Feature Analysis of Natural Sounds in the Songbird Auditory Forebrain

KAMAL SEN,[1] FRÉDÉRIC E. THEUNISSEN,[4] AND ALLISON J. DOUPE[1–3]
[1]*Sloan Center for Theoretical Neuroscience,* [2]*Department of Psychiatry, and* [3]*Department of Physiology, University of California, San Francisco 94143-0444; and* [4]*Department of Psychology, University of California, Berkeley, California 94720-1650*

Received 19 January 2001; accepted in final form 7 May 2001

Sen, Kamal, Frédéric E. Theunissen, and Allison J. Doupe. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 86: 1445–1458, 2001. Although understanding the processing of natural sounds is an important goal in auditory neuroscience, relatively little is known about the neural coding of these sounds. Recently we demonstrated that the spectral temporal receptive field (STRF), a description of the stimulus-response function of auditory neurons, could be derived from responses to arbitrary ensembles of complex sounds including vocalizations. In this study, we use this method to investigate the auditory processing of natural sounds in the birdsong system. We obtain neural responses from several regions of the songbird auditory forebrain to a large ensemble of bird songs and use these data to calculate the STRFs, which are the best linear model of the spectral-temporal features of sound to which auditory neurons respond. We find that these neurons respond to a wide variety of features in songs ranging from simple tonal components to more complex spectral-temporal structures such as frequency sweeps and multi-peaked frequency stacks. We quantify spectral and temporal characteristics of these features by extracting several parameters from the STRFs. Moreover, we assess the linearity versus nonlinearity of encoding by quantifying the quality of the predictions of the neural responses to songs obtained using the STRFs. Our results reveal successively complex functional stages of song analysis by neurons in the auditory forebrain. When we map the properties of auditory forebrain neurons, as characterized by the STRF parameters, onto conventional anatomical subdivisions of the auditory forebrain, we find that although some properties are shared across different subregions, the distribution of several parameters is suggestive of hierarchical processing.

INTRODUCTION

To understand how sounds are heard and interpreted and ultimately influence an organism's behavior, it is important to investigate the processing of natural sounds. However, little is known about the neural encoding of natural sounds. This is partly because the majority of studies have used synthetic stimuli such as white noise or tones to characterize auditory processing (for a review, see Eggermont et al. 1983c). Although these studies have provided a wealth of information on the organization of the auditory pathway and on the response characteristics of auditory neurons, it has become increasingly clear that it is difficult to use this knowledge to predict the neural responses to complex natural sounds such as vocalizations (Eggermont et al. 1983b; Theunissen et al. 2000). This is particularly problematic for characterizing high-level auditory neurons that may be optimized to analyze natural sounds. An alternative and more direct approach is to characterize auditory neurons using these sounds.

Many natural sounds are structurally complex and contain both spectral and temporal correlations (Attias and Schreiner 1997; Nelken et al. 1999; Theunissen et al. 2000). Until recently, this posed a methodological problem for the systematic characterization of the stimulus-response function of auditory neurons with natural sounds. This is because the reverse correlation method that was used to estimate the spectral-temporal receptive field (STRF) assumed a stimulus ensemble free of spectral and temporal correlations (Aertsen and Johannesma 1981; Eggermont et al. 1983a). We recently extended the STRF method to overcome this limitation by taking into account the spectral and temporal correlations present in the stimulus ensemble (Theunissen et al. 2000). Our method corrects for the spectral and temporal correlations present in sounds by performing a weighted average of the stimulus around each spike using a mathematical operation that involves a de-correlation in frequency and de-convolution in time. In this study, we apply this extended method to investigate the processing of natural sounds in the birdsong system.

The birdsong system offers several advantages for studying the processing of natural sounds. Songbirds display a remarkable ability to process auditory information (for a review of the birdsong system and behavior, see Konishi 1985). At birth, songbirds are endowed with an inborn behavioral selectivity for the sounds of their own species (Marler 1991). Auditory information plays a critical role in song learning in juvenile songbirds and in song maintenance in adult birds and is an important component of many social behaviors in songbirds. For this highly sophisticated behavioral repertoire to be possible, a wide variety of natural sounds, especially songs, must be detected and discriminated by the auditory system of songbirds. Currently, the neural basis of these behaviors is poorly understood.

Anatomical (Fortune and Margoliash 1992; Kelley and Nottebohm 1979; Vates et al. 1996) and physiological (Janata and Margoliash 1999; Langner et al. 1981; Lewicki and Arthur

Address for reprint requests: K. Sen, Dept. of Physiology, Box 0444, University of California, 513 Parnassus Ave., San Francisco, CA 94143-0444 (E-mail: kamal@phy.ucsf.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1996; Mello and Clayton 1994; Muller and Leppelsack 1985) experiments suggest that auditory forebrain areas such as field L may contribute to the ability of songbirds to detect and discriminate a wide variety of complex natural sounds. In the anatomical chain of acoustical processing stages of the avian brain, the field L region lies between the thalamic auditory relay nucleus ovoidalis (Ov) and higher-level auditory areas such as HVc and the medial portion of the caudal neostriatum (NCM) (Vates et al. 1996) (Fig. 1). This location is analogous to the location of auditory cortex in mammals. As in the primary auditory areas of many other animals, field L in zebra finches and other birds displays a tonotopic organization (Bonke et al. 1979; Gehr et al. 1999; Muller and Leppelsack 1985; Zaretsky and Konishi 1976). Based on Nissl and Golgi staining studies, the field L region has been divided into 5 subregions called L2a, L2b, L1, L3, and L (Fortune and Margoliash 1992). Neuro-anatomical tracer studies have shown that the thalamic input from Ov projects strongly to area L2a and L2b and more weakly to L1 and L3. L2a projects strongly to L1 and L3, and all field L regions project to cHV (Fig. 1).

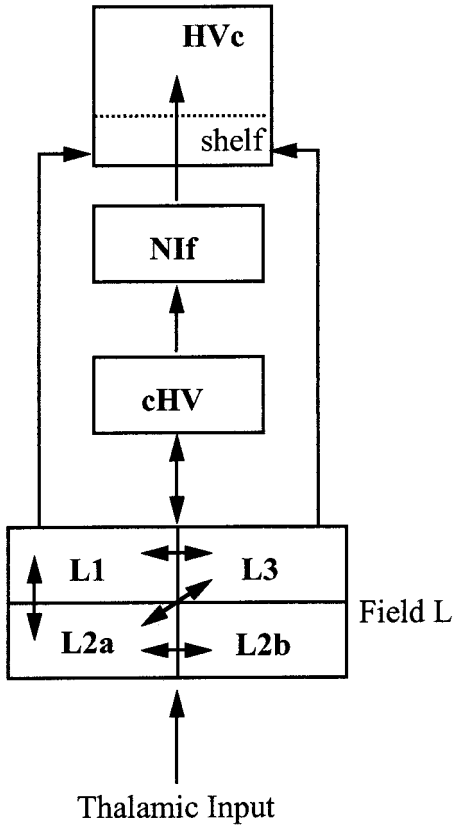

FIG. 1. Schematic of auditory forebrain connectivity. The figure illustrates the source of thalamic input to the auditory forebrain and connections between the different auditory forebrain region (Vates et al. 1996). Based on Nissl and Golgi staining studies, the field L region has been divided into five subregions called L2a, L2b, L1, L3, and L (Fortune and Margoliash 1992). Subregion L2a receives a strong thalamic projection, whereas subregions L1 and L3 receive weaker thalamic projections. L2b and L were considered as a composite region (see METHODS). Part of this composite region also receives a strong thalamic projection. Subregions in field L are reciprocally connected to other subregions in field L, as shown (↔), and to the overlying area cHV. Possible sources of auditory input to high level area HVc are directly from L1 and L3 through the underlying shelf region and indirectly from field L through areas caudal hyperstriatum ventrale (cHV) and nucleus interfacialis (NIf).

In zebra finches, the stages of auditory processing in field L and other auditory forebrain areas are also likely to contribute to the response properties of "song-selective" neurons found in high level auditory areas such as HVc, since these areas are the primary source of sensory input to HVc. Song-selective neurons, which respond more strongly to the bird's own song (BOS) than to even very similar auditory stimuli, have been well characterized in a number of studies (Margoliash 1983, 1986; Margoliash and Fortune 1992; Mooney 2000; Theunissen and Doupe 1998; Volman 1993). However, the earlier stages of auditory processing that may participate in the generation of such highly selective neurons have only begun to be explored (Janata and Margoliash 1999; Lewicki and Arthur 1996).

So far, a systematic study of the stimulus-response function of auditory forebrain neurons has not been undertaken with natural sounds. Thus several interesting questions remain to be addressed. To what features of natural sounds do auditory forebrain neurons respond? What are the characteristic spectral and temporal parameters of such features? Do the distributions of parameters indicate the emergence of increasingly complex features in the auditory forebrain? In this paper, we address these questions by obtaining the STRFs for auditory forebrain neurons using a large ensemble of conspecific songs (CONs) and extracting several parameters from the STRFs to assess multiple aspects of the processing of songs in the auditory forebrain.

METHODS

Electrophysiology

All physiological recordings were done in urethan-anesthetized adult male zebra finches in acute experiments. Extracellular waveforms were obtained using parylene-coated tungsten electrodes (resistance 1–3 MΩ) that were inserted into the neostriatum of the bird at locations that were previously marked with stereotaxic measurements. The extracellular waveforms were transformed into spike trains, using a window discriminator, by windowing the largest action potential. Waveforms from successive spikes in the window were examined on a fast time base to estimate the number of units. Cases where the waveform had a single reliable and stereotyped spike shape were classified as single units. Multiunit recordings consisted of spike waveforms that could be easily distinguished from background activity but not from each other. Single units (18/62) or small multiunit clusters consisting of two to five neurons (44/62) were recorded in this manner. We did not observe any significant differences in our results for these two groups (see RESULTS). At the end of the experiment, the bird was deeply anesthetized and transcardially perfused. The locations of the recordings were verified histologically in Nissl-stained brain sections. The location of the sites was classified into anatomical subregions of field L as described in Fortune and Margoliash (1992). We considered L and L2b as a single composite region since no clear border between these two regions was apparent. We will refer to this composite region as L2b. The data presented here were obtained from 10 birds and 62 recording sites (6 in L2a, 21 in L2b, 13 in L1, 16 in L3 and 6 in cHV). (For a more detailed description of recording methods, see Theunissen and Doupe 1998; Theunissen et al. 2000.)

