

Published in final edited form as:

Hear Res. 2011 March ; 273(1-2): 123–133. doi:10.1016/j.heares.2010.08.008.

A bird brain's view of auditory processing and perception

Katherine Nagel^{a,b}, Gunsoo Kim^a, Helen McLendon^a, and Allison Doupe^a

Katherine Nagel: Katherine_Nagel@hms.harvard.edu; Gunsoo Kim: gkim@phy.ucsf.edu; Helen McLendon: helen@phy.ucsf.edu; Allison Doupe: ajd@phy.ucsf.edu

^a Depts of Physiology and Psychiatry and Keck Center for Integrative Neuroscience, University of California, San Francisco, 513 Parnassus Ave, San Francisco, CA 94143, USA

^b Dept of Neurobiology, Harvard Medical School, 25 Shattuck St, Boston, MA 02115, USA

Abstract

By studying the primary forebrain auditory area of songbirds, field L, using a song-inspired synthetic stimulus and reverse correlation techniques, we found a surprisingly systematic organization of this area, with nearly all neurons narrowly tuned along the spectral dimension, the temporal dimension, or both; there were virtually no strongly orientation-sensitive cells, and in the areas that we recorded, cells broadly tuned in both time and frequency were rare. In addition, cells responsive to fast temporal frequencies predominated only in the field L input layer, suggesting that neurons with fast and slow responses are concentrated in different regions. Together with other songbird data and work from chicks and mammals, these findings suggest that sampling a range of temporal and spectral modulations, rather than orientation in time-frequency space, is the organizing principle of forebrain auditory sensitivity. We then examined the role of these acoustic parameters important to field L organization in a behavioral task. Birds' categorization of songs fell off rapidly when songs were altered in frequency, but, despite the temporal sensitivity of field L neurons, the same birds generalized well to songs that were significantly changed in timing. These behavioral data point out that we cannot assume that animals use the information present in particular neurons without specifically testing perception.

Keywords

spectro-temporal; timing; frequency; receptive field; songbird; field L

Introduction

Songbirds, like humans, learn their complex, highly individualized vocalizations (songs, Fig. 1A) during a hearing-dependent process early in life, and there has been much study of the brain areas (primarily in males) involved in producing these sounds (Konishi, 1985; Doupe and Kuhl, 1999; Zeigler and Marler, 2004; Mooney, 2009). However, both male and female birds also listen to songs, which are replete with information - about who a bird is, both as an individual and as a local group- and species-member, as well as about a bird's fitness (Searcy and Nowicki, 1999; Stoddard et al., 1991). In addition, both males and females have a life-long capacity to learn to recognize the songs of other individuals (for example

Correspondence to: Allison Doupe, ajd@phy.ucsf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Kroodsma et al., 1982; Nelson, 1989; Nelson and Marler, 1989; Gentner and Hulse, 1998; Vignal et al., 2008). Songbirds thus provide an excellent model both for examining how complex, natural sounds are represented in higher auditory processing areas, and for examining which aspects of these sounds matter for pattern recognition.

Consistent with the importance of sound to songbird vocal behavior, the songbird brain has long been known to contain some of the most complex and selective auditory neurons ever identified, so-called 'song-selective' neurons. These cells respond much more strongly to the bird's own song than to songs of other individuals of the same species (conspecifics) or even the bird's own song played in reverse order (Margoliash, 1983; Margoliash & Fortune, 1992; Lewicki & Konishi, 1995; Doupe, 1997; Mooney, 2000; Rosen & Mooney, 2003). Such neurons are found throughout the set of higher brain areas involved in controlling song (the 'song system'; Nottebohm et al., 1976), including the sensorimotor nucleus HVC (Fig. 1B), and are thought to be important in learning and producing song.

Such extreme selectivity is likely to be generated by a hierarchy of areas that gradually transform responses from simple to complex. Moreover, songbirds must have auditory neurons that are not restricted in their responsiveness to bird's own song, neurons that could function in the many other auditory recognition tasks that birds perform. Candidates for neurons with an intermediate level of selectivity are found in a group of high-level auditory areas immediately afferent to HVC, especially the caudomedial mesopallium (CM) and the caudomedial nidopallium (NCM; Fig. 1B; Gentner, 2004; Mello et al., 2004). Neurons in these areas respond strongly to a variety of naturalistic stimuli including conspecific songs, not just to the bird's own song (Stripling et al., 1997; Gentner and Margoliash, 2003). These areas are intriguing but still poorly understood, with cells that respond to many songs intermingled with cells that respond selectively only to features of a few songs. Furthermore, responses in these areas appear to be strikingly sensitive to recent experience, even in adult birds, adding an additional layer of complexity (Gentner and Margoliash, 2003; Mello et al., 1995; Kruse et al., 2004).

Given the complexity of the intermediate areas, for our experiments aimed at understanding how natural auditory stimuli are represented in the brain, we turned first to the primary auditory area known as field L. This region is analogous to the primary auditory cortex of mammals, and forms a gateway for auditory information to reach forebrain areas involved in song production and recognition (Wild et al., 1993; Fortune and Margoliash, 1995; Fig. 1B). In this article we first review our recent single and multi-unit studies of field L (Nagel and Doupe, 2008), using a rich song-inspired synthetic stimulus and reverse correlation techniques (Eggermont et al., 1983; Depireux et al., 2001; Kim and Rieke, 2001; Miller et al., 2002; Escabi et al., 2003; Theunissen et al., 2000), which allowed us to provide new insights into the organization of this area. Second, we describe how, armed with knowledge about which features of stimuli are important to field L neurons, we asked which features birds rely on behaviorally to decide which song they have heard in a particular task, and how these might relate to the field L representation (Nagel et al., in review).

