total of 945 cases in which we could examine this relationship (corresponding to 38 unique target syllables, each preceded by multiple sequences of a given length from 1–9 recorded across 18 multiunit sites from 6 birds). For 809/945 cases (86%) best-fit regressions had positive slopes (as in Fig. 4B,E), indi-
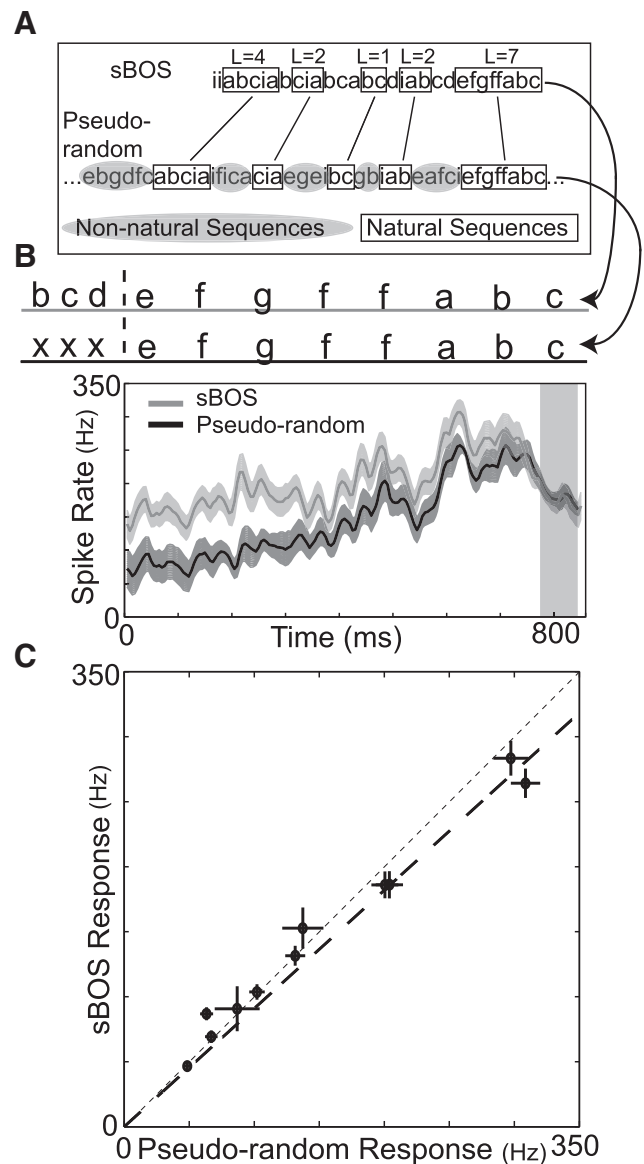
Figure 3. Pseudorandom sequences are effective auditory stimuli. *A*, Illustrative pseudorandom stimulus sequence. Pseudorandom stimuli were designed to contain naturally occurring sequences (examples in black boxes) of various lengths randomly interleaved with non-naturally occurring sequences (examples in gray ovals). Within the pseudorandom stimuli, we define a natural sequence of length L (e.g., $S^L = $ "efgffab," L = 7) preceding a target syllable (e.g., $s_0 = $ "c") as a string of L syllables sequenced as in BOS, preceded by a syllable that does not naturally occur before the first syllable [i.e., $P(x|e) = 0$]. A sample BOS segment is shown above and lines connect naturally occurring sequences that were present in both sBOS and pseudorandom stimuli. *B*, Example of auditory responses at one site to the same naturally occurring sequence ("efgffabc") that occurred in both sBOS (gray) and pseudorandom (black) stimuli (mean ± SE). In this case, the response to the last syllable (shaded region) was the same regardless of which stimulus it occurred in. *C*, Average magnitude of auditory responses to a syllable when preceded by the same sequence of 7 syllables in sBOS and pseudorandom stimuli. Each datum corresponds to one recording site and shows the mean ± SE of the responses to all syllables. Thin gray dashed line is unity. Thick black line is from regression. Responses were not different between stimuli ($p > 0.5$, Wilcoxon signed-rank test) and exhibited a strong linear relationship ($r = 0.93$, $p \leq 10^{-8}$).

