Behavioral/Systems/Cognitive

Functional Groups in the Avian Auditory System

Sarah M. N. Woolley,[1] Patrick R. Gill,[2] Thane Fremouw,[5] and Frédéric E. Theunissen[2,3,4]

[1]Department of Psychology, Columbia University, New York, New York 10027, [2]Biophysics Graduate Group, [3]Helen Wills Neuroscience Institute, and [4]Department of Psychology, University of California, Berkeley, California 94720, and [5]Department of Psychology, University of Maine, Orono, Maine 04469

Auditory perception depends on the coding and organization of the information-bearing acoustic features of sounds by auditory neurons. We report here that auditory neurons can be classified into functional groups, each of which plays a specific role in extracting distinct complex sound features. We recorded the electrophysiological responses of single auditory neurons in the songbird midbrain and forebrain to conspecific song, measured their tuning by calculating spectrotemporal receptive fields (STRFs), and classified them using multiple cluster analysis methods. Based on STRF shape, cells clustered into functional groups that divided the space of acoustical features into regions that represent cues for the fundamental acoustic percepts of pitch, timbre, and rhythm. Four major groups were found in the midbrain, and five major groups were found in the forebrain. Comparing STRFs in midbrain and forebrain neurons suggested that both inheritance and emergence of tuning properties occur as information ascends the auditory processing stream.

Introduction

Complex sounds such as vocalizations have been described as having both "contour" and "texture" (Eggermont, 2001). Sound contour is conveyed by temporal properties such as changes in amplitude over time. Sound texture includes spectral properties such as spectral pitch and spectrotemporal properties such as frequency modulation. Understanding how neurons code these aspects of complex sounds will contribute to understanding how complex, natural sounds such as vocalizations are perceived and recognized.

Neurons in the avian and mammalian auditory midbrain and forebrain show diverse tuning properties when tested with synthetic stimuli (Eggermont, 2001; Woolley and Casseday, 2004, 2005), and discrete characterizations have been used to organize auditory neurons into functional classes. These functional classes can then be assigned roles in the coding of contour or texture features. For example, in terms of specializations for amplitude modulation coding, neurons in primary auditory cortex (A1) (Lu et al., 2001) and the inferior colliculus (IC) (Escabi and Schreiner, 2002) have been divided into two distinct groups, those that synchronize firing to amplitude modulations, responding to only slower modulation rates, and those that do not synchronize but do respond to faster modulation rates. Functional specializations for textural coding have also been reported, such as the division of A1 neurons into two groups based on the number of peaks in frequency tuning (Kadia and Wang, 2003) or the distinction between frequency modulation "specialized" and "mixed" neurons in IC (Poon et al., 1992). These characterizations focus on specific tuning properties and therefore do not fully capture the complexity and heterogeneity of the observed responses of auditory midbrain and forebrain neurons (Schreiner, 1995; Eggermont, 2001), suggesting that a characterization of functional classes of cells based on combined spectrotemporal tuning properties may be informative (Depireux et al., 2001; Sen et al., 2001; Miller et al., 2002; Escabí and Read, 2003; Nagel and Doupe, 2008). Furthermore, with the exception of studies on the bat auditory system (Suga, 1989), relating tuning properties obtained with synthetic sounds to the processing of complex natural sounds such as vocalizations has been limited because the tuning properties of auditory neurons obtained from responses to synthetic sounds often do not predict the responses of the same neurons to natural sounds (Rauschecker et al., 1995; Wang et al., 1995; Bieser, 1998; Theunissen et al., 2000; Nagarajan et al., 2002; Cohen et al., 2007).

Here, we investigated the functional properties of midbrain and forebrain auditory neurons in songbirds using an approach that differs from previous studies that have described functional similarities among auditory neurons. First, we used behaviorally relevant natural sounds, birdsongs, to examine the tuning properties of auditory neurons. Second, we measured the joint spectrotemporal tuning of neurons from responses to those natural sounds by calculating the spectrotemporal receptive field (STRF) for each neuron. Third, we subjected the STRFs to a cluster analysis to determine whether cells could be classified into functional groups based on their spectrotemporal tuning properties. Fourth, we compared the spectrotemporal tuning of functional groups with the statistical distribution of the acoustic features of song, making quantitative assessments of what song features may be encoded by cells in each group. We found that auditory midbrain and forebrain neurons fall into distinct functional groups based on spectrotemporal tuning properties measured from responses to song. The comparison of spectrotemporal tuning in these functional groups and the acoustic features of songs indicated that each functional group may be specialized to code sound features that are important cues for one of three perceptual

Received May 4, 2008; revised Jan. 16, 2009; accepted Jan. 20, 2009.

This work was supported by National Institutes of Health Grants DC05087 (S.M.N.W.), MH067391 (T.F.), and DC007293, MH66990, and MH59189 (F.E.T.). We thank Y.M. for animal husbandry, B.R. for histological assistance, and Taffeta Elliott for comments on this manuscript.

Correspondence should be addressed to either of the following: Frédéric E. Theunissen, Department of Psychology and Helen Wills Neuroscience Institute, 3210 Tolman Hall, Berkeley, CA 94720, E-mail: theunissen@berkeley.edu; or Sarah M. N. Woolley, Columbia University, Department of Psychology, 1190 Amsterdam Avenue, New York, NY 10027, E-mail: sw2277@columbia.edu.

DOI:10.1523/JNEUROSCI.2042-08.2009

aspects of sound, pitch, rhythm, or timbre. Finally, we compared the functional clusters in the midbrain and forebrain and determined the degree to which tuning properties were preserved in the auditory processing stream.

Materials and Methods

Animals

Forty-two adult (>120 d of age) male zebra finches (*Taenopygia guttata*) were used. All birds were bred and raised at the University of California, Berkeley. Birds were raised in families that were housed in a large colony room. Each family was visually but not acoustically isolated from other families. Birds were given seed and water *ad libitum*, vegetables, grit, egg, calcium, baths, and natural spectrum lighting with dawn and dusk simulation. All animal procedures were approved by the Animal Care and Use Committee at the University of California at Berkeley.

Surgery

Two days before recording, a bird was anesthetized with Equithesin (0.03 ml, i.m., of the following: 0.85 g chloral hydrate, 0.21 g pentobarbital, 0.42 g $MgSO_4$, 8.6 ml propylene glycol, and 2.2 ml of 100% ethanol to a total volume of 20 ml with H_2O). The bird was then placed in a custom stereotaxic frame with ear bars and a beak holder. Lidocaine (2%) was applied to the skin overlying the skull, and a midline incision was made. A metal pin was fixed to the skull with dental cement. Ink dots were placed on the skull to indicate stereotaxic coordinates for electrode placement. The bird was then allowed to recover for 2 d.

On the day of recording, the bird was anesthetized with three injections of 20% urethane (three intramuscular injections, 30 μl each, 30 min apart) and was placed in the stereotaxic frame. The bird's head was immobilized by attaching the metal pin cemented to the bird's skull to a customized holder, mounted on the stereotaxic frame. After craniotomy, electrodes were positioned over the forebrain. For recordings in the auditory midbrain nucleus mesencephalicus lateralis, pars dorsalis (MLd), lidocaine was applied to the skin overlying the skull covering the optic lobe. A small opening was then made in the skin and skull overlying the optic tectum, dura was resected from the surface of the brain, and an electrode was positioned over the midbrain.

Electrophysiology

Neural recordings were conducted in a sound-attenuated chamber (Acoustic Systems). The bird was positioned 20 cm in front of a Bose 101 speaker so that the bird's beak was centered both horizontally and vertically with the center of the speaker cone. The output of the speaker was measured before each experiment with a Radio Shack condenser microphone (33–3013) and custom software to ensure a flat frequency response (±5 dB) from 250–8000 Hz. Sound levels were checked with a Brüel and Kjær sound level meter (RMS weighting B, fast) positioned 20 cm in front of the speaker at the bird's head. Body temperature was continuously monitored and adjusted to between 38 and 39°C using a thermistor placed under a wing and a heating blanket placed under the bird (FHC Inc.).

Recordings were obtained using epoxy-coated tungsten electrodes (0.5–7.0 MΩ; FHC Inc. or A-M Systems). Electrodes were advanced into the brain with a stepping microdrive at 0.5 μm steps (Newport). The extracellular signal was obtained with an extracellular amplifier (100× gain; high-pass f_c, 300 Hz; low-pass f_c, 5 kHz; A-M Systems), displayed on a multi-channel oscilloscope (Tektronix TDS 210), and monitored on an audio amplifier/loudspeaker (Grass Instruments AM8). Spike arrival times were obtained by thresholding the extracellular recordings with a window discriminator and were logged on a Sun Microsystems computer running custom software (32 kHz sampling rate). Neural recordings were assessed to come from single units by the following: (1) a high signal-to-noise ratio in the recordings (amplitude signal-to-noise ratio >5); (2) monitoring the shape of the triggered action potentials on a digital oscilloscope with trace storage, and (3) calculating the spike autocorrelation function *post hoc*. All spike autocorrelation functions from what we determined to be single units showed the signature depression ~0 ms from postspiking inhibition. Examination of the distribution of interspike interval (ISI) showed that the probability of finding ISIs <1 ms were 3.5 and 3% for MLd and field L, respectively (expected values for Poisson with same mean rates are 10 and 9%). Using this procedure, we recorded from single neurons with an average song-driven firing rate of 9.1 spikes/s (range, 1.2–39.4) in MLd and 10.7 spikes/s (range, 0.8–29.4) in field L. The average background rates were 0.15 spikes/s in MLd and 0.6 spikes/s in field L.

Pure tones, zebra finch songs, modulation-limited noise, and white noise were used as search stimuli. Presentation of the different songs was random within a trial. Two seconds of background spontaneous activity was recorded before the presentation of each stimulus. Song samples from 20 adult male zebra finches were presented to each neuron. Stimuli were presented at a peak intensity of 70 dB sound pressure level. A random interstimulus interval with a uniform distribution between 4 and 6 s was used. At the end of a penetration, one to three electrolytic lesions (100 μA for 5 s) were made to verify the recording sites along that penetration. Lesions were made well outside of any auditory areas unless it was the last penetration.

Histology

After the recording session, the bird was killed by overdose of Nembutal. The carcass was then transcardially perfused with 0.9% saline, followed by 3.7% Formalin in 0.025 M phosphate buffer. The skullcap was removed, and the brain was postfixed in Formalin for at least 5 d. The brain was then cryoprotected in 30% sucrose. Sagittal sections (40 μm) were cut on a freezing microtome and divided into two series. Sections were mounted on gelatin-subbed slides, and one series was stained with cresyl violet and the other with silver stain. Electrolytic lesions were visually identified, and the distance between two lesions within the same electrode track was used to calibrate depth measurements and reconstruct the location of recording sites.

