Published in final edited form as:

Eur J Neurosci. 2008 March ; 27(5): 1231–1244.

Neural encoding of auditory temporal context in a songbird basal ganglia nucleus, and its independence of birds' song experience

Satoshi Kojima and Allison J. Doupe

Keck Center for Integrative Neuroscience, Departments of Physiology and Psychiatry, University of California, San Francisco, CA 94143-0444, USA

Abstract

Some of the most complex auditory neurons known are found in the songbird forebrain, throughout the 'song system', including its basal ganglia nucleus Area X. These cells are selective for the temporal order of the bird's own song (BOS): they typically respond strongly to BOS, but more weakly when the syllable sequence of BOS is played in reverse order (roBOS), indicating that they integrate auditory information over more than single syllables. Here, studying the zebra finch Area X, we found that order selectivity strongly depends on the mean syllable duration of individual songs, decreasing markedly as this duration approaches 150–200 ms. Simply segmenting the same songs differently, creating an increase in average syllable length towards 150 ms, caused a similar decrease in order selectivity. This suggests that song neurons integrate acoustic information over a relatively limited time window, predominantly less than 150 ms. We provided further support for this by showing that a significant fraction of Area X order selectivity was accounted for by the acoustic similarity between each BOS and roBOS, measured using cross-correlation with fixed window sizes, but only when the correlation windows were at least 50 ms and no more than 200 ms long. All the same findings were evident in birds raised without tutor exposure, indicating that tutor learning has little effect on neural mechanisms underlying song temporal selectivity. Our results suggest that song-selective neurons encode much of the temporal context of song using a short, constant time window that is conserved across differences in songs, birds and learning.

Keywords

Area X; isolate; neural integration; selectivity; vocal learning

Introduction

Neural sensitivity to the temporal structure of auditory signals is important for identification and discrimination of behaviorally relevant sounds, such as species-specific vocalizations, including human speech. Some of the most temporally complex auditory neurons known are found in the vocal control system of songbirds. These cells not only respond preferentially to the bird's own song (BOS) over songs of conspecifics, but also exhibit sensitivity to the temporal features of BOS, including the order of the component song syllables (Margoliash, 1983; Margoliash & Fortune, 1992; Lewicki & Konishi, 1995; Lewicki & Arthur, 1996; Lewicki, 1996; Doupe, 1997; Solis & Doupe, 1997; Dave & Margoliash, 2000; Rosen & Mooney, 2003; Coleman & Mooney, 2004). For example, in auditory neurons of the zebra finch basal ganglia nucleus Area X, which enhances behavioral discrimination of BOS from conspecific songs (Scharff *et al.*, 1998), the normal strong response to BOS (Fig. 1A) dramatically decreases when the same song is played backward ('reverse BOS'; Fig. 1B): this

manipulation preserves the spectral components of BOS while completely altering its temporal properties. The neuron must therefore 'listen to' or integrate auditory information in song over some amount of time before firing. One way to examine the time scale of this neural integration is to reverse the song less completely, by reversing only the order of the syllables of song, thus preserving the local temporal context within individual syllables. In many song-selective neurons, this 'reverse order BOS (roBOS)' evokes a substantially weaker response than the forward BOS but often stronger than completely reversed BOS (Fig. 1C), revealing that the neuron integrates auditory information over more than single syllables but less than the entire song length.

Beyond this difference between reverse and roBOS responses, the time scale of the neural integration underlying BOS responses remains relatively poorly understood. Although previous studies have reported that song-selective neurons in the HVC nucleus (previously called the high vocal centre), which is afferent to Area X, can integrate song acoustic information over hundreds of milliseconds and multiple syllables (Margoliash, 1983; Margoliash & Fortune, 1992; Lewicki & Konishi, 1995; Lewicki & Arthur, 1996; Dave & Margoliash, 2000), this was true only for a subset of neurons and was analysed only for the neural responses to a subset of syllables in song. The overall integration time scale underlying the entire BOS response in a large, unselected population of song-selective neurons has not been investigated systematically. Moreover, although BOS selectivity is shaped by the bird's experience of its own song during song development, the contribution of song experience to neural encoding mechanisms underlying BOS selectivity is unknown.

In the present study, we analyse syllable order selectivity in Area X of adult zebra finches and provide several lines of evidence that much of the order-selective response of Area X neurons reflects an integration time window of relatively constant and surprisingly limited duration that is conserved across birds, regardless of individual differences in song structure. Furthermore, by showing similar results in birds raised without tutor song experience, we demonstrate that the neural mechanisms for auditory temporal context in Area X develop independently of birds' song experience in early life.

Materials and methods

Animals and electrophysiology

Experiments were performed in adult male zebra finches (*Taeniopygia guttata*, 151–774 days old). The care and treatment of experimental animals was reviewed and approved by the Animal Care and Use Committee at the University of California, San Francisco, and followed by NIH guidelines for the care and use of laboratory animals. Normal birds were raised in individual cages in the colony, with their parents and siblings. At post-hatch day (PHD) ~60, they were removed from their parents, and housed in cages with their brothers and/or other conspecifics. Birds to be isolated were initially raised in sound-attenuating chambers with their parents and siblings from the same clutch, until PHD 9–11. After this day the male parent was removed. At PHD ~30, the young birds were removed from the female parent and the siblings, and housed alone in individual sound-attenuating chambers.

Single-unit extracellular recordings of Area X neurons were performed in isolate and normal birds, as described by Solis & Doupe (1999) and Kojima & Doupe (2007). Briefly, the birds were anesthetized with a 20% solution of urethane (60–85 μ L, i.m.; Sigma, St Louis, MO, USA) or sedated with 0.5% diazepam (10–30 μ L, i.m.; Abbott Laboratories, North Chicago, IL, USA), secured in the stereotaxic apparatus by a head post that had been attached to the skull in a previous surgery, and placed in a sound-attenuating chamber. Extracellular activity was recorded with tungsten electrodes, and spikes were sorted online with a window discriminator (FHC, Bowdoinham, ME, USA), or offline using spike-sorting software

developed by Michael Lewicki (Caltech; Lewicki, 1994) and Brian Wright (University of California, San Francisco, USA). Electrolytic lesions were made at selected locations for reconstructing recording sites. After the experiment, birds were anesthetized with isoflurane (Baxter Health Corporation, Deerfield, IL, USA) and transcardially perfused with saline followed by 3.7% formaldehyde in 0.025 M phosphate buffer, and electrolytic lesions were identified in brain sections stained with Cresyl violet. Only neurons histologically confirmed to be in Area X were analysed.