cating increasing responses as a function of convergence probability. The distribution of slopes (Fig. 4F) was significantly >0 (diamond, mean slope = 90 Hz/unit change in C, $p < 10^{-10}$, one-sample, two-tailed t test, $n = 945$). Similar results were obtained for a restricted examination (for length 3 and 4)
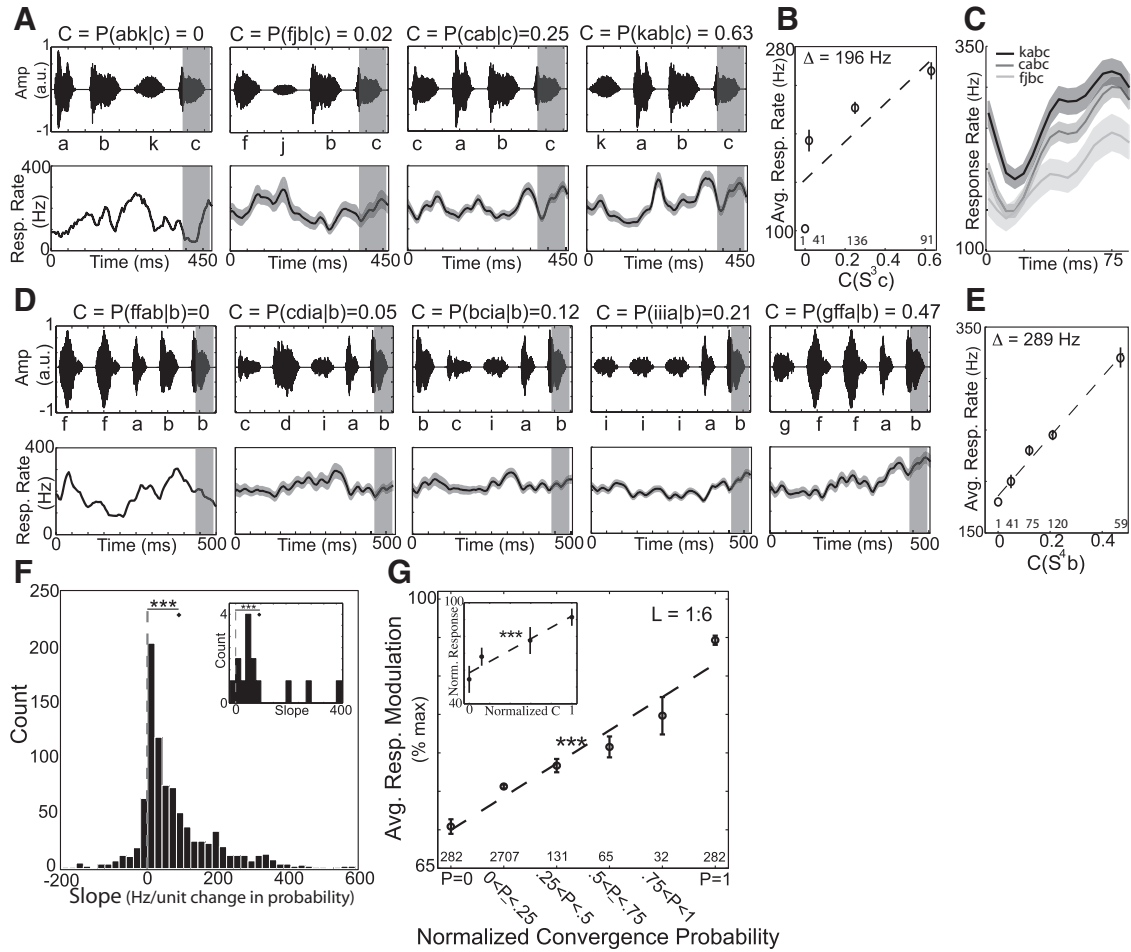
Figure 4. Convergence probability of preceding sequences modulates auditory responses to single syllables in a positive, linear manner, **A**, Average time varying responses (mean ± SE) and accompanying stimuli for length 3 sequences with varying convergence probability, followed by the target syllable "c." **B**, Average responses to the target syllable "c" (mean ± SE for shaded region in **A**) when preceded by the sequences in **A**. **C**, Response waveform for syllable "c" when preceded by the 3 naturally occurring sequences shown in **A**. **D**, **E**, Second example illustrating increasing responses to target syllable "b" as a function of convergence probability for preceding length 4 sequences. **D**, Average time varying responses (mean ± SE) and accompanying stimuli for length 4 sequences with varying convergence probability, followed by the target syllable "b." **E**, Average responses to the target syllable "b" (mean ± SE for shaded region in **D**) when preceded by the sequences in **D**. **F**, Distribution of slopes for regression of responses to target syllables against convergence probability of preceding sequences for entire dataset (***$p < 10^{-10}$, one sample t test, $n = 945$ cases). Black diamond: mean slope (90 Hz/unit change in C). Inset shows histogram of slope values for experiments on single units (***$p < 10^{-3}$, one-sample t test). Black diamond: mean = 96 Hz/unit change in probability. **G**, Response modulation as a function of convergence probability for all syllables and sequences of lengths 1–6 (mean ± SE). The slope of the regression line (dashed line) was significantly different from 0 (***$p < 10^{-3}$). Inset shows results from single units, mean ± SE, and best-fit line (***$p < 10^{-4}$). Number of unique instances of the sequences in **B**, **E**, and **G** are indicated at bottom of plots.

of 9 single-unit recordings from putative HVC inhibitory interneurons (inset, slope > 0 in 13/14 cases, average slope = 96 Hz/unit change in C, $p < 10^{-3}$, one-sample, two-tailed t test, $n = 14$ syllables). Therefore, auditory responses of HVC neurons are positively modulated by the convergence probability of the preceding sequence.

Previous recordings from HVC have shown that responses are much larger to natural than to unnatural sequence variants and that some neurons exhibit a categorical response to particular syllables (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995; Lewicki and Arthur, 1996; Prather et al., 2009; Fujimoto et al., 2011). Therefore, one might expect categorical responses to natural versus unnatural sequences with no modulation due to probability beyond that. To investigate whether the positive modulation of auditory responses is the result of graded or categorical encoding, we combined data from different recording sites, syllables, and sequence lengths (from 1–6) by normalizing both mean responses and convergence probability for a given target syllable to their maximum values for a given length (see

Materials and Methods). We grouped the normalized probabilities for each syllable and sequence length into six monotonically increasing bins and averaged normalized responses in each bin (Fig. 4G).

We found that, on average, convergence probability explained 25% of the variability in responses to individual syllables preceded by sequences of length 1–6 ($R^2 = 0.252 \pm 0.090$, mean ± SD) and responses to syllables increased as a graded, monotonic function of the convergence probability of the preceding sequences (Fig. 4G). The best-fit line to the means (black dashed line) had a slope of 22% (95% confidence interval, 14%–29%), indicating that, on average, there was a +22% modulation of auditory responses by a unit change in convergence probability. Similar results were obtained from the single-unit recordings (Fig. 4G, inset), where linear regression indicated that convergence probability explained 28% of the variation in auditory responses and modulated the magnitude of the response by 33% ($R^2 = 0.28$, slope = 0.33, $p < 10^{-4}$). These results demonstrate that auditory responses to a given syllable are positively modu-

lated in a graded fashion by the convergence probability of the preceding sequence.