Data analysis

STRF calculation. For each neuron, a normalized reverse correlation analysis was used to determine the stimulus–response relationship. This analysis yields the STRF, a dynamical model of the auditory tuning properties of a neuron that incorporates some static nonlinearities. The STRF calculation has three steps. First, the log spectrogram of the sound stimulus (e.g., a sample of song) is cross-correlated with the average time-varying response to that stimulus obtained by averaging across the 10 trials to obtain the spike-triggered average. Second, the spike-triggered average is normalized by the correlations in the stimulus. Third, a regularization-cross validation procedure is used to effectively minimize the number of parameters that are fitted in the STRF estimation. Once the STRF is obtained, it is validated using data that were not used in the STRF calculation. A song spectrogram is convolved with the STRF to yield a predicted response to that song. The predicted response is then compared with the actual response to that song. The similarity between the predicted response and the actual response, measured using noise-corrected correlation coefficients after smoothing the peristimulus time histogram (PSTH) with an 11 ms Hanning window (see Tables 2, 3), provides the measure of how well the STRF captures the tuning of a neuron. Neurons for which the STRF gave poor predictions were excluded from the analysis. A poor prediction was defined as having $r < 0.2$ or a mutual information between predicted and actual responses <1.2 bit/s. This STRF methodology has been described in detail previously (Theunissen et al., 2000, 2001; Hsu et al., 2004a; Gill et al., 2006; Woolley et al., 2006). STRF estimation and validations were done using STRFPAK, a Matlab toolbox developed by the Theunissen and Gallant laboratories at University of California, Berkeley (http://strfpak.berkeley.edu).

Cluster analysis. The clustering procedure involved three steps: (1) determining the pairwise similarity of the STRFs; (2) generating a sorted list of similar STRFs; and (3) determining the boundaries on this list to make clusters.

The pairwise similarity was measured by estimating the correlation coefficient between the two STRFs after allowing shifts in frequency and in latency (restricted to 10 ms). This shift allows STRFs with the same overall shape but different best frequency or latency to be clustered as similar. The correlation coefficient is the dot product of the STRFs divided by the square root of the products of the power of the STRFs and falls between −1 and +1. This measure is called the similarity index (SI) and has been used in other studies to quantify STRF similarities (DeAngelis et al., 1999; Escabi and Schreiner, 2002).

Given all pairwise similarities between STRFs, the next step in clustering was to sort STRFs with similar shapes into groups. For this sorting, we used a general genetic algorithm. The genetic algorithm is effective to avoid local minima in this optimization problem. As for all optimization problems, the sorting procedure requires the definition of an objective function that is used to quantify the "goodness" of the sort order; the objective function takes the pairwise STRF similarities and an STRF ordering as inputs and returns how well that ordering puts STRFs with similar receptive fields together.

We used an *ad hoc* objective function designed to capture both local and global similarity:

$$P = \sum_{d-1}^{6} \left(|(S_d\, M) - M| + 10|(S_d\, \bar{M}) - \bar{M}| \right).$$

Here, P is the total penalty, M is the matrix of sorted similarities, $\bar{M}$ is a matrix with a 1 for every entry in $M > 0.6$ and a 0 otherwise, and S_d is the column shift operator (i.e., the operator taking the matrix [A B C D] and returning [B C D A] where {A, B, C, D} are columns) applied d times. Because M is symmetric, the choice of column shifts (and not row shifts) is arbitrary. The second term (with $\bar{M}$) penalizes sorting schemes in which large groups of neurons that are highly similar (with SI > 0.6) are not grouped together. The first term overlays a slight refining force on this coarse characterization by shifting neurons with more similar SIs together. Because the diagonal of M is constrained to be 1, sorting schemes that minimize P also tend to push clusters toward the diagonal. The choice of the SI threshold of 0.6 to define "highly similar" neurons was determined by examining the histogram of SIs and finding a value that would separate a significant fraction of neurons with high similarity from those that have typical positive SIs. Setting this threshold to a lower value would encourage low coefficients to be mixed with higher coefficients, thus grouping STRFs with dissimilar features together. Setting this threshold higher would result in the threshold term having little effect on the final order. We calculated P for local shifts of up to 6 to bias our search toward finding clusters with at least a few members but not to continue favoring ever-greater numbers of neurons in the same cluster. Note, however, that the shift is performed recursively as part of the genetic algorithm: the final solution will result in an ordered matrix with much larger shifts for any particular row.

The genetic algorithm was performed using Matlab. We also investigated alternative clustering algorithms. Instead of using the SI, STRFs can be clustered based on sets of filter parameters extracted from the STRFs, such as temporal and spectral bandwidth and latency. That approach gives similar results (supplemental Fig. 2, available at www.jneurosci.org as supplemental material) but is sensitive to the subjective choice of relevant STRF parameters. Using the SI, STRFs can be clustered based on complex shape similarities that are only captured with a high number of yet undefined STRF parameters. We also tried different clustering methods with the SI as a distance measure such as k means and agglomerative clustering, but these gave results that were less intuitive. k means clustering led to a nearly uniform distribution of cluster sizes, in which populous large clusters were split into two. Agglomerative clustering gave better results than k means but emphasized local structure at the expense of global structure (grouping only a few nearly identical STRFs together).

Once the neurons were ordered such that similar STRFs were adjacent, we partitioned the clusters using the sorted array of SIs. The partitioning was guided by a statistical analysis based on bootstrapping, and the final boundaries were modified slightly to enforce a lumping bias (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). The bootstrapping technique involved estimating the distribution of pairwise SIs for groups of three neurons given the null hypothesis that there is no higher-than-second-order structure to the pairwise SIs of STRFs. For this estimate, we performed the sorting algorithm on a shuffled set of pairwise SIs, in which the values were the same as for our real data but their locations were randomized. In this randomized dataset, the distribution of pairwise similarities of any two neighboring neurons after clustering is the same as in the actual data. However, the pairwise similarities of the third neuron in a cluster have a random distribution and give us a measure of the similarity that could be expected by chance in the least similar pair of neurons in a cluster of three neurons. In other words, in the randomized dataset, the off-diagonal elements of the sorted matrix of STRF similarities have the same distribution as the real data, but the off–off diagonal gives us the values expected by chance for all other pairwise comparisons given that no higher-order structure exists in the SI matrix. We repeated the randomizing and clustering 1000 times to build a distribution of such probabilities. We then made plots of significance level for the sorted similarity matrix of the real data using a threshold value of $p = 0.225$. This value is higher than the typical values used for α (e.g., $p = 0.01$), but it should be remembered that, for any cluster with more than three members, the probabilities (given independence) would multiply. For example, in a cluster of four members, there would be three distinct values beyond the off-diagonal. The chance that these three are all significant would then be $(0.225)^3 = 0.014$. The significance level would be even smaller for larger clusters. We considered only clusters including four or more cells. The significance plot is shown in supplemental Figure 1 (available at www.jneurosci.org as supplemental material). We used this plot to determine major cluster boundaries. Additional visual examination of all the STRFs within each cluster justified our procedure in the sense that we did not find any outliers within a group and that the unclassified cells were sufficiently different not to be included (supplemental Figs. 4, 5, available at www.jneurosci.org as supplemental material). We also used this statistical analysis to guide, albeit more subjectively, the subdivision of the major clusters in field L into subgroups, each of which was mostly composed of neurons with significant pairwise SIs.

Frequency and temporal tuning measured from STRFs

To extract the spectral and temporal tuning properties of neurons from the STRF, we fitted each filter with a product of Gabor functions (Qiu et al., 2003):

$$\mathrm{STRF}(t,f) \approx AH(t) \cdot G(f),$$

where

$$H(t) = e^{-0.5[(t-t_0)/\sigma_t]^2} \cdot \cos(2\pi \cdot \Omega_t\,(t - t_0) + P_t),$$

and

$$G(f) = e^{-0.5[(f-f_0)/\sigma_f]^2} \cdot \cos(2\pi \cdot \Omega_f\,(f - f_0) + P_f).$$

The initial decomposition into separable time and frequency functions is obtained using singular value decomposition of matrices (SVD). The fit of the parameters describing the Gabor function is performed by minimizing the mean square error between the H and G functions obtained from the SVD and those obtained from the Gabor model. The fitted parameters for the temporal function are as follows: t_0, the temporal latency (in seconds); σ_t, the temporal bandwidth (in seconds); Ω_t, the best temporal modulation frequency (in hertz); and P_t, the temporal phase. Similarly, the fitted parameters for the spectral function are as follows: f_0, the best frequency (in hertz); σ_f, the spectral bandwidth (in hertz); Ω_f, the best spectral modulation frequency (in cycles/hertz); and P_f, the spectral phase. Here, we analyzed the spectral and temporal bandwidths.

Results

Functional groups in midbrain

The STRFs for 110 of 137 (80%) MLd neurons clustered into four functional groups (Fig. 1). The two largest groups were narrowband-temporal neurons (NB-T) neurons (25 of 110), which are characterized by sharp temporal and spectral tuning, and broadband (BB) neurons (74 of 110), which are characterized by a similarly sharp temporal tuning but broad spectral tuning. The NB-T and BB neurons in the midbrain were also char-

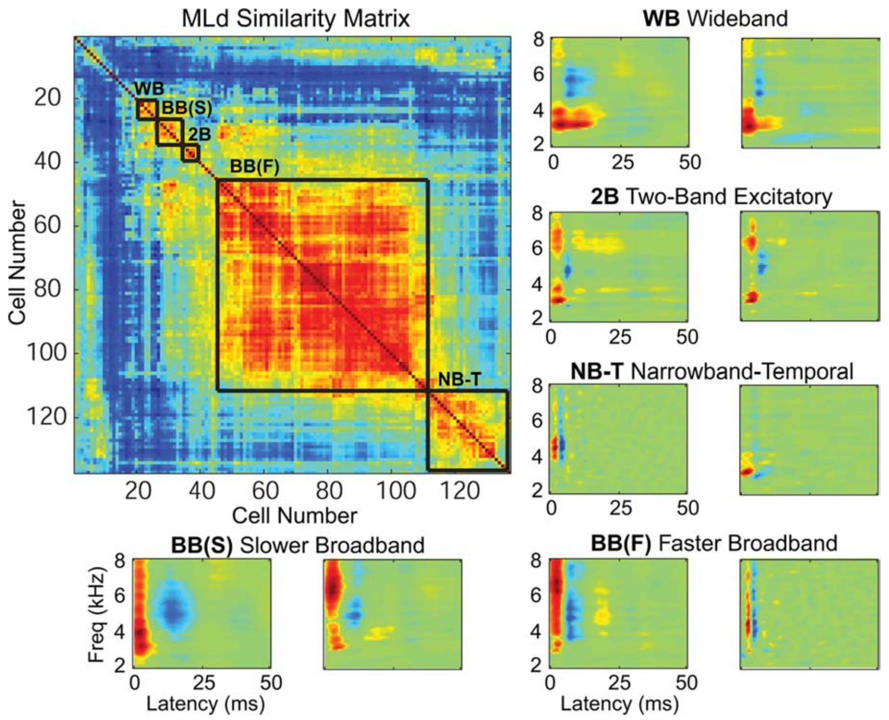

Figure 1. The functional groups in the auditory midbrain region MLd. The top left shows the matrix of the similarity indices between each pairwise comparison after the cells were sorted using the genetic algorithm and objective function. The neurons clustered into four groups based on STRF shape. The two larger groups are BB and NB-T. Broadband neurons separated into two non-adjacent subgroups in the similarity matrix: a larger, faster group [BB(F)] and a smaller, slower group [BB(S)]. Two other small clusters, WB and 2B, were found. The STRFs at opposing edges of each group are shown to illustrate the type and range of STRFs found in each cluster. The STRFs for all neurons in MLd are shown in supplemental Figure 4 (available at www.jneurosci.org as supplemental material).

acterized by temporally contiguous excitatory and inhibitory STRF components, indicating that these neurons were maximally excited by the absence of sound followed by a sound; they were excited by sound onsets. As described below, some NB neurons in the forebrain did not have this temporal inhibition. Instead, they had spectral inhibitory sidebands. To distinguish these two types of NB neurons in the forebrain, we used the labels NB-temporal (NB-T) and NB-spectral (NB-S); all NB neurons in the midbrain are therefore labeled NB-T. Two smaller functional groups were apparent. Six neurons had STRFs showing both wide spectral and temporal tuning and were labeled wideband (WB). These neurons were not intermediates of the neurons classified as NB-T and the slower BB neurons for two reasons. First, although the spectral bandwidth of WB neurons was between that of NB and BB neurons, their temporal widths were found at the tail (i.e., showed long integration times) of the distributions of temporal widths for NB and BB neurons (see below). Second, WB neurons showed a different arrangement of inhibitory and excitatory regions; WB neurons have spectral inhibitory sideband, whereas NB-T and BB neurons have temporal inhibitory bands. Finally, five neurons had complex STRFs that were characterized by two distinct excitatory regions followed in time by an intermediate inhibitory area. These neurons were labeled as two-band (2B). Some of the neurons classified as WB also had two excitatory peaks along the spectral dimension, but, in WB neurons, the second peak was weaker and the temporal bandwidth was much longer. Given the small number in each group, however, it is possible that these two groups correspond to a continuum.

