

## Chapter 2 : Neuronal ensemble coding of birdsong

*The work described in this chapter was done in collaboration with Michale Fee (Bell Labs)*

**In many models of neural population coding, similar sensory or motor states are represented in the brain by similar neural ensembles (Georgopoulos et al., 1999; Lewis and Kristan, 1998; Wilson and McNaughton, 1993). To explore this issue, we measured the activity of large numbers of single neurons in the pre-motor nucleus RA of the singing zebra finch. During singing, individual RA neurons generate precise bursts of action potential spikes (Yu and Margoliash, 1996). We found that highly similar song elements were typically produced by stereotyped but uncorrelated ensembles of RA neurons. Small changes in acoustic structure between two syllables were frequently subserved by entirely different ensembles of RA neurons. Furthermore, the neural activity in RA often changed on a time scale an order of magnitude faster than the acoustic structure in the song, causing constant acoustic outputs to be produced by sequences of rapidly changing neural ensembles. Thus there are dynamics internal to the song control system that occur reliably and precisely yet are not directly correlated with structure in the song itself. These data represent a new type of neural code underlying complex learned behaviors.**

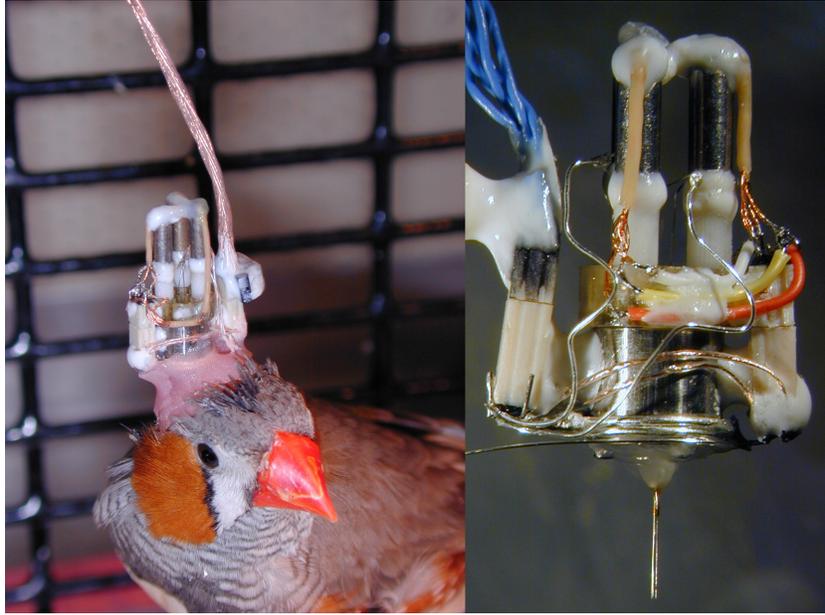
### 2.1 Introduction

Zebra finch song contains spectral and temporal structure over a wide range of time scales. The individual sound elements of the song (syllables) contain distinct spectral features and are produced in a fixed sequence known as a motif. Each time the zebra finch sings, the motif is repeated a variable number of times. This learned behavior is remarkable in its stability and precision; once the song is mastered, the bird will generate it identically for the remainder of his life (Marler, 1970). A set of discrete brain nuclei known collectively as the song system underlie this complex behavior, and within this system, the forebrain pre-motor

nucleus RA (robust nucleus of the archistriatum) controls the instantaneous temporal and spectral structure in the song (Vu, 1994, Chi and Margoliash, 2001). In this chapter we explore the relation between the activity of different RA neurons and their coordination in the production of song structure.

Early recordings in RA demonstrated that neural activity increased substantially before singing began, indicating that RA has a pre-motor role in song production (McCasland, 1987). More detailed experiments in singing birds by Yu and Margoliash (1996) demonstrated that individual RA neurons fire sequences of rapid and highly stereotyped bursts of action potentials at specific time points in the song, and these bursts are timed with submillisecond precision. Stimulation in RA of the awake nonsinging bird generates sounds that contains learned acoustic structure (Vicario and Simpson, 1995), and stimulation of RA during song produces distortions in the syllable being generated at the moment of stimulation (Vu et al., 1994). RA is thus a forebrain nucleus specialized for complex, learned vocal control.

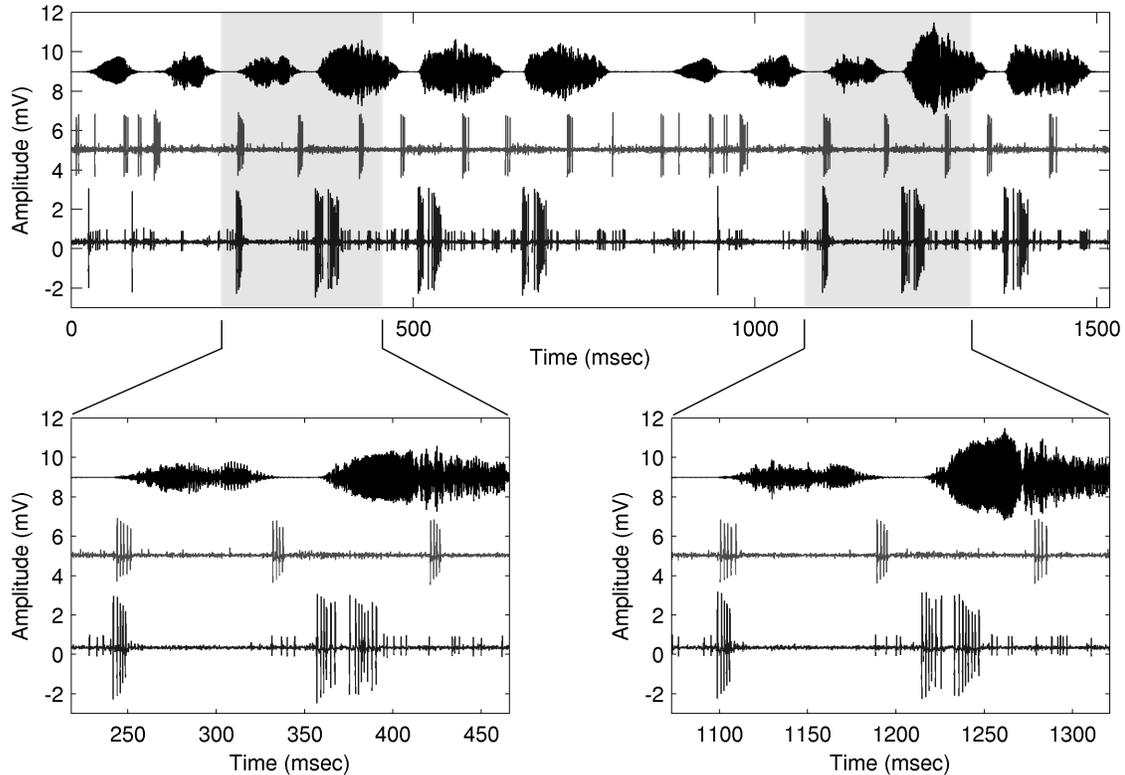
RA is approximately 700  $\mu\text{m}$  in diameter, and contains two classes of neurons: projection neurons, whose axons exit the nucleus, and interneurons, that only make local GABAergic connections within RA (Spiro et al., 1999). The projection neurons receive input from HVC (Nottebohm et al., 1976), and both classes of neurons receive input from the anterior forebrain pathway known to be involved in song learning and maintenance. The interneurons provide fast inhibition to projection neurons and are capable of synchronizing projection neurons that are not directly connected to each other (Spiro et al., 1999). RA is organized myotopically, with respect to the syringeal muscles (Vicario and Nottebohm, 1988). It projects to the tracheosyringeal portion of the hypoglossal nucleus, whose motorneurons project directly to the syrinx, as well as to dorsal regions of the thalamus involved in respiratory control (Vicario, 1991). This anatomical information, and the physiological data discussed in the previous paragraph, summarize our current understanding of RA's place in the neural control system used to generate song. Many pieces of the puzzle are clearly missing. In particular, the neural code used by ensembles of interacting neurons in RA to generate specific types of acoustic structure has not yet been explored. To address this question, we have made the first measurements of the detailed structure of the RA neuronal population activity during singing.