Convergence probability explains more of the sequence-induced variation in auditory responses than other probabilistic characterizations

We found that convergence probability could account for more of the sequence-induced variation in auditory responses than a variety of alternative probabilistic characterizations. In particular, divergence probability has been the primary focus of prior experimental and theoretical studies of sequencing in songbirds (Okanoya and Yamaguchi, 1997; Woolley and Rubel, 1997; Nakamura and Okanoya, 2004; Sakata and Brainard, 2006; Nishikawa et al., 2008; Sakata and Brainard, 2008; Jin, 2009; Fiete et al., 2010; Katahira et al., 2011; Warren et al., 2012). Therefore, we first examined the degree to which auditory responses were modulated by convergence probability compared with divergence probability. We found that auditory responses were only weakly modulated by divergence probability. The best linear fit of auditory responses as a function of divergence probability indicates that responses are modulated by only 8% over a unity change in divergence probability (Fig. 5A, gray dashed line; data for convergence probability are replotted in black for comparison). This is significantly less than the response modulation due to convergence probability (22%, $p \leq 0.05$), and was not significantly different from 0 (95% confidence interval of slope for divergence probability: 0%–16%). Moreover, divergence probability explained less of the variation in responses than did convergence probability (R^2 for divergence probability $= 0.152 \pm 0.049$, mean $\pm$ SD vs R^2 for convergence probability $= 0.252 \pm 0.090$, mean $\pm$ SD). Furthermore, a serial regression analysis revealed that divergence probability accounted for only 5% of the residual variability that could not be accounted for by convergence probability. Therefore, convergence probability provided a better account of sequence-induced modulations of auditory responses than did divergence probability.

We additionally used a general linear model (GLM; see Materials and Methods) to compare how well convergence probability, divergence probability, and several other probabilistic characterizations of sequence statistics could explain responses to target syllables. The GLM provides a single framework that can be used to compare the relative explanatory value of distinct probabilistic characterizations. This analysis quantifies how well the responses to a target syllable could be explained by a weighted sum of the transition probabilities between each pair of syllables within the sequences preceding the target syllable (see Materials and Methods). We quantified how well the model explained response variations (for sequences of length $L = 1–9$) using the correlation coefficient (R value). Figure 5B displays the distribution of R values for models using convergence probabilities (C) and divergence probabilities (D) to characterize the pairwise transitions between syllables. Consistent with our previous analyses, the correlation coefficients for the GLM were significantly greater when convergence probability rather than di-
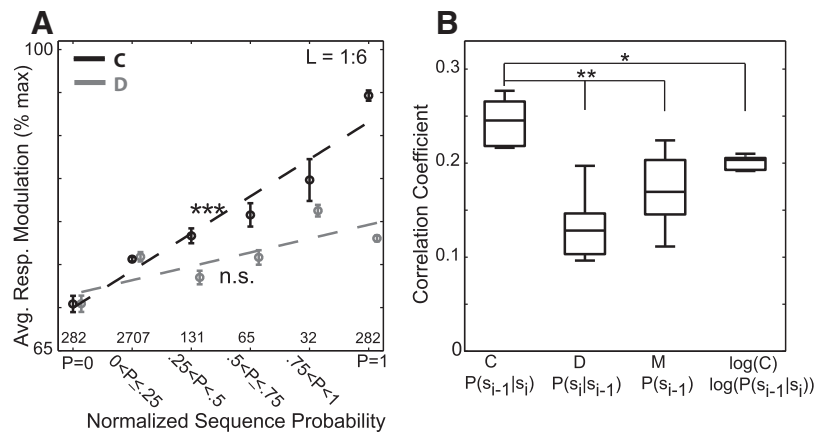

Figure 5. Convergence probability explains more variation in auditory responses than other probabilistic characterizations of sequencing. **A**, Response modulation as a function of convergence probability (black) and divergence probability (gray) for all syllables and sequences of lengths 1–6, as well as the best-fit lines from linear regression (***$p < 10^{-3}$; nonsignificant: $p > 0.05$; H_0: slope $= 0$). **B**, Correlation coefficients resulting from fitting a GLM to auditory responses using four different probabilistic characterizations of sequence statistics as regressors. (*$p < 0.05$, **$p < 0.01$, Wilcoxon signed-rank test, with Bonferroni correction, $n = 9$ for all distributions). Boxes indicate median, 25th, and 75th percentiles, whiskers extend to ± 2.7 SD, ~95% confidence intervals.

vergence probability was used to describe the probabilities of transitions ($p < 0.01$, Wilcoxon signed-rank test, $n = 9$ lengths).

We also used GLMs to test how well two other statistical characterizations of sequencing could account for variation in HVC auditory responses. First, we tested how well the log of the convergence probability between pairwise transitions could explain variation in responses. This value [$\log_{10}(C)$] was tested based on previous findings from primates suggesting that the nervous system may preferentially represent the logarithm of the probability of sequential sensory events that lead to reward (Yang and Shadlen, 2007; Pouget et al., 2013). Second, we tested how well the probability with which individual syllables in the preceding sequence were produced could account for variation in responses to the target syllables. This value [M] was tested because both theory and experiments in other systems indicate that the overall frequency of occurrence of a stimulus can shape the strength of responses to that stimulus (Barlow, 1961; Machens et al., 2005). Again, each of these probabilities provided significantly less explanatory power than did convergence probability (Fig. 5B, *$p < 0.05$, **$p < 0.01$, Wilcoxon signed-rank test, Bonferroni corrected for $m = 3$ comparisons).