Based on the sample of 110 neurons, we estimated the relative proportions of broadband and narrowband units in MLd. Among neurons that clustered well, 23 ± 8% (at the 95% confidence level) were NB-T and 67 ± 9% were BB. Cells from the other two groups were too few to estimate their relative proportions. Twenty-seven (of 137, or ~20 ± 7%) neurons did not fall into a functional group. Most of these cells showed unique spectrotemporal tuning patterns that did not match that of other cells.

The NB and BB neurons were subdivided into subgroups to (1) describe the range of observed tuning, (2) compare the tuning in MLd and field L, and (3) relate spectrotemporal tuning to the extraction of acoustic features important for analyzing song and mediating the percepts of pitch, timbre, and rhythm. This subgroup classification is subjective and does not imply the existence of separate functional subclusters; the tuning across subgroups varies continuously, although we did use discontinuities in our data (see Materials and Methods) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material) to set subgroup boundaries. In MLd, we further classified BB neurons into a large subgroup of faster broadband neurons [BB(F)] (66 of 74) and a smaller subgroup of slow broadband neurons [BB(S)] (8 of 74). Although the mode of the distribution for the temporal width of BB neurons was similar to those of the NB neurons, the distribution for BB neurons exhibited a long tail for long integration times; a small number of BB neurons had much longer temporal widths than did the other STRFs in the cluster. These slower neurons [BB(S)] formed a separate group in the clustering matrix, although it is not clear that they make a discontinuous group from the BB(F) neurons; the tuning property distributions (see Figs. 3, 4) of all BB neurons are unimodal, and there is relatively high similarity between the BB(S) neurons and the BB(F) neurons in the off-diagonal elements of the similarity matrix of Figure 1. The STRFs for the entire dataset (organized according to functional group) are shown in supplemental Figure 4 (available at www.jneurosci.org as supplemental material).

Functional groups in the forebrain

In the forebrain, 105 of 137 (77%) neurons clustered into five major groups (Fig. 2). The diversity of tuning patterns in field L was larger than in MLd. We found two groups in field L that were not observed in MLd, “hybrid” and “offset,” and the spectrotemporal tuning properties of neurons within the NB and BB clusters were more varied in field L than in MLd. This difference between the midbrain and the forebrain is shown in the similarity matrices of MLd (Fig. 1) and field L (Fig. 2) and by comparing all the STRFs for each area shown in supplemental Figures 4 and 5 (available at www.jneurosci.org as supplemental material). As in the midbrain, the two major functional groups in field L are NB and BB neurons. Thirty-five of 105 clustered field L neurons were in the NB group. Forty-one of the 105 neurons clustered into the BB group. As in MLd, the broadband neurons in field L were distinct in spectral tuning from the narrowband neurons. This property can be seen in the histogram of spectral tuning properties in Figure 3: all NB neurons are concentrated below one narrow peak in the spectral distribution at small bandwidths, whereas BB neurons occupy a broad second peak for higher bandwidths.

A group of WB (12 of 105) neurons, characterized by slow temporal tuning and intermediate spectral bandwidth, was also found in field L. As in MLd, these neurons have no (or weak) inhibition after the excitation, distinguishing them from the slower BB neurons. In WB neurons, inhibition is found along the spectral dimension. In field L, some of the WB neurons showed a larger spectral bandwidth than those observed in MLd [compare WB STRFs in MLd and field L in Figs. 5, 6 and in supplemental Figs. 4, 5 (available at www.jneurosci.org as supplemental material)].

Two functional groups were observed in field L but not in MLd. A small group of clearly distinct neurons (7 of 105) were broadband offset neurons (Off), with the excitation after inhibition. The final group showed the properties of both slow NB-T and fast or medium BB neurons combined and were thus called hybrid (Hy) (10 of 105 neurons). In Hy neurons, the tuning for lower frequencies is similar to that of the slow NB-T neurons, whereas the tuning for higher frequencies is similar to the fast BB neurons. Because these neurons showed combined (i.e., summed) responses rather than intermediate responses, we classified them as a separate group. The neurons showing two excitatory regions (2B neurons) observed in MLd were not found in field L.

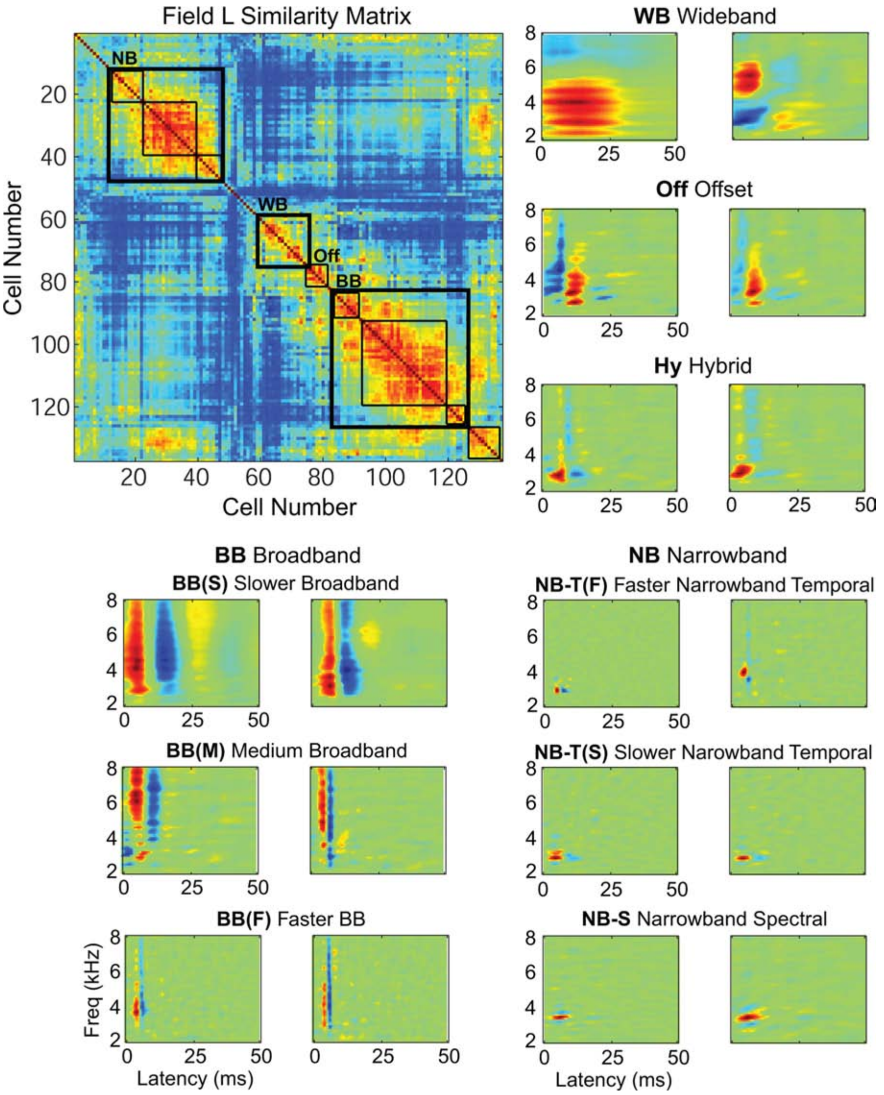

Figure 2. The functional groups in the auditory forebrain region field L. The top left shows the matrix of the similarity indices between each pairwise comparison after the cells were sorted using the genetic algorithm and objective function. The neurons clustered into five groups. The two larger groups are BB and the NB. The BB group was subdivided into BB(S), BB(M), and BB(F) neurons. The NB group was subdivided into NB-T, NB-T(S), and NB-S neurons. As in MLd, a group of WB was also observed. Two smaller groups were also identified and labeled as Off and Hy neurons. The STRFs of neurons at opposing edges of each group are shown to illustrate the type and range of STRFs found in each cluster. The STRFs for all the neurons in field L are shown in supplemental Figure 5 (available at www.jneurosci.org as supplemental material).

Based on 105 classified neurons, we can estimate the relative proportions of BB and NB units in field L. Among neurons that clustered, 33 ± 9% (at the 95% confidence level) of neurons are NB and 39 ± 9% of neurons are BB. The other three groups had too few cells for a reliable proportion estimate. In summary, in field L, approximately one-third of the clustered neurons showed NB tuning, slightly more than one-third showed BB tuning, and another one-third was divided into three groups of approximately equal size.

As for MLd, it was useful for descriptive purposes to further subdivide the NB and BB neurons into subgroups. The STRFs for a majority of the NB neurons in field L (27 of 35) showed inhibitory regions immediately after excitation, as was the case for all the NB cells in MLd, and were also labeled NB-T. In field L, we further subclassified the NB-T neurons into fast (10/27 neurons) [NB-T(F)] and slow (17 of 27) [NB-T(S)] neurons. We also found NB neurons in field L (but not in MLd) that had sharp excitatory spectral tuning and equally sharp inhibitory side bands. This subset (8 of 35) of NB neurons may therefore be useful to encode sharp spectral derivatives, such as spectral edges found in harmonic stacks. For this reason, we labeled these neurons NB-S, to designate narrowband spectral tuning. Most of the NB-S neurons lacked the inhibitory region that followed the excitatory region in time, as was typical for the NB-T neurons.