**Figure 2.1.** Miniature motorized microdrive for chronic recording in small animals. The microdrive weighs 1.5 g and contains three independent, motorized electrodes which can be positioned remotely.

## 2.2 Results

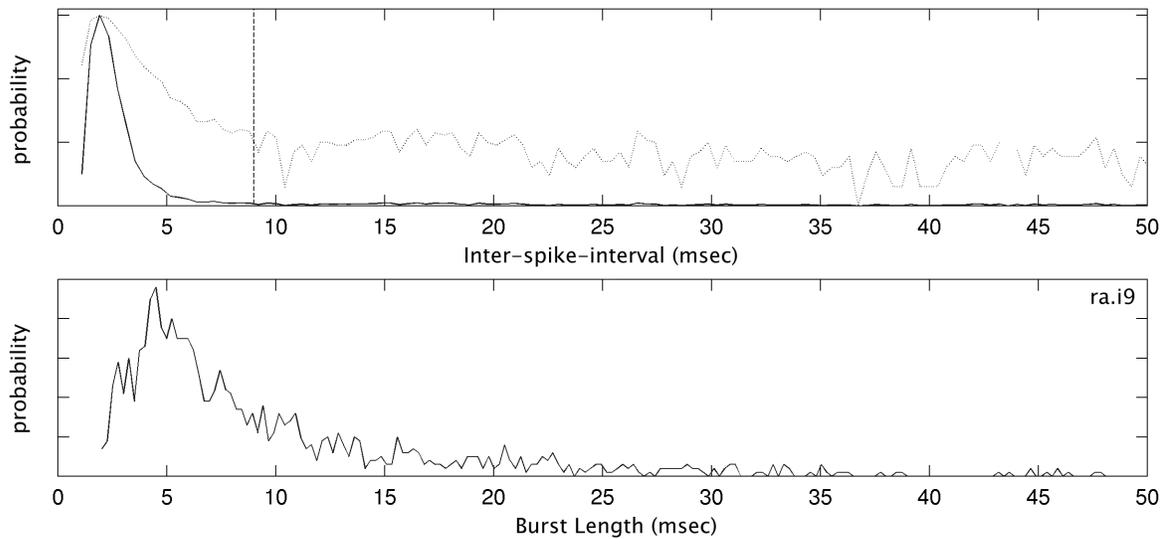
Using a newly developed 3-channel motorized microdrive (Fee and Leonardo, 2001; Figure 2.1), approximately 90 single neurons were recorded in four singing zebra finches. Pre-song, these neurons had periodic firing rates of  $\sim 20$ -30 spikes/sec, but were not synchronized with each other. During each motif of song, the baseline firing patterns were replaced by a bursting regime in which each neuron fired a stereotyped sequence of spike bursts (Figure 2.2). As has been shown previously, these bursts were locked to acoustic structure in the song with millisecond precision (Chi and Margoliash, 2001), and were locked to each other with submillisecond precision (Yu and Margoliash, 1996). We calculated the inter-spike-interval (ISI) histograms obtained during song for each neuron, and found they were narrowly peaked at 2.5 msec (Figure 2.3a). This sharp peak in the ISI histogram reflects the separation between inter-spike-intervals within bursts and inter-spike-intervals between bursts. The transition point between these two modes of activity can be seen from the ISI histogram to occur at  $\sim 8$  msec. Using this as a threshold, we decomposed each spike train into a series of bursts (ISIs  $< 8$  msec) and intervals between bursts (ISIs  $\geq 8$  msec). Based



**Figure 2.2.** Simultaneous recording of three RA neurons in the singing bird. Top panel: two motifs of song are shown along with three simultaneously recorded RA neurons (one neuron on the first electrode, and two separable neurons on the second electrode). Bottom panel: RA neurons generate action potential bursts precisely locked to each other and to structure in the song. Each time the same syllables are repeated from motif to motif of song, virtually identical spike patterns are generated by the neurons.

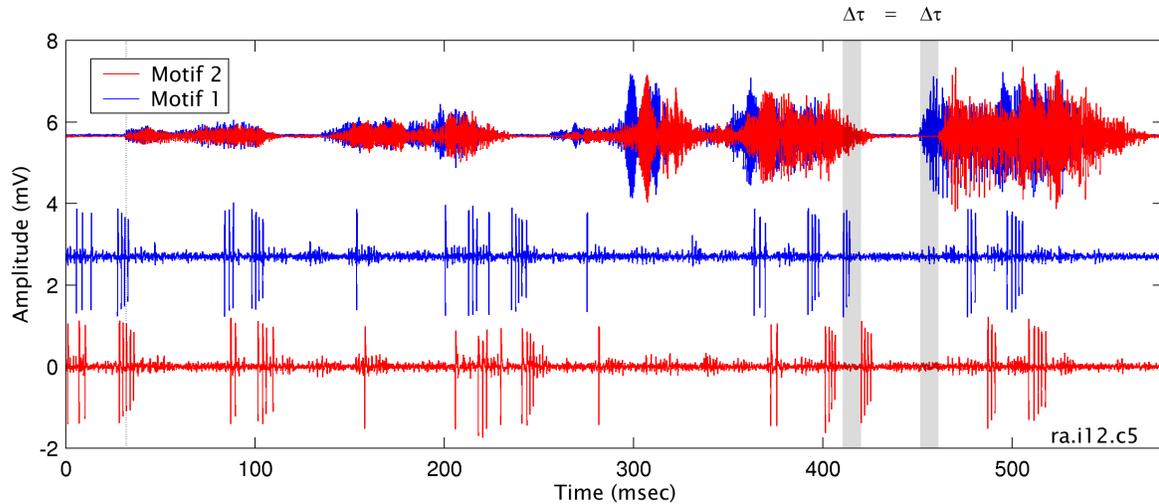
on this transformation, we found that the average length of a spike burst in RA was 10 msec. This reflects the number of short ISIs that tended to occur in series.