The organization of time and frequency encoding in field L

Stimulus design

Studies using simple tonal stimuli have identified multiple tonotopic maps in the field L complex (Scheich et al., 1979; Heil and Scheich, 1985; Gehr et al., 1999; Terleph et al., 2006), while anatomical studies have identified three layers (L1, L2, and L3) that differ in their connectivity and cytoarchitecture (Fortune and Margoliash, 1992). Layer L2 receives thalamic input, while L1 and L3 project to higher auditory areas (Wild et al., 1993; Vates et al., 1996). Although stimuli such as tones or noise bursts often drive field L cells well, these

neurons are also extremely responsive to conspecific songs. Songs and simpler synthetic stimuli derived from them, in combination with reverse correlation methods, have been very useful in fleshing out the properties of field L and surrounding regions (Woolley et al., 2005, 2006, 2009; Sen et al, 2001, Hsu et al., 2004, Nagel and Doupe, 2006, 2008).

To systematically study field L we wanted a stimulus that maintained some of the spectro-temporal complexity of song while nonetheless being simple and mathematically well-behaved enough to use with reverse correlation approaches. These methods, in which the neural response to the input is used to estimate the neuron's receptive field, require a stimulus that is free of correlations (which generate artifactual receptive fields) or at least a stimulus in which correlations can be mathematically corrected for (Eggermont et al., 1983; Ringach and Shapley, 2004; Theunissen et al., 2000).

We therefore designed a stimulus that sampled a wide range of time and frequency combinations in an unbiased way. The stimulus consisted of 32 logarithmically spaced frequency bands (Figure 1C, column 1), each modulated independently by a different time-varying amplitude envelope (column 2). These frequency bands were then summed to produce the final signal (column 3). The amplitude envelopes were designed such that the log amplitude of each band was a random Gaussian noise signal with an exponential distribution of frequencies. Frequency bands were Gaussian in log frequency and overlapped by one standard deviation. Full details of the stimulus construction can be found in Nagel and Doupe (2008). As a result of its construction, the frequency content of our stimulus varied randomly and smoothly in time. This can be seen in the spectrogram of the stimulus (Figure 1C, third column, second panel), which shows the intensity of sound at each frequency as a function of time. Frequency peaks in the stimulus are broader and last longer than those in white noise (a segment of white noise is shown for comparison in the bottom panel of the third column, Fig. 1C). Local correlations such as those in our stimulus are found in most natural sounds, including song, speech, and environmental noise (Singh and Theunissen, 2003). They enabled our stimulus to drive field L neurons more effectively and reliably than white noise.

We obtained neurons' spectro-temporal receptive fields (STRFs) by playing 50–100 trials of our stimulus to each cell and calculating the average spectrogram preceding a spike, then “decorrelating” the resulting spike-triggered average (STA). STRFs estimated from true natural stimuli (Sen et al., 2001; Machens et al., 2004; Ringach et al., 2002) can be distorted by high-order stimulus correlations (Sharpee et al., 2006). Our stimulus had a known correlation structure, ensuring that the influence of such correlations on our STRFs could be removed through decorrelation. Decorrelation generally sharpened the shape of the STRF without significantly altering its features. All of our analyses, with the exception of the multi-unit mapping study described later, were performed on decorrelated STRFs.

Types of STRFs

We obtained significant STRFs from more than 80% of the 81 single units we recorded in five birds. Half of the stimulus trials we played to each cell contained unique random Gaussian noise signals in each frequency band, to broadly sample the space of time-frequency combinations, while half contained repeats of the same (randomly assigned) noise segment. Unique segments were used to estimate the STRF for each neuron, while responses to the repeated segments were used to test how well the calculated STRF could predict the actual response of a neuron (using both the STRF and an estimate of non-linear aspects of the neuron's response, see Nagel and Doupe, 2008). The goodness of predictions was assessed by measuring the correlation coefficient between predicted and actual STRFs, so that perfect prediction of the actual firing using the STRF would give a value of 1. Across our entire set of cells, these correlation coefficients averaged 0.52 \pm 0.14, but could range

as high as 0.8 (see Figure 2 for individual examples from the range of correlation coefficients).

Although the shapes of the STRFs we recorded were diverse, three patterns emerged repeatedly (illustrated in Figure 2). Figure 2A shows a cell that is tightly tuned in both time and spectrum. It has a single central positive peak, 3.3 ms wide in time, and 0.3 octaves wide in frequency (widths were obtained from a fitting procedure described below, and in detail in Nagel and Doupe, 2008). This peak is flanked by prominent negative sidebands in both frequency and time, giving the cell strong sensitivity to both spectral and temporal features. As expected from its temporal sensitivity, the PSTH of this cell's response to repeated trials (Figure 2A, right panel, black line) shows rapid fluctuations over a 100 ms interval.

Figure 2B shows a cell selective for temporal features. It has positive and negative subfields arranged sequentially in time; the highest positive peak is just 2.6 ms wide, but extends over 0.7 octaves in frequency. Finally, Figure 2C shows a cell tuned for spectral features. It has a single long positive peak that extends over 13.4 ms in time, but is constrained to less than 0.3 octaves in frequency. This peak is flanked by negative spectral sidebands, giving it sensitivity to spectral differences or modulations.