Stimuli

An ensemble of 20 conspecific songs, previously used in Theunissen et al. (2000), was used to obtain neural responses in the auditory forebrain of each bird. The same set of conspecific songs was used in

all our experiments. For each bird, we added the BOS to this ensemble giving a total of 21 songs. Stimuli were played at a peak intensity of 80 dB SPL and randomly interleaved to obtain 10 trials of responses to each song in the ensemble. The average song duration was 2.1 s. All of the songs, including BOS, were used to compute robust estimates of stimulus ensemble properties such as the power spectrum and autocorrelation matrix, as previously described in Theunissen et al. (2000). We did not separately characterize the STRFs in response to the BOS in this study, since a reliable estimate of the STRF (see following text) requires much more data than we had for the BOS alone. Moreover, this calculation would also lead to significant methodological difficulties, because the BOS alone samples only a very small part of stimulus space (Theunissen et al. 2000). To compare the responses of neurons to BOS versus the other songs in the ensemble, we used the d′ measure of selectivity, previously used to quantify song selectivity in other areas of the song system (Janata and Margoliash 1999; Theunissen and Doupe 1998). We did not detect a difference in response to BOS over the other songs ($P = 0.4$, 1 sample sign test). Our observation is consistent with previous studies, which have found the majority of field L neurons to be unselective for the BOS compared with other conspecific songs or manipulations of the BOS such as reversed BOS and syllable order reversed BOS (Janata and Margoliash 1999; Lewicki and Arthur 1996).

STRF calculation

A detailed description of the calculation of STRFs from natural sounds can be found in Theunissen et al. (2000). This is briefly summarized here. We used an invertible spectrographic representation of sound in which sound is first decomposed by passing it through a set of Gaussian filters of 250 Hz width (SD) spanning center frequencies between 250 and 8,000 Hz. The sound is then represented by a set of functions of time $s_{\{i\}}(t)$, where $s_i(t)$ is taken to be the log of the amplitude envelope of the signal in the frequency band i. The STRF is defined as the multi-dimensional linear Volterra filter $h_{\{i\}}(t)$ such that

$$r_{\text{pre}}(t) = \sum_{i=1}^{\text{nf}} \int h_i(\tau) s_i(t - \tau) \mathrm{d}\tau$$

where $r_{\text{pre}}(t)$ is the predicted firing rate and nf is the total number of frequency bands. $h_{\{i\}}(t)$ is found by requiring that $r_{\text{pre}}(t)$ be as close to possible as $r_{\text{est}}(t)$, the estimated firing rate obtained from a peristimulus time histogram (PSTH). In the frequency domain, the solution for the set of $h_{\{i\}}$ for each frequency w, can be written in vector notation as

$$\vec{H}_w = A_w^{-1} \cdot \vec{C}_w$$

where A_w is the stimulus autocorrelation matrix and C_w is the cross-correlation between the spike trains and the stimulus amplitudes in each band. The normalization of the cross-correlation by the stimulus autocorrelation matrix corrects for the spectral-temporal correlations in the stimulus. A detailed description of the numerical parameters used in our calculations can be found in Theunissen et al. (2000). (For a recent extension of this method to visual neurons, see Theunissen et al. 2001.)

To determine the significance of regions in the STRFs obtained, we used a jackknife resampling method where STRFs were calculated for multiple subsets of the conspecific song ensemble that were obtained by deleting one song at a time from the complete ensemble. The variance for each spectral-temporal bin in the STRF estimate was calculated from this set of STRFs using the jackknife formula

$$\text{Var} = \frac{n-1}{n} \sum_j [\theta(j) - \theta(.)]^2$$

where $\theta(j)$ is the value when the jth song is deleted, $\theta(.)$ is the average of all $\theta(j)$, and n is the number of songs in the ensemble. The standard error was obtained by taking the square root of the variance.

Figure 2*A* shows the raw STRF obtained from a site in L2a, and Fig. 2*B* shows the jackknife standard error for this STRF.

To display the significant part of the STRF, we first estimated the noise level in the raw STRFs using a singular value decomposition (SVD) technique. The SVD decomposes the STRF into a weighted sum of a number of terms, each of which is an outer product of a function of time and a function of frequency. The weights corresponding to each of the terms are the singular values obtained from the SVD. For an ideal, completely noise-free STRF the nonzero singular values can be used to reconstruct the STRF without any loss of information. In practice, due to noise in the estimation of the STRF, the singular values do not drop abruptly to zero but tail off gradually. We therefore compared the SVD obtained from a window (width, 100 ms) containing all of the structure in the STRF to the SVD obtained from a window representing noise (a 100-ms window from the acausal portion of the STRF corresponding to stimulus following spikes). The singular values obtained from the raw STRF that exceeded the maximal singular value obtained from the noise were used to reconstruct the STRF (Fig. 2). We found that this method effectively filtered out the noise in the raw STRFs. Then, to illustrate the significance of the different regions of the STRF, we show the contours for one and two times the significance level superimposed on these reconstructed STRFs. As a conservative estimate, we defined the significance level to be the maximal jackknife standard error for the STRF.

Parameters describing STRFs

We obtained several parameters from each STRF characterizing its temporal and spectral properties. Similar parameters have been obtained from STRFs in the auditory (Depireux et al. 2001; Hermes et al. 1981, 1982; Keller and Takahashi 2000; Kim and Young 1994) as well as visual (Cai et al. 1997) domains. The time to peak (T_{peak}) was defined as the time to the absolute maximal value of the STRF. We also used the STRF to directly estimate the temporal characteristics of each neuron's processing of amplitude envelopes of songs. We call this parameter the best modulation frequency (BMF). To obtain the BMF, we took a slice through the maximal value of the STRF along the temporal dimension and obtained the peak of the power spectral density of this slice (Fig. 6*D*). The power spectral density was estimated using a fast Fourier transform with a Hanning window. As defined here, this measure may differ from the conventional BMF, which is obtained from neural responses to simple amplitude modulated tone bursts, using a range of AM frequencies. To quantify more spectral characteristics of neural responses, we took a slice through the maximal value of the STRF along the frequency dimension to obtain the peak frequency (CF) and a width at half-maximum (W). We used a quality factor, defined as the peak frequency divided by the width, $Q = \text{CF}/W$, as a measure of sharpness of spectral tuning of the largest spectral peak. The excitatory and inhibitory peak amplitudes were the maximal and minimal values of the STRF, respectively.

We also used the SVD of the STRF to assess the degree of the time-frequency separability of the STRF. Similar methods have been used in the visual system to describe the space-time inseparability of the spatio-temporal receptive fields of visual neurons (De Valois and Cottaris 1998; Jagadeesh et al. 1997; Kontsevich 1995) and the frequency-time inseparability of auditory neurons (Depireux et al. 2001). By definition, a separable STRF can be expressed as a single product of a function of time and a function of frequency. Thus for an ideal separable STRF, only one of the singular values obtained from the SVD should be nonzero. An index of separability could therefore be defined as the magnitude of the leading singular value relative to

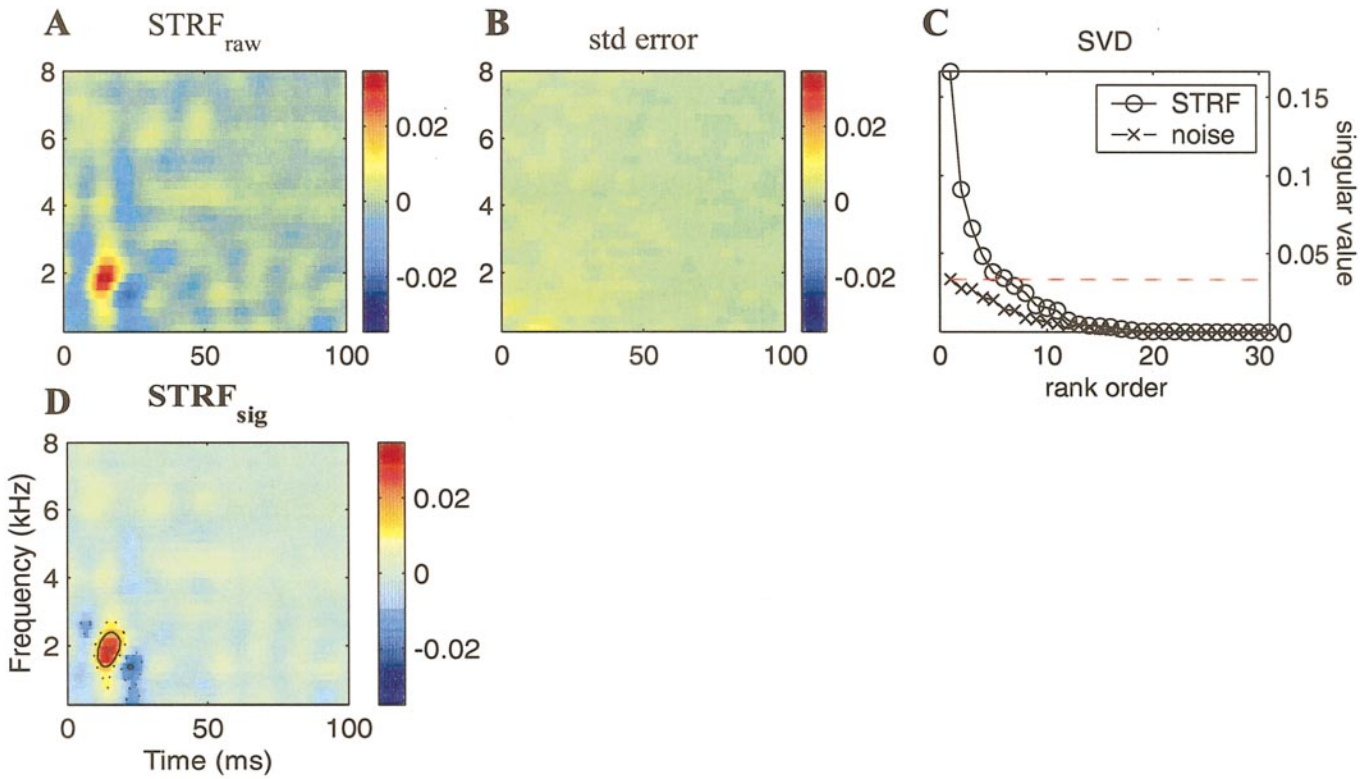

FIG. 2. Example of a spectral temporal receptive field (STRF) and significance calculation. *A*: the raw STRF in subregion L2a (site 14.10; the number before the decimal indicates the bird, and the number after the decimal the neuronal site) obtained from the neural responses to a large ensemble of conspecific songs. *B*: the SE of the STRF estimate calculated using a jackknife resampling method (see METHODS). *C*: the singular values obtained from the STRF (○) and the noise (×; see METHODS). *D*: the STRF reconstructed using singular values above the noise level shown in the - - - in *C* (see METHODS). To display the level of significance for the different regions of the STRF, we show the contours corresponding to 1 and 2 times the significance level (- - - and —, respectively; see METHODS for definition) superimposed on the STRFs and plot the regions below significance in lighter colors.

the sum of all the singular values. To avoid the effects of the noise tail in the singular values in assessing the separability of the STRFs, we defined a separability index SI as follows

$$SI = \frac{s_1}{\sum_{i=1}^{n-1} s_i}$$

where s_i is the ith singular value with all singular values measured relative to the singular value s_n corresponding to the noise level. We chose $n = 4$ because in all cases the first three singular values accounted well for the STRF structure.