The precise burst timing seen in RA spike trains suggests that patterns of RA neuronal activity may play an important role in the generation of song. Although most of the neurons in the data set were recorded as simultaneous pairs or triplets on different electrodes, these groups of neurons were too small to make any inferences about the concerted neural activity of the entire nucleus. However, the spectral structure in zebra finch song is reproduced very accurately each time the bird sings, and this suggested a method for aligning the firing patterns of RA neurons that were not recorded simultaneously. By lining up common spectral features produced each time the bird sang, we were able to align the spike trains of neurons recorded at different times onto a common time axis. The lengths of the



**Figure 2.3.** Inter-spike-interval and burst-width distribution for RA neurons. The frequent bursts generated by RA neurons produce a sharp peak at  $\sim 2.5$  msec in the ISI histogram (top panel; dotted line is the log probability). At  $\sim 8$  msec, there is a transition between the two types of ISIs produced by RA neurons, those within bursts and those between bursts. Using this value as a threshold, each spike train was decomposed into a sequence of bursts and burst intervals. From this, a histogram of burst lengths was estimated for the population of RA neurons recorded in each bird. The burst length distribution is peaked at  $\sim 6$  msec, with a mean length of 10 msec (bottom panel).

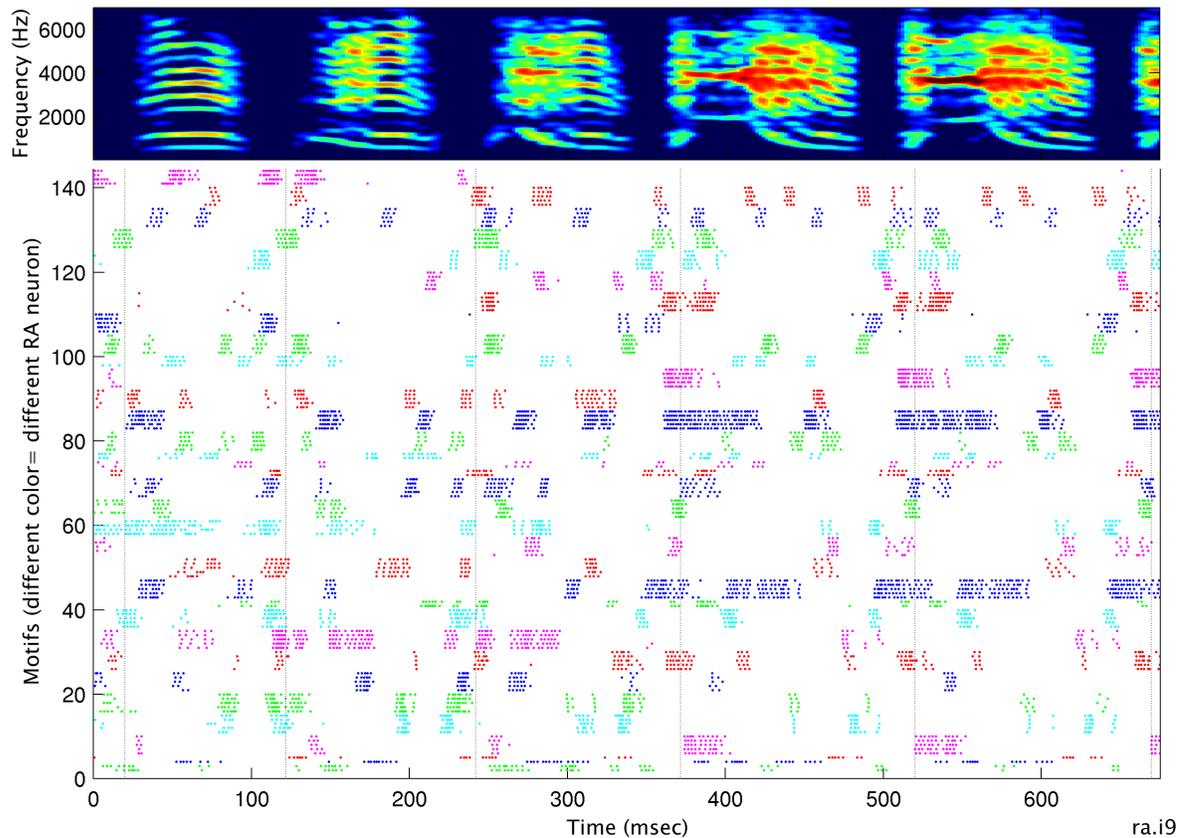
individual syllables varied independently of each other from motif to motif of song, and were randomly stretched and compressed by approximately 5% (Figure 2.4). By estimating the magnitude of this acoustic time-warping and then compensating for it appropriately in each syllable that the bird sang (see Methods), we were able to remove a large portion of the temporal variability inherent in zebra finch song. The spike trains of the simultaneously recorded neurons were then projected onto the aligned time axes of their associated song syllables. Because all of the songs were aligned to a common reference, this procedure mapped the spike trains of all of the sequentially recorded neurons onto a common time axis, allowing them to be analyzed as a group. Because the spike trains of individual RA neurons were time locked to the bird's song (Chi and Margoliash, 2001), the time locked structure between different neurons was revealed by aligning them on a time axis based on the location of precisely reproduced acoustic features in the song (Figure 2.5). As a consequence, we



**Figure 2.4.** Effect of acoustic time-warping on song motifs and neural recordings. Top trace: two motifs of song, aligned to the onset of the first syllable. Middle and bottom traces: activity of a simultaneously recorded RA neuron during each song motif. The lengths of individual syllables vary from motif to motif of song, making alignment to a single reference point inaccurate; at time points far from the reference there is considerable alignment jitter. This jitter is exactly matched by variability in the onsets of bursts in the neural recordings.