Distribution of STRF types

To describe the distribution of STRF shapes in our population quantitatively, we fit each STRF to a bivariate Mexican hat model (see Nagel and Doupe, 2008 for details). This model is able to capture many aspects of each STRF, including its preferred frequency and latency, its widths in spectrum and time, and its selectivity for spectral features, temporal features, or both. To examine the entire distribution, we plotted the fitted spectral versus temporal widths for all cells. This revealed a striking L-shaped distribution (Figure 3A; qualitatively similar although slightly noisier results were obtained by measuring the half-width of temporal and spectral cross-sections through the peak of each STRF). All but two cells were less than 0.6 octaves wide in spectrum and/or less than 7 ms wide in time. The locations of the example cells from Figure 2 are given by blue (spectro-temporal), green (temporal), and red (spectral) squares, and represent the hinge and the two arms of the distribution, respectively. The distribution of cells along the two axes suggests that all STRFs are narrowly tuned in at least one dimension while integrating over a range of different times and bandwidths in their other (nontuned) dimension. Cells that integrate broadly over both dimensions were rare.

We also observed very few STRFs showing strong orientation in time-frequency space and hence selectivity for upward or downward frequency sweeps (Figure 3B). We quantified orientation selectivity by fitting the STRFs to a slightly more complex version of our model that included an orientation parameter (see Nagel and Doupe, 2008; similar results were obtained using a direct measure of symmetry of the spectro-temporal modulation spectra of the STRFs). Most cells had orientation indices near zero (standard deviation = 8.5°), confirming that most cells in our population were un-oriented, and closely aligned to the temporal and/or spectral axes.

Anatomical distribution of STRFs

Because field L is composed of several subregions (Fortune and Margoliash, 1992), including the thalamo-recipient area L2 input and its output areas L1 and L3, we examined whether the different STRF types were localized to different regions within field L. To do so, we performed multiunit mapping studies in birds sedated with diazepam or urethane. In each experiment, we advanced a four- or eight-electrode linear array in steps of 100–200

mm through the field L complex, and recorded single or multiunit responses to our stimulus on all channels at each depth. After making marker lesions and perfusing, we prepared transverse sections of each brain, and stained alternate sections with Nissl and with an antibody against the cannabinoid receptor CB1 that selectively labels the input area L2 (Soderstrom et al., 2004; Fig. 4A).

In particular, we examined the temporal tuning of sites across the layers of field L by calculating the temporal best modulation frequency (temporal BMF, measured from the modulation spectrum of the STA, see Nagel and Doupe, 2008 for details) of STAs obtained from sorted single-unit or multiunit activity at each site. BMF is a measure of temporal width that can be calculated directly from raw STAs without fitting. Figure 4B shows temporal BMF as a function of recording location in a representative experiment. Recording sites with high temporal BMFs (narrow temporal tuning) were constrained to a narrow region of each penetration that corresponds to L2. Recording sites with low temporal BMFs (broad temporal tuning) were more common above and below the region of fast cells, in L1 and L3. This finding was consistent across all birds in which we mapped field L.

Summary and discussion of field L recording

In summary, in the zebra finch auditory forebrain we observed a highly structured distribution of STRF shapes along the axes of spectrum and time. STRFs were generally narrowly tuned along the spectral axis, temporal axis, or both. Strongly oriented sweep-selective cells, and broadly tuned cells that would act as overall sound level detectors, were rare in the forebrain population that we sampled. A recent study in zebra finches examining single unit responses to songs rather than to our naturalistic stimulus (Woolley et al., 2009) found evidence for functional groups of STRFs largely similar to those we describe here, except for differences in timing attributable to the use of a more complex stimulus, and evidence for a slightly larger number of broadly tuned neurons, although these were still rare. This distribution of response properties may be related to the structure of many natural sounds—including both zebra finch song and human speech—which are dominated by pure temporal and pure spectral features, and contain comparatively few strongly oriented spectro-temporal sweeps (Singh and Theunissen, 2003; Woolley et al., 2005, 2009). Moreover, we found that cells with different STRFS also appear to obey a structured anatomical distribution. Cells tuned to fast temporal frequencies dominated in L2, while cells that integrated over longer time intervals were concentrated in fields L1 and L3. This may be a general phenomenon in the avian auditory forebrain: studies in chick using simpler stimuli also suggested an orderly spatial arrangement of spectral and temporal bandwidths and of latencies in field L, with differences in temporal resolution between L2 and the upper layers similar to those in songbirds (Heil and Scheich, 1991a, b, 1992; Heil et al., 1992). Although much remains to be investigated, the gradient of receptive field structure along an anatomical axis in field L raises the provocative possibility that an initial layer of cells tightly tuned for both time and frequency gives rise to output layers still highly frequency-tuned, but with much broader temporal tuning.

Influence of time and frequency cues on zebra finch categorization of songs

While much progress is being made in studying central neural representations of complex stimuli in field L and elsewhere, the relationship between these representations and animals' recognition and classification behavior remains less well understood. Many physiological studies assume that animals classify stimuli as we do (Hung et al. 2005, Kreiman et al. 2006), or explicitly train animals on categories chosen by experimenters (Freedman et al. 2001). In addition, many investigations have focused on the visual system, using patterns

that are relatively stable over time. In contrast, acoustic patterns vary in time, leading to distinct temporal activity patterns. While several studies have demonstrated that these temporal patterns of spiking carry significant auditory information (Narayan et al. 2006), their role in acoustic pattern recognition has not been tested behaviorally.