Stimuli

The songs of both isolate and normal birds were recorded as described by Solis & Doupe (1999), at 2–9 days before the neurophysiological recordings. Prior to the final song recording, isolate songs were recorded 2–3 times at intervals of 1 week–4 months to make sure that the songs were crystallized. Once songs were crystallized, a song of the type that was most frequently produced was chosen by visual inspection of the spectrograms and by listening to the songs. Most BOS stimuli had several introductory notes followed by two song ‘motifs’ (the basic unit of zebra finch song, a stereotyped sequence of syllables). Once single-unit recordings were obtained in Area X, the BOS stimulus was played at different peak volumes (75, 65, 55, 45 dB) and the optimal stimulus intensity was determined at most sites. Syllable order selectivity of the neuron was then examined by playing roBOS (see below), as well as forward BOS at that intensity (usually 65 dB). The stimuli were presented in a random, interleaved fashion. The intertrial interval was randomly varied within a range of 2–5 s. An effort was made to present each neuron with 15–30 trials of each stimulus type.

The roBOS was made by segmenting BOS at syllable boundaries and reversing the syllable sequence, while maintaining the temporal order within individual syllables (Fig. 1A and C). In most birds (six of 10 isolate birds; 18 of 21 normal birds), syllable boundaries were defined (by inspection of song oscillograms, which display sound pressure vs time) as points where the song envelope fell to background level. For the songs of the other seven birds (four isolate and three normal birds), we developed a more automated and quantitative approach to segmentation: syllable/note boundaries were defined as points where the song's envelope crossed a threshold, which was set as 5–10 times the RMS amplitude of the background noise. Then, silent intervals shorter than 2 ms were discarded, and the notes before and after the short intervals were combined into one syllable. For straightforward songs with clearly defined syllables, this method resulted in the same segmentation as the visual inspection of oscillograms. However, this method also allowed us to systematically assess the effects of different segmentations of the same song, especially when syllable boundaries were less clear. We made different versions of roBOS, with varying degrees of song reversal, by using different lengths of the minimum intersyllable interval (2–20 ms) in defining syllable boundaries before the sequence was reversed (see Fig. 6).

Neural responsiveness and selectivity

The response strength (RS) of a neuron to a stimulus was calculated as the difference between the firing rate during the stimulus and the background rate. The period of stimulus presentation was offset by estimates of the latency determined in prior studies (Solis & Doupe, 1999). The background firing rate was defined as the mean firing rate of a 2-s period preceding stimulus presentation. For each neuron, the RS to a stimulus was measured for each trial and then averaged across trials to obtain a mean RS.

We analysed further only neurons that were selective for BOS over conspecific song, although this was the vast majority of auditory neurons in Area X (67/76 in 19 normal birds, 44/46 in 10 isolates; Kojima & Doupe, 2007). The selectivity of an individual neuron for the syllable sequence of BOS was then examined by comparing responses to BOS and roBOS. The degree

of syllable order selectivity was quantified based on roBOS response relative to BOS response, using the selectivity index (SI), which compares the mean RS to each stimulus in ratio form (Volman, 1996; Doupe, 1997; Solis & Doupe, 1997):

$$SI = \frac{\overline{RS_{BOS}}}{\overline{RS_{BOS}} + \overline{RS_{roBOS}}}.$$

SI will approach 1.0 if BOS response is much stronger than roBOS response, and will be close to 0.5 if BOS response is comparable to roBOS response. When a particular $\overline{RS_{roBOS}}$ had a negative value, it was considered to be zero for calculating SI, and thus the SI would be 1.0 regardless of $\overline{RS_{roBOS}}$.

Comparisons of order selectivity to song structure

To explore neural mechanisms underlying syllable order selectivity in Area X neurons, we compared the SI to aspects of song temporal structure such as the mean syllable duration of each song, and also assessed the effect on SI values of altering syllable definitions so that the mean syllable duration was longer. The mean syllable duration was calculated from all syllables within each BOS stimulus that we used in the electrophysiological experiments.

In addition, in order to test whether we could account for order selectivity with a simple neural mechanism that was suggested by our experimental results, we compared neural order selectivity to the acoustic similarity between BOS and roBOS, which we calculated assuming the same neural mechanism. Our experimental data suggested that Area X neurons integrate acoustic information over a time window of limited duration in order to respond to BOS. We therefore defined the acoustic similarity measure to compare a window of fixed size slid sequentially across the songs, in the same way a model neuron with a constant integration time would evaluate the song. If such a measure assesses song in a manner close enough to the neural mechanism underlying syllable order selectivity, then its value for each bird should be correlated with that bird's syllable order selectivity.

Our acoustic similarity measure, which we call 'average BOS–roBOS acoustic similarity', was based on cross-correlation between segments of BOS and roBOS spectrograms. Power-spectrograms ranging from 250 to 8000 Hz were made from songs that were presented in the electrophysiological experiments, using 8-ms FFT windows slid in 2-ms increments; those parameters were chosen because they make the song's time/frequency contours clearly visible in the spectrogram. The FFT window size is also comparable to that used in the widely used song analysis program, the Sound Analysis Pro (Tchernichovski *et al.*, 2000). Just as a neuron evaluates likely acoustic information in roBOS by capturing short sound segments starting at song onset, using a particular integration time window, a short time window of fixed duration was set at the onset of roBOS (see Fig. 7B). Because a BOS-tuned neuron could potentially respond to the sound segment in the roBOS window if the segment was acoustically similar to any sound segments of similar duration in BOS, the Pearson's correlation coefficients (C) were calculated between the roBOS segment and all the possible BOS segments captured with a time window of the same duration. The correlation coefficient C was calculated as follows:

$$C = \frac{\sum_x \sum_y \{ (f(x, y) - \overline{f(x, y)}) \cdot (g(x, y) - \overline{g(x, y)}) \}}{\sqrt{\sum_x \sum_y (f(x, y) - \overline{f(x, y)})^2 \cdot \sum_x \sum_y (g(x, y) - \overline{g(x, y)})^2}},$$

where $f(x, y)$ and $g(x, y)$ are the matrices of two spectrograms compared. In order to create a measure of maximum possible similarity, only the maximum value of these comparisons ($C_{\max}$, equivalent to the cross-correlation coefficient between the roBOS and the whole BOS, solid arrow in Fig. 7B) was noted.

Next, just as a neuron would evaluate roBOS from the song onset to the end, the roBOS window was moved toward the end of the roBOS in 2-ms increments (i.e. every time frame in the spectrogram), and at each step the sound in the segment of roBOS was assessed for similarity to BOS. The $C_{\max}$ (the maximum value of correlation coefficients calculated between each roBOS segment and all possible BOS segments) was noted for each 2-ms step (e.g. dashed arrow in Fig. 7B). Finally, all the maximum correlations $C_{\max}$ were averaged as the ‘average BOS–roBOS acoustic similarity’.

Because we did not have *a priori* information about the window size for which the average BOS–roBOS similarity might best explain the actual syllable order selectivity, we calculated the acoustic similarity multiple times, each with a different comparison window size (10, 20, 50, 100, 200 or 500 ms), for all the songs that were presented in the electrophysiological experiments. We then compared each of these acoustic similarity measures to the actual neuronal responses to those songs. If the neural integration time window of Area X neurons is relatively constant throughout song, and similar in duration to the window size of a particular acoustic similarity measure, there should be a correlation between order selectivity and that similarity measure.