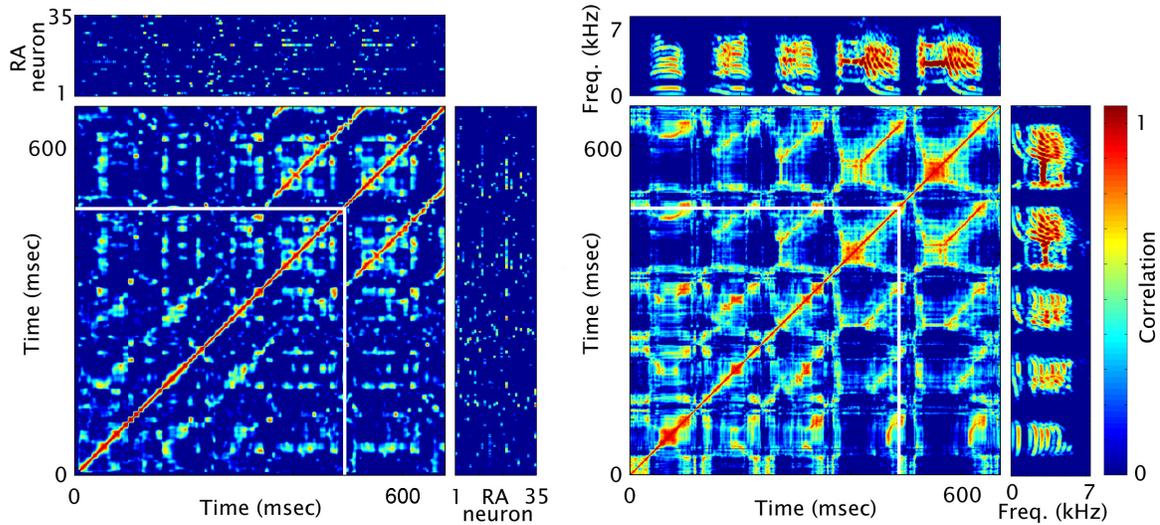
were able to analyze the collective activity of 20-40 neurons, instead of the 2-3 that we could record simultaneously.

Examination of the song-aligned neural data for each bird revealed that each time point in the bird's song could be characterized by approximately 15% of the population of recorded neurons that were currently bursting. Each pattern of neural activity formed a short-lived neural ensemble. Every time the bird sang, the neurons in RA sequenced through the same set of patterns (Figure 2.5), and the neural ensemble consequently evolved in a predictable manner. How similar were the patterns of neural activity that were generated at different time points in the bird's song? We addressed this question by constructing a two-time correlation matrix, such that any point  $[t_1, t_2]$  in the matrix was defined as the correlation between the patterns of neural activity occurring at times  $t_1$  and  $t_2$  in the song (see Methods). The correlation was normalized to be 1 if the ensembles were identical in membership and firing rate. Figure 2.6a shows the correlation matrix for the RA population shown in Figure 2.5. The matrix diagonal represents the correlation of a neural ensemble to itself, and is always one. Off-diagonal values of the matrix are symmetric, ranging from 1



**Figure 2.5.** RA neuronal ensemble activity during singing (35 neurons). The activities of the individual neurons have been converted into spike train rasters. Each row represents one motif of the song shown in the spectrogram at the top of the figure. Different colored spike trains represent the activities of different RA neurons which have all been song-aligned to a common time axis. Each RA neuron reliably produces a unique pattern of bursts. Over the course of the entire motif, different neuronal ensembles turn on and off in an intricate pattern.

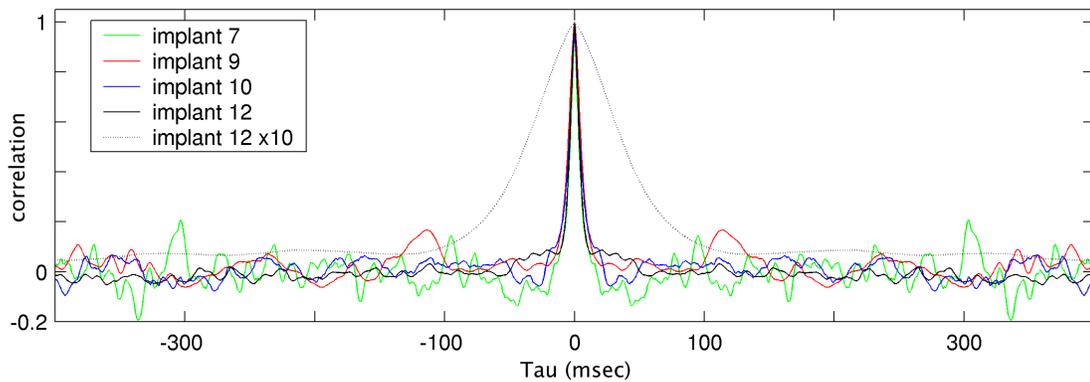
(identical neural ensembles) to zero (uncorrelated neural ensembles), and occasionally to values less than zero. Negative correlations indicate two neural ensembles that shared a common set of neurons that increased their average firing rate in one group and decreased it in the second group. Repeated sequences of neural activity across the population of RA neurons appear in the matrix as continuous lines of high correlation. The average of the neural correlation matrix along its rows represents the average correlation of each neural ensemble to all other neural ensembles, and measures how similar different neural ensembles were to each other. Interestingly, we found the average correlation between different neural



**Figure 2.6.** Neural and song correlation matrices. Left panel: Each point in the neural correlation matrix represents the correlation between two patterns of RA neural activity. The matrix of average firing rate activity for each of the 35 neurons in the recorded population is shown along each axis of the neural correlation matrix. Right panel: The song correlation matrix. The time-frequency spectrogram, from which sound correlations were calculated, is shown along each axis of the song correlation matrix. The white lines in the song and neural correlation matrix separate the unique part of the song from a repeated syllable which the bird stuttered. During the stuttering of this syllable, a repeated sequence of patterns of neural activity can be seen to occur in the neural correlation matrix (the red line parallel to the diagonal). A similar track of high correlation activity in the song correlation matrix shows the time course of the repeated sound.

ensembles in RA to be zero. Each pattern of neural activity in RA tended to be unique, and occurred at only one time point in the song. Thus over the course of the song, different sub-populations of neurons transiently turned on and off in an intricate pattern.