Prediction of responses

The method for obtaining a prediction of neural responses using the STRF is described in detail in Theunissen et al. (2000) and only briefly summarized here. The predicted firing rate was obtained by convolving the STRF with the stimulus and rectifying and scaling the result to minimize the squared error between the predicted rate and the firing rates estimated from the actual data. To obtain the predicted firing rate for each song, we used the STRF calculated from all songs in the ensemble except for the song used to generate the stimulus-response data being tested. We quantified the quality of the prediction by calculating the cross-correlation coefficient (CC) between the predicted and estimated firing rates. The measured firing rate was obtained by smoothing the PSTH (but not the predicted firing rate) with a Hanning window that gave the maximal CC. We corrected the CC for bias and obtained the standard error for the CC using a jackknife resampling method.

RESULTS

The goal of this study was to investigate the processing of natural sounds in the songbird auditory forebrain. We began by systematically characterizing the stimulus-response function of auditory forebrain neurons in response to natural sounds. We obtained STRFs from the responses of auditory forebrain neurons in adult male zebra finches to a large ensemble of zebra finch songs. These STRFs show the spectral-temporal features of songs to which auditory forebrain neurons respond and describe the optimal linear component of the response to songs. By extracting a variety of parameters from the STRFs, we were able to quantify several aspects of the processing of natural sounds in the auditory forebrain. First, we obtained multiple STRF parameters to describe the spectral and temporal properties of features important to forebrain auditory neurons. Second, we characterized the spectral-temporal separability of the STRFs. Third, we assessed the linearity versus nonlinearity of the neuronal encoding of songs by quantifying the quality of response predictions obtained from the STRF model. Finally, to begin to assess the relationship between functional properties of auditory forebrain neurons and conventional anatomical subdivisions of auditory areas, we examined how the STRF parameters in our data set mapped onto different subregions of the auditory forebrain.

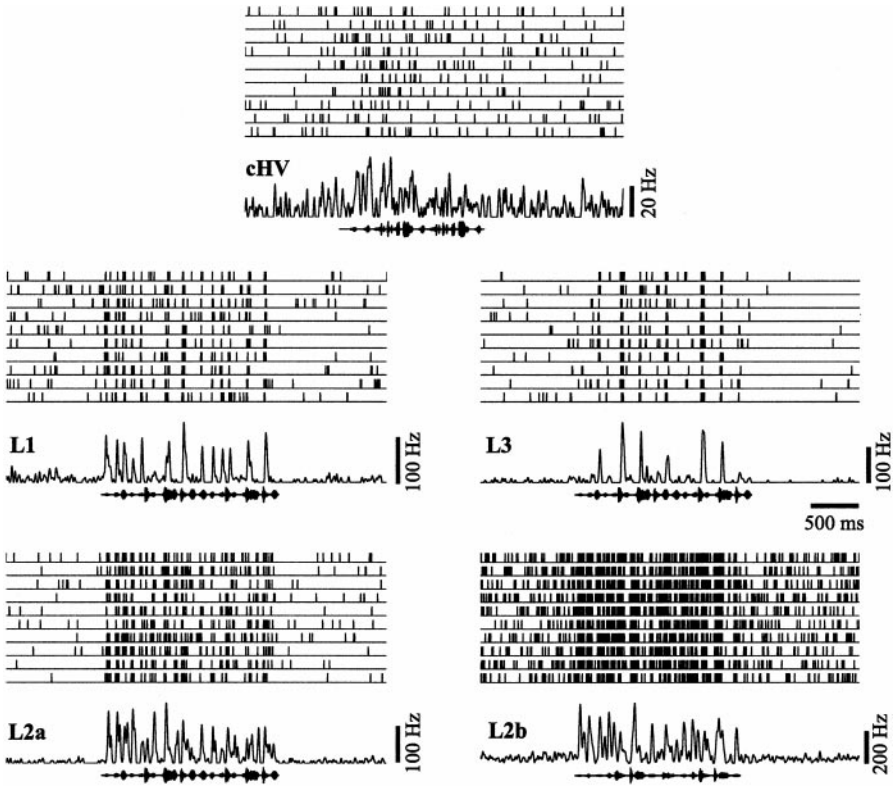

FIG. 3. Neural responses from the auditory forebrain. The figure shows examples of spike rasters and peristimulus time histograms (PSTHs) of field L and cHV neurons in response to zebra finch songs. The oscillograms of the songs, a representation of the sound pressure level as a function of time, are shown below the PSTH. *Bottom*: responses from sites in subregions L2a (site 21.2, song zfa-24.5) and L2b (site 20.4, song zfa-25.6) in field L. *Middle*: responses from sites in L1 (site 18.6, song zfa-24.5) and L3 (site 14.12, song zfa-24.5). *Top*: responses from region cHV (site 27.2A, song uc-11). The examples from L2a, L2b, and L1 are data from single units, and the examples from L3 and cHV are data from small clusters of neurons (see METHODS). We also compared the mean firing rates across the subregions (although here a caveat is that our data included both single units and small clusters of units; see METHODS and RESULTS). Overall, subregions in field L had relatively high mean firing rates above the background firing rate, with L2a being the highest (15 ± 4 spikes/s), followed by subregions L1 (11 ± 4 spikes/s), L2b (9 ± 2 spikes/s), and L3 (8 ± 2 spikes/s). In comparison, mean rates were lower in area cHV (3 ± 1 spikes/s). The difference in the mean firing rates between areas was not statistically significant.

Neural responses

We obtained neural responses from throughout the auditory forebrain including subregions L2a, L2b, L1, and L3 of field L and the overlying region of cHV. Figure 3 illustrates examples of the trial-by-trial and average neural responses, one from each of the five subregions. As can be seen, the sites in all subregions of field L responded strongly to songs. The response in the site from cHV was weaker and more variable in comparison to field L. The average firing rate in the auditory forebrain was 9 ± 1 (SE) spikes/s.

STRF

Songs have a highly complex spectral-temporal structure including strong time-varying correlations across different frequencies. Consequently, as illustrated in Fig. 4, *A* and *B*, it is difficult to assess to what spectral-temporal features of songs neurons respond, simply by comparing the song, in its spectrographic representation (Fig. 4*B*), and the neural response (Fig. 4*A*). The STRF method addresses this difficulty by analyzing the stimulus preceding each neural response, for many stimuli and many spikes, and calculating what weightings of the spectral and temporal components of the stimuli produce the best linear estimate of the actual neural response [for the mathematical definition of the STRF used in this paper, see METHODS and Theunissen et al. (2000); for discussions on the interpretation of the STRF, see Eggermont et al. (1983c); Klein et al. (2000); Theunissen et al. (2000, 2001)]. The resulting STRF can be thought of as a filter that characterizes the linear component of the stimulus-response function of auditory forebrain neurons and that can reveal the features of song critical to the neuronal response. The relationship between the STRF and the neural response to a particular stimulus can be seen by sliding a window (Fig. 4*B*) containing the time-reversed STRF (Fig. 4*C*) over the stimulus and obtaining a moment to moment prediction of the response. In each window, the stimulus is weighted by the overlapping part of the STRF, point-wise at the corresponding time and frequency, and the results from all points in the window are summed to obtain the predicted response (Fig. 4*D*). Mathematically, this is performing a convolution operation. Intuitively, the time-reversed STRF can therefore be thought of as the most effective stimulus that could drive this neuron, if the neuron was completely linear. In this example drawn from our data from region L2a, the STRF, which has a relatively simple structure, provides a good prediction of the neural response to a very complex auditory stimulus (the goodness of the linear STRF model is quantified and discussed in *Linearity versus nonlinearity*).

Feature analysis of songs by auditory forebrain neurons

Figure 5 shows 15 examples of STRFs obtained from the auditory forebrain (3 from each of the different subregions in field L and cHV), which illustrate the range of STRFs we observed in our data. As can be seen, the STRFs in Fig. 5, *A* and *D* (subregions L2a and L2b, respectively), indicate sensitivity to a simple, narrowband component of song. In contrast, much more complex features are observed in some other examples (Fig. 5, *F*, *I*, *K*, *L*, and *O*). For instance, the STRF in Fig. 5*L* (subregion L3) shows an excitatory-inhibitory component that reverses in time, and the STRF in Fig. 5*O* (subregion cHV) shows a multi-peaked frequency stack. Figure 8*D* (subregion L3) shows another STRF with a complex feature, a frequency sweep. A further observation that can be made from Fig. 5 is the difference in the time-scales of the STRFs. The STRF in Fig. 5*A* from L2a has a short delay and width. In

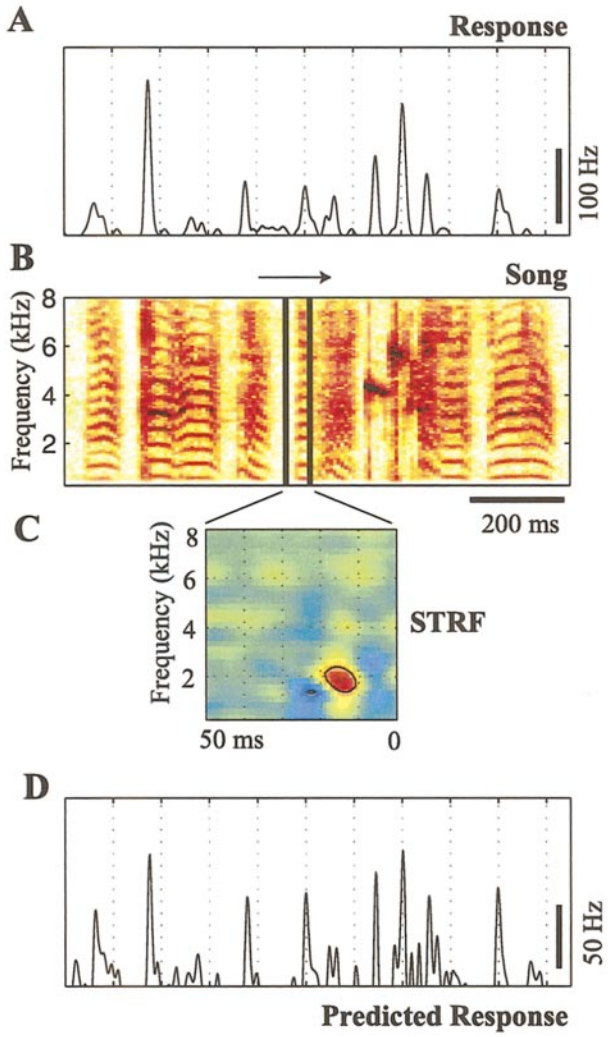

FIG. 4. Illustration of the STRF model. The STRF can be thought of as a linear filter that specifies how spectral and temporal components of the stimulus are weighted to produce the response. *B* shows a spectrographic representation of a section of 1 of the songs in our ensemble. A window (shown in the black rectangle) containing the time-reversed STRF (4C) is slid over the song. The overlapping parts of the song and the STRF are multiplied pointwise and summed together to obtain the prediction of the response to the song (see METHODS). The prediction (*D;* after rectification) can be compared with the actual response (*A*) to assess the goodness of the STRF model (see section on linearity vs. nonlinearity in RESULTS and Fig. 9).

contrast, the STRFs in Fig. 5, *L* and *O*, from L3 and cHV show longer delays and are extended over much longer durations. Collectively, these STRFs illustrate the variety of ways in which songs are analyzed by neurons in the auditory forebrain and the wide range of time scales associated with this analysis. In the following sections, we quantify some of these qualitative observations by examining a variety of parameters describing different aspects of the STRFs.