We found more variability within the BB cluster in field L than in MLd (compare the similarity matrices in Figs. 1, 2). To describe some of this variability, we subdivided the BB group into BB(S) (8 of 41 neurons), medium [BB(M)] (27 of 41 neurons), and BB(F) (6 of 41 neurons) subgroups. The boundaries of these subgroups were visually set based on discontinuities in the similarity matrix (see Materials and Methods) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material), and the neurons were labeled along the slow–medium–fast continuum after observing that the biggest difference between the subgroups was in temporal tuning (see below) (Fig. 3). As for MLd, this subclassification is somewhat arbitrary because the group of BB neurons shows a unimodal distribution of temporal tuning properties (Fig. 3). The subclassification was done because it describes the range of temporal properties observed, which allowed

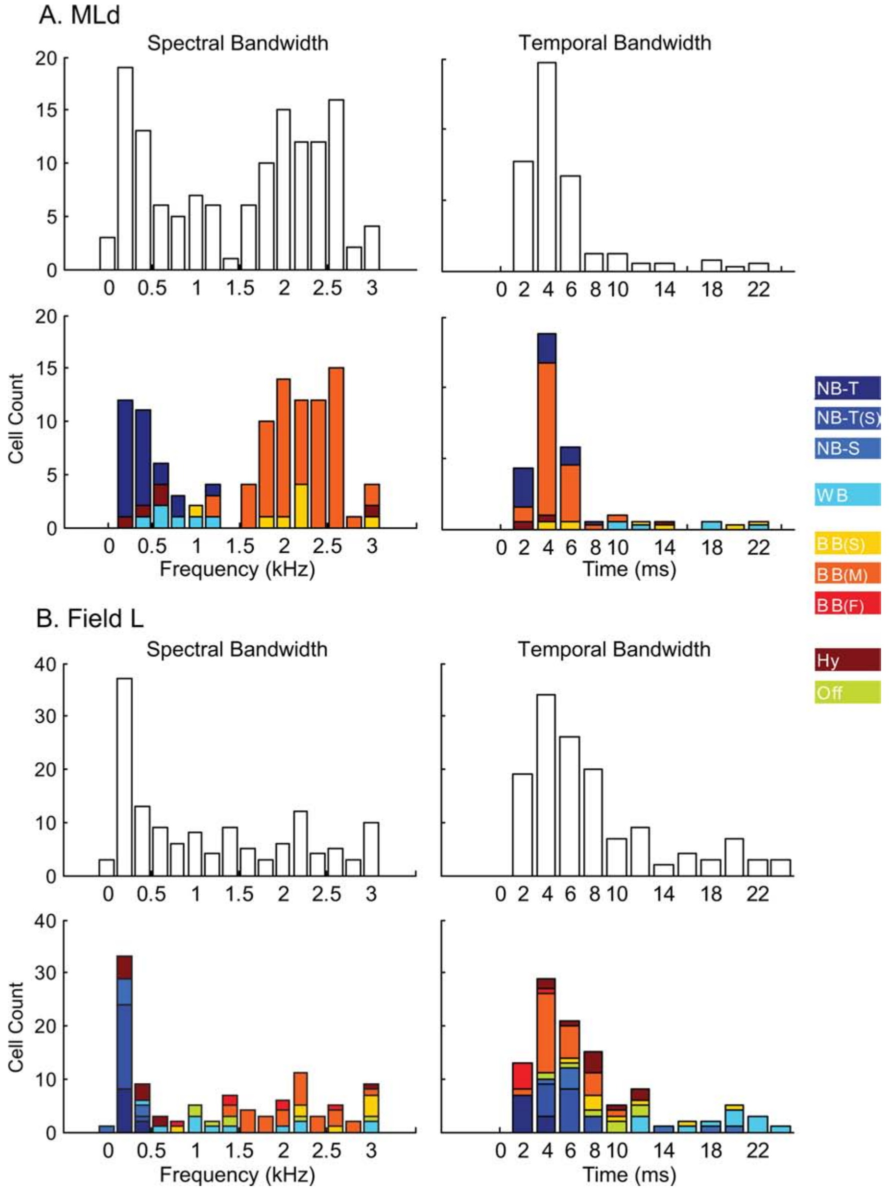

Figure 3. Distribution of temporal and spectral STRF widths for all neurons and across functional groups for MLd (***A***) and field L (***B***). The temporal width of the STRF was defined as the width of the Gaussian in a Gabor fit of the temporal profile of the STRF obtained by SVD. This parameter measures the extent of the STRF in time or the integration time of the neuron. The spectral shape of the STRF was fitted by a Gabor, and the width of the Gaussian was used. This width measures the frequency tuning bandwidth of the neuron and includes both excitatory and inhibitory regions. The top panels in ***A*** and ***B*** show histograms of the spectral bandwidth (spectral sigma) and temporal bandwidth (temporal sigma) for all neurons. The bottom panels show the distribution for all neurons that clustered into one of the functional groups shown in Figures 1 and 2.

us to compare responses in the midbrain and forebrain and to investigate the sound features that can be encoded by the fast, medium, and slow BB neurons.

Spectral and temporal tuning properties of functional groups

After the clustering and functional classification based on STRF shape, we extracted the separate spectral and temporal tuning properties of each neuron and examined their distributions in each functional group. The spectral and temporal properties were measured by fitting the separable portion of the STRF obtained by singular value decomposition with a product of Gabor functions (see Materials and Methods) (Qiu et al., 2003). The width of the temporal or spectral tuning was then quantified by the σ (the SD parameter) of the Gaussian in the Gabor function. We also estimated the best spectral modulation and best temporal modulation frequencies for each neuron by obtaining the modulation transfer function (MTF) for each STRF. Figure 3 shows the histograms of the distribution of σ for the entire dataset of STRFs and for each functional group. Figure 4 shows scatter plots of the joint distributions of spectral and temporal bandwidths and spectral and temporal best modulation frequencies.

Figure 3, *A* and *B* (top left panels), shows the distributions of spectral σ bandwidth for MLd (*A*) and field L (*B*) STRFs. In both brain regions, spectral tuning exhibits a bimodal distribution with a mode at sharp bandwidths (200–400 Hz) and a mode at larger bandwidths (2–2.5 kHz). Figure 3, *A* and *B* (bottom left panels), show the distributions of spectral bandwidths for clustered cells in MLd and L, color coded according to functional group. The two large clusters of neurons in field L and MLd, the NB (darker blues) and BB (yellow–orange) neurons, occupy extremes or modes of the distribution of spectral bandwidths. In MLd, the average bandwidth was 400 Hz for the NB-T neurons and 2.2 kHz for the BB neurons. This difference is highly significant (two-sample *t* test, $p < 10^{-6}$; $t_{(97)} = 18.4$). In field L, the average spectral bandwidth was 200 Hz for the NB neurons and 2.2 kHz for the BB neurons, also a highly significant difference ($p < 10^{-6}$; $t_{(74)} = 17$). The wideband neurons (light blue) had intermediate spectral bandwidths: an average of 750 Hz in MLd and 2.7 kHz in field L. Note, however, that the σ spectral bandwidth of the WB neurons includes both the excitatory and inhibitory region because these occur simultaneously in time. WB neurons are also characterized by long integration times. As mentioned above, the WB neurons are not intermediates of the BB and NB neurons; they compose a separate group, with intermediate spectral bandwidth and the longest temporal bandwidth. This separate group of WB neurons is best seen in the scatter plots of joint spectrotemporal properties (Fig. 4). The offset neurons in field L (light green) also have broadband spectral tuning with a distribution of bandwidths that overlaps with that of the BB and WB neurons (mean spectral $\sigma = 1.6$ kHz). The complex spectral tuning of the hybrid and two-band neurons could not be well captured with single Gabor functions.

Contrary to spectral tuning, the distribution of temporal bandwidth was unimodal but characterized by a very long tail toward longer integration times (Fig. 3, right panels). The temporal bandwidth distributions of the two large groups of neurons, BB and NB, overlapped significantly with small to insignificant differences in mean. In MLd, the mean temporal σ bandwidth was 3.7 ms for NB-T neurons and 5.3 ms for BB neurons [BB(F)

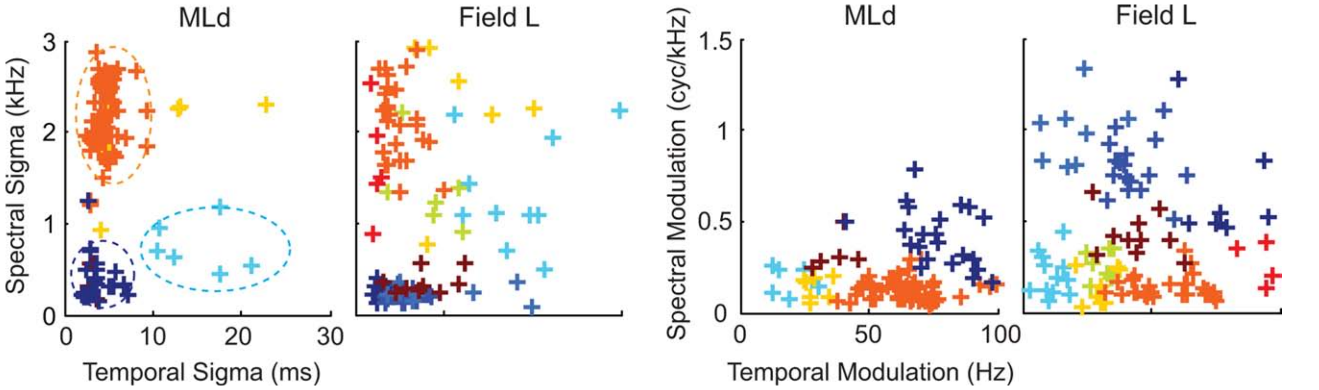

Figure 4. The left plots show the joint spectral and temporal bandwidth distribution for all clustered neurons in MLd (left) and in field L (right) in scatter plots of temporal bandwidth versus spectral bandwidth. The right panels show the scatter plots for the best spectral and temporal modulation frequency. Each cross corresponds to one neuron. The color of the cross indicates the functional group using the same color scheme as in Figure 3. As in Figure 3, the temporal and spectral bandwidth of the STRF was obtained from an SVD decomposition followed by a Gabor fit. The points corresponding to the broadband, narrowband, and wideband neurons cluster in different regions.

and BB(S) combined], a small but statistically significant difference ($p = 0.02$; $t_{(97)} = 2.37$). In field L, the mean temporal σ bandwidth was 5.9 ms for NB neurons [NB-T(F), NB-T(S), and NB-S] and 5.8 ms for BB neurons [BB(F), BB(M), and BB(S)]. These measures were statistically indistinguishable ($p = 0.94$; $t_{(74)} = 0.08$). The WB neurons had larger temporal bandwidths and were therefore only found in the tail of the distribution; the mean temporal σ bandwidth was 15 ms for MLd WB neurons and 18.8 ms for field L WB neurons. As expected, in the two large groups of NB and BB neurons, the slow and fast subgroups also separated along the temporal dimension.

Both in MLd and in field L, we found BB neurons with much longer temporal bandwidths relative to the median, and, as described above, it was useful for descriptive purposes to subclassify these neurons into a subset of BB(S). The temporal bandwidths of the BB(S) neurons are all found in the tail of the distribution and are similar to those of the WB neurons. The mean temporal σ bandwidth of BB(S) neurons was 11.0 ms in MLd and 11.1 ms in field L [vs 4.5 ms for BB(F) neurons in MLd and 4.5 ms for the BB(F) and BB(M) neurons in field L]. The NB-S neurons in field L also had wider temporal bandwidths than the NB-T: 9.9 vs 4.7 ms for the NB-T neurons in L and 3.7 ms for the NB-T neurons in MLd. The mean temporal bandwidth for all field L neurons (8 ms) was significantly longer than that of MLd neurons (5.3 ms; $p < 0.0001$; $t_{(272)} = 4.46$). This difference is relatively small compared with the much larger differences in integration times that have been reported in the analogous regions in the mammalian auditory system (Ter-Mikaelian et al., 2007).

We also analyzed the relationship between stimulus driven firing rate and cell group in MLd and field L. No systematic relationship was found; low and high firing neurons were found in all major functional groups (supplemental Fig. 3, available at www.jneurosci.org as supplemental material). There was a weak positive correlation between temporal bandwidth and mean firing rates in MLd ($r = 0.35$; $p < 10^{-3}$) but not in field L.