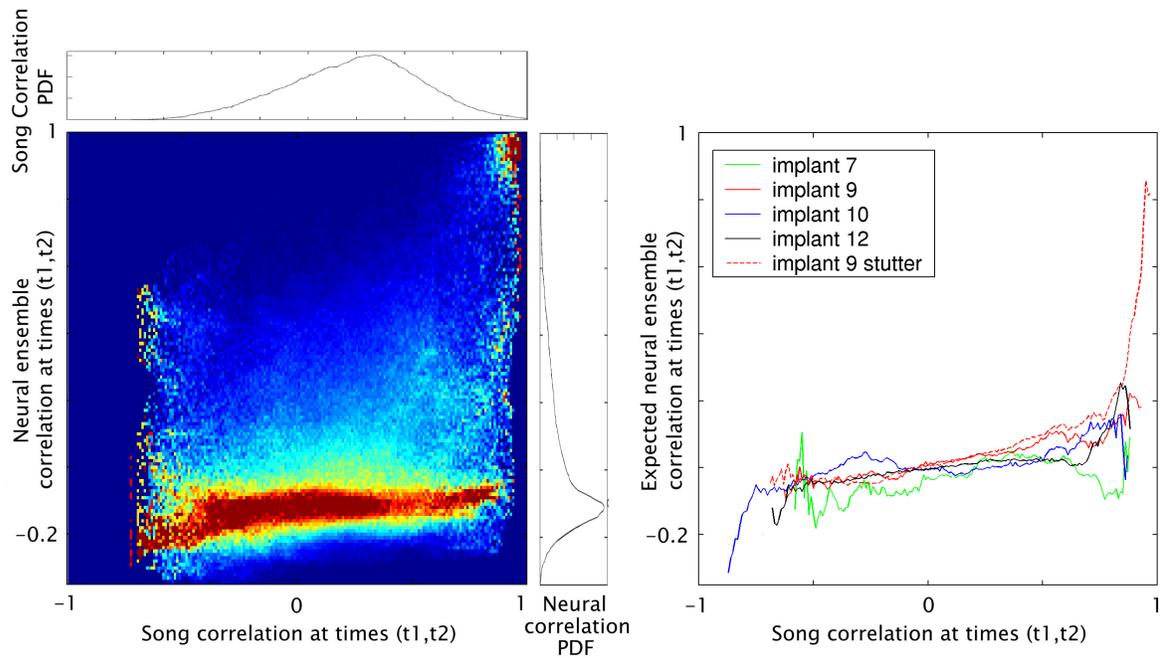
We measured how long patterns of RA neural activity lasted by averaging the correlation matrix along its paradiagonals. This is the neural ensemble auto-correlation over the entire duration of the song (Figure 2.7), and is a measure of the persistence time each neural ensemble before it was replaced by the next pattern. The auto-correlation had a 10 msec width, indicating that patterns of neural activity tended to be stable for 10 msec. By adding random time offsets to the firing profile of each neuron, and then recomputing the neural correlation matrix and auto-correlation, we were able to measure how much the precise alignment of different neurons relative to each other contributed to the persistence time of a neural ensemble. For example, if bursts tended to line up with each other (onset to



**Figure 2.7.** Neural ensemble auto-correlation. Patterns of neural activity tended to last, on average, 10 msec, as can be seen from the width of the auto-correlation peak.

onset and offset to offset), this could substantially increase the correlation time of that pattern of neural activity. The auto-correlation of the randomized neural data should then be significantly narrower than that of the unrandomized neural data, since the synchronization between neurons has been removed. However, we found that the auto-correlation width of the randomized neural data was also 10 msec. Thus, the patterns of neural activity in RA last for 10 msec only because of the distribution of burst widths for RA neurons (which have a mean of 10 msec), and not because of any substantial degree of synchronization between RA neurons.

The importance of the time scale of activity in RA and the similarity between different patterns of neural activity can only be fully appreciated by considering their relation to the spectral structure they generated in the song. Each segment of song that occurred in a small window of time was decomposed into its sound frequency spectrum. A song correlation matrix was then constructed in a manner similar to that of the neural correlation matrix, so that each element in the matrix represented the correlation between the sounds produced at two different time points. The song correlation matrix resembles that of the neural correlation matrix (Figure 2.6a), but it contains noticeably more areas of high correlation (Figure 2.6b). This observation was confirmed by visual inspection of the time-frequency spectrogram of the bird's song: although in general individual syllables are not repeated within a single motif, there are frequently repeated syllable-subsequences, stretches of sound 20-50 msec in length that recur multiple times.

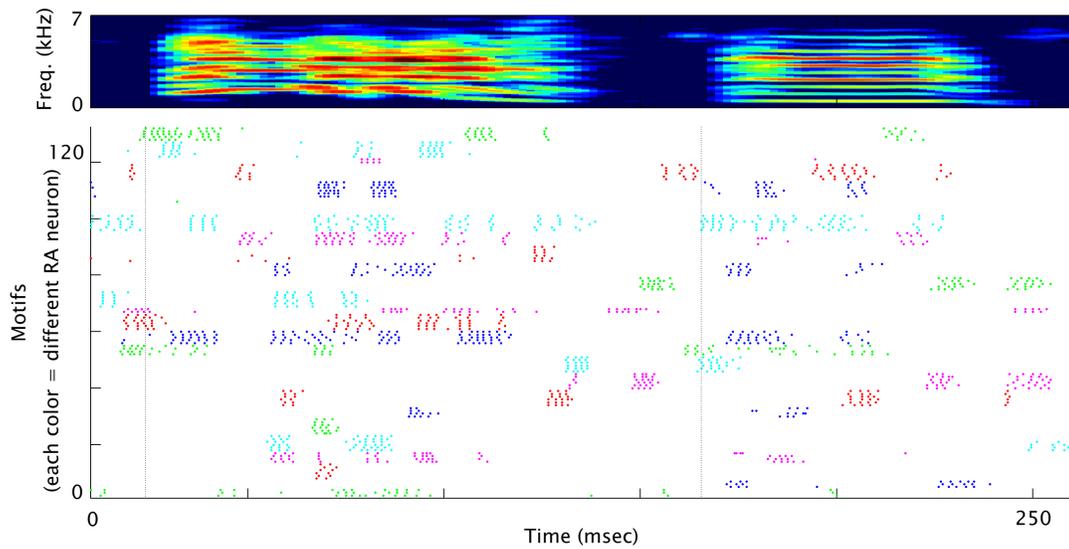


**Figure 2.8.** Conditional probability distribution for neural ensemble correlation. Left panel: each column represents the distribution of neural ensemble correlations, given the level of song correlation indicated on the x-axis. There is an abrupt transition, in which sounds that are nearly identical (correlations  $> 0.90$ ) are generated by highly similar patterns of neural activity, but below this threshold there is no longer any relation between the similarity of songs and the similarity of the neural ensembles which generated them. The bright spot in the top right corner of the matrix represents the correlation of neural ensembles generating the stuttered syllable shown in Figure 2.6. Right panel: expected neural ensemble correlation given the level of song correlation indicated on the x-axis. Note that in general, sounds with similar spectral structure are represented by completely different, uncorrelated patterns of neural activity, unless the two sounds are virtually identical as in the stuttering of a syllable (red dashed line).