STRF parameters

To characterize some of the spectral-temporal properties of the STRFs and to quantify the differences between the STRFs in different subregions, we first extracted several simple parameters from each STRF. Such parameters have previously been used to characterize the response of auditory neurons to simple sounds such as white noise or tone pips. However, we extracted these parameters directly from the STRFs obtained with natural sounds to quantify several aspects of the processing of these sounds. This was an important step since we had previously observed that the STRFs obtained from natural sounds could be dramatically different from the STRFs obtained from simple stimuli for auditory forebrain neurons (Theunissen et al. 2000).

Figure 6 illustrates the parameters for a particular STRF in subregion L2a. As illustrated in the figure, the parameters are obtained from the spectral and temporal slices of the STRF taken along its maximal value. We obtained the time to peak (T_{peak}), which is a measure of delay between the stimulus and response (Fig. 6*B*); the Q factor, defined as the ratio of the best frequency to the width at half-maximum, which is a measure of the sharpness of spectral tuning of the largest spectral peak (Fig. 6*C*); the BMF, which is defined as the frequency corresponding to the peak of the power spectral density and is a measure of the frequency of amplitude modulations (AM) in songs that drive neurons best (Fig. 6*D*); and the ratio of the excitatory and inhibitory peak amplitudes of the STRF (see METHODS for definitions). Figure 5 can be used to illustrate how the values of these parameters correspond to the particular STRF from which they were obtained. For example, Fig. 5*A* shows an STRF that has a short delay with a T_{peak} of 11 ms, whereas the STRF shown in Fig. 5*N* has a much longer delay with a T_{peak} of 55 ms. The STRFs in Fig. 5, *A* and *L*, have BMF values of 70 and 10 Hz, indicating preferences for relatively high and low modulation frequencies, respectively. An example of an STRF that has relatively sharp spectral tuning with a *Q* value of 3.2 is shown in Fig. 5*B*, whereas Fig. 5*C* shows an STRF that has a more broadly tuned spectral peak with a *Q* value of 0.71.

Figure 7 shows the distribution of these parameters for the auditory forebrain and quantifies the diversity of processing of songs in the auditory forebrain, confirming our qualitative observations in Fig. 5. Although T_{peak} (Fig. 7*A*) ranged from 7 to 55 ms, the majority of the sites we examined fell into an intermediate range, consistent with the location of the auditory forebrain between the auditory thalamus and HVc. The distribution of the values for BMF (Fig. 7*B*) shows that the majority of sites (~90%) in our data set preferred relatively lower frequency AM in songs (<30 Hz). Almost half the sites in our data (~48%) had a *Q* factor close to 1 (between 0.5 and 1.5), indicating that for many sites the width of the largest spectral peak of the STRF was comparable to the peak frequency (Fig. 7*C*; also see Fig. 6 and METHODS). The ratio of excitatory and inhibitory peaks of the STRFs (*E-I* ratio; Fig. 7*D*) was distributed around a peak value at 1.3, indicating an approximate balance between the relative magnitudes of the excitatory and inhibitory peaks within a range around this value.

Our data consisted of both single units as well as small clusters of units (see METHODS). Although, in theory, if individual neurons close to each other differed markedly, complex STRFs could be created simply by the simultaneous recording of single units with different properties, we saw no evidence suggesting that this was occurring. The range of complexity of STRFs from single units was similar to that seen with the small clusters (examples of STRFs obtained from single units are shown in Figs. 5, *A*, *D*, *G*, *K*, and *L*, and 8*A*). Moreover, we did not observe a significant difference between single units and

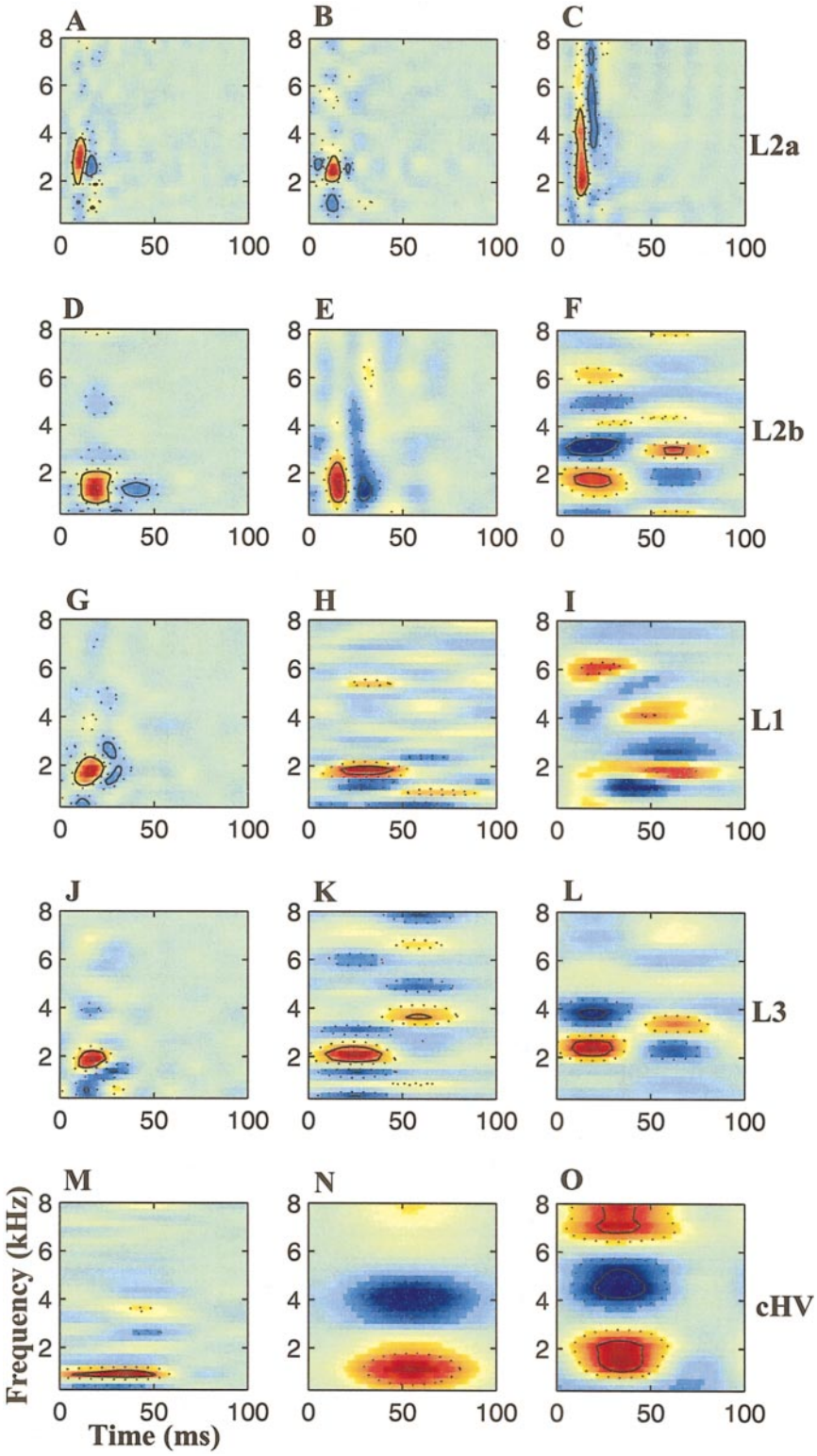

FIG. 5. Examples of STRFs from different regions in the auditory forebrain. 3 examples from each subregion are shown. *A–C*: examples from subregion L2a (sites 21.2, 14.11, and 27.4A). *D–F*: examples from subregion L2b (sites 20.4, 26.2B, and 23.3B). *G–I*: examples from subregion L1 (sites 18.6, 18.7, and 14.9). *J–L*: examples from subregion L3 (sites 25.2, 27.5B, and 27.4B). *M–O*: examples from subregion cHV (sites 14.2, 27.1A, and 27.2A). The examples in *A*, *D*, *G*, *K*, and *L* were from single units (see METHODS). The STRFs display the spectral-temporal features of songs to which auditory forebrain neurons respond. This figure illustrates the diverse range of features to which neurons responded in the auditory forebrain. These range from simple features showing narrowband components of song, as in the examples in *A* from subregion L2a and *D* from subregion L2b, to more complex multi-peaked features as in *L* from subregion L3 and *O* from subregion cHV. The figure also shows the wide range of time scales of the features that can be found in the auditory forebrain.

clusters for any STRF parameter ($P = 0.8$ for T_{peak}; $P = 0.9$ for BMF and Q; $P = 0.5$ for *E-I* ratio; Wilcoxon rank sum test).