Finally, we investigated whether responses across syllables varied as a function of how frequently the syllable occurred in the bird's repertoire and how frequently the syllable and preceding sequence of length L occurred jointly. These quantities were chosen because they provide additional tests of the possibility that the frequency with which a stimulus is experienced will influence the strength of the response to that stimulus (Barlow, 1961; Simoncelli and Olshausen, 2001; Machens et al., 2005). They also are of potential significance because they are related to the divergence probability and convergence probability by Bayes theorem (see Materials and Methods). We found that neither the marginal probability of target syllables nor the joint probability of target syllables and preceding sequences accounted for significant amounts of response variation across syllables (see Materials and Methods, Analysis of effects of marginal probabilities on auditory responses).
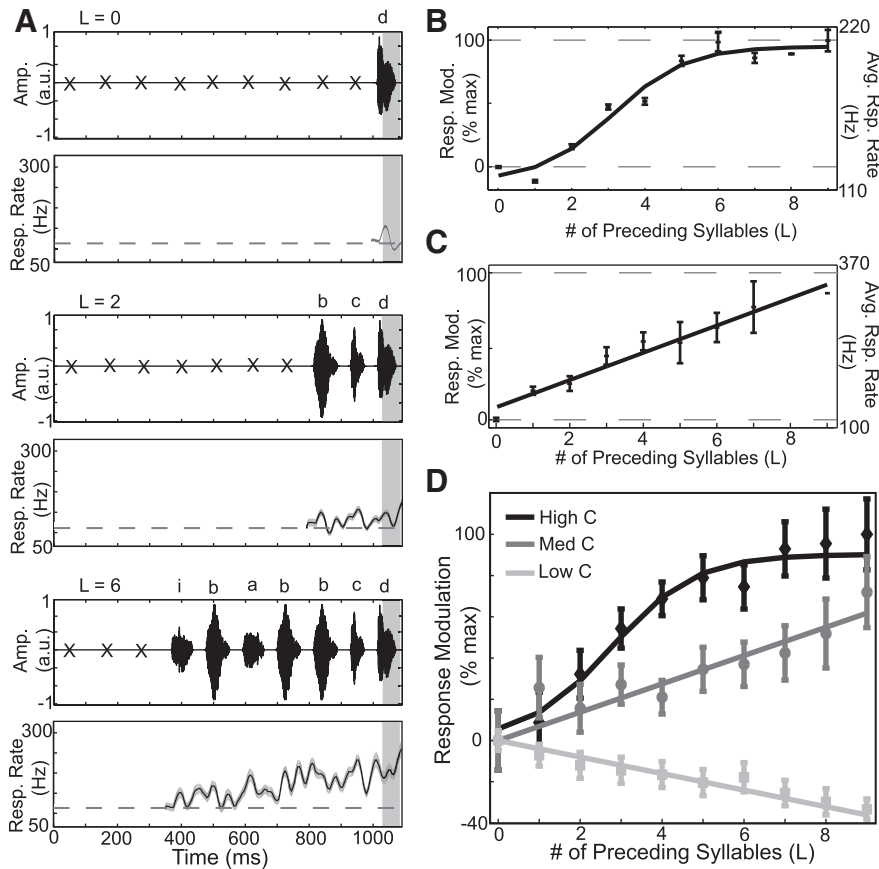
Figure 6. Integration of auditory responses extends over many syllables and is dependent on convergence probability. *A*, *B*, Example of response integration for high-probability sequences. *A*, Series of acoustic stimuli (*x*: non-naturally sequenced syllables, conditioned on subsequent sequence) and resulting average response (mean ± SE) for sequences of length $L = 0$ ("*d*"), $L = 2$ ("*bcd*"), and $L = 6$ ("*ibabbcd*") ending in the same syllable ($s_0 = $ "*d*"). Dashed black line in response plots is the mean activity during the entire pseudorandom stimulus. *B*, Average (± SE) response modulation (left ordinate) and response rate (right ordinate) for "*d*" and best-fitting sigmoid function. *C*, Example of response integration for medium-probability sequences. Average (± SE) response modulation (left ordinate) and response rate (right ordinate) for the last syllable and best linear fit. *D*, Average response modulation for sequences of high probability (black diamonds with sigmoidal fit), medium probability (dark gray circles with linear fit), and low probability (light gray squares with linear fit). Data are presented as means ± SE.

Temporal integration for auditory responses extends over many syllables and depends on convergence probability

Our finding that responses to a given syllable depend on the convergence probability of preceding sequences implies that there is some form of temporal integration for HVC auditory responses. Therefore, we next used responses to sequences extracted from pseudorandom stimuli to measure systematically the temporal extent of integration and how integration depends on convergence probability. We measured responses to the 38 unique syllables in our dataset as a function of the length and probability of preceding sequences. To examine how integration depends on convergence probability, we separately measured integration over sequences with low, medium, and high convergence probabilities [low: $C = 0$; medium: $0 < C < $ max (C); high: $C = $ max (C)].

Figure 6, *A* and *B*, shows an example of temporal integration over sequences with high convergence probabilities. Here, we measured how responses to the target syllable "*d*" ($s_0 = $ "*d*") varied as a function of the length of the most common preceding sequences. For length 0, we identified in the pseudorandom stimuli each occurrence of syllable "*d*" when it was preceded by a non-naturally occur-

ring syllable transition ($L = 0$, top). Corresponding neural responses were extracted and averaged ($L = 0$, bottom). For length 2, the most common sequence preceding "*d*" was "*bc*." Therefore, we identified in the pseudorandom stimuli each occurrence of the sequence "*xbcd*" (where "*xb*" is an explicitly unnatural transition, $L = 2$, top) and extracted and averaged the corresponding neural activity ($L = 2$, bottom). Similarly, for length 6, the most common sequence preceding "*d*" was "*ibabbc*" and we therefore identified each occurrence of the sequence "*xibabbcd*" ($L = 6$, top) and extracted and averaged corresponding neural activity ($L = 6$, bottom). These examples illustrate that as more naturally occurring syllable transitions preceded the target syllable "*d*," the elicited response (shaded area) progressively increased. The sequence integration function in Figure 6B summarizes how the response to syllable "*d*" was modulated as the length of the most common preceding sequence was increased from 0 to 9. In this example, integration extended over ~6 syllables (i.e., responses reached asymptote after ~6 syllables).