Results

Two questions regarding syllable order selectivity in the zebra finch Area X are as follows. (i) How do neurons encode song temporal structure? In particular, for how long on average do Area X neurons integrate auditory information in song? (ii) Is syllable order selectivity influenced by experience, in particular sensory acquisition of tutor song? We first analysed syllable order selectivity in normal birds, to address the first question, and then compared the neural properties of these birds with those of birds raised without tutor song exposure (isolate birds) to address the second question.

Syllable order selectivity of Area X is highly variable across different birds

Consistent with previous reports both in the Area X and the upstream nucleus HVC (Margoliash, 1983; Margoliash & Fortune, 1992; Lewicki & Konishi, 1995; Lewicki & Arthur, 1996; Doupe, 1997; Rosen & Mooney, 2000, 2003; Coleman & Mooney, 2004), Area X neurons in our normal birds showed sensitivity to the sequence of syllables in BOS (67 neurons from 19 birds). Most neurons responded to BOS more strongly than to roBOS (Fig. 2A), exhibiting substantial syllable order selectivity. When we quantified the syllable order selectivity with a SI, which will approach 1.0 if a neuron has high order selectivity and will be close to 0.5 if it has low order selectivity (see Materials and methods), most neurons had SI values larger than 0.5 (Fig. 2B), indicating their preference for forward BOS over roBOS. The degree of syllable order selectivity was, however, highly variable across neurons. The SI varied from 0.5 (i.e. no order selectivity) to 1.0 (i.e. high order selectivity) and was distributed widely across that range.

What causes the highly variable degree of syllable order selectivity across different neurons? Syllable order selectivity is calculated from neural responses to forward and reversed versions of BOS. Because BOS differs between different birds, the degree of order selectivity may vary across birds depending on the acoustic structure of their song. Alternatively, because Area X is known to have more than one type of neuron (Farries & Perkel, 2002), different types of neurons may have different degrees of syllable order selectivity, even in the same birds. To

test these hypotheses, we examined the variability of syllable order selectivity across different birds as well as within individual birds.

We found that the degree of syllable order selectivity in Area X neurons was highly variable across individual birds. Figure 3A illustrates the extremes. The bird ‘normal 15’ had Area X neurons that responded to roBOS as strongly as to forward BOS, thus exhibiting very low order selectivity (see also Fig. 3B). In contrast, many Area X neurons of ‘normal 16’ responded well to forward BOS but much more weakly to roBOS, so that they showed high order selectivity. Within individual birds, however, the variance of order selectivity was relatively small: all Area X neurons in normal 15 had SI values smaller than 0.7, showing little preference for BOS over roBOS, while many Area X neurons of normal 16 had much higher SI values, indicating a strong preference for forward BOS over roBOS (Fig. 3B). This trend for smaller variance in syllable order selectivity within each bird and higher variance across individual birds was observed in Area X of most birds that we tested and, accordingly, the variance across individual birds was significantly larger than the variance within individual birds (Fig. 3B; $P < 0.0001$, one-way ANOVA across different birds). These results suggest that the highly variable degree of syllable order selectivity is not due to the heterogeneity of neurons within individual birds, but could be due to bird-dependent factors such as the acoustic structure of BOS.

The degree of syllable order selectivity is correlated with syllable duration

The highly variable degree of syllable order selectivity across different birds raises the question of what song features might cause this. Syllable order selectivity results from a neuron integrating auditory information across syllables and, thus, the degree of order selectivity could depend on aspects of the temporal structure of song, for instance the lengths of song syllables. To investigate this possibility, we examined syllable order selectivity in individual birds in relation to their song syllable duration.

We found that the degree of syllable order selectivity in Area X neurons was inversely correlated with the mean syllable duration of each bird's song (Fig. 4). That is, a song with longer syllables on average elicited stronger roBOS responses, resulting in lower syllable order selectivity. In general, roBOS became as effective as BOS at eliciting responses when average syllable duration approached 150–200 ms (note where the regression line crosses the x -axis in Fig. 4). We also observed a similar phenomenon when we did not take the variability of neurons within birds into account, but simply compared mean syllable duration with the mean SI value of all neurons sampled within a nucleus in individual birds ($r = -0.663$, $P < 0.002$). These results suggest that the different degrees of syllable order selectivity across birds reflect features of the temporal structure of song, such as mean syllable duration.

Possible neural mechanisms underlying the negative correlation between syllable order selectivity and mean syllable duration

What is the neural basis for this negative correlation between syllable order selectivity and mean syllable duration? One possible explanation is that Area X neurons integrate auditory information in BOS over a limited time window, so that syllable order selectivity is mainly determined by the relative time scales of the neural integration time window and syllable duration. Figure 5A explains this hypothesis, as follows.

If we assume that Area X neurons integrate auditory information in song using an integration time window that is limited in duration (e.g. 100 ms) and relatively constant throughout song and even across different birds, zebra finch song syllables (normally 20–200 ms duration) will be either shorter or longer than the integration time window depending on their duration. If syllables are longer than the neural integration time window, Area X neurons will be integrating auditory information for a shorter period of time than the individual syllable durations when

they respond to BOS. Because reversing the order of syllables maintains the local temporal structure within individual syllables, the neurons will still be able to receive enough information from the syllables, even in roBOS, to generate a response as strong as that in forward BOS (Fig. 5A, top). Conversely, if individual syllables are shorter than the neural integration time window, that is, if a neuron needs to integrate acoustic information for longer than individual syllables in order to generate firing after the syllables in BOS, reversing the order of those syllables will place an acoustic change within the integration time window. Then those syllables in roBOS will evoke a much weaker response than those in forward BOS (Fig. 5A, bottom). Thus, responsiveness of Area X neurons to roBOS will be determined mainly by the lengths of syllables contained in song. Many zebra finch songs do not have just long or short syllables, but instead have a mixture of syllable durations (Fig. 5A, middle). Even for such songs, any syllables longer than the integration time window will evoke a response in roBOS as strong as those in forward BOS, while any syllables shorter than the time window will evoke a much weaker response in roBOS, resulting in a compound response to the entire roBOS stimulus, with a combination of strong and weak firing. Therefore, the overall neural response to roBOS (i.e. mean firing rate during roBOS) will vary depending on the proportion of long and short syllables in the song, and should still be positively correlated with mean syllable duration. Because roBOS response is inversely proportional to syllable order selectivity (SI) by definition (see Materials and methods), the relationship between roBOS response and mean syllable duration will result in a negative correlation between syllable order selectivity and mean syllable duration (right column in Fig. 5A).

Although this hypothesis can explain the variable degree of syllable order selectivity of Area X neurons and its negative dependence on mean syllable duration of individual songs, there are different factors that could also influence order selectivity. For instance, some songs have repeats of acoustically similar syllables or of similar syllable pairs (Fig. 5B). Because these syllables provide very similar acoustic information to neurons during both forward and roBOS, regardless of their duration, they will lower the order selectivity of the neurons independently of the neural integration time. Although this factor seems unlikely to explain by itself the negative correlation between order selectivity and mean syllable duration that we found (Fig. 4), the syllable repetitions could contribute to this negative correlation to some extent. For example, if the syllables or syllable pairs repeated in a motif tended to be long, many songs with syllable repeats would have both relatively long mean syllable duration and low order selectivity.