Separability versus inseparability

A parameter that describes the complexity of STRFs is the degree of separability in time and frequency. Separable features can be described as a product of a spectral and temporal function, whereas inseparable features cannot be described in this simple manner. Using the singular value decomposition technique (SVD; see METHODS and Fig. 2), we analyzed the separability of song-derived STRFs and defined a separability index (SI) ranging from 0 to 1, with 1 indicating a fully separable STRF. We observed both separable and inseparable

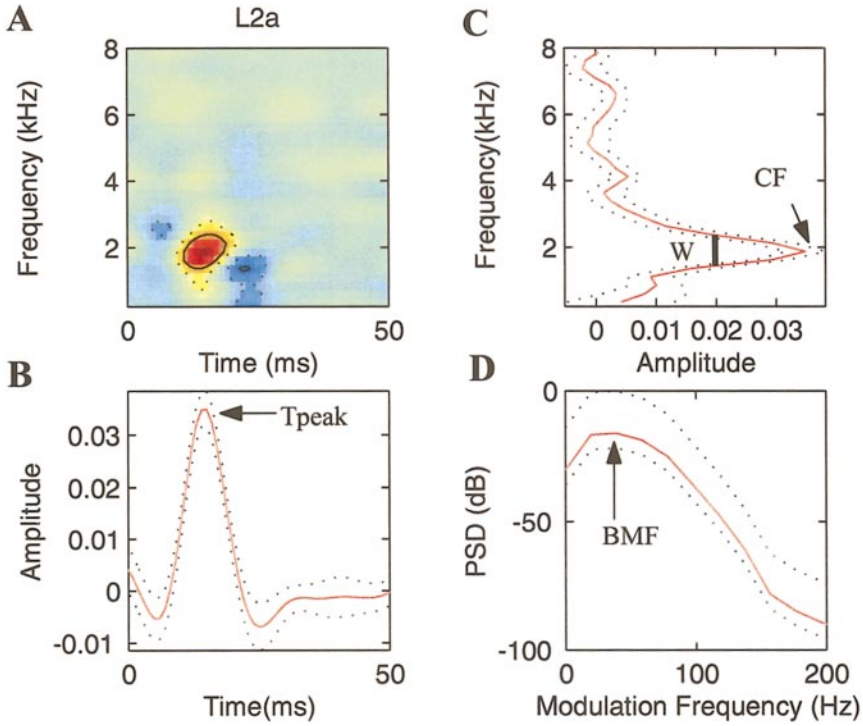

FIG. 6. Illustration of parameters of the STRFs. *A*: the STRF from Fig. 2. *B*: illustration of the time to peak (T_{peak}) parameter, which was obtained from a slice along the temporal axis passing through the maximal point of the STRF. *C*: illustration of the peak (CF) and width (*W*) obtained from a slice along the spectral axis passing through the maximal point of the STRF. The quality factor, *Q*, which is a measure of the sharpness of spectral tuning, is defined as CF/*W*. *D*: the power-spectral density (PSD) of the temporal slice in *B*. The frequency corresponding to the peak of the PSD indicated (↑) is defined as the best modulation frequency (BMF).

STRFs in the auditory forebrain. Figure 8*A* shows an example of an approximately separable STRF from subregion L2b. Figure 8*B* shows the STRF obtained using only the first component of the SVD of this STRF, and Fig. 8*C* shows the difference between this first component and the full STRF. As can be seen, the first component accounts for most of the structure of the full STRF, and thus this STRF is separable. This STRF had a SI of 0.91. In contrast, Fig. 8, *D–F*, shows an example of an inseparable STRF from subregion L3, which contains a frequency sweep. Unlike the separable STRF in Fig. 8, *A–C*, the difference (Fig. 8*F*) between the leading component and the full STRF is much larger in this case, and this STRF had a SI of 0.53. Figure 5, *A* and *K*, shows additional examples of STRFs with relatively high and low separability indices (SI = 0.82 and 0.52, respectively). Figure 8*G* shows the broad distribution of SIs obtained from the auditory forebrain for our entire data set. We did not observe a significant difference between the SI distributions of single units and small clusters of units in our data ($P = 0.3$).

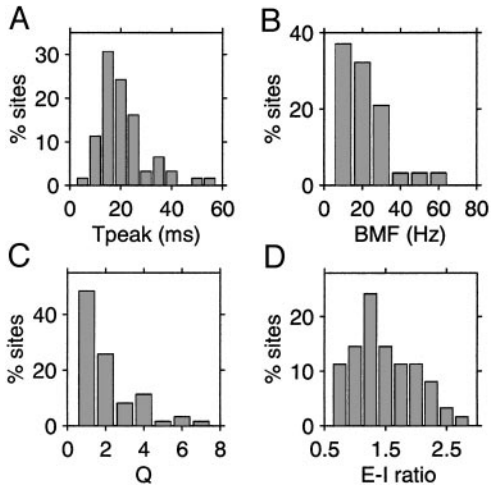

FIG. 7. Distribution of parameters of the STRFs in the auditory forebrain. *A*: distribution of T_{peak}. *B*: distribution of BMF. *C*: distribution of *Q*. *D*: distribution of the ratio of excitatory and inhibitory peaks (*E-I* ratio). See Fig. 6 and METHODS for definitions of parameters and Table 1 for a summary.

Linearity versus nonlinearity

The STRF is a linear model in that it describes only the linear component of the neural encoding of the stimulus. Thus one can use the quality of the predictions of the neuronal responses obtained from the linear STRF model to assess the linearity or nonlinearity of the neural encoding of the stimulus. We used the STRFs to obtain predictions of the neuronal responses to songs (see METHODS) and quantified the quality of the prediction by the correlation coefficient (CC) between an estimation of the deterministic part of the actual response and the response predicted by the STRF (see METHODS) (see also Theunissen et al. 2000). Figure 9*A* shows the estimated response from the actual data (*top*) and predicted response (*bottom*) using the STRF shown on the right of the traces, to a section of the stimulus ensemble for a site in L2a. For this site, a relatively good prediction could be obtained (CC = 0.68), indicating that a substantial component of the encoding of this site was linear. However, this linear component varied over a wide range for our data set (range of CCs: 0.07–0.72), indicating both relatively linear as well as nonlinear encoding of songs. Illustrative examples with different values of CC are shown in Fig. 9, *B–E*. These examples illustrate the range of performance of the linear STRF model in being able to predict the neural response. For example, in Fig. 9, *A* and *B*, the timing and widths as well as the relative amplitudes of the peaks and troughs in the responses appear to be well predicted. Figure 9*C* shows an example where the timing and width of the responses are still relatively well predicted but the STRF fails to capture the relative amplitudes of the peaks and troughs in

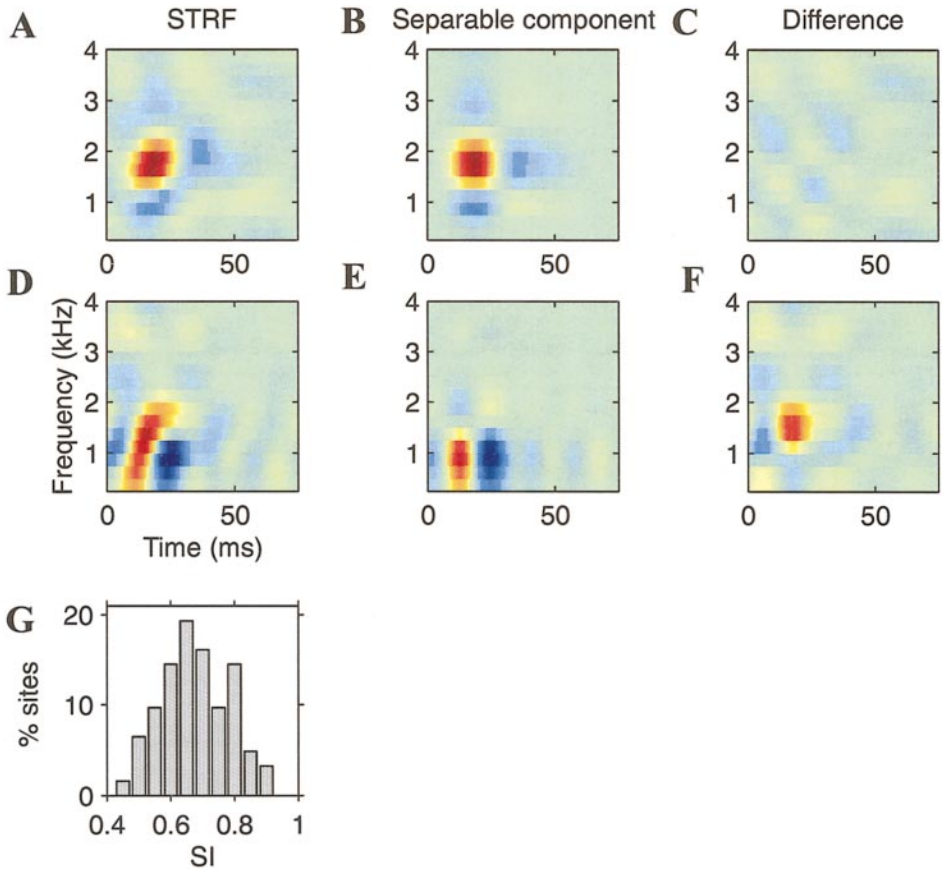

FIG. 8. Separable and inseparable STRFs. *A*: an example of a separable STRF from subregion L2b (site 20.8, single unit). *B*: the separable component of the STRF, which is obtained from the leading term in the singular value decomposition of the STRF and corresponds to the outer product of the function of time and the function of frequency associated with the largest singular value (see METHODS). *C*: the difference between the original and the separable component is shown. This STRF had a separability index (SI; see METHODS) of 0.91. *D–F*: same sequence of plots for an inseparable STRF from subregion L3 (site 25.8) which had a SI of 0.53. *G*: distribution of separability indices in the auditory forebrain.

the response. Figure 9, *D* and *E*, illustrates examples where the STRF makes errors in predicting the timing and width of the response peaks and troughs as well. Figure 9*F* shows the distribution of CCs obtained for our entire data set. A comparison of the distribution of CCs for single units and small clusters of units did not indicate a significant difference ($P = 0.7$). Table 1 summarizes the values of all the parameters we obtained from the STRFs.

We also investigated whether the parameters obtained from the STRF were correlated with each other. We examined the pair-wise scatter plots of the parameters for all pairs (data not shown), calculated the correlation coefficient between parameters and the significance of the correlation coefficient (Fisher's r to z test). We found that many of the parameters were significantly correlated with each other. These values are summarized in Table 2. In particular, T_{peak} and BMF, T_{peak} and CC, BMF and CC, and *E-I* ratio and CC were strongly correlated. This suggests that short latency responses, short integration times, and a preponderance of excitation tended to co-occur with increased linearity.

Mapping STRF parameters onto anatomical subregions

To begin to investigate the relationship between the functional properties of auditory forebrain neurons, as indicated by the STRF parameters, and conventional anatomical subdivisions of the auditory forebrain, we compared the STRF parameters in our data across the different subregions of the auditory forebrain: L2a, L2b, L1, L3, and cHV. Figure 10*A* shows the mean and the inter-quartile range of values for the parameter T_{peak} across the different subregions. We observed a significant difference in the mean T_{peak} across the subregions ($P = 0.024$, $F = 3.0$, ANOVA, see figure legend for further statistics). T_{peak} was shortest in L2a (mean $T_{peak} = 14$ ms) and longest in cHV (31 ms) with subregions L2b (20 ms), L1 (21 ms), and L3 (22 ms) showing intermediate values. This pattern indicates the timing for the processing of songs in the different subregions. On average neurons in the thalamo-recipient area L2a responded fastest followed by the subsequent areas. The range of T_{peak} in regions L1, L3, and cHV was larger compared with the range in L2a, indicating a more heterogeneous distribution in these areas ($P = 0.002$, 0.001, and 0.0006 for L2a vs. the other areas, respectively, F test with Bonferroni correction; see Table 1 for SEs and ranges of values). We did not observe significant differences in heterogeneity between the remaining areas.