Because songbirds recognize both groups and individuals based on their song, they provide a useful model for studying acoustic pattern recognition (Kroodsma et al., 1982; Stoddard et al., 1991; Nelson, 1989; Gentner and Hulse, 1989, Vignal et al., 2008). A bird's song is a high-dimensional stimulus not unlike a human voice, with structure that can be parameterized in many ways. Yet each song rendition is unique: its volume will depend on the distance between singer and listener; background noises - often other songs - will corrupt the signal (Zahn 1996), and the pitch (fundamental frequency) and duration of individual notes will vary slightly (Kao and Brainard 2006; Glaze and Troyer 2006). To recognize a male bird by its song, then, females must be able both to discriminate between the songs of different individuals and to generalize across variations in a single bird's song. Although the importance of timing cues, frequency, or both sets of acoustic parameters to field L neurons is clear, the effect of these cues on behavioral responses of zebra finches has received comparatively little attention. For instance, how much variation in pitch or duration can birds tolerate before a song becomes unrecognizable? Questions such as these are critical to understanding how the representation of song features or parameters by neurons in the zebra finch brain relates to birds' perception of these features and to their behavioral responses to song.

A behavioral assay of song recognition

To begin to address such questions, we studied behavioral categorization of different songs (Beecher et al. 1994; Gentner and Hulse 1998; this task has also been called 'generalization of a learned discrimination'). In this operant task, birds are taught to categorize two sets of songs obtained from two different conspecific individuals, but the birds are not at all constrained in which parameters of the songs they use to make the categorization. In this sense the task differs greatly from other psychophysical tasks in which a single stimulus parameter is varied and the subject is therefore intentionally focused on this parameter. After the categorization is learned, presentation of song continues, but birds are then intermittently tested with novel 'probe' songs, which are modified in a particular parameter. How the bird categorizes the modified songs can then reveal which aspects of the original songs were critical for categorization. Probe stimuli are rewarded without respect to the bird's choice, ensuring that the bird cannot not learn the "correct" answer for these stimuli from the reward pattern. Performance significantly above chance on a probe stimulus is considered evidence for 'generalization' to that probe, i.e. although that probe stimulus is different, that particular difference does not cause it to be re-classified as belonging to a different song category.

In the version of this behavioral categorization assay that we used here, after training female zebra finches to classify a large set of songs as belonging to one of two males, we systematically altered those songs in pitch and duration, and used them as probe stimuli, presented as 10–15% of all trials. In this way we could ask which parameters govern whether a female zebra finch classifies a novel song as belonging to a familiar individual.

Classification training and stimuli

A trial began when the female hopped on a central "song perch" that faced a speaker (Fig. 5A). Following this hop, the computer selected a stimulus from a database of songs from two individuals, described below. The female had to decide which male the song belonged to, and to report her choice by hopping on one of two response perches, located to the right (for male A) and to the left (for male B) of the song perch. If she classified the song

correctly, a feeder located under the song perch was raised for 2–5 seconds, allowing her access to food. If she classified the song incorrectly, no new trials could be initiated for a time-out period of 20–30 seconds. To make a response, females had up to 6 seconds after song playback ended and were allowed to respond at any point during song playback; playback was halted as soon as a response was made. If no response was made within 6 seconds of the song ending, the trial was scored as having no response, and a new trial could be initiated by hopping again on the song perch.

Birds required one to two weeks of training to acquire this task. Training consisted of several stages, in which the bird was first habituated to the operant cage, and learned to hop on the perches using songs that did not include any from the two individuals used for classification trials. Then, in “discrimination mode,” the female began classification trials. The habituation song set was replaced with a database of training songs, and trials were rewarded (with food) or punished (with a time-out) depending on which response perch the female chose. Females generally learned the classification task (defined as greater than 75% correct, in a task where chance responding is 50%) within 2–5 days of beginning classification trials (Fig. 5C). After birds reached criterion in their discrimination behavior, the overall reward rate for correct trials was lowered to 75–95%. Performance was very good, with over 80 percent of songs classified correctly.

The training database consisted of 100 song bouts, 50 from one male (“male A”, Fig. 1B), and 50 from another (“male B”, Fig. 5B). Each bout consisted of 1–6 repeated “motifs” (see blue bars above the spectrograms in Fig. 1B), and was drawn from both directed (sung to a female) and undirected (sung alone) recording sessions. We used this large database to ensure that females learned to associate each perch with an individual, and not with a single rendition of his song. The song bouts of the two males were chosen to have similar distributions of overall duration, RMS volume, and overall power spectrum (see Nagel et al., in review, for details), but had quite distinct syllable structure and timing.

Classification of pitch-shifted songs

To examine the role of pitch and duration in classifying songs of individuals, we presented trained birds with probe stimuli consisting of songs in which either pitch or duration were altered, and observed how the bird classified these novel stimuli. As described earlier, probe stimuli are rewarded regardless of the bird’s choice of perch, so that there is no information from the reward pattern about how these stimuli might be classified.

We used a phase vocoder algorithm to independently manipulate the overall pitch (fundamental frequency) and duration of training songs (Flanagan and Golden 1966). To alter the duration of a song without changing its pitch, the vocoder broke it into short overlapping segments, and took the Fourier transform of each segment, yielding a magnitude and phase for each frequency. We used segments of 256 samples, or 10.5 msec at 25kHz, overlapping by 128 samples. This produced a function of time and frequency similar to the spectrogram often used to visualize an auditory stimulus (such as that shown in Fig. 4B) but with a phase value as well as a magnitude for each point. We then chose a new time scale for the song, and interpolated the magnitude and phase of the signal at the new time points based on the magnitude and phase at nearby times. Finally, the vocoder reassembled the stretched or compressed song from the interpolated spectrogram. This process yielded a song that was slower or faster than the original but whose frequency information was largely unchanged.

To alter the pitch of a song without changing its duration, we first used the above algorithm to stretch or compress the song in time. We then resampled it back to its original length

yielding a song with a higher (for stretched songs) or lower (for compressed songs) pitch but with the same duration and temporal pattern of syllables as the original.