Altering the definitions of song syllables can change syllable order selectivity

We wondered how well the hypothesis of a limited neural integration time could explain the negative correlation between syllable order selectivity and mean syllable duration, and how much other factors such as repeated syllables might contribute to the negative correlation (see above). To address this question, we examined the effect of different syllable segmentations of the same song on syllable order selectivity and mean syllable duration. We presented birds with two different versions of roBOS, with different syllable definitions and thus different mean syllable durations, and measured the syllable order selectivity of Area X neurons. If the negative correlation between syllable order selectivity and mean syllable duration reflects the relative time scales of a limited neural integration time and syllable duration, changing the mean syllable duration by simply changing the syllable definitions will lead to different degrees of syllable order selectivity, depending on the mean duration of the newly defined syllables. Thus, Area X neurons should show a decrease in order selectivity as we cause the mean syllable duration to become longer, resulting in a negative correlation between syllable order selectivity and mean syllable duration even in the same neuron. In contrast, if other factors such as repeated syllables play critical roles in determining the syllable order selectivity of Area X neurons,

neurons will not necessarily show a decreased degree of syllable order selectivity when mean syllable duration is increased by changing syllable definition.

Syllables in zebra finch songs are generally defined as segments of sound separated by silent intervals, but because the intervals are often very short (< 5 ms; for example, the interval between 'e' and 'f' in Fig. 6B), in many cases the sound segments before and after the short interval can be defined either as two separated syllables ('e' and 'f') or as a single syllable ('C') composed of two subsyllables, depending on the experimenters. We segmented the song in Fig. 6A in two different ways, 'a-i' and 'A-E' (Fig. 6B), using different lengths of minimum intersyllable intervals (see Materials and methods), and generated two different versions of roBOS (Fig. 6C): roBOS1 is generated from finely segmented BOS (syllables 'ihgfedcba') and roBOS2 is generated from more grossly segmented BOS (syllables 'EDCBA').

The peristimulus time histograms (PSTH) in Fig. 6C show an example of Area X neuronal responses to BOS and the two versions of roBOS. The roBOS1 evoked much weaker responses than BOS, while roBOS2 evoked responses as strong as BOS, resulting in higher syllable order selectivity for BOS vs roBOS1 than for BOS vs roBOS2. Because roBOS1 had a shorter mean syllable duration (83.1 ms) than roBOS2 (138.5 ms), this result shows that syllable order selectivity is negatively correlated with mean syllable duration even for the same song, simply segmented differently (filled circles in Fig. 6D). This negative correlation, together with the strong roBOS2 response, comparable to the BOS response, supports the idea that the neuron has an integration time for generating the BOS response that is relatively limited in duration throughout the song, regardless of the acoustic structure of syllables, and mostly shorter than the mean syllable duration of roBOS2. A similar negative correlation (a decrease in syllable order selectivity after the artificial increase in mean syllable duration) was observed in most neurons in all the birds that we tested (17/19 neurons from three birds; Fig. 6D), and the SI in many cells approached 0.5 when mean syllable duration was about 150 ms. Thus, this manipulation of syllable length further supported our hypothesis of a limited neural integration time, across different neurons and different birds.

Syllable order selectivity of Area X neurons may reflect acoustic similarity over a limited neural integration time

All the neurons analysed here are selective for BOS, responding strongly to BOS but not to completely reversed BOS or conspecific songs (Kojima & Doupe, 2007). If such a song-selective neuron responds to roBOS as well as it does to BOS, this means that it has detected similar acoustic segments within the two songs. Thus, the degree of syllable order selectivity should reflect the acoustic similarity that the neuron detects between forward and roBOS. Given this fact, the correlation between syllable order selectivity and mean syllable duration (Fig. 4) can be taken apart into two different correlations (Fig. 7A): (i) syllable order selectivity is correlated with the acoustic similarity that neurons detect between BOS and roBOS; and (ii) the acoustic similarity is in turn correlated with mean syllable duration. We have so far investigated possible neural underpinnings of syllable order selectivity by focusing on the relationship of order selectivity to mean syllable duration. However, because BOS-roBOS acoustic similarity is the more direct determinant of syllable order selectivity, we next estimated the BOS-roBOS acoustic similarity and examined its relationship to syllable order selectivity (arrow 1 in Fig. 7A), as well as that to mean syllable duration (arrow 2), in order to further test the hypothesis of a limited integration time in Area X neurons.

Because the hypothesis assumes that Area X neurons integrate acoustic information over a limited time window throughout song, and across neurons and birds (Fig. 5A), we estimated song acoustic similarity under the same assumptions – that is, our similarity measure compares segments of forward and reverse order songs using a time window that is constant in duration throughout song and across birds (see Materials and methods for details). We compared the

values of this measure for each song to the same bird's neuronal order selectivity in order to examine the relationship between the acoustic similarity and syllable order selectivity (arrow 1 in Fig. 7A). If such a measure assesses acoustic similarity in a manner that resembles the neural mechanisms underlying syllable order selectivity, then it should be correlated with syllable order selectivity – that is, songs with more similar segments in forward and reversed versions should be associated with neurons with lower order selectivity. Moreover, this measure has a variable – the size of the time window used for the acoustic comparison – that can be used to estimate the time scale of neural integration that Area X neurons might employ. Our results so far show that the roBOS becomes as effective as BOS at eliciting responses when average syllable duration is about 150–200 ms (Figs 4 and 6D), indicating that most neurons integrate auditory information over shorter times than that in responding to BOS. By exploring a range of acoustic similarity window sizes, including both below and above 150 ms, and assessing which window duration leads to the strongest correlation between forward–reverse song acoustic similarity and order selectivity, we could provide more direct estimates of the average time scale of neural integration, for all neurons.

Our acoustic similarity measure (see Fig. 7B and Materials and methods for details), which we call ‘average BOS–roBOS acoustic similarity’, is based on cross-correlation between the roBOS and BOS spectrograms. This measure makes the simple assumption that responsiveness to a sound segment in roBOS is determined by the acoustic similarity between this roBOS segment and the most similar segment in BOS. It therefore calculates a correlation coefficient between the first roBOS segment (chosen to be of a particular, fixed duration, i.e. the acoustic comparison window) and the most similar segment of the same duration in BOS (i.e. a cross-correlation coefficient between the roBOS segment and the whole BOS), then repeats this procedure for all subsequent roBOS windows in the song (shifting the window 2 ms across the song for each repeat), and averages the results. This provides a measure of the maximum possible similarity of roBOS to forward BOS, for a given comparison window duration. In order to estimate what the duration of the relevant acoustic comparison window might be for the neurons, we calculated the acoustic similarity between forward and reversed songs multiple times, each time with a different, fixed acoustic comparison window size, ranging from 10 ms to 500 ms.