Across our dataset, for high-probability sequences, integration extended over ~7 syllables. The average integration function (see Materials and Methods) for high-probability sequences is plotted in Figure 6D (black diamonds, mean ± SE; solid line, sigmoid fit, $R^2 = 0.22$, $p \leq 0.01$, $n = 31$ syllables, for which sequence lengths were well sampled; see Materials and Methods). Response modulation was significantly different from the fit asymptote for length 6, but not for length 7 (serial Z test of response modulation against estimated asymptotic values from fit sigmoid, $p < 0.05$, see Materials and Methods), indicating that integration extended over ~7 syllables. Each syllable plus adjacent gap in our Bengalese finch songs is ~100 ms in duration. Therefore, for sequences with high convergence probability, auditory responses integrated over ~700 ms.

For medium-probability sequences, the pattern of temporal integration was strikingly different from that for high-probability sequences. An example integration curve for a medium-probability sequence is displayed in Figure 6C; here, responses increased in a linear manner and did not asymptote. On average, medium probability sequences gave rise to a similar pattern of response integration (Fig. 6D, dark gray circles: solid line, linear fit, $R^2 = 0.05$, $p \leq 0.05$, $n = 34$ syllables). As with high-probability sequences, increasing the length of medium-probability sequences also caused increasing response modulation; however, the rate at which responses increased with increasing sequence length (the slope or "gain" of integration) was lower than for high-probability sequences. Furthermore, response modulation continued to increase over the entire range of sequence lengths included in our stimuli (up to length 9) without reaching apparent asymptote. These data indicate that for naturally occurring sequences with medium convergence probability, the gain of the integration process (i.e., slope of integration function) is less

than that for high convergence probability sequences. Moreover, the temporal extent of integration for medium probability sequences (≥ 9 syllables, >900 ms) was greater than that observed for high-probability sequences (7 syllables, ~ 700 ms).

In contrast to the positive integration observed with natural sequences (i.e., $C > 0$), for sequences with low convergence probability (i.e., no naturally occurring transitions, $C = 0$), the integration process consisted of a low gain suppression of auditory responses (Fig. 6D, light gray squares: solid line, linear fit, $R^2 = 0.36$, $p < 0.01$, $n = 37$ syllables). This indicates that as a target syllable is preceded by progressively longer sequences of non-naturally occurring transitions, responses to the target syllable decline from their "baseline" values at length 0. Again, for the low-probability sequences, the extent of temporal integration was also >9 syllables (lack of asymptote for decrease in response modulation). These results demonstrate that HVC auditory responses integrate over long temporal extents and that the sign, gain, and temporal extent of integration depend on convergence probability.

The preceding analyses examine how the responses to target syllables are modulated as a function of the length and probability of preceding sequences. This approach treats the sequences that precede target syllables as single "chunks" of song. However, these analyses do not test directly how the convergence probability of transitions at varying locations within these sequences influence responses to the target syllable. To address this issue, we used a GLM to measure how pairwise convergence probability at different transition locations in the preceding sequence modulated auditory responses to the last syllable in the sequence (see Materials and Methods). The weights assigned by the model (β_i) for various transition locations (i) describe the effect of the probability at that transition location on the response.

In Figure 7A, we plot the linear weights, β_i (mean $\pm$ SE, $n = 38$ syllables), for convergence probability as a function of the transition location (i). We found that the mean weights were all positive, indicating that increasing the convergence probability at all transition locations resulted in increases in responses to the last syllable in the sequence. Furthermore, the average weights derived from convergence probability exhibited a fairly systematic decrease as a function of increasing transition location preceding the target syllable. However, it is noteworthy that the assigned distributions of weights are significantly >0 ($p < 10^{-2}$, t test, $n = 38$, Bonferroni corrected for m $= 9$ comparisons) until the ninth preceding transition, indicating that the pairwise convergence probabilities at distant transitions affect auditory responses to individual syllables. Therefore, we found that more distant transitions generally affected auditory responses to a lesser degree, with the influence of probability effectively disappearing by the ninth preceding transition.

These results from pseudorandom stimuli predict that HVC auditory responses to BOS stimuli should increase for several hundred milliseconds after song onset. In Figure 7B we present the average (black, mean $\pm$ SE, $n = 47$ sites) baseline normalized auditory response evoked by the first 1.5 s of sBOS stimuli (same normalization as in Fig. 2B). The dark gray line corresponds to the best fitting sigmoid function, with the estimated asymptotic
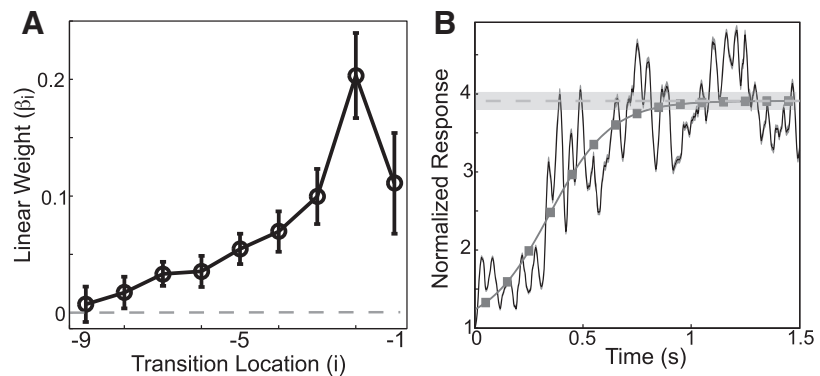
Figure 7. Evidence for integration from GLM analysis and sBOS responses. **A**, Auditory responses depend on convergence probability at distant transition locations. Linear weights (β_i) describing modulations of auditory responses due to pairwise convergence probability as a function of the location of the transition (i). Data are presented as means $\pm$ SE, $n = 38$ for all. **B**, Long-time dependent integration in sBOS responses. The average baseline normalized auditory response evoked by the first 1.5 s of sBOS stimuli (black, mean $\pm$ SE, $n = 47$). The dark gray line corresponds to the best-fitting sigmoid function, with the estimated asymptotic value demarcated in light gray ($\pm$ 1 SD of estimated value). Dark gray boxes demarcate the average approximate location of syllable centers.