Anatomical organization of functional groups in field L

The anatomical distribution of neurons from different functional groups within the subregions of field L is shown in Table 1. Two results were found. First, the three larger groups (BB, WB, and NB) were equally represented in all areas (χ^2 test, df = 6; $p =$ 0.39). However, all offset neurons were found in L2b and most hybrid neurons were found in L2a (83%). Because of this large uneven distribution of offset and hybrid neurons, a χ^2 analysis that included all five major functional groups rejected the null hypothesis of equal distribution (χ^2 test, df = 12; $p = 0.007$). The χ^2 analysis for the NB subgroups also showed a pattern of anatomical distribution; the NB-T neurons (fast and slow) were not found in L1 but were found in higher than expected numbers in L2, L2b, and L3, whereas NB-S neurons were found in L1, L2a, and L2b and in lower than expected numbers in L3 (χ^2 test, df = 6; $p = 0.03$). Fast, medium, and slow BB neurons were distributed evenly among the different subregions (χ^2 test, df = 6; $p = 0.27$). A larger sample size would be required to make these results more conclusive.

Table 1. Distribution of cells in different functional groups across the different subregions of field L

	L1	L2a	L2b	L3
NB	2 (3.1)	11 (11)	8 (11.4)	10 (5.5)
WB	1 (0.9)	3 (3.2)	3 (3.3)	2 (1.6)
BB	4 (2.6)	9 (9.2)	11 (9.5)	2 (4.6)
Offset	0 (0.7)	0 (2.5)	7 (2.6)	0 (1.2)
Hybrid	1 (0.6)	5 (2.1)	0 (2.2)	0 (1.1)

Neurons were excluded from this analysis if they fell very close to subregion boundaries. The number in parentheses shows the expected value given the null hypothesis that neurons belonging to each functional group are found in equal proportion in each region.

Coding of sound features

The roles of different functional groups and subgroups in coding the acoustic features of song and their potential roles in mediating different percepts were investigated by examining the population code of each functional subgroup (Figs. 5, 6). Here, the population code was defined as the sum of the neural responses in each subgroup. This population code was studied in three ways: (1) calculating the average STRF for the subgroup; (2) comparing the predicted responses given by the average STRFs with actual responses and relating the responses to the acoustic features of song; and (3) calculating the ensemble modulation transfer function (eMTF) and relating it to the modulation power spectrum (MPS) of zebra finch song (Woolley et al., 2005).

The average STRF was obtained by averaging after aligning each STRF with the most typical STRF of the subgroup. The most typical STRF was the STRF that was the most similar to all the STRFs in the subgroup, as determined by cross-correlation. The alignment was done by shifting the best frequency and latency of the STRF. Average STRFs are shown in the left columns of Figures 5 and 6. The actual population response was obtained by sum-

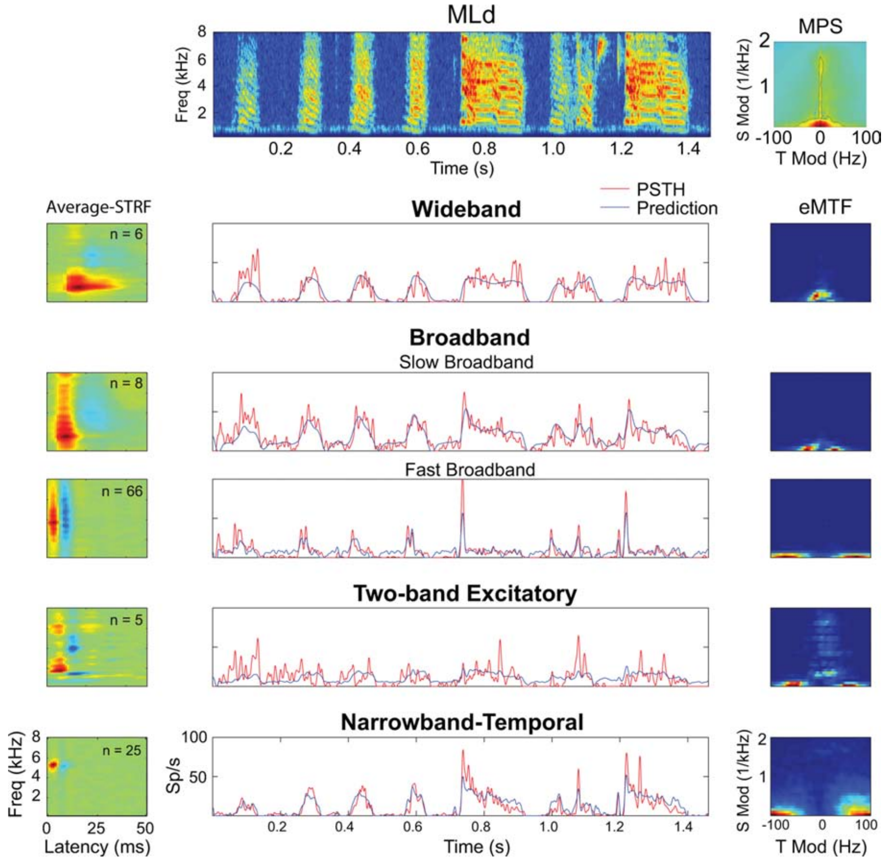

Figure 5. Population responses, average STRFs, and eMTFs for all neurons in each MLd functional group. The average STRFs shown on the left are obtained by averaging the individual STRFs after adjusting for differences in latency and best frequency. The *n* shown in each average STRF shows the number of neurons in each group. The plots in the middle panels show the average predictions (blue) and the average PSTHs (red) for all neurons in each group in response to the song shown as a spectrogram above. The average prediction is obtained by averaging the predictions obtained from each STRF, not by using the average STRF. The right plots show the eMTFs obtained by averaging the MTF of each STRF in the cluster. The MPS of zebra finch song is shown to the right of the spectrogram for comparison.

ming the responses of all neurons in each subgroup to a song. The predicted population response was obtained by summing the predicted response obtained by convolving the STRF of each neuron with the spectrogram of the song. Actual and predicted population responses are shown as PSTHs in the middle panels of Figures 4–6, under the song spectrograms. The match between the prediction and the actual responses and the similarity between the response and the amplitude envelope of the song were quantified by cross-correlation (Tables 2, 3).

The MTF shows the gain of the STRF in terms of spectral and temporal modulations (Depireux et al., 2001; Miller et al., 2002; Woolley et al., 2005). Ensemble MTFs are obtained by calculating the MTF for each STRF and then averaging those MTFs for all cells in a group. To quantify the spectrotemporal modulations in the songs, we calculated the MPS: the power of the log spectrogram for joint spectral and temporal frequency modulations (Singh and Theunissen, 2003). The eMTF of each average STRF was then compared with the MPS of zebra finch song to evaluate the match between the neural tuning and the spectrotemporal features of song (Woolley et al., 2005).

Neurons with fast and medium BB tuning code fast amplitude transients and features important for rhythm. These cells encode syllable onset with a high degree of temporal precision as illustrated in their responses to the example song shown in Figure 5 (third row) and Figure 6 (eighth and ninth rows), generating a precise neural representation of the temporal sequence of the syllables. These acoustic cues convey the tempo or rhythm of song. As indicated by the average STRF shapes that show little spectral tuning and fast temporal tuning, the eMTFs of the BB(M) and BB(F) neurons show very low spectral modulation tuning and bandpass temporal modulation tuning, encoding temporal modulations up to ~100 Hz.

Examining the responses to song addresses the issue of why some neurons are tuned to very fast temporal modulations (50 to 100 Hz) when sounds with such high repetition rates could be perceived as continuous (Klump and Okanoya, 1991; Dent et al., 2002). The perceived temporal patterns in vocalizations occur at much lower temporal modulation frequencies, corresponding to the sequence of syllables. In zebra finch song, the syllable rate is an average of 7.5 Hz. The onsets of sounds such as syllables are very fast. Neurons with high temporal precision and low spectral selectivity function to reliably and accurately encode syllable onsets and show intersyllable intervals that match syllable rates. There are small but important differences between the responses of BB(F) and BB(M) neurons. As illustrated in Figure 6 (compare seventh and eighth row), the responses of BB(F) neurons in field L are mostly limited to syllable onsets, whereas the BB(M) neurons may also detect the onsets of different notes in a syllable. BB neurons can therefore not only mark the onset of each syllable but also the fast transitions within syllables.

Off neurons code silent intervals and therefore song rhythm. The few Off neurons found in field L encode the ends of syllables with high firing rates (Fig. 6, row 9). The responses of these neurons therefore mirror the responses of BB neurons to onsets. Together, the two types of responses code syllable onset and offset. In music perception, the coordinated responses of BB and Off neurons could extract tempo and distinguish short notes (as in a staccato) from long notes (as in a tenuto). Offset neurons had longer integration times and therefore lower temporal modulation tuning peaks than did the BB(M) and BB(F) neurons. One functional explanation for this difference may be that syllable onsets are much sharper than decays. Thus, longer integration times may be useful for coding syllable offset.

Slower BB and WB neurons code amplitude envelope and timbre. These cells exhibited sustained responses to syllables, after the shape of the amplitude envelope of the sound (Figs. 5, 6). Such sustained responses are indicated by the long integration times observed in the average STRFs. This long integration time can be contrasted with the much shorter integration time observed in the average STRFs for BB(M) and BB(F) neurons. This difference in temporal tuning is also observed in the eMTF; peak gain is found at progressively lower temporal modulation frequencies when comparing BB(F), BB(M), BB(S), and WB neurons. Similar to the BB(F) and BB(M) neurons, the BB(S) and WB cells have relatively broad spectral tuning and therefore respond to most sounds. The combination of these two properties makes BB(S) and WB neurons good envelope extractors. Small numbers of BB(S) or WB neurons can create an accurate representation of the amplitude envelope (Tables 2, 3). In MLd, we found a correlation of 0.87 between the average population response of five BB(S) neurons and the sound envelope. The correlation between the responses of BB(F) neurons, which follow only amplitude onsets, was 0.58. Amplitude envelope coding differed significantly among functional groups; the mean *r* values for the four functional groups for which we had five or more neurons were statistically different (balanced ANOVA, $F_{(3,71)} = 95$; $p < 10^{-6}$). In field L, five WB neurons yield a representation of the sound envelope that has an $r = 0.85$, whereas five BB(M) have $r = 0.57$. As in MLd, the mean *r* values between actual responses and the sound envelopes were significantly different across functional groups (balanced ANOVA, $F_{(7,137)} = 121$; $p < 10^{-6}$).

The eMTF of the BB(S) group shows tuning for low spectral modulations and bandpass tuning for temporal modulations between ~5 and ~30 Hz. Although the tuning of these neurons is comparatively "slow," BB(S) neurons encode modulation changes in the envelope that would be considered to be very fast perceptually (up to 30 Hz) for birds (Klump and Okanoya, 1991; Dent et al., 2002) and humans (Krumbholz et al., 2000). When fast modulations affect the amplitude envelope shape, these smaller and faster envelope fluctuations are perceived as textural qualities or timbre; the "attack" and "decay" times for musical notes convey timbre (Risset and Wessel, 1999; Caclin et al., 2005). These amplitude fluctuations are well represented in the population response of BB(S) and WB neurons.