The average values of BMF (Fig. 10*B*), showed a significant difference across subregions ($P = 0.006$, $F = 4.1$). We observed a preference for high modulation frequencies in L2a (mean BMF = 38 Hz) compared with lower modulation frequencies in L2b (21 Hz), L1 (22 Hz), L3 (15 Hz), and cHV (17 Hz). The inverse of the BMF parameter can be thought of as a characteristic time scale of integration of songs. The average values of this parameter indicated a short time scale of integration for sites in L2a (26 ms), followed by L1 (46 ms) and L2b (48 ms), cHV (59 ms), and L3 (67 ms).

A comparison of the Q factor (see METHODS and Fig. 10*C*) did not show a significant difference ($P = 0.17$) across subregions. Thus on average the features obtained from the different subregions were comparable in the sharpness of spectral tuning of the largest spectral peak (see Table 1).

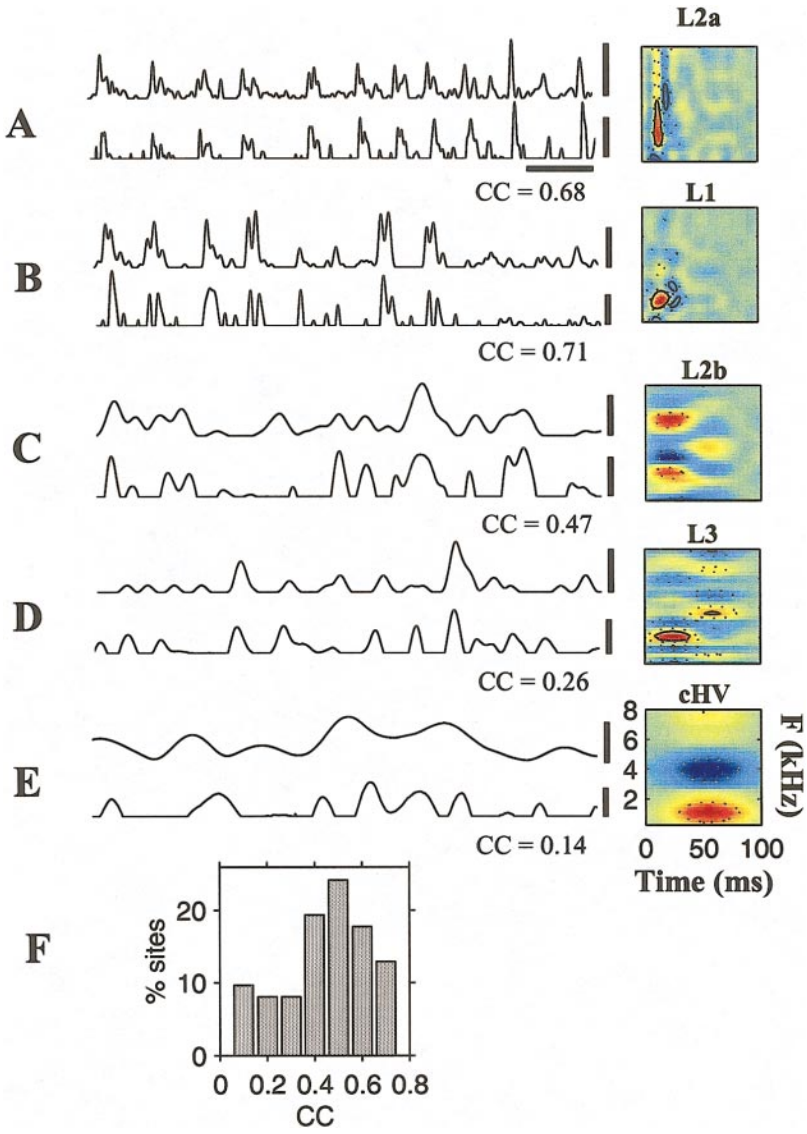

FIG. 9. Prediction of neural responses. In the examples shown, the *top trace* illustrates the estimated PSTH for the neural responses from a site and the *bottom trace* illustrates the predicted PSTH obtained from the STRF for this site for the same segment of the data. The STRF used to obtain the prediction is displayed on the right of each set of traces. One example from each subregion is shown, spanning a range of correlation coefficients (CCs) obtained between the actual and predicted responses for our entire data set. *A*: example from L2a (site 26.5A). The scale bar for the time axis in this and all other traces is 200 ms. The scale bars for the actual and predicted responses are 200 and 100 Hz, respectively. The CC between the estimated PSTH and the predicted PSTH for this site was 0.68. *B*: example from L1 (site 18.6; single unit; scale bars: 100 and 50 Hz) with CC = 0.70. *C*: example from L2b (site 23.2A; scale bars: 20 and 10 Hz) with CC = 0.47. *D*: example from L3 (site 27.5B, single unit, scale bars 20 and 10 Hz) with CC = 0.26. *E*: example from cHV (site 27.1A, scale bars 5 and 10 Hz) with CC = 0.14. *F*: distribution of CCs in the auditory forebrain.

We compared the magnitudes of peak excitatory and inhibitory STRF amplitudes in each of the auditory areas. As can be seen by comparing Fig. 10, *E* and *F*, the ratios of the excitatory to inhibitory peak in the different subregions were approximately equal ($P = 0.7$; see Table 1) even though both the excitatory and inhibitory amplitudes varied significantly across the subregions ($P = 0.02$).

When we examined the SI for different subregions (Fig. 10*D*), we found that although the subregions L2b, L1, L3, and cHV contained the sites that were the most spectral temporally inseparable, there was no statistically significant difference in the mean SI across the different subregions ($P = 0.8$; see Table 1 for values).

Figure 10*G* shows the mean values for the CC across the different subregions. These values indicate a significant difference across subregions ($P = 0.023$, $F = 3.1$) with the CCs being highest in L2a (mean CC = 0.63) and significantly different from the CCs in all the other regions, followed by L1

TABLE 1. *Summary of parameters obtained from STRFs across different subregions*

	L2a	L2b	L1	L3	cHV
T_{peak}, ms	14 ± 1 (11–16)	20 ± 1 (12–40)	21 ± 3 (12–41)	22 ± 3 (7–50)	31 ± 6 (17–55)
BMF, Hz	38 ± 9 (20–70)	21 ± 2 (10–50)	22 ± 4 (5–50)	15 ± 2 (5–30)	17 ± 4 (5–30)
Q	1.6 ± 0.4 (0.7–3.2)	1.8 ± 0.3 (0.8–5.9)	2.8 ± 0.5 (1.1–6.1)	2.5 ± 0.5 (0.4–7.8)	1.3 ± 0.3 (0.6–2.8)
E-I ratio	1.7 ± 0.1 (1.4–2.1)	1.4 ± 0.1 (0.7–2.6)	1.6 ± 0.2 (0.7–2.5)	1.4 ± 0.1 (0.7–2.3)	1.3 ± 0.2 (0.7–2.3)
SI	0.71 ± 0.03 (0.65–0.82)	0.68 ± 0.03 (0.47–0.91)	0.70 ± 0.03 (0.58–0.91)	0.66 ± 0.03 (0.49–0.83)	0.66 ± 0.04 (0.56–0.84)
CC	0.63 ± 0.02 (0.58–0.68)	0.44 ± 0.04 (0.11–0.70)	0.48 ± 0.05 (0.16–0.72)	0.37 ± 0.05 (0.07–0.64)	0.37 ± 0.06 (0.14–0.53)

The table shows the mean ± SE and ranges (in parentheses) for the parameters: time to peak (T_{peak}), best modulation frequency (BMF), quality factor of the largest spectral peak (*Q*), the ratio of the excitatory to inhibitory peak (*E-I* ratio), separability index (SI), and correlation coefficient (CC) between the estimated response from the actual data and the prediction of the response obtained from the spectral temporal receptive field (STRF; see METHODS and RESULTS for definition of parameters).

TABLE 2. *Correlations between STRF parameters*

	BMF	Q	E-I ratio	SI	CC
T_{peak}	−0.46*	0.05	−0.37*	−0.08	−0.44*
BMF		−0.25	0.28*	0.25	0.64*
Q			−0.09	−0.09	−0.25
E-I ratio				0.22	0.49*
SI					0.15

The table shows the correlation coefficient between each pair of parameters for our entire data set. The asterisks indicate the correlation coefficients that were significant ($P < 0.05$).

(0.48), L2b (0.44), L3 (0.37), and cHV (0.37). Although the sample size is small, the range of CCs in L2a was also significantly smaller compared with L2b, L1, and L3, indicating a more heterogeneous distribution in these regions ($P = 0.002$, 0.004, 0.002, respectively; see Table 1 for ranges). There were no significant differences in heterogeneity between the regions L2b, L1, L3, and cHV. These results suggest a difference in the nonlinear component of the neural encoding of songs in different regions, with region L2a showing relatively linear encoding of songs and subsequent areas showing linear as well as nonlinear encoding of songs.

DISCUSSION

An important goal of auditory neuroscience is to understand the processing of natural sounds by auditory neurons, which may have evolved to efficiently encode these sounds (Attias and Schreiner 1998; Rieke et al. 1995) and which respond much more strongly to such sounds in higher-level auditory areas (Margoliash 1983; Rauschecker et al. 1995; Theunissen et al. 2000; Wang et al. 1995). However, due to the complexity of natural sounds such as human speech and birdsong, it has been difficult to obtain the stimulus-response properties of auditory neurons with such sounds using conventional methods. Previously, the STRF approach has been successfully employed to characterize the responses of auditory neurons to synthetic sounds (deCharms et al. 1998; Depireux et al. 2001; Eggermont et al. 1983a,c; Escabi et al. 1998; Keller and Takahashi 2000; Klein et al. 2000; Kowalski et al. 1996a,b). In this study, we used our recent extension of the STRF approach (Theunissen and Doupe 1998; Theunissen et al. 2000, 2001) to analyze the processing of natural sounds in the songbird auditory forebrain.

In the few physiological studies that have been done to date with small sets of natural vocalizations and complex synthetic sounds, auditory neurons in field L were found to be quite diverse, ranging from broadly responsive to selective (Langner et al. 1981; Muller and Leppelsack 1985; Scheich et al. 1979; Uno et al. 1991). However, these studies could not identify the components of the stimuli responsible for the neuronal response. Our approach here was to use the extended STRF method to investigate directly the features of songs to which auditory forebrain neurons responded.

Our results revealed a diverse range of processing of songs in the auditory forebrain with some neurons responding to simple tonal components of songs and others responding to more complex spectral-temporal structures such as frequency sweeps and multi-peaked frequency stacks. We quantified multiple aspects of the processing of songs in the auditory forebrain by extracting several parameters from the STRFs. Using the parameter T_{peak}, we characterized the timing of responses in the auditory forebrain. The range of values indicated both fast and relatively slower processing of song features.