value demarcated in light gray ($\pm$ 1 SD of estimated value). As predicted from the auditory responses to a given syllable presented in the pseudorandom stimuli, on average, auditory responses to sBOS stimuli increased for several hundred milliseconds before reaching an asymptote after ~ 1 s (~ 10 syllables).

Discussion

Encoding of convergence probability

Previous studies using zebra finches have shown that auditory responses in HVC are sensitive to sequence (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995; Lewicki and Arthur, 1996). However, zebra finches exhibit little sequence variation, so in these studies, except for BOS, the stimuli were never heard by birds. Indeed, for stereotyped sequences typical of zebra finches, divergence and convergence probabilities are identical ($D = C^T = 1$ for all natural transitions). Therefore, they provide insufficient sequence variability to differentiate encoding of different probabilistic characterizations. In contrast, Bengalese finches have much more variation in syllable sequencing. This allowed us to test how responses to sequences varied as a function of several different characterizations of the probability with which they were produced.

Because birds produce songs that unfold in time according to divergence probability, it has been hypothesized that neural populations in HVC corresponding to different syllables are coupled by "synaptic chains" with weights that are proportional to divergence probability (Nishikawa et al., 2008; Jin, 2009). Furthermore, some HVC neurons exhibit sensory-motor mirroring (Prather et al., 2008) such that playback of a given syllable might activate the same neurons that are engaged during production of that syllable. Therefore, activity during playback of a sequence of syllables could propagate along the same synaptic chain that is engaged during production of that sequence. In this case, neuronal activity in HVC in response to a given syllable might be positively modulated by the accumulated divergence probability of the preceding sequence.

In contrast to divergence probability, convergence probability reflects the temporal association of events that might naturally become engrained by postsynaptic Hebbian plasticity mechanisms that operate widely in the nervous system (Hebb, 1959; Miller et al., 1989; Dan and Poo, 2004; Bouchard et al., 2012). We

found that responses to a syllable increased monotonically with the convergence probability of preceding sequences (Fig. 4G). Moreover, several alternative probabilistic characterizations (including divergence probability) explained significantly less variation in auditory responses than did convergence probability (Fig. 5). It is possible that other experimental approaches would detect a more salient encoding of divergence probability or other probabilistic characterizations in HVC. Indeed, an encoding of divergence probability must in some sense be present in the brain and may be present in HVC, because the nervous system generates behavior that reflects divergence probability. However, in our auditory playback experiments, HVC neurons most robustly encoded the convergence probability of produced syllable sequences.

Probability-dependent sequence integration

We found that auditory responses in HVC can integrate over at least nine syllables, and this integration depends on the convergence probability of preceding sequences: convergence probability modulated the sign, gain, and temporal extent of integration (Figs. 6, 7). This probabilistic integration is reminiscent of neural integration in primate parietal cortex during sensory discrimination tasks (Yang and Shadlen, 2007; Pouget et al., 2013). For songbirds, such integration could serve a role in perception and discrimination of songs, which may be mediated in part by HVC (Gentner et al., 2000).

Because auditory responses to target syllables integrated over more syllables when preceded by medium-probability sequences than by high-probability sequences, the integrative limits observed for high-probability sequences likely result from a limit of the firing rate of neurons, not from a limit of the "memory capacity" of the circuit *per se*. We found that responses to syllables were approximately linearly proportional to the convergence probability of preceding sequences (Fig. 4) and integration appeared to be linear up to a threshold (Fig. 6). These findings suggest a model in which a linear sequence integrator, the inputs of which are weighted by convergence probability (positive weights for $C > 0$, negative weights for $C = 0$), is constrained by biophysical saturating nonlinearities on neural firing.

Previous examinations of integration in zebra finches report shorter integration times than we found for Bengalese finches (Dave and Margoliash, 2000; Kojima and Doupe, 2008). Our results suggest this difference can be explained by the high degree of stereotypy of zebra finch song. The correspondingly high values of convergence probability predict a steep slope for integration in the zebra finch, causing firing rate saturation after just a few syllables. Conversely, species with highly variable sequences might exhibit even longer integration times than those observed for the Bengalese finch.

In contrast to our results, previous examination of sequence tuning in Bengalese finches using pairwise sequences did not find HVC responses that were positively correlated with sequence probability (Nishikawa et al., 2008). However, this previous study examined divergence probability, not convergence probability. In addition, the differences between integration functions for high- and medium-probability sequences that we observed (Fig. 6D) predict that responses to individual syllables may not be easily distinguishable between these two categories when only preceded by one syllable.