WB neurons are on average tuned for slower temporal modulations than are BB(S) neurons, as seen in the eMTF and in the average STRFs for each group. In addition, WB neurons show spectral bandpass tuning, whereas the SBB units show spectral low-pass tuning. Therefore, WB cells can follow the amplitude envelope of song and discriminate between sounds with high- and low-frequency content. This is best illustrated by the WB neurons in field L with narrower spectral bandwidth. As shown in Figure 6 (fifth row, WB-EI), these neurons respond strongly to the first note of the long syllable in the song (repeated twice, at ~0.8 and 1.3 s). This note is characterized by more energy in the higher-frequency range. Thus, the responses of WB neurons can also code the coarse spectral envelopes of syllables, an important acoustic feature for timbre perception in music (Risset and Wessel, 1999; Caclin et al., 2005) as well as for formant perception in speech (Elliott and Theunissen, 2009).

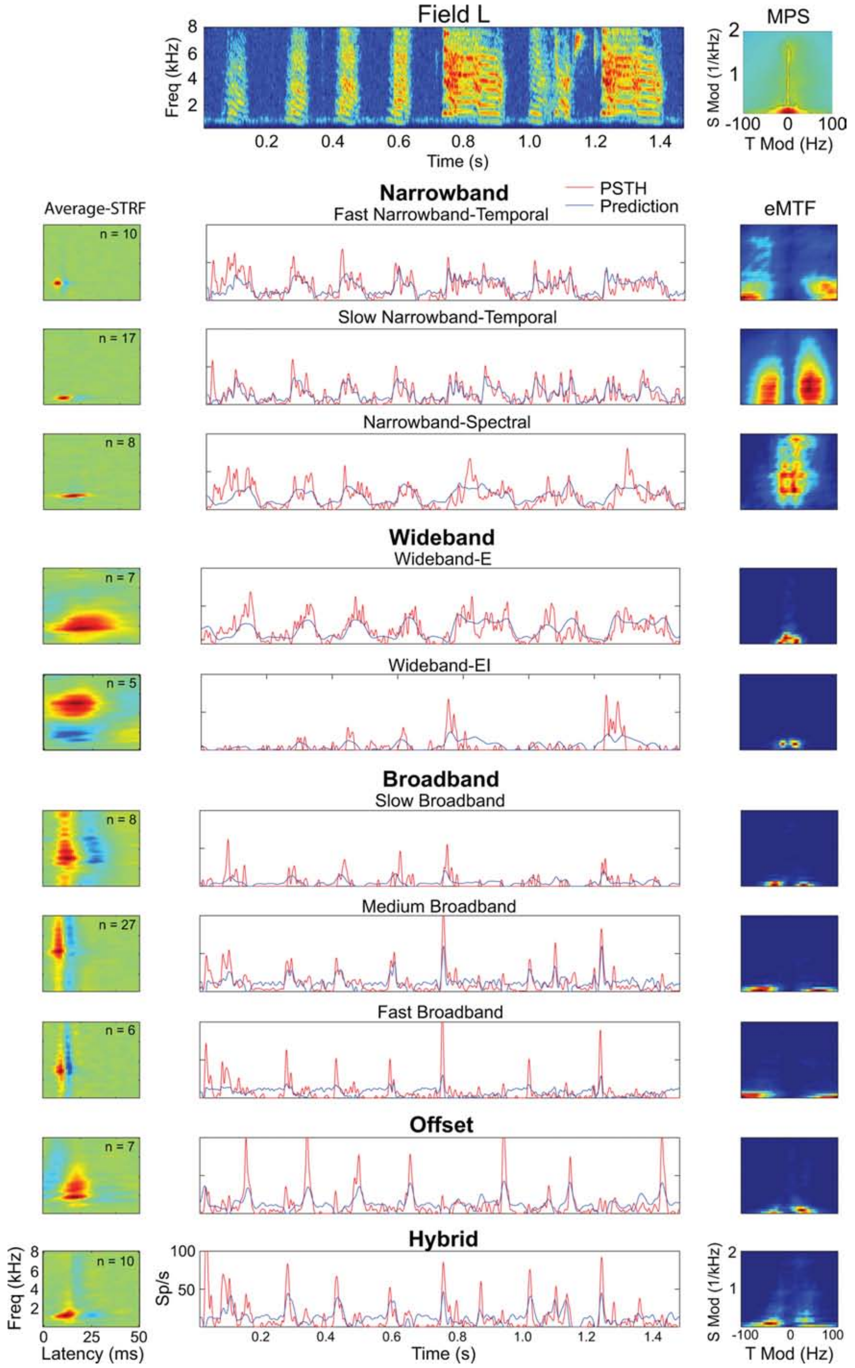

Figure 6. Population responses, average STRFs, and eMTFs for all neurons in each of the field L functional groups. The layout is the same as in Figure 5. Predictions and population responses for narrowband, wideband, broadband, offset, and hybrid functional groups are shown.

NB-T neurons may function as building blocks for additional processing. These cells show sharp spectral and temporal selectivity; their average STRF is narrow in time and frequency. The eMTF for these cells is therefore broad in temporal and spectral modulation tuning. The STRFs of NB-T neurons also have well defined inhibition that immediately follows the excitation. The responses of NB-T neurons show sharp frequency tuning and a temporal response characterized by a strong onset followed by a much weaker sustained response. In response to song, single NB-T neurons can selectively respond to individual notes and therefore act as feature detectors. When the responses of many

Table 2. The average (SD) for the correlation coefficient between predicted and actual responses and between actual responses and the amplitude envelope of the sound for neurons in MLd classified into the five functional groups

Group®	Prediction	Pop. Pred. ($n = 5$)	Pop. Pred. ($n = 10$)	Act versus Env ($n = 5$)	Act versus Env ($n = 10$)
NB-T	0.67 (0.12)	0.72 (0.06)	0.78 (0.04)	0.75 (0.06)	0.80 (0.05)
BB(F)	0.54 (0.16)	0.56 (0.08)	0.57 (0.08)	0.58 (0.09)	0.62 (0.08)
BB(S)	0.60 (0.15)	0.66 (0.07)		0.87 (0.02)	
WB	0.58 (0.21)	0.68 (0.05)		0.83 (0.02)	
2B	0.45 (0.10)				

The columns show the predicted and the actual responses for single neurons (Prediction), between the population predictions and the population PSTH for 5 neurons and 10 neurons (Pop. Pred.), and between the population PSTH and the sound amplitude envelope for 5 and 10 neurons (Act vs Env). Before calculating the correlation coefficient, the PSTH is smoothed with an 11 ms Hanning window. The correlation coefficient for single neurons is corrected by the noise in the estimation of the STRF as described by Hsu et al. (2004a). All predictions are estimated with data not used in the estimation of the STRF. Cells with no values correspond to cases in which there were not enough neurons in that category.

Table 3. Same as Table 1 but for neurons in field L

Group	Prediction	Pop. Pred. ($n = 5$)	Pop. Pred. ($n = 10$)	Act versus Env ($n = 5$)	Act versus Env ($n = 10$)
NB-T	0.51 (0.16)	0.53 (0.08)	0.55 (0.09)	0.77 (0.03)	0.80 (0.02)
NB-T(S)	0.63 (0.27)	0.63 (0.08)	0.72 (0.03)	0.70 (0.07)	0.78 (0.03)
NB-S	0.55 (0.14)	0.47 (0.11)		0.79 (0.05)	
BB(S)	0.41 (0.12)				
BB(M)	0.45 (0.15)	0.47 (0.10)	0.52 (0.10)	0.57 (0.06)	0.59 (0.05)
BB(F)	0.37 (0.17)	0.38 (0.07)		0.50 (0.04)	
WB	0.43 (0.15)	0.48 (0.08)		0.85 (0.02)	
Offset	0.50 (0.13)	0.59 (0.07)		0.63 (0.04)	
Hybrid	0.43 (0.14)	0.47 (0.07)		0.57 (0.02)	

For abbreviations, see Table 2.

Table 4. Classification of the response properties of neurons according to functional grouping, the major contour versus texture grouping proposed by Eggermont (2001), and in terms of their potential roles in mediating auditory sound qualities important in human auditory perception

Functional group	Contour/texture	Perceptual quality
BB(F) and BB(M)	Contour	Rhythm
BB(S)	Contour	Timbre
WB	Contour and texture	Timbre
NB-T	Contour and texture	General
NB-S	Contour and texture	Spectral pitch

NB-T neurons are summed, as shown in Figures 5 and 6, the resulting population response follows the amplitude envelope of the sound (Tables 2, 3). Combinations of NB-T neurons could be used to build other responses such as those seen in the other functional groups.

NB-S neurons code the slower harmonic sounds and could be used to extract spectral pitch. In contrast to the NB-T neurons, the NB-S neurons found in field L show inhibition that flanks the excitation along the frequency axis rather than the temporal axis. One strong feature is the overall sharpness of both the excitatory and inhibitory regions along the frequency axis; the overall frequency bandwidth (including the inhibitory side band) is similar to the NB-T neurons, with a mode ~200 Hz. However, NB-S neurons have a longer integration time than the NB-T neurons (Figs. 3, 4). Because of this longer integration time and lack of temporal inhibition, the response of NB-S neurons to song is more sustained than for the NB-T or the faster BB neurons [BB(F) and BB(M)]. Also because of their sharp spectral inhibition, the peak response is evoked by notes that show a high degree of harmonic structure rather than at the onsets of syllables.

The eMTF of NB-S neurons is nearly orthogonal to the tuning observed for BB neurons: NB-S cells are tuned exclusively for slow temporal and high spectral modulations (along the y-axis in the MTF), whereas BB neurons are tuned to higher temporal and low spectral modulations (along the x-axis in the eMTF). The areas of high gain in the eMTF for the NB-S neurons correspond to the areas in the modulation spectrum of sound for which the energy from the relatively long harmonic stacks in song is strong; the best spectral modulation frequencies in NB-S neurons were between 0.8 and 1.4 cycles/kHz, corresponding to spectral pitch between 750 and 1250 Hz. As shown in Figure 7, the best frequencies for these neurons were always higher than these spectral pitch frequencies. These neurons could encode pitch fundamentals in the range of 750 and 1250 Hz by being maximally excited by the first, second, or third harmonic. Because zebra finch song has more power in the frequency range above 1.5 kHz than below (Hsu et al., 2004b), frequency tuning for the higher harmonics could be advantageous. In humans, spectral pitch is maximally sensitive for energy in the third to fifth harmonics for pitch fundamentals below 400 Hz, but these harmonics are also found below 2000 Hz (Ritsma, 1967). In summary, the spectral modulation tuning of NB-S neurons is appropriate for the detection of harmonic stacks in zebra finch song by detecting spectral modulations corresponding to a fundamental between 750 and 1250 Hz. By analogy, one could envision a similar mechanism to compute spectral pitch in the mammalian auditory system, with the caveat that the best frequency and spectral modulations should match the behavioral and psychophysical data for those species.

Although the majority of STRFs were classified as NB, WB, or BB, we found neurons that shared similar STRFs but were not included in the larger groups. These STRFs illustrate some of the more complex responses that are found in higher auditory regions and whose functions are not well understood. The 2B neurons in MLd showed primary excitatory peaks ~2 and 6 kHz, which are multiples of each other. Their responses to song are phasic, but, as with the NB-S neurons, the peak responses are strongest for particular harmonic sounds rather than for onsets. These neurons may play a role in the recognition of sounds with specific harmonic structure and could therefore also participate in the pitch perception of spectrally complex sounds. The Hy neurons were characterized by major excitation in the low frequencies and weaker broadband excitation followed by a stronger broadband inhibition. The Hy neurons were also good onset detectors. Similar to the BB neurons, the responses of neurons with Hy tuning also showed peaks at note transitions within a syllable.