The degree of average BOS–roBOS acoustic similarity (the abscissa in Fig. 7C) depended strongly on the length of the segments being compared. When the comparison window was short (and shorter than most syllables, < 20 ms), the similarity measure was high for all songs, and covered only a small range of values across different songs, regardless of the individual songs' mean syllable durations (green dots in Fig. 7C). This is presumably because most such short roBOS segments found a matching segment in the forward song and generated correlation coefficients close to 1.0. In contrast, when the comparison window was longer (~100 ms), the similarity measure had a wider and lower range of values (red dots in Fig. 7C), likely because a considerable fraction of roBOS windows of this length will include more than one syllable placed in reverse order. Such segments will not find a matching segment (a high cross-correlation value) in the forward song. Similarity measured using this window size thus also depends on how well the comparison window size corresponds to the durations of the syllables in the song being compared and, accordingly, is significantly positively correlated with the mean syllable duration of the individual songs (Fig. 7C and D). When the comparison window was longer than most syllables in any bird (500 ms), the similarity measure was low for all songs (blue dots in Fig. 7C), probably because most roBOS windows included multiple syllables placed in reverse order. These similarity values were also not significantly correlated with mean syllable duration (Fig. 7D). Thus, when song segments are compared using an intermediate range of comparison time windows (Fig. 7D), average BOS–roBOS acoustic similarity is positively correlated with mean syllable duration, supporting the relationship between song similarity and syllable length hypothesized in Fig. 7A (arrow 2).

When we then compared the song similarity measures to the syllable order selectivity of Area X neurons, we found that we could account for a significant fraction of order selectivity, but again only for a certain range of acoustic comparisons. The average BOS–roBOS acoustic similarities calculated using windows with intermediate durations (50–200 ms) were significantly negatively correlated with syllable order selectivity (Fig. 8A and B). That is, neurons in birds with songs of high average BOS–roBOS acoustic similarity (measured in 50–200-ms windows) tended to have low degrees of syllable order selectivity. With short (< 50 ms) and long (> 200 ms) windows, however, no significant correlations were observed. This indicates that the average BOS–roBOS acoustic similarities calculated using the intermediate time windows best account for the syllable order selectivity of those neurons, demonstrating the correlation indicated with the arrow 1 in Fig. 7A. This result, together with the correlation between the acoustic similarity and mean syllable duration (arrow 2 in Fig. 7A), therefore corroborates the idea that Area X neurons have relatively limited integration times throughout song and across different neurons and birds, and suggests that the average integration time is on the order of 100 ms. This agrees well with the marked loss of order selectivity we observed when mean syllable duration was closer to 150–200 ms (Figs 4 and 6D), which indicates that the neural integration time windows are mostly less than that length.

Together, the results of all our experimental manipulations and analyses suggest that much of the syllable order selectivity of Area X neurons can be explained by a relatively simple neural mechanism in which neurons integrate over a limited time window throughout song, on the order of 100 ms duration, regardless of the differences in acoustic structure between individual songs.

Area X mechanisms for encoding song temporal context in isolate birds were comparable to those in normal birds

Song-selective neurons with temporal context sensitivity have been hypothesized to be involved in song learning: their response to a sequence of syllables, at least in the song motor nucleus RA (the robust nucleus of the arcopallium), appears to be a prediction of premotor activity for the next syllable (Dave & Margoliash, 2000), which could be very useful in learning the full song sequence. Because songbirds acquire their song by first memorizing the song of an adult individual (the tutor) and then matching their vocalizations to the tutor song model, the sequence-predicting activity of song neurons may reflect sensory learning of tutor song. Consistent with this possibility, many neurons in the songbird anterior forebrain pathway, including Area X, develop selectivity for the temporal order of tutor song (Solis & Doupe, 1997, 1999; Yazaki-Sugiyama & Mooney, 2004). To address the issue of whether learned tutor song information contributes to the development of song coding mechanisms, we raised zebra finches without exposure to tutor song (isolate birds). We then examined neural encoding of BOS temporal context in isolate Area X neurons, and compared the properties of these neurons with those of normal birds.

All the isolate birds ($n = 10$) whose neural activity we recorded developed stereotyped songs, which often included abnormal acoustic features such as broadband noise notes, upward sweeps, an abundance of call-like notes (harmonic stacks), and abnormally long and short syllables and intervals. The isolate songs shared no song notes with their father's song, except for generic call-like notes. The spectrograms and statistical analyses of these songs have been published elsewhere (Kojima & Doupe, 2007).

Despite the abnormal acoustic properties of the isolate birds' songs, we found, using the same series of analyses of syllable order selectivity as in normal birds, that the neural encoding of song temporal structure in Area X of isolate birds was comparable to that of normal birds (44 neurons in 10 isolate birds). The results of those analyses in isolate birds are as follows (Fig. 9). (i) The degree of syllable order selectivity was highly variable across neurons (Fig. 9A and

B). The SI was distributed widely from 0.5 (non-selective) to 1.0 (highly selective), and neither the mean nor the variance of SI values were significantly different from those in normal birds (unpaired *t*-test for means, $P = 0.23$; *F*-test for equality of variances, $P = 0.17$; see Fig. 2 for normal birds' data). (ii) Syllable order selectivity was negatively correlated with the mean syllable duration of individual songs (Fig. 9C; also see Fig. 4 for comparison with normal birds' data). As in normal birds, the roBOS became as effective as BOS at eliciting responses when average syllable duration approached 150–200 ms, suggesting that Area X neurons in isolate birds integrate auditory information over a limited time window, shorter than 150 ms. (iii) This idea was supported by a decrease in syllable order selectivity after an artificial increase of mean syllable duration, by changing syllable definition, in most isolate Area X neurons that we tested (Fig. 9D and 16/17 neurons in four isolate birds; also see Fig. 6 for comparison with normal birds' data). (iv) The idea of a limited integration time window was further corroborated by direct comparison between syllable order selectivity and the average BOS–roBOS acoustic similarity, calculated using comparison windows of fixed duration. As in normal birds, the average BOS–roBOS acoustic similarity in isolate birds was significantly correlated both with the mean syllable duration of individual songs and with the syllable order selectivity of Area X neurons, but only when it was calculated using time windows of intermediate durations (50 and 100 ms; Fig. 9E and F; also see Figs 7D and 8B for comparison with normal birds' data). This indicates that a neural mechanism with a relatively short, constant integration time window (of 50–100 ms duration) can account for much of syllable order selectivity in Area X neurons of isolate birds.

These results in isolate birds, highly comparable to those in normal birds, demonstrate that the neuronal mechanisms for temporal selectivity can develop without learned tutor song experience. Together with the conserved time scale of temporal context integration across marked differences in song acoustic structure among normal birds, these findings illustrate the independence of temporal coding in Area X neurons from the bird's experience during song development.