The precise mechanisms that enable long-time scale integration of auditory responses remain to be determined. However, it is noteworthy that, during song production, there are also long-time scale interactions: the probabilities associated with a given

transition during song production can depend on the precise sequence of preceding syllables (i.e., not first-order Markovian; Nakamura and Okanoya, 2004; Jin and Kozhevnikov, 2011; Katahira et al., 2011; Warren et al., 2012). Such long-timescale dependencies in both song production and HVC auditory responses may result from overlapping mechanisms. These mechanisms likely involve the interaction of local circuit processing within HVC with activity propagating across the song-system via internal feedback loops through the brainstem (Ashmore et al., 2005; Long and Fee, 2008; Wang et al., 2008).

Sensory-motor mapping for sequential behaviors

In principle, encoding of convergence probability could result from Hebbian plasticity operating on sequential patterns of purely sensory, or purely motor, activity (Bouchard et al., 2012). However, for sensory-motor structures such as HVC, an intriguing possibility is that the encoding of convergence probability reflects the association of sensory activity arising from previous actions with motor activity associated with the production of current actions. This might arise because, during the course of normal behavior, the brain experiences the probabilistic association between sensory feedback signals from previous actions and premotor activity controlling the current action (Wiener, 1948; Todorov and Jordan, 2002; Sakata and Brainard, 2006, 2008). Therefore, associational learning rules such as Hebbian plasticity could shape sensory-motor circuits to reflect the probabilistic correspondence between premotor activity controlling current actions and sensory feedback resultant from previous actions. Under these conditions, postsynaptic Hebbian plasticity can shape synaptic weight distributions to reflect convergence probabilities (Bouchard et al., 2012), resulting in an encoding of convergence probability in auditory responses. Indeed, in Bengalese finches, HVC neurons are responsive to sound during both singing and quiescence, with the magnitude of responses to BOS being similar to that observed here (Sakata and Brainard, 2008). Therefore, one explanation for our results is that auditory tuning of neurons in the song system is derived from the individual's unique feedback statistics experienced during singing.

Many behaviors, like birdsong, are subserved by sensory-motor circuits that both control the sequencing of actions and receive sensory feedback arising from those actions (Dave and Margoliash, 2000; Heyes, 2001; Sakata and Brainard, 2008; Edwards et al., 2010). Indeed, some sensory-motor neurons exhibit "mirrored" motor activation and sensory responses, with activity that is modulated during both the performance of an action and during sensory perception of that action (Dave and Margoliash, 2000; Rizzolatti and Craighero, 2004; Wilson et al., 2004; Prather et al., 2008; Sakata and Brainard, 2008; Edwards et al., 2010). Sensory feedback arising from an action is necessarily delayed relative to premotor activity that generates the action (Fig. 1E; Wiener, 1948; Brainard and Doupe, 2000; Troyer and Doupe, 2000a, 2000b; Todorov and Jordan, 2002; Sakata and Brainard, 2006, 2008). These considerations have several implications for the role of feedback in sequence production and the mapping between sensory feedback and ongoing motor activity. First, during sequence production, sensory feedback from the previous action may contribute to the activity generating the next action in the sequence; that is, sensory feedback may contribute to the propagation of activity from one sequence element to the next and thus influence sequence probabilities (Lashley, 1951; Grossberg, 1986; Sakata and Brainard, 2006, 2008; Hanuschkin et al.,

2011). Second, because sensory feedback from previous actions is mapped onto motor activity for current actions, the motor field of mirror neurons should be delayed relative to their sensory field. This interpretation suggests that for birdsong, the sensory-motor "mirroring" that has been described previously (Dave and Margoliash, 2000; Prather et al., 2008) may reflect the mapping of auditory responses to previous syllables onto the motor activity for current syllables.

The results described here for HVC may apply broadly to other sensory-motor circuits; both sensory-motor delays and Hebbian plasticity are ubiquitous in the nervous system (Wiener, 1948; Grossberg, 1986; Troyer and Doupe, 2000a,b; Todorov and Jordan, 2002; Dan and Poo, 2004) and these could interact to engrain convergence probabilities associated with repeated sensory-motor associations. Therefore, an encoding of convergence probability may be a widespread feature of sensory-motor circuits that control sequential behaviors and of neural systems in general.

References

Ashmore RC, Wild JM, Schmidt MF (2005) Brainstem and forebrain contributions to the generation of learned motor behaviors for song. J Neurosci 25:8543–8554. CrossRef Medline

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory communication (Rosenblith WA, ed), pp 217–234. Cambridge, MA: MIT.

Bauer EE, Coleman MJ, Roberts TF, Roy A, Prather JF, Mooney R (2008) A synaptic basis for auditory-vocal integration in the songbird. J Neurosci 28:1509–1522. CrossRef Medline

Bouchard KE, Ganguly S, Brainard MS (2012) Hebbian mechanisms underlying the learning of Markov sequence probabilities. Soc Neurosci Abstr 38:846.16.

Brainard MS, Doupe AJ (2000) Auditory feedback in learning and maintenance of vocal behaviour. Nat Rev Neurosci 1:31–40. CrossRef Medline

Dan Y, Poo MM (2004) Spike timing-dependent plasticity of neural circuits. Neuron 44:23–30. CrossRef Medline

Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. Science 290:812–816. CrossRef Medline

Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 22:567–631. CrossRef Medline

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50:291–301. CrossRef Medline

Fee MS, Kozhevnikov AA, Hahnloser RH (2004) Neural mechanisms of vocal sequence generation in the songbird. Ann N Y Acad Sci 1016:153–170. CrossRef Medline

Fiete IR, Senn W, Wang CZ, Hahnloser RH (2010) Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron 65:563–576. CrossRef Medline

Fujimoto H, Hasegawa T, Watanabe D (2011) Neural coding of syntactic structure in learned vocalizations in the songbird. J Neurosci 31:10023–10033. CrossRef Medline

Gentner TQ, Hulse SH, Bentley GE, Ball GF (2000) Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. J Neurobiol 42:117–133. CrossRef Medline

Grossberg S (1986) The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In: Pattern recognition by humans and machines, Vol. 1 (Schwab EC, Nussbaum HC, eds), pp 187–294. New York: Academic.

Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419:65–70. CrossRef Medline

Hampton CM, Sakata JT, Brainard MS (2009) An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song. J Neurophysiol 101:3235–3245. CrossRef Medline

Hanuschkin A, Diesmann M, Morrison A (2011) A reafferent and feed-forward model of song syntax generation in the Bengalese finch. J Comput Neurosci 31:509–532. CrossRef Medline

Hebb DO (1959) The organization of behavior: a neuropsychological theory. New York: Wiley.

Heyes C (2001) Causes and consequences of imitation. Trends Cogn Sci 5:253–261. CrossRef Medline

Jin DZ (2009) Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. Phys Rev E Stat Nonlin Soft Matter Phys 80:051902. Medline

Jin DZ, Kozhevnikov AA (2011) A compact statistical model of the song syntax in Bengalese finch. PLoS Comput Biol 7:e1001108. CrossRef Medline

Katahira K, Suzuki K, Okanoya K, Okada M (2011) Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. PLoS One 6:e24516. CrossRef Medline

Kojima S, Doupe AJ (2008) Neural encoding of auditory temporal context in a songbird basal ganglia nucleus, and its independence of birds' song experience. Eur J Neurosci 27:1231–1244. CrossRef Medline

Lashley KS (1951) The problem of serial order in behavior. In: Cerebral mechanisms in behavior (Jeffress LA, ed), pp 112–136. New York: Wiley.

Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. J Neurosci 16:6987–6998. Medline

Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. Proc Natl Acad Sci U S A 92:5582–5586. CrossRef Medline

Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature 456:189–194. CrossRef Medline

Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. Nature 468:394–399. CrossRef Medline

Machens CK, Gollisch T, Kolesnikova O, Herz AV (2005) Testing the efficiency of sensory coding with optimal stimulus ensembles. Neuron 47:447–456. CrossRef Medline

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

McCasland JS (1987) Neuronal control of bird song production. J Neurosci 7:23–39. Medline

Miller KD, Keller JB, Stryker MP (1989) Ocular dominance column development: analysis and simulation. Science 245:605–615. CrossRef Medline

Mooney R, Prather JF (2005) The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. J Neurosci 25:1952–1964. CrossRef Medline

Nakamura KZ, Okanoya K (2004) Neural correlates of song complexity in Bengalese finch high vocal center. Neuroreport 15:1359–1363. CrossRef Medline

Nishikawa J, Okada M, Okanoya K (2008) Population coding of song element sequence in the Bengalese finch HVC. Eur J Neurosci 27:3273–3283. CrossRef Medline

Nottebohm F, Stokes TM, Leonard CM (1976) Central control of song in the canary, *Serinus canarius*. J Comp Neurol 165:457–486. Medline

Okanoya K, Yamaguchi A (1997) Adult Bengalese finches (*Lonchura striata* var. *domestica*) require real-time auditory feedback to produce normal song syntax. J Neurobiol 33:343–356. CrossRef Medline

Pouget A, Beck JM, Ma WJ, Latham PE (2013) Probabilistic brains: knowns and unknowns. Nat Neurosci 16:1170–1178. CrossRef Medline

Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature 451:305–310. CrossRef Medline

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R (2009) Neural correlates of categorical perception in learned vocal communication. Nat Neurosci 12:221–228. CrossRef Medline

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192. CrossRef Medline

Rosen MJ, Mooney R (2006) Synaptic interactions underlying song-selectivity in the avian nucleus HVC revealed by dual intracellular recordings. J Neurophysiol 95:1158–1175. CrossRef Medline

Sakata JT, Brainard MS (2006) Real-time contributions of auditory feedback to avian vocal motor control. J Neurosci 26:9619–9628. CrossRef Medline

Sakata JT, Brainard MS (2008) Online contributions of auditory feedback to

neural activity in avian song control circuitry. J Neurosci 28:11378–11390. CrossRef Medline

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216. CrossRef Medline

Sutter ML, Margoliash D (1994) Global synchronous response to autogenous song in zebra finch HVc. J Neurophysiol 72:2105–2123. Medline

Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. Nat Neurosci 5:1226–1235. CrossRef Medline

Troyer TW, Doupe AJ (2000a) An associational model of birdsong sensorimotor learning II. Temporal hierarchies and the learning of song sequence. J Neurophysiol 84:1224–1239. Medline

Troyer TW, Doupe AJ (2000b) An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. J Neurophysiol 84:1204–1223. Medline

Vu ET, Mazurek ME, Kuo YC (1994) Identification of a forebrain motor programming network for the learned song of zebra finches. J Neurosci 14:6924–6934. Medline

Wang CZ, Herbst JA, Keller GB, Hahnloser RH (2008) Rapid interhemispheric switching during vocal production in a songbird. PLoS Biol 6:e250. CrossRef Medline

Warren TL, Charlesworth JD, Tumer EC, Brainard MS (2012) Variable sequencing is actively maintained in a well learned motor skill. J Neurosci 32:15414–15425. CrossRef Medline

Wiener N (1948) Cybernetics or control and communication in the animal and the machine. Cambridge, MA: MIT.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702. CrossRef Medline

Woolley SM, Rubel EW (1997) Bengalese finches Lonchura striata domestica depend upon auditory feedback for the maintenance of adult song. J Neurosci 17:6380–6390. Medline

Yang T, Shadlen MN (2007) Probabilistic reasoning by neurons. Nature 447:1075–1080. CrossRef Medline

Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. Science 273:1871–1875. CrossRef Medline