The pitch and duration of individual zebra finch notes are known to vary slightly across song renditions, by about 1–4% (Kao et al. 2006, Glaze et al. 2006, Cooper and Goller 2006). In a categorization task such variation within the normal range for individual birds should not affect the bird's classification of an individual song, while variation beyond that range might or might not, depending on the importance of the parameter to recognition and classification. We therefore created songs with durations and pitches increased and decreased by 0, 1, 2, 4, 8, 16, 32, and 64%. This resulted in durations ranging from 1/1.64 or 61% to 164% of the original song length, and pitches ranging from 0.71 octaves lower to 0.71 octaves higher than the original pitch. The frequency content of pitch-shifted probes was systematically altered but their temporal pattern--including the sequence of harmonic and noisy notes-- was not. Conversely, for duration-altered songs, the songs were proportionally stretched or compressed in overall duration while the relative durations of syllables and intervals were maintained, and the frequency content of each song was unchanged.

The phase vocoder algorithm will of necessity slightly alter the temporal fine structure of song (because of phase smearing). It is increasingly recognized that temporal fine structure of sounds is important to speech recognition (e.g. Sheft et al., 2008), and birds are also known to be capable of very precise temporal resolution (Dooling et al., 2002; Lohr et al., 2006). However, in this study the frequency- and duration-shifted songs were constructed using a very similar manipulation. We therefore think it unlikely that any systematic differences in categorization between these two sets of manipulated songs are due to differences in the songs' temporal fine structure, although the role of such cues in song categorization remains to be directly studied.

Figure 6A shows a representative female's response to pitch-shifted probe songs. The blue bar represents the mean \pm one standard deviation across days of her performance on control (normally rewarded) trials. For small pitch shifts (less than 0.05 octaves) her performance on probes overlapped with her control performance. However, her performance fell off rapidly for shifts larger than those naturally observed (\pm 3%, indicated by dashed lines), and she performed at chance for pitch shifts of 0.2 octaves or greater. Similar results were seen in all birds tested ($n=5$) on these probes, indicating that pitch strongly determined how songs were classified in this task.

Classification of songs with altered duration

Figure 6B shows the performance of a bird in response to versions of the training stimuli that were stretched or compressed in time, for the same example bird as shown in Figure 6A. 100% duration indicates no change. The blue bar represents the bird's performance on normally rewarded control trials, and the dashed vertical lines represent the range of natural variation (\pm 4%). In contrast to her behavior on pitch-shifted probes, this female responded well above chance to all duration probes, including those compressed or expanded well beyond the range of natural variability. For duration changes up to 32%, her performance overlapped with her control performance. For a 64% change in duration, her performance dropped but remained above chance. The same result was observed for all birds tested, indicating that syllables and intervals within a broad range of durations are sufficient for correct classification of songs from different individuals.

Summary and discussion of behavioral analyses

In summary, we found that, in this task, birds relied much more on frequency than on duration to categorize modified songs as belonging to one of two individuals. Although

these cues vary by similar amounts in naturally occurring songs (1–3% for pitch, 1–4% for duration), females showed much narrower tolerance for changes in fundamental frequency than for changes in duration. Females generalized over a wider range than the 0.5% just noticeable difference for pitch measured in most avian species (Dooling et al. 2000). However, performance on pitch-shifted stimuli rapidly fell to chance once stimuli fell outside of the range of natural pitch variation. In contrast, despite birds' demonstrated ability to discriminate temporal manipulations of sounds (Dooling et al. 2002; Lohr et al., 2006), females performed significantly above chance over a more than 2-fold range of song and syllable durations.

These results did not match any simple predictions about how changes in frequency and duration should affect classification of individual songs. From the perspective of auditory neurons, which are highly sensitive to these parameters, one might expect that any change would strongly affect classification. Similarly, the bird's experience of songs has taught it that pitch and duration vary only slightly in individual songs, suggesting that both these cues should be important in this task. Instead, we found that in this specific task, although the temporal cues also contain a great deal of information, birds relied much more on spectral than on duration cues to categorize modified songs as belonging to one of two individuals. It is likely, however, that this is task-dependent, and that birds could use the temporal information if asked to do so.

These results parallel aspects of human speaker identification. Like zebra finches, humans depend heavily on pitch cues to identify speakers (O'Shaughnessy, 1986), but not to identify phonemes (Avenidaño et al. 2004). Word duration can be important as well (Walden, 1978), but is less reliable than spectral features. Thus, the acoustic parameters used by these songbirds to identify individuals may be closer to those used by human beings to identify speakers rather than individual words.

These findings on pitch and duration have important consequences for understanding how neural responses to song may be related to perceptual behavior. Neurons in field L respond to song with precisely time-locked spikes (Woolley et al. 2006), whose temporal pattern provides significant information about which song was heard (Narayan et al. 2006). Several papers have therefore suggested that these temporal patterns of spiking are critical for birds' discrimination and recognition of songs (Wang et al. 2007; Larson et al. 2009). Temporal patterns in field L, however, would also be expected to change markedly as a song is expanded or contracted in time. The ability of birds to generalize across large changes in duration calls into question any simple model using field L's temporal spiking patterns for classifying individuals by their songs.

An alternate model of song classification could be based on the particular population of cells activated by song, rather than their temporal pattern of firing. Since many field L neurons are tightly tuned in frequency, the subset of field L neurons that fires during the course of a song will depend on the spectral content of that song, and will thus vary between songs of different individuals. Shifting the fundamental frequency of a particular song should greatly change the population of neurons activated for that song, making this population quite different from that responding during the original song. Consistent with the possibility that the population of neurons activated matters, in our categorization task, females' poor behavioral tolerance for pitch-shifted songs showed remarkable similarity to the spectral tuning widths of individual field L neurons: the half-width of the pitch tuning curves was 0.11–0.22 octaves (8–16%), while the half-widths of field L neurons had a minimum of 0.13 octaves, and a median of 0.33 octaves (Nagel and Doupe 2008). This suggests that classification depends more on which field L cells are active at a given time than the precise pattern of activity in those cells.