Discussion

In the present study, we showed that syllable order selectivity in Area X neurons is strongly dependent on the mean syllable duration of individual songs. This dependency was found even when syllable boundaries were defined in different ways in the same songs. These results show that Area X neurons' response to roBOS depends on how long a stretch of BOS is maintained in roBOS, and suggest that these cells integrate acoustic information in song over a limited time window. The idea of a general, song-independent neural mechanism for order selectivity was further corroborated by measuring the acoustic similarity between BOS and roBOS using a range of sliding time windows designed to represent neural integration times, and analysing which window duration gave the best prediction of the actual neural order selectivity in each bird. Moreover, all these neural properties were also found in birds raised without tutor song exposure, indicating that the learned tutor song memory contributes very little to the development of neural coding for song sequence. Together, these results suggest that Area X neurons encode much of the temporal context of song using a relatively limited and constant time window conserved across different song temporal structure, different birds and different learning conditions.

Strong dependency of syllable order selectivity on mean syllable duration and implied neural mechanisms in Area X

In a number of previous studies throughout the song system, order selectivity of song neurons has been used as a measure of sensitivity to temporal context, and has been compared between birds, brain areas and developmental stages (Lewicki & Arthur, 1996; Doupe, 1997; Solis &

Doupe, 1997; Rosen & Mooney, 2003; Coleman & Mooney, 2004; Nakamura & Okanoya, 2004). These investigations suggest that in the anterior forebrain pathway and the upstream nucleus HVC of adult birds, a proportion, although by no means all, of song-selective neurons have significant syllable order selectivity. However, our results showing the strong dependence of syllable order selectivity on mean syllable duration indicate that syllable order selectivity measures may be highly biased by individual song structure. This bias will be most critical when syllable order selectivity is compared across individual birds: for example, neurons recorded from a bird with long syllables are likely to have lower syllable order selectivity compared with cells from a bird with short syllables, even if they are recorded from the same brain areas in birds at the same age. Thus, studies assessing the fraction of order-selective neurons, or differences in the degree of order selectivity across song system areas or development, using neurons recorded from different birds, need to take into account the temporal structure of individual songs.

Our data also clearly show that syllable order selectivity can vary depending on how syllables are defined by an experimenter. The ‘syllable’ is normally defined as a distinct sound element that is separated from the other elements by a silent period of a certain length, but in some songs this definition can depend on the experimenter. In zebra finch songs in particular, syllable boundaries are often less clear compared with songs of other species, such as Bengalese finches or white crowned sparrows, and therefore syllables can easily be defined in different ways. Such ambiguous syllables are found more frequently in isolate birds' songs (Kojima & Doupe, 2007), and alerted us to the effect of different syllable definitions on syllable order selectivity. In both isolate and normal birds, we were then able to show that intentionally different syllable definitions in the same song lead to different mean syllable durations, and that syllable order selectivity of each bird's neurons varies depending on those durations. This finding not only demonstrates the importance of the experimenter's syllable definitions in analyses of syllable order selectivity, but also strongly suggests the idea that Area X neurons integrate acoustic information over a surprisingly limited time window. Moreover, the systematic relationship between syllable duration and order selectivity within and across birds indicates that order selectivity is not ‘tuned’ to each bird's particular song temporal structure. Such tuning was possible, given the importance of BOS experience in shaping song-selective neurons (Margoliash, 1983; Volman, 1993; Doupe, 1997; Yazaki-Sugiyama & Mooney, 2004; Roy & Mooney, 2007): with respect to order selectivity, neuronal properties might have reflected each BOS, with birds with a sequence of longer duration syllables showing longer neural integration times than birds with short syllables. Instead, our observations suggest that song neurons integrate over a similar time window in different birds, regardless of their song structure.

Estimate of neuronal integration time underlying BOS responses

There are only a small number of prior studies that considered the integration time scale of song-selective neurons in zebra finches (Margoliash, 1983; Margoliash & Fortune, 1992; Lewicki & Konishi, 1995; Lewicki & Arthur, 1996; Dave & Margoliash, 2000). Although these are often summarized by saying that song-selective neurons integrate over hundreds of milliseconds (ranging as high as 250–340 ms) and multiple syllables, in fact this is only true for a subset of order-selective neurons. In the particular studies in which syllables or syllable pairs were examined, some neurons responded even to single syllables played in isolation, and a majority of order-selective cells had strong responses to syllable pairs (the duration of individual syllables is often less than 100 ms), which is consistent with our results from Area X neurons. Moreover, in most of the prior investigations, the analysis of neural integration time involved playing only the particular syllables that evoked a response peak, and was further limited to the small subset of HVC neurons that exhibit syllable combination sensitivity, that is, that show a non-linear increase in response to syllables played together compared with sum of the responses to the same syllables played in isolation. The integration times derived from

these neurons may therefore be biased towards the high end. In contrast, our estimate of integration time scale is based on overall responses during the entire BOS stimulus (i.e. mean firing rate), recorded from the entire population of BOS-selective neurons in Area X, and therefore represents an average integration time scale of the neurons. Thus, although for some song-responsive neurons the maximal temporal context-selective response may require integration across many hundreds of milliseconds and many syllables, our results indicate that a surprising amount of order selectivity, in many cells, reflects integration across a much smaller time window, approximately 100 ms on average.

These estimates of neuronal integration time speak to possible underlying cellular mechanisms. Lewicki & Konishi (1995) examined intracellular mechanisms for neuronal sensitivity to syllable order and combination selectivity in HVC. They demonstrated that HVC neurons encode information about song temporal context using a variety of mechanisms, including inhibitory and excitatory postsynaptic potentials with a range of different time courses, and burst-firing non-linearity. A rebound potential following syllable-triggered inhibition was also hypothesized to contribute to temporal context sensitivity of HVC neurons in a previous extracellular study (Margoliash, 1983). While these mechanisms are likely to be combined in various ways and to contribute to variations in temporal context sensitivity for different syllables and in different neurons, our data suggest that there could be a general computation, such as temporal summation across a small number of syllables, underlying much of the neuronal coding of temporal context of song.

The relatively short integration time observed here also raises the questions of how it relates to integration times of the auditory areas afferent to the song system, and how much of it is inherited from these more purely sensory processing areas. In the primary auditory areas of the zebra finch, the integration time scale of neurons has been investigated by examining receptive fields in response to songs or other temporal varying stimuli (Sen *et al.*, 2001; Woolley *et al.*, 2005; Nagel & Doupe, 2006). The temporal frequency of amplitude modulations in sound that drives neurons best, called the best modulation frequency (BMF), can be calculated from these receptive fields, and the inverse of the BMF gives one indication of the time scale of neural integration. All of these studies show that auditory neurons in field L and in its projection target, the caudal mesopallium, have average values of this parameter ranging from approximately 20 to 100 ms. Similar time scales of integration have been observed in auditory cortex of cats and marmosets (Schreiner & Urbas, 1988; Wang *et al.*, 1995; Joris *et al.*, 2004). Although the BMF is a different measure than our estimate of neural integration time in Area X, the longer integration time in Area X than in primary auditory areas is consistent with the vocal control nuclei receiving the output of an auditory processing hierarchy. At the same time, although Area X neurons, and song system neurons generally, have much more complex selectivity than field L neurons (Lewicki & Arthur, 1996), their integration time is not dramatically longer than that in field L, suggesting that this aspect of their temporal responsiveness is only slightly changed from the auditory forebrain.