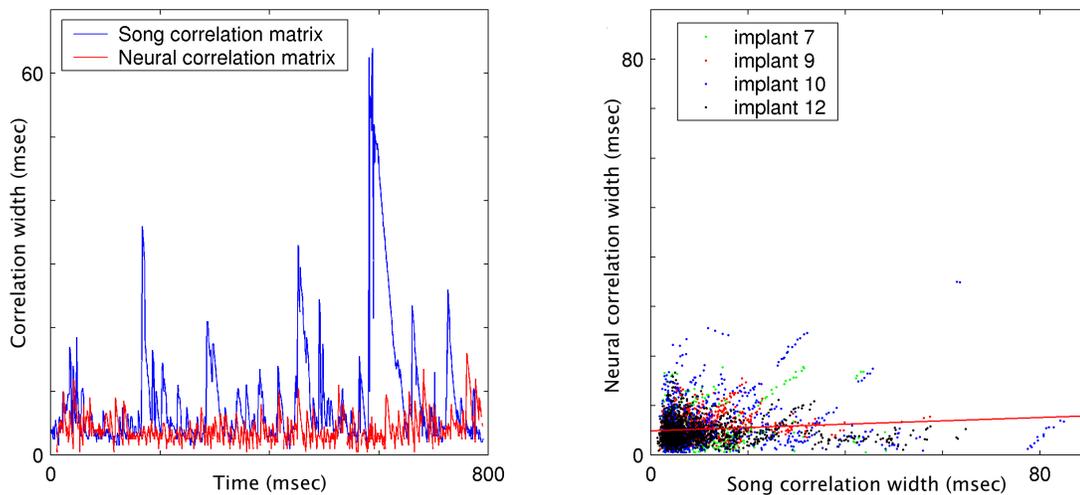
Given two highly similar sounds, the similarity of the patterns of neural activity that generated them may be examined by finding the corresponding locations in the song and neural correlation matrices (after a latency correction; see Methods). We quantified the relation between neural pattern structure and song spectral structure by estimating the conditional probability distribution for the correlation between two neural ensembles, given the level of correlation between the two sounds they generated. For example, consider all of the locations in the song correlation matrix that have a correlation between 0.95 and 0.96, indicating that they are highly similar sound pairs. The corresponding locations in the neural correlation matrix form a distribution of neural ensemble correlations that reflects how

similar the patterns of neural activity were that generated these similar sounds. A two-dimensional histogram of conditional probability distributions was obtained by calculating each neural ensemble correlation distribution as the song correlation stepped from  $-1$  to  $1$  (Figure 2.8a). Each column of the image shows the distribution of neural ensemble correlations, given a song correlation in a particular interval (e.g., 0.95 to 0.96). Figure 2.8b shows the means of the two-dimensional histogram; that is, the expected neural ensemble correlation as a function of song correlation. For virtually identical sounds, such as the stuttering of a syllable (Figure 2a, b) or the repetition of syllables in a second motif of song, the neural ensembles tended to be highly correlated. However, for song correlations less than 0.9, the neural ensembles became completely uncorrelated, despite the strong similarities that were still present between the sounds. In other words, although the bird was producing many similar sounds at different time points in the song, he generated each of them using entirely different ensembles of RA neurons. It is quite surprising that small variations in the spectral similarity of two sounds were not accompanied by small variations in the RA neural activity that generated them.

We used the correlation matrices to quantify the relation between the time scale of the spectral structure in the song and the time scale of the neural ensemble activity in RA. Some syllables, like chirps, contain rapidly changing time-dependent spectral structure, whereas others, like harmonic stacks, contain relatively constant spectral structure for their entire duration (Figure 2.9). Measurements of the properties of the vocal organs and their associated muscles (Fee et al., 1998; Goller and Suthers, 1995; Goller and Suthers, 1996) have shown that, in general, a constant sound like a harmonic stack is produced by a fixed syringeal configuration, whereas a rapidly changing sound is produced by a changing syringeal configuration. We analyzed the time scale of correlations in syllable spectral structure and neural ensemble structure by estimating the width of the correlation matrix diagonals (see Methods). The width of the correlation matrix diagonal reflects the length of time over which a sound or pattern of neural activity was constant. (Figure 2.10a). A scatter plot of neural ensemble widths as a function of corresponding song widths (Figure 2.10b) shows that song widths occurred over a range of values a factor of ten larger than those of the neural ensembles. A linear regression across the combined data sets of all four birds showed a correlation value of 0.04, and a slope of 0.08 ( $p < 0.0001$ ). The time scale of the neural



**Figure 2.9.** Rapidly and slowly changing sounds in zebra finch song. Top panel: time-frequency spectrogram for a complex syllable with fast transitions in spectral structure, and simple syllable with relatively constant spectral structure. Bottom panel: aligned spike train rasters; as in figure 2.5, different colored spike trains denote different RA neurons.



**Figure 2.10.** Constant acoustic structure is produced by rapidly changing neural ensembles. The width at which the correlation matrix diagonals dropped below a threshold was used to estimate the time scale of constant activity in the neural and song correlation matrices. Left panel: Width of the song correlation matrix (blue) and neural correlation matrix (red) as a function of time in the song, using a correlation threshold of 0.6. Right panel: scatter plot of neural correlation widths as a function of song correlation widths. Although the time scale of constant acoustic structure in the song was as long as 100 msec, the time scale over which patterns of neural activity in RA remained constant was only ~10 msec.

activity in RA was therefore not well correlated with that of the spectral structure of individual syllables, which ranged from ~10-150 msec, but was instead constant at a width of approximately 10 msec. This has considerable implications for the downstream decoding mechanisms that convert RA neural activity into sound.