Another important temporal parameter of complex sounds, such as speech and birdsong, is the modulation in the amplitude envelope of sounds. Complex sounds typically contain a broad range of modulation frequencies. The BMF parameter, extracted from the STRF, allowed us to characterize the preferred modulation frequency for auditory forebrain neurons and showed that, as a group, auditory forebrain neurons could

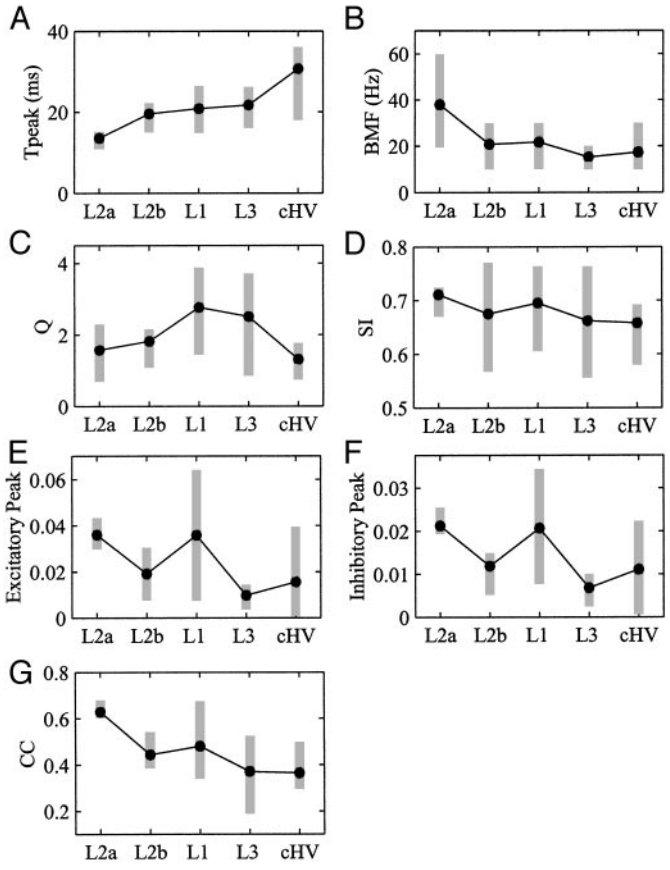

FIG. 10. Distribution of STRF parameters for the different regions in the auditory forebrain. Figure shows the mean values and the inter-quartile range of values (from the 25th to 75th percentile) of the distribution (▨) for different parameters of the STRF for each subregion (see Fig. 6 and METHODS for definitions). The parameters for the respective subregions are plotted in the order: L2a, L2b, L1, L3, and cHV for all the parameters. The differences in the mean values of the parameters across different regions were found to be statistically significant in *A*, *B*, and *E–G* (see RESULTS). *A*: T_{peak} of the STRF, which is a measure of the delay between the stimulus and response. Multiple pair-wise comparisons (post hoc Fisher's test) indicated the following regions to be significantly different: L2a/cHV, L2b/cHV, L1/cHV, and L3/cHV. *B*: the best modulation frequency (BMF), which is a measure of the preferred frequency of temporal modulations in the amplitude envelope of songs. The inverse of this frequency is a measure of the characteristic time scale of integration. The regions that showed pair-wise significant differences were: L2a/L2b, L2a/L1, L2a/L3, and L2a/cHV. *C*: the quality factor (*Q*), which is a measure of the sharpness of spectral tuning. *D*: distribution of separability indices across different regions of the auditory forebrain. *E* and *F*: the excitatory and inhibitory peak amplitudes. The regions that showed pair-wise significant differences for both the excitatory and inhibitory peak amplitudes were: L1/L2b, L1/L3, and L2a/L3. *G*: correlation coefficients between predicted and actual responses across different regions in the auditory forebrain. The regions that showed a pair-wise significant difference were: L2a/L2b, L2a/L3, and L2a/cHV.

encode a broad range of AM frequencies. The majority of the neurons, however, preferred lower modulation frequencies, approximately matching the dominant range of modulation frequencies found in songs (Theunissen et al. 2000). By itself, the value of the BMF does not necessarily imply sharp bandpass tuning to an AM frequency corresponding to the BMF. It is, nevertheless, a useful indicator of the AM frequency in songs that is most effective in driving neurons. In addition, the value of the BMF parameter obtained here could be different from the conventional BMF parameter obtained with simple amplitude modulated tone bursts because, as we found in our previous study, many auditory forebrain neurons show different stimulus-response properties when probed with natural versus synthetic sounds (Theunissen et al. 2000). The inverse of the BMF parameter also gave us an indication of the time scale of integration for auditory forebrain neurons. High-level auditory neurons displaying context dependent phenomena such as combination-sensitivity have often been found to integrate their inputs over a relatively long duration (Lewicki and Arthur 1996; Margoliash 1983; Margoliash and Fortune 1992; Ohlemiller et al. 1996). In our data, the time scales of the features to which neurons responded in some of the auditory forebrain regions were surprisingly long, in some cases showing integration times on the order of 100 ms. Integration of input over such a long duration could contribute to the known sensitivity of some field L neurons to combinations of song syllables as well as to the selectivity for BOS seen in high level auditory areas (Lewicki and Arthur 1996).

The quality of the predictions of neural responses obtained from the STRF model, as assessed by the CC, indicated the presence of both relatively linear as well as more nonlinear encoding of songs in the auditory forebrain. Here, it is important to point out that, although we were able to estimate the magnitude of the nonlinear component of the stimulus-response function by assessing the quality of predictions obtained from the linear STRF model, this model could not provide any information about the exact nature of the nonlinearity. We have previously shown that part of the nonlinearity across different stimulus ensembles can be described by constructing separate STRFs for each stimulus ensemble (Theunissen et al. 2000). This is analogous to constructing a piece-wise linear approximation of a nonlinear function. However, describing the residual nonlinearities within a particular stimulus ensemble remains an important challenge for current methods in auditory neuroscience. In principle, one could include higher-order terms in the Volterra expansion describing the stimulus-response relationship. However, estimating these terms and interpreting their biophysical significance is quite difficult. Examination of the linear prediction showed several types of errors. In some cases, the timing and width of responses was well predicted but the amplitude was not. In such cases, it may be possible to improve the prediction by incorporating a static nonlinearity in the model for predicting responses. In other cases, errors occurred in predicting the timing and width of responses as well, suggesting dynamic nonlinearities. Such nonlinearities could arise from underlying nonlinear cellular and synaptic processes such as adaptation, facilitation, and depression. Further elucidation of the nonlinearities may require modeling them based on a detailed description of such underlying biophysical mechanisms or developing new methods that describe such nonlinearities.

The auditory forebrain showed narrowly as well as broadly tuned STRFs, suggesting that neurons in this region analyze songs at a variety of spectral resolutions. Analysis over a range of spectral resolutions is thought to be a prominent principle of the organization of mammalian auditory cortex as well (Schreiner et al. 2000).

We found that the ratio of the excitatory and inhibitory peaks of the STRFs was approximately balanced in the auditory forebrain, which may reflect properties of the local circuitry in the auditory forebrain. Models of auditory neurons have suggested how neural responses can be shaped by the local excitatory and inhibitory circuitry (Nelken and Young 1997; Shamma 1989). STRFs with excitatory and inhibitory regions could be the result of such excitatory and inhibitory interactions. Such a balance of excitatory and inhibitory regions, organized in an appropriate way in the time-frequency domain, could result in more temporally phasic and/or more spectrally selective responses. For example, in cases in which the excitatory region precedes the inhibitory region, response would be initiated by the activation of the excitatory region but subsequently terminated or attenuated by the activation of the inhibitory region, thus producing a more temporally phasic response. One possible way to directly investigate the relation between the STRF and the local excitatory and inhibitory circuitry in the auditory forebrain would be to manipulate the amounts of inhibition or excitation in these areas and examine the resultant changes in the STRFs.

We observed both separable and inseparable STRFs in the auditory forebrain. Neurons with inseparable STRFs could be used to detect spectral temporal structures of sound that change with time, such as frequency sweeps, analogous to direction selective neurons found in the visual system. Such STRFs might be important in the analysis of songs, since frequency sweeps are prominent in many zebra finch songs. In the visual system, a simple model for motion-sensitive neurons was proposed, in which two spatio-temporally separable receptive fields combine in quadrature to produce a spatio-temporally inseparable receptive field (Adelson and Bergen 1985; Watson and Ahumada 1985). In the auditory system, a similar principle could apply in the spectral-temporal domain. Thus the inseparable STRFs found in the auditory forebrain could be generated by combining inputs from the separable STRFs in the same or previous regions.

The preceding discussion highlights the diversity of the auditory forebrain in the distribution of STRF parameters, reflecting the range of complexity we observed in the STRFs. In our data, we also observed that some parameters indicative of more complex processing tended to co-occur. For example, neurons with long time scales of integration also tended to have more nonlinear encoding properties, indicating that some neurons found in the auditory forebrain could be jointly complex in multiple attributes. Thus several functional stages of song processing, ranging from simple to quite complex, appear to occur within the auditory forebrain and suggest that the auditory forebrain may be involved in the analysis of many different aspects of song structure. The resultant multiple representations of songs, of varying complexity and time scales, could together provide useful information to higher level auditory areas that are likely to be involved in the perception of highly complex, behaviorally relevant stimuli.

Mapping STRF parameters

A problem of great interest in the study of auditory systems has been to understand the organization of auditory maps of different parameters of sounds. To begin to look for patterns in the mapping of functional properties of auditory forebrain neurons onto conventional anatomical subregions of the auditory forebrain, we compared the STRF parameters across the different subregions. Clearly, more data will be required for a complete analysis of the different subregions, especially subregions such as L2a and cHV, where we had a relatively small number of neurons. This is even more important for subregion cHV, which was quite heterogeneous in the distribution of STRF parameters, unlike L2a. Nevertheless we observed several significant and suggestive trends in our data.

A comparison of the parameter T_{peak} across the different regions revealed a significant difference in the timing for the processing of songs in the auditory forebrain, with L2a responding fastest, followed by L2b, L1, and L3, and then cHV, which had the slowest responses of all the areas studied here. This pattern is consistent with the known anatomical connectivity in the auditory forebrain (see Fig. 1) (see also Vates et al. 1996).

When we compared the time scales of integration in different regions of the auditory forebrain, we found that L2a showed relatively short integration time scales compared with regions L2b, L1, L3, and cHV. A similar increase in the time scale of integration, as indicated by the best modulation frequency, has also been observed in successive areas of the auditory cortex of cats (Schreiner and Urbas 1988).