To illustrate how neurons in different groups and subgroups capture different sound features, and how these features correspond to different percepts (summarized in Table 4), we filtered an exemplar zebra finch song by the ensemble MTF for a particular neuron subgroup. The example zebra finch song was filtered by the eMTF of BB(F) neurons, BB(S) neurons, WB neurons, and NB-S neurons. The resulting spectrograms are shown in Figure 8, and the corresponding sounds are online. In the sound filtered by the eMTF of the BB(F) neurons, the rhythm of the song is evident. In the filtered sound using the eMTFs of the BB(S) and the WB neurons, the noisy timbre quality of the zebra finch song is emphasized. In the sound filtered by the eMTF of the NB-S neurons,

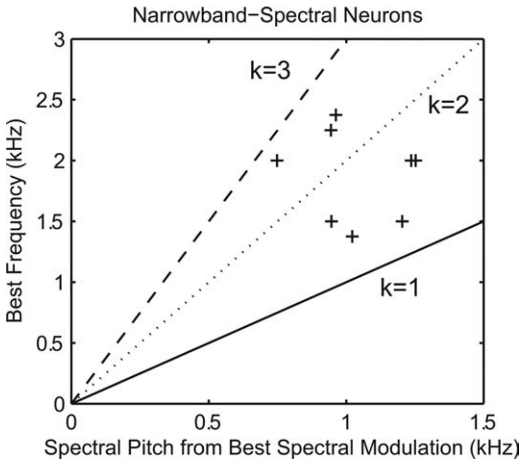

Figure 7. NB-S neurons and spectral pitch. The frequency of the pitch that is obtained from the best spectral modulation frequency (by taking its inverse) is plotted against the best frequency of the neuron obtained from its tuning curve. The best frequency of the neurons is always above the fundamental pitch frequency showing that neurons are responding to the higher harmonics.

the sections of the tonal aspects of the song, the harmonic stacks, are emphasized.

Discussion

A central approach to understanding the neural underpinnings of auditory perception has been to investigate how different information-bearing parameters of sounds can be represented by distinct auditory response types found in auditory neurons (Schreiner et al., 2000; Eggermont, 2001; Shamma, 2001; King and Schnupp, 2007; Wang, 2007). Here, we provide evidence for such a division of labor in the form of functional groups in the avian auditory system. Three major groups emerged: BB, NB, and WB. These groups share some similarities with those found in the mammalian system. The spectral tuning of the WB and NB-S neurons is reminiscent of the sideband inhibition found in some mammalian auditory neurons such as type III neurons in the dorsal cochlear nucleus (Young and Brownell, 1976) and type A neurons in cortex (Shamma and Symmes, 1985). NB-T neurons show responses that are similar to those of primary auditory nerves fibers (Rhode and Greenberg, 1992). Nagel and Doupe (2008) also found that the STRFs of neurons recorded in field L of awake zebra finches could be classified into different types. Neurons that were narrowband in time with temporally contiguous excitation and inhibition, slower neurons that were narrowly tuned in frequency with inhibitory sidebands, and more complex neurons with both spectral and temporal sidebands were found. The first group of neurons overlaps well with the neurons that we classified as NB-T and BB (F or M). The second group of neurons corresponds well to the NB-S neurons described here. Nagel and Doupe did not separate the NB-T from the BB neurons but reported a range of bandwidths (measured in octaves) for the temporal neurons that is consistent with the range of bandwidths (measured in kilohertz) found here. Because spectrotemporal tuning is stimulus dependent in many neurons (Woolley et al., 2006), the STRFs described here are expected to show broader spectral tuning than the STRFs described by Nagel and Doupe; they studied responses to noise, whereas the responses to song were analyzed here. We also found neurons that were broad in both spectral and temporal dimensions (WB neurons). These neurons were not described by Nagel and Doupe and were rarely observed in this study. The two studies are therefore primarily consistent in their descriptions of the major functional cell types in field L.

We analyzed the sound features that are encoded by each functional group, their role in song processing, and their potential roles in auditory perception. In terms of song processing, we found that a large number of neurons are well suited for coding the onsets and offsets of song syllables. The importance of temporal coding in audition has been shown by neurophysiology (Joris et al., 2004; Nagel and Doupe, 2006), in songbird auditory perception (Okanoya and Dooling, 1990; Cynx, 1993), and in human speech perception (Shannon et al., 1995). The temporally precise responses of fast BB neurons could also be important for sound localization (Konishi, 2003) and scene analysis computations, such as echo suppression by precedence (Litovsky et al., 1999).

Zebra finches have been shown in behavioral (Lohr and Dooling, 1998) and neurophysiological (Theunissen and Doupe, 1998) experiments to have a high degree of sensitivity to spectral changes in songs. Neurons in the NB-S group, which can encode spectral bands with high precision, could function in the detection of such changes. Neural tuning for temporal and spectral features of songs may be integrated to achieve the highly accurate discrimination of conspecific songs that has been observed in zebra finches (Cynx, 1993).

The distinct functional groups described here code acoustic features that are important for three major percepts in audition: rhythm, timbre, and pitch. The acoustic features that are essential for these three percepts can be found in different regions of a representation of acoustic space referred to as spectrotemporal modulation space, and the tuning of the functional groups described here map onto distinct regions of modulation space, with little overlap (Fig. 9). Pitch perception depends on features that occur either in harmonic sounds (spectral pitch) or via fast repetitions in inharmonic broadband sounds (periodicity pitch). Sound features for spectral pitch are found close to the *y*-axis of the MPS and at intermediate spectral modulation frequencies. The information for periodicity pitch can be found at either higher temporal modulations along the *x*-axis or higher spectral modulations along the *y*-axis (depending on the time–frequency scale used for estimating the spectrogram and the frequency of the pitch). Rhythm is conveyed by the rate of the modulations of the overall amplitude envelope of the sound, a sound feature that is dependent on its power along the *x*-axis on the MPS at low temporal modulation frequencies. Timbre depends on the shapes of the amplitude envelope and the spectral envelope. The shapes of amplitude and spectral envelopes are reflected as power in the joint low spectrotemporal modulation region of the MPS. Very low temporal modulations are perceived as intensity changes and therefore do not contribute to timbre. Figure 9 (left) shows the partition of the modulation space across the perceptual features of rhythm, timbre, and pitch.

Some of the functional groups showed tuning that was concentrated in the regions of the MPS that correspond to different perceptual features. BB(F) and BB(M) neurons code high temporal modulation frequencies, which are correlated with the lower frequencies that one could consider to be rhythm of song (see Results). These neurons could therefore be important for the representation of rhythm, in addition to other auditory tasks. BB(S) and WB neurons encode the details of the amplitude envelope, an important feature for timbre. In addition, WB neurons perform a coarse spectral analysis, extracting the low spectral modulation frequencies that are important for timbre and formant identification in speech. We found that NB-T and NB-S

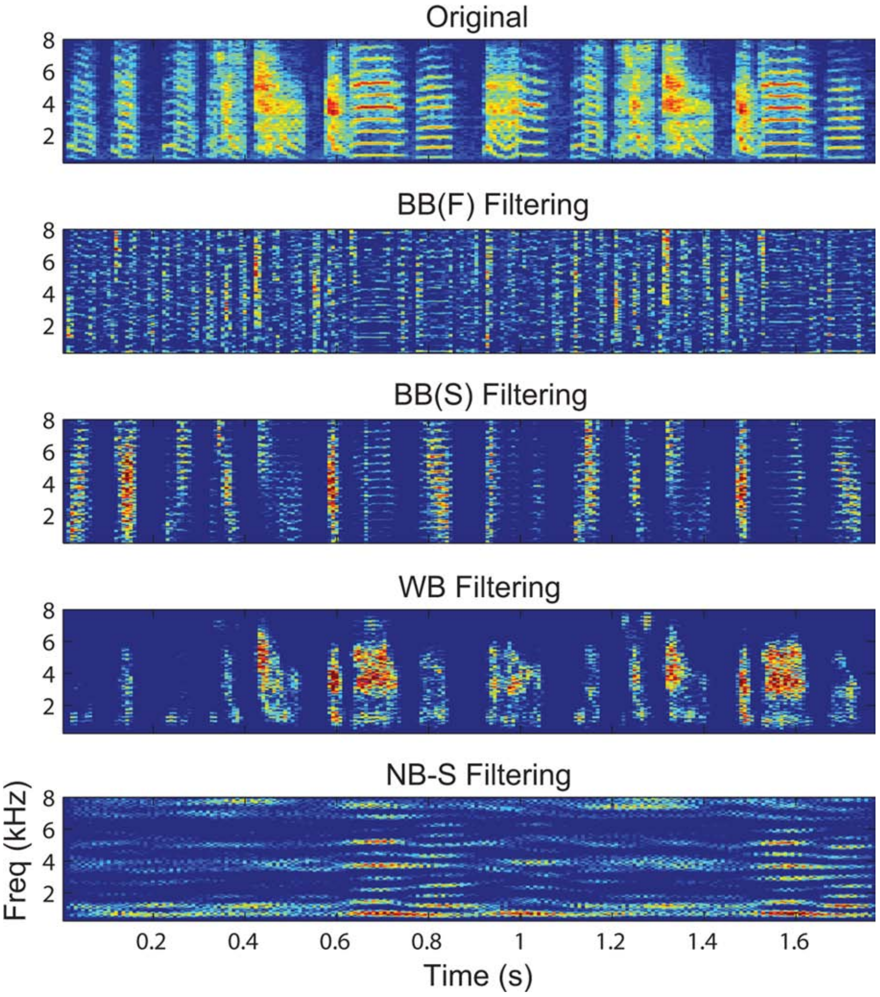

Figure 8. To illustrate the acoustic features in song that are extracted by some of the functional groups, an example zebra finch song was filtered by the eMTF of BB(F) neurons, BB(S) neurons, WB neurons, and NB-S neurons. The figure shows the spectrogram of the original song and filtered versions of the song. The original song can be heard online as song.wav, and the filtered songs can be heard as song_BBF.wav, song_BBS.wav, song_WB.wav, and song_NBS.wav (Audio 1–5, available at www.jneurosci.org as supplemental material).

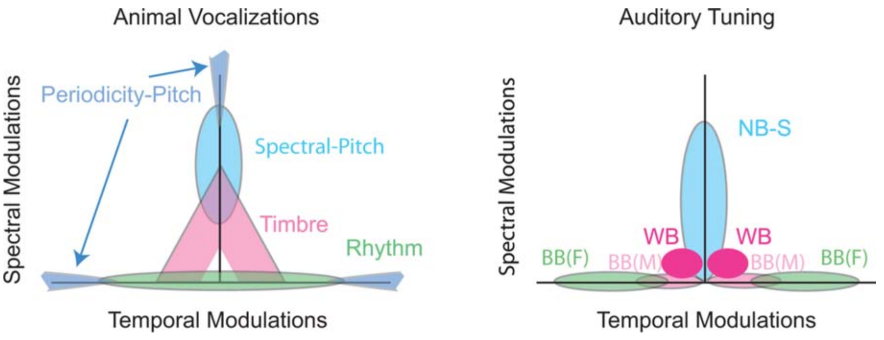

Figure 9. Illustration of the modulation power spectrum of vocalizations (left) and the modulation transfer function of avian auditory neurons (right). The spectrotemporal modulations that convey significant information for the percepts of rhythm, timbre, and pitch are color coded in the left. Acoustic features that carry information for periodicity pitch can be found at the faster temporal modulations and/or the faster spectral modulations. Whether they are represented as temporal or spectral modulations depends on the frequency of the pitch and on the scale of the time–frequency decomposition that is used to calculate the modulations (Singh and Theunissen, 2003). The spectrotemporal modulation tuning of the neurons in the four functional groups are coded with matching colors to indicate their potential importance in the working model for each of the perceptual aspects.

neurons had distinct responses, with NB-S neurons being more sensitive to longer and more harmonic structures. These neurons could therefore function to extract the pitch of complex harmonic sounds by coding the short time–frequency spectrum. This coding of spectral pitch using place information could complement coding for temporal pitch based on tuning to fast amplitude modulation (Hose et al., 1987; Langner, 1992).