Clearly, much remains to be learned about the properties of auditory neurons in field L and beyond, and how they allow songbirds to recognize complex patterns. Indeed, the bird's behavior in our task better matches the relative insensitivity to overall song duration characteristic of song-selective neurons in the high-level nucleus HVC (Margoliash, 1983; Margoliash and Fortune, 1992). It remains unclear, however, how important HVC is to categorization of songs of conspecific birds (Brenowitz, 1991; Gentner et al., 2000; MacDougall-Shackleton et al., 1998). Regardless, our behavioral data testing the role of parameters known to be important for auditory neurons suggest that animals group stimuli in unexpected ways, and teach us that simply identifying the signals present in neurons does not tell us what information is used for perception. Neurophysiological studies often assume that we know which tasks the brain solves, but the results here remind us that relating neural activity to perceptual behavior will require detailed and quantitative assessments of behavior as well as of neural activity.

Overall summary and future directions

Songbirds have proven to be a surprisingly informative model for the study of primary auditory forebrain areas. Our work and that of others (Woolley et al., 2009) suggests that songbird field L breaks down the analysis of complex sounds into the detection of three basic kinds of modulations: temporal, spectral and spectro-temporal, which may also be segregated in particular areas of the brain.

It remains to be seen whether this striking organization is unique to songbirds, or whether a similar underlying organization across all birds or even all vertebrates is simply more obvious in these auditory specialists, where we know a great deal about the behaviorally most relevant stimuli and can use them to help dissect the system. Consistent with the latter possibility, earlier work on the chicken field L also suggested a functional subdivision of field L, including a mapping of temporal aspects of frequency-modulated signals (Heil and Scheich, 1991a,b, 1992; Heil et al., 1992). There is certainly precedent for 'neuroethological' models like songbirds providing insights that prove to have some generality, for instance the studies of sound localization in barn owls (Konishi, 1990, 2000; see also Koppl, this volume).

The question of whether there is a similar organization in mammals is unclear. Studies of anesthetized auditory cortex in untrained cats have found that, as in the songbird work, most neurons are not selective for oriented frequency sweeps (Miller et al., 2002). However, that study found no systematic relationship between the spectral and temporal tuning properties of A1 neurons, while we found a strong trade-off between temporal and spectral selectivity. These differences may arise from the structure of the avian forebrain, which contains many fast-firing cells that are able to follow rapid modulations in the stimulus, while the mammalian auditory cortex responds more slowly (Miller et al., 2002; Depireux et al., 2001; Lu et al., 2001). The differences may also depend on recording conditions, since most previous studies have measured STRFs under pentobarbital anesthesia, while we recorded from unanesthetized animals. Anesthesia can profoundly influence the temporal dynamics of cortical auditory responses (Wang et al., 2005). It may also be important to attend to laminar differences in mammals when making across-species comparisons; a recent study has suggested a hierarchy of tuning within cat A1, with the input layer largely unoriented as in birds, and neurons with joint tuning for spectrum and time beginning to emerge in the upper and lower cortical layers (Atencio et al., 2009). Another intriguing possibility is that the marked spectro-temporal tradeoff, in mammals as in birds, will be most obvious and highly developed in vocal learners, i.e. in humans - this awaits further study. Inspired by the visual system, mammalian A1 has been modeled with receptive fields that evenly span a range of time-frequency orientations, similar to the uniform sampling of 2D spatial orientations in the

visual cortex (Chi et al., 2005). However, the same group reported that A1 neurons showed mostly unoriented tuning properties (Depireux et al., 2001; Simon et al., 2007). Together with data from these mammalian studies, the findings from songbirds suggest that sampling a range of temporal and spectral modulations, rather than orientation in time-frequency space, may be the organizing principle of forebrain auditory sensitivity.

Zebra finch song, like human speech, is characterized by a relative lack of frequency-modulated sounds, but that is not true of the songs of many other songbirds (nor, in the mammals, of many bats, for instance). This suggests that it could be very informative to record field L of birds with different song types, to see whether the type of song produced influences the organization of primary auditory areas. In mammals, early exposure to frequency-modulated sweeps exposure can enhance sensitivity to these sounds (Insanally et al., 2009). More generally, this raises the intriguing question of how the organization seen in songbirds develops, and how much it depends on experience. In starlings, field L becomes grossly disorganized when birds are raised in acoustic isolation (Cousillas et al., 2006, 2008), but it is not yet clear whether this reflects social or auditory effects of isolation. Preliminary studies in zebra finches also suggest that experience is crucial (Amin and Theunissen, 2008). If so, it will be important to determine whether there are critical periods for the effects of experience, and whether and how the effects of pathological experience can be remediated.

Songbirds are also proving to be a very useful model for studying acoustic pattern recognition (e.g. Kroodsma and Miller, 1996; Kroodsma et al, 1982; Gentner and Hulse, 1989 ; Vignal et al., 2008). Again this is in part because of our clear understanding of the behaviorally relevant signals, and the ease in training birds in these tasks that are so ethologically appropriate for them.