The quality of the predictions of neural responses obtained from the STRF model, as assessed by the CC between the estimated and predicted response, also varied significantly across the auditory forebrain regions. CCs were highest in area L2a, followed by L1, L2b, L3, and cHV. This difference is suggestive of an increase in the nonlinear component of the neural encoding of songs from L2a to L1, L2b, L3, and cHV, respectively. Such an increase in nonlinearity could be reflective of preparatory stages of processing for the generation of highly nonlinear properties such as BOS selectivity, seen later in the auditory pathway (Janata and Margoliash 1999; Margoliash 1983; Margoliash and Fortune 1992; Theunissen and Doupe 1998). An increase in the nonlinear component of the processing of sensory stimuli in successive stages of a sensory system has also been reported in the electric fish system (Gabbiani et al. 1996).

The rough mapping of the STRF parameters discussed in the preceding text onto conventional anatomical subdivisions of the auditory forebrain is suggestive of hierarchical processing, with the thalamo-recipient area L2a showing parameters characteristic of simpler processing and subsequent areas revealing the gradual emergence of more complex processing properties. However, other observations indicate that the auditory forebrain may not be organized in a strictly serial hierarchy. Not all STRF parameters varied significantly across the different subregions. For instance, the separability index did not reveal a significant difference among the subregions. It remains possible, however, that qualitatively different types of inseparability occur in the different subregions. The subregions also shared properties such as the sharpness of spectral tuning and the ratio of excitatory to inhibitory peaks. Thus instead of being organized in a strictly serial hierarchy, the auditory forebrain may be organized in a more elaborate way, performing both serial and parallel processing of auditory information. The known, extensive interconnectivity between the anatomical subregions of the auditory forebrain also supports this idea (Vates et al. 1996). Thus the complex processing properties we observed could arise via a combination of hierarchical and parallel processing in the network of auditory forebrain subregions. The intrinsic circuitry within each of the subregions may also play a role in the emergence of this complexity.

Overall our data are consistent with L2a being the major input region of the auditory forebrain, responding to relatively simple features of complex sounds with short delays, short integration times and more linear processing. Surprisingly, area L2b often showed complex STRFs, even though it is anatomically described as an early auditory area similar to L2a. There are several possible explanations for this finding. Although L2b receives direct thalamic input, the parts of Ov that project to L2b and L2a are distinct, thus potentially contributing to the differences in the response properties of these two areas (Vates et al. 1996). Second, in this study, area L2b was defined to include area L, thus making it a much larger composite region. Since the inputs to area L have not been described in detail so far, it remains possible that the strongest sources of inputs to parts of this composite region are from other auditory forebrain regions and not directly from the thalamus, which could lead to more complex response properties. Our results suggest a gradual emergence of more complex features, longer delays and integration times, and nonlinear processing properties in the auditory forebrain subsequent to area L2a. As auditory forebrain areas begin to be probed in much more detail it is likely that additional differences between the subregions of the auditory forebrain will be identified. The stages of processing in these areas are likely to contribute both to the generation of song selective neurons found in higher-level areas in the songbird brain as well as to the detection and discrimination of a wide variety of natural sounds behaviorally relevant to songbirds.

We thank M. Brainard, M. Escabi, and C. Schreiner for comments and discussion on an earlier version of the manuscript; R. Kimpo, C. Roddey, G. Carrillo, and A. Arteseros for technical assistance; and two anonymous referees for critical comments on the manuscript.

This work was supported by research grants from the Alfred P. Sloan Foundation (to K. Sen, F. E. Theunissen, and A. J. Doupe) and the National Institute of Neurological Disorders and Stroke (NS-34835 to A. J. Doupe).

REFERENCES

ADELSON EH AND BERGEN JR. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2: 284–289, 1985.

AERTSEN AM AND JOHANNESMA PI. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol Cybern* 42: 133–143, 1981.

ATTIAS H AND SCHREINER CE. Temporal low-order statistics of natural sounds. *Adv Neural Inform Process* 9: 27–33, 1997.

ATTIAS H AND SCHREINER CE. Coding of naturalistic stimuli by auditory midbrain neurons. *Adv Neural Inform Process* 10: 103–109, 1998.

BONKE D, SCHEICH H, AND LANGNER G. Responsiveness of units in the auditory neostriatum of the guinea fowl (*numida meleagris*) to species-specific calls and synthetic stimuli. I. Tonotopy and functional zones of field L. *J Comp Physiol* 132: 243–255, 1979.

CAI D, DEANGELIS GC, AND FREEMAN RD. Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *J Neurophysiol* 78: 1045–1061, 1997.

De Valois RL and Cottaris NP. Inputs to directionally selective simple cells in macaque striate cortex. *Proc Nat Acad Sci USA* 95: 14488–14493, 1998.
deCharms RC, Blake DT, and Merzenich MM. Optimizing sound features for cortical neurons [see comments]. *Science* 280: 1439–1443, 1998.
Depireux DA, Simon JZ, Klein DJ, and Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85: 1220–1234, 2001.
Eggermont JJ, Aertsen AM, and Johannesma PI. Quantitative characterisation procedure for auditory neurons based on the spectro-temporal receptive field. *Hear Res* 10: 167–190, 1983a.
Eggermont JJ, Aertsen AM, and Johannesma PI. Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectro-temporal receptive field. *Hear Res* 10: 191–202, 1983b.
Eggermont JJ, Johannesma PM, and Aertsen AM. Reverse-correlation methods in auditory research. *Q Rev Biophys* 16: 341–414, 1983c.
Escabi MA, Schreiner CE, and Miller LM. Dynamic time-frequency processing in the cat midbrain, thalamus, and auditory cortex: spectro-temporal receptive fields obtained using dynamic ripple stimulation. *Soc Neurosci Abstr* 24: 1879, 1998.
Fortune ES and Margoliash D. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *J Comp Neurol* 325: 388–404, 1992.
Gabbiani F, Metzner W, Wessel R, and Koch C. From stimulus encoding to feature extraction in weakly electric fish [see comments]. *Nature* 384: 564–567, 1996.
Gehr DD, Capsius B, Grabner P, Gahr M, and Leppelsack HJ. Functional organisation of the field-L-complex of adult male zebra finches. *Neuroreport* 10: 375–380, 1999.
Hermes DJ, Aertsen AM, Johannesma PI, and Eggermont JJ. Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (*Rana temporaria* L.) investigated with noise stimuli. *Hear Res* 5: 147–178, 1981.
Hermes DJ, Eggermont JJ, Aertsen AM, and Johannesma PI. Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (*Rana temporaria* L.) investigated with tonal stimuli. *Hear Res* 6: 103–126, 1982.
Jagadeesh B, Wheat HS, Kontsevich LL, Tyler CW, and Ferster D. Direction selectivity of synaptic potentials in simple cells of the cat visual cortex. *J Neurophysiol* 78: 2772–2789, 1997.
Janata P and Margoliash D. Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. *J Neurosci* 19: 5108–5118, 1999.
Keller CH and Takahashi TT. Representation of temporal features of complex sounds by the discharge patterns of neurons in the owl's inferior colliculus. *J Neurophysiol* 84: 2638–2650, 2000.
Kelley DB and Nottebohm F. Projections of a telencephalic auditory nucleus-field L in the canary. *J Comp Neurol* 183: 455–469, 1979.
Kim PJ and Young ED. Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers. *J Acoust Soc Am* 95: 410–422, 1994.
Klein DJ, Depireux DA, Simon JZ, and Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9: 85–111, 2000.
Konishi M. Birdsong: from behavior to neuron. *Annu Rev Neurosci* 8: 125–170, 1985.
Kontsevich LL. The nature of the inputs to cortical motion detectors. *Vision Res* 35: 2785–2793, 1995.
Kowalski N, Depireux DA, and Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76: 3503–3523, 1996a.
Kowalski N, Depireux DA, and Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *J Neurophysiol* 76: 3524–3534, 1996b.
Langner G, Bonke D, and Scheich H. Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. *Exp Brain Res* 43: 11–24, 1981.
Lewicki MS and Arthur BJ. Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16: 6987–6998, 1996.
Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 3: 1039–1057, 1983.
Margoliash D. Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci* 6: 1643–1661, 1986.
Margoliash D and Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J Neurosci* 12: 4309–4326, 1992.
Marler P. Song-learning behavior: the interface with neuroethology. *Trends Neurosci* 14: 199–206, 1991.
Mello CV and Clayton DF. Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *J Neurosci* 14: 6652–6666, 1994.
Mooney R. Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. *J Neurosci* 20: 5420–5436, 2000.
Muller CM and Leppelsack HJ. Feature extraction and tonotopic organization in the avian auditory forebrain. *Exp Brain Res* 59: 587–599, 1985.
Nelken I, Rotman Y, and Bar Yosef O. Responses of auditory-cortex neurons to structural features of natural sounds [see comments]. *Nature* 397: 154–157, 1999.
Nelken I and Young ED. Linear and non-linear spectral integration in type IV neurons of the dorsal cochlear nucleus. I. Regions of linear interaction. *J Neurophysiol* 78: 790–799, 1997.
Ohlemiller KK, Kanwal JS, and Suga N. Facilitative responses to species-specific calls in cortical FM-FM neurons of the mustached bat. *Neuroreport* 7: 1749–1755, 1996.
Rauschecker JP, Tian B, and Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111–114, 1995.
Rieke F, Bodnar DA, and Bialek W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc R Soc Lond B Biol Sci* 262: 259–265, 1995.
Scheich H, Langner G, and Bonke D. Responsiveness of units in the auditory neostriatum of the guinea fowl (numida meleagris) to species-specific calls and synthetic stimuli. II. Discrimination of Iambus-like calls. *J Comp Physiol* 132: 257–276, 1979.
Schreiner CE, Read HL, and Sutter ML. Modular organization of frequency integration in primary auditory cortex. *Annu Rev Neurosci* 23: 501–529, 2000.
Schreiner CE and Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 32: 49–63, 1988.
Shamma S. Spatial and temporal processing in central auditory networks. In: *Methods in Neuronal Modelling*, edited by Koch C and Segev I. Cambridge, MA: MIT Press, 1989.
Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, and Gallant J. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. In: *Network: Computation in Neural Systems.* In press.
Theunissen FE and Doupe AJ. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *J Neurosci* 18: 3786–3802, 1998.
Theunissen FE, Sen K, and Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315–2331, 2000.
Uno M, Ohno Y, Yamada T, and Miyamoto K. Neural coding of speech sound in the telencephalic auditory area of the mynah bird. *J Comp Physiol* 169: 231–239, 1991.
Vates GE, Broome BM, Mello CV, and Nottebohm F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J Comp Neurol* 366: 613–642, 1996.
Volman SF. Development of neural selectivity for birdsong during vocal learning. *J Neurosci* 13: 4737–4747, 1993.
Wang X, Merzenich MM, Beitel R, and Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685–2706, 1995.
Watson AB and Ahumada AJ. Model of human visual-motion sensing. *J Opt Soc Am A* 2: 322–341, 1985.
Zaretsky MD and Konishi M. Tonotopic organization in the avian telencephalon. *Brain Res* 111: 167–171, 1976.