A remaining question is whether the functional types observed here correspond to different anatomical types. A strong link between structure and function has been shown in the cochlear nucleus (Young, 1998) and retinal (Puchalla et al., 2005) neurons. Recently, it has also been shown that excitatory and inhibitory neurons in the mammalian auditory cortex have different STRFs (Atencio and Schreiner, 2008). Although we did not observe any firing rate differences across functional groups, it is possible that neurons in different groups correspond to the different cell types that have been described in the avian auditory forebrain (Saini and Leppelsack, 1981; Pinaud and Mello, 2007; Nagel and Doupe, 2008).

We investigated whether there was a correspondence between functional groups and the anatomically defined subregions of field L (Fortune and Margoliash, 1992; Vates et al., 1996). The principal finding is that all offset neurons were in subregion L2b and that most hybrid neurons were in L2a. We also found that L3 contained a high percentage of NB-T neurons, whereas L1 contained many NB-S neurons. Thus, L1 and L3 may function to extract different acoustic features from complex sounds. This finding is consistent with a recent functional magnetic resonance imaging study (Boumans et al., 2007). It is also possible that functional subregions that differ from those defined using histology and anatomical connectivity also exist in the avian forebrain (Gehr et al., 1999; Cousillas et al., 2005). A larger dataset will be required to complete the analysis of the functional anatomy of field L.

The comparison of tuning in midbrain and forebrain neurons suggests that, as auditory information progresses through the auditory processing stream, both the preservation of tuning and the generation of novel, more complex tuning occurs. Similar findings have been reported for the ascending mammalian auditory pathway (Eggermont, 2001; Miller et al., 2001; Chechik et al., 2006) and visual pathway (Mahon and De Valois, 2001; David et al., 2006). Although the reasons for both preserving and generating new responses remain unclear, an interesting hypothesis is based on the idea that the cohesive percept of an auditory object occurs in the cortex (Eggermont, 2001). The cohesive percept is composed of an ensemble of individual perceptual features, each requiring a separate computation. It is possible that some of the relevant acoustic features for perception are coded early in the auditory processing stream. In this case, the basic but perceptually relevant features represented in subcortical structures would be transmitted to the cortical areas with little or no modification. Our data are consistent with a similar hypothesis. The recognition (and possibly the subjective perceptual experience) of complex sounds such as those produced by conspecific song could arise in the forebrain and require that all information about distinct acoustic features be present at that level. The midbrain coding of acoustic features that are relevant for complex sound recognition would therefore be preserved in the auditory forebrain.

References

Atencio CA, Schreiner CE (2008) Spectrotemporal processing differences between auditory cortical fast-spiking and regular-spiking neurons. J Neurosci 28:3897–3910.

Bieser A (1998) Processing of twitter-call fundamental frequencies in insula and auditory cortex of squirrel monkeys. Exp Brain Res 122:139–148.

Boumans T, Theunissen FE, Poirier C, Van Der Linden A (2007) Neural representation of spectral and temporal features of song in the auditory forebrain of zebra finches as revealed by functional MRI. Eur J Neurosci 26:2613–2626.

Caclin A, McAdams S, Smith BK, Winsberg S (2005) Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J Acoust Soc Am 118:471–482.

Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tishby N, Nelken I (2006) Reduction of information redundancy in the ascending auditory pathway. Neuron 51:359–368.

Cohen YE, Theunissen F, Russ BE, Gill P (2007) Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. J Neurophysiol 97:1470–1484.

Cousillas H, Leppelsack HJ, Leppelsack E, Richard JP, Mathelier M, Hausberger M (2005) Functional organization of the forebrain auditory centres of the European starling: a study based on natural sounds. Hear Res 207:10–21.

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

David SV, Hayden BY, Gallant JL (2006) Spectral receptive field properties explain shape selectivity in area V4. J Neurophysiol 96:3492–3505.

DeAngelis GC, Ghose GM, Ohzawa I, Freeman RD (1999) Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. J Neurosci 19:4046–4064.

Dent ML, Klump GM, Schwenzfeier C (2002) Temporal modulation transfer functions in the barn owl (*Tyto alba*). J Comp Physiol A Neuroethol Sens Neural Behav Physiol 187:937–943.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85:1220–1234.

Eggermont JJ (2001) Between sound and perception: reviewing the search for a neural code. Hear Res 157:1–42.

Elliott T, Theunissen FE (2009) The modulation transfer function for speech intelligibility. PLoS Comp Biol, in press.

Escabí MA, Read HL (2003) Representation of spectrotemporal sound information in the ascending auditory pathway. Biol Cybern 89:350–362.

Escabi MA, Schreiner CE (2002) Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. J Neurosci 22:4114–4131.

Fortune ES, Margoliash D (1992) Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). J Comp Neurol 325:388–404.

Gehr DD, Capsius B, Gräbner P, Gahr M, Leppelsack HJ (1999) Functional organisation of the field-L-complex of adult male zebra finches. Neuroreport 10:375–380.

Gill P, Zhang J, Woolley SM, Fremouw T, Theunissen FE (2006) Sound representation methods for spectro-temporal receptive field estimation. J Comput Neurosci 21:5–20.

Hose B, Langner G, Scheich H (1987) Topographic representation of periodicities in the forebrain of the mynah bird: one map for pitch and rhythm? Brain Res 422:367–373.

Hsu A, Borst A, Theunissen FE (2004a) Quantifying variability in neural responses and its application for the validation of model predictions. Network 15:91–109.

Hsu A, Woolley SM, Fremouw TE, Theunissen FE (2004b) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. J Neurosci 24:9201–9211.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577.

Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multipeaked tuning characteristics. J Neurophysiol 89:1603–1622.

King AJ, Schnupp JW (2007) The auditory cortex. Curr Biol 17:R236–R239.

Klump GM, Okanoya K (1991) Temporal modulation transfer functions in the European starling (*Sturnus vulgaris*). I. Psychophysical modulation detection thresholds. Hear Res 52:1–11.

Konishi M (2003) Coding of auditory space. Annu Rev Neurosci 26:31–55.

Krumbholz K, Patterson RD, Pressnitzer D (2000) The lower limit of pitch as determined by rate discrimination. J Acoust Soc Am 108:1170–1180.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115–142.

Litovsky RY, Colburn HS, Yost WA, Guzman SJ (1999) The precedence effect. J Acoust Soc Am 106:1633–1654.

Lohr B, Dooling RJ (1998) Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). J Comp Psychol 112:36–47.

Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131–1138.

Mahon LE, De Valois RL (2001) Cartesian and non-Cartesian responses in LGN, V1, and V2 cells. Vis Neurosci 18:973–981.

Miller LM, Escabí MA, Read HL, Schreiner CE (2001) Functional convergence of response properties in the auditory thalamocortical system. Neuron 32:151–160.

Miller LM, Escabí MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87:516–527.

Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J Neurophysiol 87:1723–1737.

Nagel KI, Doupe AJ (2006) Temporal processing and adaptation in the songbird auditory forebrain. Neuron 51:845–859.

Nagel KI, Doupe AJ (2008) Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. Neuron 58:938–955.

Okanoya K, Dooling RJ (1990) Temporal integration in zebra finches (*Poephila guttata*). J Acoust Soc Am 87:2782–2784.

Pinaud R, Mello CV (2007) GABA immunoreactivity in auditory and song control brain areas of zebra finches. J Chem Neuroanat 34:1–21.

Poon PW, Chen X, Cheung YM (1992) Differences in FM response correlate with morphology of neurons in the rat inferior colliculus. Exp Brain Res 91:94–104.

Puchalla JL, Schneidman E, Harris RA, Berry MJ (2005) Redundancy in the population code of the retina. Neuron 46:493–504.

Qiu A, Schreiner CE, Escabí MA (2003) Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. J Neurophysiol 90:456–476.

Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111–114.

Rhode W, Greenberg S (1992) Physiology of the cochlear nuclei. In: The mammalian auditory pathway: neurophysiology (Popper AN, Fay RR, eds), pp 94–152. New York: Springer.

Risset JC, Wessel D (1999) Exploration of timbre by analysis synthesis. In: The psychology of music (Deutsch D, ed), pp 113–169. San Diego: Academic.

Ritsma RJ (1967) Frequencies dominant in the perception of the pitch of complex sounds. J Acoust Soc Am 42:191–198.

Saini KD, Leppelsack HJ (1981) Cell types of the auditory caudomedial neostriatum of the starling. J Comp Neurol 198:209–229.

Schreiner CE (1995) Order and disorder in auditory cortical maps. Curr Opin Neurobiol 5:489–496.

Schreiner CE, Read HL, Sutter ML (2000) Modular organization of frequency integration in primary auditory cortex. Annu Rev Neurosci 23:501–529.

Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. J Neurophysiol 86:1445–1458.

Shamma S (2001) On the role of space and time in auditory processing. Trends Cogn Sci 5:340–348.

Shamma SA, Symmes D (1985) Patterns of inhibition in auditory cortical cells in awake squirrel monkeys. Hear Res 19:1–13.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am 114:3394–3411.

Suga N (1989) Principles of auditory information-processing derived from neuroethology. J Exp Biol 146:277–286.

Ter-Mikaelian M, Sanes DH, Semple MN (2007) Transformation of temporal properties between auditory midbrain and cortex in the awake Mongolian gerbil. J Neurosci 27:6091–6102.

Theunissen FE, Doupe AJ (1998) Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. J Neurosci 18:3786–3802.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315–2331.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network 12:289–316.

Vates GE, Broome BM, Mello CV, Nottebohm F (1996) Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taenopygia guttata*). J Comp Neurol 366:613–642.

Wang X (2007) Neural coding strategies in auditory cortex. Hear Res 229:81–93.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685–2706.

Woolley SM, Casseday JH (2004) Response properties of single neurons in the zebra finch auditory midbrain: response patterns, frequency coding, intensity coding, and spike latencies. J Neurophysiol 91:136–151.

Woolley SM, Casseday JH (2005) Processing of modulated sounds in the zebra finch auditory midbrain: responses to noise, frequency sweeps, and sinusoidal amplitude modulations. J Neurophysiol 94:1143–1157.

Woolley SM, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci 8:1371–1379.

Woolley SM, Gill PR, Theunissen FE (2006) Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. J Neurosci 26:2499–2512.

Young ED (1998) Cochlear nucleus. In: The synaptic organization of the brain (Sheperd GM, ed), pp 121–157. New York.: Oxford UP.

Young ED, Brownell WE (1976) Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats. J Neurophysiol 39:282–300.