### **2.3 Discussion**

How could rapidly changing, uncorrelated neural activity in RA produce constant, correlated vocal outputs? The anatomy of the song control system makes such a transformation conceptually straightforward. Motorneurons, which receive input directly from RA, could integrate the short burst sequences into analog control signals of arbitrary length and amplitude. In this model of the system, burst timing between neurons is the dominant feature. Consistent with this, the results of our analysis are unchanged regardless of whether we use analog or binary firing rates to define the RA neural ensemble activity. There is a tremendous convergence between RA and the vocal organs: approximately 10000 RA neurons [ref], of which 15% are active at any time, project, via the tracheosyringeal nerve, to approximately 10 syringeal muscles (Goller and Suthers, 1996). This suggests there is substantial redundancy between RA and the syrinx. Our analysis is consistent with this idea, as many different RA subpopulations are capable of producing the same vocal output. RA is organized myotopically, with respect to the syringeal muscles (Vicario and Nottebohm, 1988).

Recent experimental results in HVc (Hahnloser et al, 2002) have shed considerable light on the upstream mechanisms that may drive the uncorrelated patterns of neural activity that we observe in RA. During singing, the HVc neurons that project to RA fire at only a single time point in each motif. Each RA neuron, in contrast, tends to fire at 10-15 specific time points in the song, because it receives input from many HVc neurons, as well as from other RA projection neurons. Given that each HVc neuron fires only once during each motif of song, the random projections from HVc neurons to different subpopulations of RA neurons could account for the lack of correlation between RA neural ensembles at different time points in the song. Because there is no correlation in HVc activity over the course of a single motif of song, there is no correlation in RA activity. As the HVc projection neurons

also generate brief bursts of spikes, there is a short time scale of neural activity present in HVC that could drive the short time scale of activity we observe in RA.

Our description of the patterns of neural activity generated in RA during singing also provides an explanation for some recent behavioral observations on the process of song learning. Tchernikovski et al. (1999) have shown that rather than learning a few sound primitives (like harmonic stacks, frequency downsweeps, and whistles) and then refining these into mature syllables occurring in a particular sequence, extremely different syllables can in fact emerge from the same primitives. For example, consider a scenario in which the zebra finch is learning three sequential syllables, a harmonic stack (H), followed by a frequency downsweep (F), followed by a harmonic stack (H). When the bird begins the song learning process, he generates a sequence of sounds that randomly begin as F-H-H. The bird must learn both the detailed spectral structure of the three syllables and their proper order (H-F-H). Now, in one model of vocal control, the zebra finch could refine the spectra of each syllable and then switch their orders so they followed the proper sequence. In this scheme, each syllable is generated from a primitive that is similar to its final form. What Tchernikovski et al. (1999) observed, however, was that the zebra finch slowly transformed each initial syllable into its final target, simultaneously learning both the spectral structure and the sequence order. There was no direct correspondence between the final structure of a syllable and its starting point. Our results clarify the neural basis of the results of this learning process. There is no representation of syllable primitives in RA, such that different sounds of the same type (like harmonic stacks with different fundamental frequencies) are generated by correlated patterns of neural activity with subtle differences between them. Instead, the zebra finch learns to activate a different RA neural ensemble at each time point in the bird's song, independent of what acoustic structure is being produced. These different neural ensembles are slowly refined until they generate the desired spectral structure. Similar sounds are thus subserved by entirely different patterns of neural activity.

The tuning of single neuron responses to sensory or motor features (Georgopoulos et al., 1986; Lewis and Kristan, 1998; Wilson and McNaughton, 1993; Bialek et al., 1991) has dominated the classical view of neural coding. Within these systems, smooth variations in a stimulus feature typically produce smooth variations in the neural activity that represents it. For example, the direction of reach of a monkey can be decoded by a vector average of the

activities of individual M1 neurons with respect to their preferred directional tuning (Georgopoulos et al., 1986). A small change in the direction of reach will, according to the model, produce a small change in the activity and membership of the M1 neural ensemble. A second property of these types of neural representation is that adding more neurons, appropriately distributed across the encoded feature space, to the population estimate of the stimulus will generally produce a steady increase in the precision of the estimate. For example, in rat hippocampus, increasing the number of place cells used to estimate the rats spatial position produces a steady increase in the spatial resolution of the estimate (Wilson and McNaughton, 1993).

Our results describe a very different type of neural representation, one in which there is no systematic mapping between the distributed activity of the network and the vocal output of the bird. In contrast to neurons in the systems described above, RA neurons show no consistent tuning to a single acoustic feature across the set of time points in the song when they are active. Furthermore, small changes in spectral structure are subserved by the activation of completely different subpopulations of RA neurons, despite the strong correlations that are known to be present in the vocal muscles during the production of similar sounds. Finally, the addition of neurons to the RA neuronal ensemble does not yield a steady increase in predictability about what type of sound is being generated. One neuron may produce ten bursts during the song motif, each timed to a completely different type of sound. The knowledge that this neuron was active reduces the uncertainty about what sound is being produced to a set of ten different possibilities. Adding a second active neuron to the estimate reduces the uncertainty about what sound is being generated to the occasions when the two neurons fired together. This may reduce the initial ten locations to three, but there is not a corresponding increase in correlation between the remaining sound types. It is only when the neural ensemble has sufficient neurons that it becomes uniquely defined that accurate prediction of song structure is possible. Most models of neural population coding follow a sum-of-the-parts principle, in which the individual neurons provide considerable information about what is being represented, and additional neurons provide progressively more information. The neural code used within RA is substantially different from these models, and individual neurons provide little unambiguous information about song structure. These observations suggest that within the neural control system of RA, network interactions

are the dominant feature, and it is only by reading out sufficient activity to represent the entire state of the network that decoding is possible. The spike patterns we observe in RA thus represent a new type of neural code underlying complex learned behaviors

## 2.4 Methods

Birds were housed in custom designed plexiglass cages, and had unlimited access to food and water. All birds used in the experiment were male zebra finches (adults with crystallized songs). Females zebra finches were housed in similar, separate plexiglass cages, and were presented to the males upon isolation of one or more RA neurons (so-called directed song).

**Neurophysiology.** Birds were anesthetized with 1-2% isoflurane and nucleus RA was identified with an extracellular targeting electrode based on stereotaxic coordinates and electrophysiological activity. Upon identification of the center of RA, a three electrode miniature motorized microdrive was cemented onto the skull using the procedure described in Fee and Leonardo (2001). The microdrive weighs 1.5 grams, and contains three independently controlled motors, allowing each electrode to be remotely positioned extracellularly with 0.5  $\mu\text{m}$  spatial resolution. The electrode tips were implanted  $\sim 700$   $\mu\text{m}$  above RA. Electrodes were made from 80  $\mu\text{m}$  tungsten wire, insulated with parylene, and had  $\sim 3$  M $\Omega$  impedance (5-10  $\mu\text{m}$  tips; Microprobe, Inc.). Birds were allowed to recover for sufficient time so that they were singing reliably upon presentation of a female bird ( $\sim 1$ -2 days). During each day of recording, a custom modified Sutter MP-285 microdrive controller was used to position the electrodes in nucleus RA and record their depths. Singing was normal in birds with implanted drives; songs produced after the implant are identical in structure to those produced before the implant.