Implications of temporal context sensitivity for song learning

Song selectivity emerges during development, and reflects experience of BOS. That is, neurons in juvenile birds require auditory experience of the birds' immature song in order to begin encoding the song and to develop selectivity for it (Volman, 1993; Solis & Doupe, 1999). However, our data on integration time windows from both normal and isolate birds suggest that there is little influence of song auditory experience on the neural mechanisms that are used for encoding BOS temporal context. In both types of birds, syllable order selectivity correlates with mean syllable duration, and despite the often abnormal temporal structure of isolate song (Price, 1979; Eales, 1985; Kojima & Doupe, 2007), the integration time of Area X neurons does not appear to be shaped by individual differences in song temporal structure, but instead

is similar across isolate birds, just as in normally reared ones. These results suggest that, when song system neurons start building temporal selectivity in response to auditory experience of song, they use basic, 'hard-wired' mechanisms to do so; these mechanisms seem to be independent of birds' learning of song, and may reflect genetically encoded neural circuitry in the song system or even the auditory areas afferent to it.

Acknowledgements

This work was supported by National Institutes of Health Grants MH055987 and NS34835 (to A.J.D.), and a Research Fellowship (No. 7202) from the Japan Society for the Promotion of Science for Young Scientists (to S.K.). We thank A. Arteseros for expert histological assistance.

References

- Coleman MJ, Mooney R. Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J Neurosci* 2004;24:7251–7265. [PubMed: 15317851]
- Dave AS, Margoliash D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 2000;290:812–816. [PubMed: 11052946]
- Doupe AJ. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J Neurosci* 1997;17:1147–1167. [PubMed: 8994068]
- Eales LA. Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Anim Behav* 1985;33:1293–1300.
- Farries MA, Perkel DJ. A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. *J Neurosci* 2002;22:3776–3787. [PubMed: 11978853]
- Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. *Physiol Rev* 2004;84:541–577. [PubMed: 15044682]
- Kojima S, Doupe AJ. Song selectivity in a pallial-basal ganglia song circuit of zebra finches raised without tutor song exposure. *J Neurophysiol* 2007;98:2099–2109. [PubMed: 17625059]
- Lewicki MS. Bayesian modelling and classification of neural signals. *Neural Comp* 1994;6:1005–1030.
- Lewicki MS. Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. *J Neurosci* 1996;16:5855–5863. [PubMed: 8795637]
- Lewicki MS, Arthur BJ. Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 1996;16:6987–6998. [PubMed: 8824335]
- Lewicki M, Konishi M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci USA* 1995;92:5582–5586. [PubMed: 7777552]
- Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 1983;3:1039–1057. [PubMed: 6842281]
- Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 1992;12:4309–4326. [PubMed: 1432096]
- Nagel KI, Doupe AJ. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* 2006;51:845–859. [PubMed: 16982428]
- Nakamura KZ, Okanoya K. Neural correlates of song complexity in Bengalese finch high vocal center. *Neuroreport* 2004;15:1359–1363. [PubMed: 15167566]
- Price PH. Developmental determinants of structure in zebra finch song. *J Comp Phys Psych* 1979;93:268–277.
- Rosen MJ, Mooney R. Intrinsic and extrinsic contributions to auditory selectivity in a song nucleus critical for vocal plasticity. *J Neurosci* 2000;20:5437–5448. [PubMed: 10884327]
- Rosen MJ, Mooney R. Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron* 2003;39:177–194. [PubMed: 12848941]
- Roy A, Mooney R. Auditory plasticity in a basal ganglia-forebrain pathway during decrystallization of adult birdsong. *J Neurosci* 2007;27:6374–6387. [PubMed: 17567798]

- Scharff C, Nottebohm F, Cynx J. Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J Neurobiol* 1998;36:81–90. [PubMed: 9658340]
- Schreiner CE, Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 1988;32:49–63. [PubMed: 3350774]
- Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 2001;86:1445–1458. [PubMed: 11535690]
- Solis MM, Doupe AJ. Anterior forebrain neurons develop selectivity by an intermediate stage of birdsong learning. *J Neurosci* 1997;17:6447–6462. [PubMed: 9236252]
- Solis MM, Doupe AJ. Contributions of tutor and bird's own song experience to neural selectivity in the songbird anterior forebrain. *J Neurosci* 1999;19:4559–4584. [PubMed: 10341255]
- Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. A procedure for an automated measurement of song similarity. *Anim Behav* 2000;59:1167–1176. [PubMed: 10877896]
- Volman SF. Development of neural selectivity for birdsong during vocal learning. *J Neurosci* 1993;13:4737–4747. [PubMed: 8229196]
- Volman SF. Quantitative assessment of song-selectivity in the zebra finch 'high vocal center'. *J Comp Physiol [A]* 1996;178:849–862.
- Wang X, Merzenich MM, Beitel R, Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 1995;74:2685–2706. [PubMed: 8747224]
- Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 2005;8:1371–1379. [PubMed: 16136039]
- Yazaki-Sugiyama Y, Mooney R. Sequential learning from multiple tutors and serial retuning of auditory neurons in a brain area important to birdsong learning. *J Neurophysiol* 2004;92:2771–2788.

Abbreviations

BMF	best modulation frequency
BOS	bird's own song
HVC	high vocal centre
PHD	post-hatch day
PSTH	peristimulus time histogram
roBOS	reverse order bird's own song
RS	response strength
SI	selectivity index

Fig. 1.

Examples of auditory responses to normal (forward) bird's own song (BOS). (A) Temporally reversed BOS (reverse BOS; B), and BOS in which syllable order was reversed [reverse order BOS (roBOS); C], in a single Area X neuron from a normally raised bird. (A–C) Each left panel shows sonograms (top; plots of frequency vs time, with amplitude indicated by the darkness of the trace) and oscillograms (bottom; plots of time-varying sound pressure) of the song stimulus. Each syllable is identified by a letter on the top; two motifs ('a'–'e') and preceding introductory notes ('i') are shown. Note that the roBOS maintains the temporal order within individual syllables but reverses the syllable sequence in a song. Each right panel shows a PSTH of neural activity (30-ms bins; top) and an oscillogram of the song (bottom) presented in the left panel; 15 trials of each song were presented. Note that roBOS evoked responses weaker than BOS but stronger than reverse BOS.

Fig. 2.

Syllable order selectivity of Area X neurons in normal birds. (A) The mean response strength (RS) to BOS and roBOS in individual neurons are plotted. The diagonal line indicates where cells would lie if the mean values of RS to the two stimuli were equal. (B) Syllable order selectivity was quantified with a selectivity index (SI) calculated from the mean RS to BOS and roBOS in individual neurons (see Materials and methods). SI values of individual neurons are distributed widely from 0.5 (non-selective) to 1.0 (highly selective), indicating highly variable syllable order selectivity across different neurons.

Fig. 3.