Our finding that birds can generalize across large differences in song duration but not frequency in a categorization task refines our questions about field L's role in perception. In particular, this pattern of generalization suggests that in classifying songs, the particular population of field L neurons activated by a song is more important for deciding which song it was than is the neurons' temporal pattern of firing. However, it is likely that this generalization is task-dependent, and it will be informative to see how birds generalize to different tasks. Most revealing of all, with respect to the relationship between neural representation and behavior, will be to record from field L while birds are performing behavioral tasks. The emerging evidence on experience-dependence of field L also raises the question of the effects of disruption of experience on perception. Studying both field L and behavior in birds with altered experience could be informative regarding how these relate to each other.

In addition, although it is important to link the information available at each level of the auditory hierarchy to perception, it is also clear that there are many auditory areas beyond field L that are likely to function critically in perception and that remain to be studied. Notwithstanding, knowledge about the organization of lower levels as well as about behavioral constraints can greatly aid in dissecting such higher, more complex levels. Finally, auditory studies in songbirds provide yet another example, like the many others in this volume, of how a comparative approach to studying the auditory system can be very revealing, and how it has the potential to expose quite general rules about neural signals and how they are involved in generating behavior.

Acknowledgments

The work described in this paper was supported by NIH grants NS34835, DC04975, and MH055987 to AJD, and a Howard Hughes Medical Institute fellowship to KN. All animal experiments were approved by the University of California, San Francisco IACUC.

Abbreviations used

HVC	acronym for a high-level vocal control nucleus, used as a proper name
STRF	spectro-temporal receptive field
PSTH	peri-stimulus time histogram
STA	spike-triggered average
BMF	best modulation frequency
RMS	root mean square

References

- Amin N, Theunissen FE. Selectivity for natural sounds in the songbird auditory forebrain is strongly shaped by the acoustic environment. *Society for Neuroscience Abstract*. 2008:99.8.
- Atencio CA, Sharpee TO, Schreiner CE. Hierarchical computation in the canonical auditory cortical circuit. *Proc Natl Acad Sci U S A*. 2009 Nov 16. [Epub ahead of print].
- Avedaño, C.; Deng, L.; Hermansky, H.; Gold, B. *Speech Processing in the Auditory System*. New York, NY: Springer-Verlag; 2004. The analysis and representation of speech.
- Beecher MD, Campbell SE, Stoddard PK. Correlation of song learning and territory establishment strategies in the song sparrow. *Proc Natl Acad Sci U S A*. 1994 Feb 15; 91(4):1450–4. [PubMed: 11607460]
- Brenowitz EA. Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science*. 1991; 251(4991):303–5. [PubMed: 1987645]
- Cooper BG, Goller F. Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *J Neurophysiol*. 2006 Jun; 95(6):3798–809. Epub 2006 Mar 22. [PubMed: 16554509]
- Cousillas H, George I, Henry L, Richard JP, Hausberger M. Linking social and vocal brains: could social segregation prevent a proper development of a central auditory area in a female songbird? *PLoS One*. 2008; 3(5):e2194. [PubMed: 18493580]
- Cousillas H, George I, Mathelier M, Richard JP, Henry L, Hausberger M. Social experience influences the development of a central auditory area. *Naturwissenschaften*. 2006 Dec; 93(12):588–96. Epub 2006 Aug 25. [PubMed: 16932952]
- Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectrotemporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol*. 2001; 85:1220–1234. [PubMed: 11247991]
- Dooling, RJ.; Lohr, B.; Dent, ML. *Hearing in Birds and Reptiles*. In: Dooling, RJ.; Fay, RR.; Popper, AN., editors. *Comparative Hearing*. New York, NY: Springer-Verlag; 2000.
- Dooling RJ, Leek MR, Gleich O, Dent ML. Auditory temporal resolution in birds: discrimination of harmonic complexes. *J Acoust Soc Am*. 2002; 112(2):748–59. [PubMed: 12186054]
- Doupe AJ. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J Neurosci*. 1997; 17(3):1147–67. [PubMed: 8994068]
- Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci*. 1999; 22:567–631. Review. [PubMed: 10202549]
- Eggermont JJ, Johannesma PM, Aertsen AM. Reverse-correlation methods in auditory research. *Q Rev Biophys*. 1983; 16:341–414. [PubMed: 6366861]