Extracellular recordings of single neurons of up to 10 mV in amplitude were isolated on individual electrodes; isolation quality was equivalent to that seen in a head-fixed anesthetized animal. Single neuron recordings were verified with two methods. First, individual spike waveforms of 1.5 msec in length were extracted from each raw electrode signal using a 3x RMS threshold, and then interpolated by a factor of 10 to remove sampling jitter. This produced a matrix of aligned spike waveforms. We calculated the singular value decomposition (SVD) of this matrix. The eigenvectors associated with the largest two eigenvalues of the SVD represent a subspace of the matrix, which contains most of the

variability of the spike waveforms. We projected the spike waveforms onto this two-dimensional subspace (a great reduction from the original 60 dimensional space) and looked for well-defined clusters of points. These clusters represent well-isolated neurons. As we are isolating single neurons, this clustering process was not used for spike-sorting, but rather as an reliable denoising mechanism to automatically remove the occasional contamination from a second neuron or electrical artifact. Single neuron isolations were further verified by confirming the presence of a spike refractory period in the inter-spike-interval histogram. The locations of the electrodes within RA were verified histologically at the conclusion of the experiment. Neural and acoustic data were collected by custom designed software written in Labview (National Instruments).

RA is organized myotopically, with respect to the syringeal muscles (Vicario and Nottebohm, 1988). It is possible that uncorrelated sets of neurons within a single myotopic projection (from RA to the hypoglossal nucleus) generate similar vocal outputs. However, the electrodes in our microdrive were spaced sufficiently that they sampled, relatively uniformly, the entire volume of nucleus RA. Within the population of neurons we recorded in each bird, some cells came from the same area of RA, and others from different areas. Because of this, the redundancy of any single subregion of RA cannot account for our observation that uncorrelated neural ensembles generate similar vocal outputs.

**Song alignment.** Variability from acoustic time-warping can be removed from the data by using highly resolved spectral features in different song syllables as clock points, and then linearly stretching and compressing common segments of sound in different motifs to be the same length. This piecewise-linear time-warping is based only on the song, and is completely independent of the spike trains. The partial derivative of the song spectrogram over time contains sharply peaked transitions at the points when the spectrum of the song changes abruptly (e.g., the beginnings and ends of syllables), that are suitable for use as alignment points. For each motif, we construct a vector of times corresponding to the occurrences of common spectral features, based on the alignment of the spectral derivatives of each syllable to a set of syllable templates. Adjacent time points in this vector define an interval of sound in the song. By stretching and compressing the corresponding time intervals in different motifs to be a common length, the variability from acoustic time-warping can be normalized out of all of the recorded motifs. Each spike train is then

projected onto the warped time-axis of its corresponding acoustic signal. The alignment templates are representative features for each syllable.

**Neural ensemble correlations.** A neural ensemble was defined as the spatial pattern of activity across the recorded population of neurons at a particular time point in the song. Each neural ensemble was represented by a vector of instantaneous analog firing rates for each neuron in the recorded population. Let  $N_i$  be the neural ensemble activity vector observed at time  $t_i$ , and  $N_j$  be the neural ensemble activity vector observed at time  $t_j$ . We defined the correlation between the two ensembles as

$$C(t_i, t_j) = \frac{N_i \cdot N_j}{\|N_i\| \cdot \|N_j\|}$$

The mean was subtracted from each neural ensemble activity vector prior to calculating its correlation to other ensembles, to ensure the measured correlations reflect coordinated fluctuations in the two ensembles and not common DC offsets. Correlations between two null ensembles (times at which no neurons in the population fired) were excluded from the analysis as they contained no data. We also used the Hamming distance to measure the neural ensemble correlations, without change in results. The neural ensemble activity was estimated in a 0.5 msec window in 0.5 msec steps.

**Song correlations.** Direct multitapered estimates of the power spectrum of each song were calculated with a sliding 8 msec window, using a step size of 0.5 msec, and a time-bandwidth product of  $NW = 2$  (Thomson, 1982; Tchernichovski et al., 2000). Sound correlations were calculated with the dot-product metric described above. However, strong correlations between different time points in the song were produced by the average power-spectrum of the syrinx and the oscillatory temporal pattern of syllables and silent intervals. These correlations obscured the fluctuations in fine spectral structure. Because of this, we first normalized out of the song the dominant spectral and temporal modes (essentially the average power spectra and the envelope).

**Conditional distributions.** In order to compare the song and neural correlation matrices, we first compensated for the 10 msec latency in the neural-to-song response (A. Kozhevnikov and M. Fee, unpublished measurements), and then smoothed the neural data with the same window function used to estimate the time-frequency spectrum of the song (8 msec). This ensured that the neural and acoustic data had the same spatial and temporal resolution. This

analysis was restricted to nonlocal effects – that is, events that occurred sufficiently off-diagonal that they were not influenced by the average song or neural ensemble width. Local syllable correlations were considered in the correlation width analysis.

The weakness of any analysis relating sound correlations to neural activity in the song control system is that there is no simple linear relation between acoustic structure in the song and the control parameters used by the bird to generate that structure. For example, the syrinx is known to be highly nonlinear in its response to linear control signals (Fee et al., 1998). Further, it is likely that the control parameters in RA are represented in motor coordinates, rather than auditory ones, since RA projects to motoneurons. However, despite these considerations, there are clear correlations in the activity of the syringeal muscles when similar sounds are generated (Goller and Suthers, 1996). Thus, regardless of the exact nature of the control parameters represented in RA, the main point of our analysis is that there is no correlation in the neural ensemble activity over time, despite the strong correlations in activity in the vocal control areas downstream of RA.

**Correlation width analysis.** For each song correlation matrix, we calculated a contour line with respect to the matrix diagonal. This contour line represents how much time elapsed before a particular sound at time  $t$  dropped below a correlation threshold. The same threshold was then used to calculate the time-varying width of the neural ensemble correlation matrix. A correlation threshold of 0.6 was found empirically to accurately reflect the length of stationary acoustic structure in the bird's song. The value of the minimum correlation widths seen in the scatter plot (Figure 4c) is a function of this correlation threshold. Large thresholds produce smaller correlation widths as the contour line becomes a tighter fit to the matrix diagonal (having zero width at a threshold of 1). However, our results (the slope of the line relating neural and song widths) are robust over a large range of threshold values, and are not affected by changes in the y-intercept of the scatter plot.