Highly variable degrees of syllable order selectivity in Area X neurons. (A) Single-unit recordings of an Area X neuron with low syllable order selectivity from one bird, 'normal 15' (top), and of an Area X neuron with high syllable order selectivity from a different bird, 'normal 16' (bottom). The PSTHs show the neural responses to 20 presentations of forward BOS and roBOS. The normal 15 neuron had responses to roBOS that were as strong as that to BOS, whereas the normal 16 neuron had weak responses to roBOS compared with BOS. (B) The degree of syllable order selectivity in Area X neurons recorded from individual birds including normal 15 and normal 16. The circles each illustrate a neuron; each number on the abscissa refers to an individual bird; dotted horizontal line represents selectivity index (SI) = 0.5 (no order selectivity). Only data from birds in which more than two neurons were recorded are shown. The degree of syllable order selectivity is significantly more variable across different birds than across different neurons within individual birds ($P < 0.0001$, one-way ANOVA across different birds).

Fig. 4.

Scatter plot compares syllable order selectivity with mean syllable duration of song in Area X neurons (top); oscillograms of two example songs are shown below, one with relatively long (middle) and one with short (bottom) mean syllable duration. The dots each illustrate a neuron; lines indicate the linear least squares fit through the data.

Fig. 5.

Possible neural coding mechanisms underlying syllable order selectivity. In each row, schematic drawings of simplified amplitude envelopes of BOS (left column) and its roBOS (middle column) are shown on the bottoms, and the expected neural responses (PSTHs) to those stimuli are shown on the top. The degree of syllable order selectivity obtained from the neural responses is indicated in the right column. BOS stimulus starts with three introductory notes, which are followed by simplified syllables labeled with lower case letters. (A) A model with a neural integration time window with fixed duration. Dotted rectangles represent integration time windows with fixed duration. Note that a song with syllables longer than the neural integration time window leads to low syllable order selectivity (top) and vice versa (bottom), while song with a mixture of syllable lengths is likely to have intermediate order selectivity (middle). (B) A song that contains similar syllables can also lead to low order selectivity, regardless of neural integration time.

Fig. 6. Syllable order selectivity with different syllable definitions. (A) Oscillogram of a song of a bird 'normal 17'. Introductory notes and the subsequent two motifs (dashed rectangles) are shown. (B) Two different segmentations of the first motif of the song. Sonogram (top) and oscillogram (middle) of the first motif are shown on an expanded time scale. The dashed lines indicate the boundaries of song segments defined as points where the song's envelope crosses a threshold, which was set as five times the RMS amplitude of the background noise. The silent intervals shorter than 2 ms were discarded, and the notes before and after the interval were combined into one syllable. These syllables defined in this way are labeled as 'a'–'i' in the bottom. The syllables labeled 'A'–'E' were defined by discarding longer intervals (< 20 ms). (C) Auditory responses of a single Area X neuron to BOS and roBOS with different syllable definitions. This neuron showed strong responses to the roBOS that was made using syllables 'A'–'E' (roBOS2), as well as to forward BOS, while it responded much more weakly to the

roBOS made by syllables ‘a’–‘i’ (roBOS1). The song oscillograms and the syllable labels are shown underneath the PSTHs; 20 trials of each song were presented. (D) The selectivity indices (SI) of all the Area X neurons recorded from three birds, calculated from the responses to BOS and the two different versions of roBOS (roBOS1 and roBOS2) created using the same approach as shown in (C). Two circles connected with a line illustrate each neuron; filled circles with a thick line represent the neuron whose responses are shown in (C). Seventeen out of 19 neurons showed a decrease in syllable order selectivity when mean syllable duration was artificially increased by changing syllable definition.

Fig. 7.

Acoustic similarity between forward BOS and roBOS measured using sliding windows is positively correlated with mean syllable duration. (A) Possible relationships between syllable order selectivity, acoustic similarity between BOS and roBOS, and mean syllable duration. The relationship indicated by arrow 2 will be examined in (B) and (D), and that by arrow 1 in Fig. 8. (B) Schematic showing the calculation of the average BOS–roBOS acoustic similarity in a bird ‘normal 14’. The acoustic similarities were obtained from comparisons of forward BOS and roBOS spectrograms using time windows of fixed duration sliding through both songs (red rectangles on the spectrograms, see Materials and methods). Example time windows (100 ms duration) on roBOS and their best matching windows on BOS are shown with the cross-

correlation coefficients; solid and dashed arrows indicate strong and weaker correlations, respectively. To find the window size that best explains neural selectivity, the acoustic similarity measure was calculated multiple times, each time with a different, fixed acoustic comparison window size, ranging from 10 ms to 500 ms, and then compared with syllable order selectivity of Area X neurons (see Fig. 8). (C) The average BOS–roBOS acoustic similarities calculated with 10-, 100- and 500-ms windows (green, red and blue dots) are plotted against mean syllable duration of individual birds. The dots each illustrate a bird. The lines indicate the linear least squares fit through the data with each window size; the correlation coefficients (r) are shown for each analysis. (D) Correlation coefficients (r) between average BOS–roBOS acoustic similarity and mean syllable duration plotted as a function of the window size used to estimate the similarity. The filled circles indicate those correlations that are statistically significant ($P < 0.05$). Note that the acoustic similarity measure calculated with 50–200-ms windows was significantly correlated with mean syllable duration.

Fig. 8.

Average BOS-roBOS acoustic similarity can account for syllable order selectivity. (A) The average BOS-roBOS acoustic similarities calculated with 10-, 100- and 500-ms windows (green, red and blue dots) are plotted against the syllable order selectivity (SI) of individual Area X neurons. The dots each illustrate a neuron. Lines indicate the linear least squares fit through the data, and the correlation coefficient (r) is shown for each analysis. (B) Correlation coefficient (r) between syllable order selectivity (SI) and average BOS-roBOS acoustic similarity plotted as a function of the window size used to estimate the similarity. The filled circles indicate those correlations that are statistically significant ($P < 0.05$). Note that the acoustic similarity measure was significantly and negatively correlated with syllable order selectivity when the similarity measure was calculated with intermediate window durations.

Fig. 9.

Area X neural properties in isolate birds, revealed with the same analyses that were used in normal birds, are comparable to those in normal birds. (A) The mean response strength (RS) to BOS and roBOS in all Area X neurons recorded from isolate birds. Conventions are as in Fig. 2A. (B) Highly variable selectivity index (SI) values in isolates' Area X neurons. Conventions are as in Fig. 2B. (C) Negative correlation between syllable order selectivity and mean syllable duration of individual songs. Conventions are as in Fig. 4. (D) Decrease in syllable order selectivity after artificial increase of mean syllable duration by changing syllable definitions using the same approach as shown in Fig. 6. Conventions are as in Fig. 6D. (E) Correlation coefficient (r) between average BOS–roBOS acoustic similarity and mean syllable duration plotted as a function of the window size used to estimate the similarity. The filled circles indicate those correlations that are statistically significant ($P < 0.05$). Conventions are as in Fig. 7D. (F) Correlation coefficients (r) between syllable order selectivity and average BOS–roBOS acoustic similarity are plotted as a function of the window size used to estimate the similarity. The filled circles indicate those correlations that are statistically significant ($P < 0.05$). Conventions are as in Fig. 8B.