- Escabi MA, Miller LM, Read HL, Schreiner CE. Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J Neurosci*. 2003; 23:11489–504. [PubMed: 14684853]
- Flanagan RM, Golden M. Phase Vocoder. *Bell System Technical Journal*. 1966; 1493:1509.
- Fortune ES, Margoliash D. Parallel pathways and convergence onto HVC and adjacent neostriatum of adult zebra finches (*Taeniopygia guttata*). *J Comp Neurol*. 1995; 360(3):413–41. [PubMed: 8543649]
- Fortune ES, Margoliash D. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taeniopygia guttata*). *J Comp Neurol*. 1992; 325(3):388–404. [PubMed: 1447407]
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*. 2001; 291(5502):312–6. [PubMed: 11209083]
- Gehr DD, Capsius B, Grabner P, Gahr M, Leppelsack HJ. Functional organisation of the field-L-complex of adult male zebra finches. *Neuroreport*. 1999; 10(2):375–80. [PubMed: 10203338]
- Gentner TQ. Neural systems for individual song recognition in adult birds. *Ann NY Acad Sci*. 2004; 1016:282–302. [PubMed: 15313781]
- Gentner TQ, Hulse SH. Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Anim Behav*. 1998 Sep; 56(3):579–594. [PubMed: 9784206]
- Gentner TQ, Hulse SH, Bentley GE, Ball GF. Individual vocal recognition and the effect of partial lesions to HVC on discrimination, learning, and categorization of conspecific song in adult songbirds. *J Neurobiol*. 2000; 42(1):117–33. [PubMed: 10623906]
- Gentner TQ, Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature*. 2003; 424(6949):669–74. [PubMed: 12904792]
- George I, Cousillas H, Vernier B, Richard JP, Henry L, Mathelier M, Lengagne T, Hausberger M. Sound processing in the auditory-cortex homologue of songbirds: functional organization and developmental issues. *J Physiol Paris*. 2004; 98(4–6):385–94. [PubMed: 16293403]
- Glaze CM, Troyer TW. Temporal structure in zebra finch song: implications for motor coding. *J Neurosci*. 2006 Jan 18; 26(3):991–1005. [PubMed: 16421319]
- Heil P, Scheich H. Functional organization of the avian auditory cortex analogue. I. Topographic representation of iso-intensity bandwidth. *Brain Res*. 1991a; 539(1):110–20. [PubMed: 2015496]
- Heil P, Scheich H. Functional organization of the avian auditory cortex analogue. II. Topographic distribution of latency. *Brain Res*. 1991b; 539(1):121–5. [PubMed: 2015497]
- Heil P, Langner G, Scheich H. Processing of frequency-modulated stimuli in the chick auditory cortex analogue: evidence for topographic representations and possible mechanisms of rate and directional sensitivity. *J Comp Physiol A*. 1992; 171(5):583–600. [PubMed: 1494138]
- Heil P, Scheich H. Spatial representation of frequency-modulated signals in the tonotopically organized auditory cortex analogue of the chick. *J Comp Neurol*. 1992; 322(4):548–65. [PubMed: 1401249]
- Heil P, Scheich H. Quantitative analysis and two-dimensional reconstruction of the tonotopic organization of the auditory field L in the chick from 2-deoxyglucose data. *Exp Brain Res*. 1985; 58(3):532–43. [PubMed: 4007092]
- Hsu A, Woolley SM, Fremouw TE, Theunissen FE. Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *J Neurosci*. 2004; 24(41):9201–11. [PubMed: 15483139]
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. *Science*. 2005; 310(5749):863–6. [PubMed: 16272124]
- Insanally MN, Köver H, Kim H, Bao S. Feature-dependent sensitive periods in the development of complex sound representation. *J Neurosci*. 2009; 29(17):5456–62. [PubMed: 19403813]
- Kao MH, Brainard MS. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J Neurophysiol*. 2006 Sep; 96(3):1441–55. Epub 2006 May 24. [PubMed: 16723412]
- Kim KJ, Rieke F. Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J Neurosci*. 2001 Jan 1; 21(1):287–99. [PubMed: 11150346]

- Konishi M. Birdsong: from behavior to neuron. *Annu Rev Neurosci.* 1985; 8:125–70. Review. [PubMed: 3885827]
- Konishi M. The neural algorithm for sound localization in the owl. *Harvey Lect 1990–1991.* 1990; 86:47–64. Review.
- Konishi M. Study of sound localization by owls and its relevance to humans. *Comp Biochem Physiol A Mol Integr Physiol.* 2000; 126(4):459–69. [PubMed: 10989338]
- Kreiman G, Hung CP, Kraskov A, Quiroga RQ, Poggio T, DiCarlo JJ. Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron.* 2006; 49(3):433–45. [PubMed: 16446146]
- Kroodsma, DE.; Miller, EH.; Ouellet, H. *Acoustic Communication in Birds*, editors. New York, NY: Academic Press; 1982.
- Kruse AA, Stripling R, Clayton DF. Context-specific habituation of the zenk gene response to song in adult zebra finches. *Neurobiol Learn Mem.* 2004 Sep; 82(2):99–108. [PubMed: 15341795]
- Larson E, Billimoria CP, Sen K. A biologically plausible computational model for auditory object recognition. *J Neurophysiol.* 2009; 101(1):323–31. [PubMed: 18987124]
- Lewicki MS, Konishi M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci U S A.* 1995 Jun 6; 92(12):5582–6. [PubMed: 7777552]
- Lohr B, Dooling RJ, Bartone S. The discrimination of temporal fine structure in call-like harmonic sounds by birds. *J Comp Psychol.* 2006; 120(3):239–51. [PubMed: 16893261]
- Lu T, Liang L, Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci.* 2001; 4:1131–1138. [PubMed: 11593234]
- MacDougall-Shackleton SA, Hulse SH, Ball GF. Neural bases of song preferences in female zebra finches (*Taeniopygia guttata*). *Neuroreport.* 1998; 9(13):3047–52. [PubMed: 9804314]
- Machens CK, Wehr MS, Zador AM. Linearity of cortical receptive fields measured with natural sounds. *J Neurosci.* 2004; 24(5):1089–100. [PubMed: 14762127]
- Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci.* 1983 May; 3(5):1039–57. [PubMed: 6842281]
- Margoliash D, Fortune ES. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci.* 1992; 12(11):4309–26. [PubMed: 1432096]
- Mello C, Nottebohm F, Clayton D. Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *J Neurosci.* 1995 Oct; 15(10):6919–25. [PubMed: 7472448]
- Mello CV, Vehlo TA, Pinaud R. Song-induced gene expression: a window on song auditory processing and perception. *Ann NY Acad Sci.* 2004; 1016:77–108. [PubMed: 15313771]
- Miller LM, Escabi MA, Read HL, Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol.* 2002; 87:516–527. [PubMed: 11784767]
- Miller LM, Escabi MA, Read HL, Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol.* 2002; 87:516–27. [PubMed: 11784767]
- Mooney R. Neurobiology of song learning. *Curr Opin Neurobiol.* 2009 Nov 3. [Epub ahead of print].
- Mooney R. Different subthreshold mechanisms underlie song selectivity in identified HVC neurons of the zebra finch. *J Neurosci.* 2000; 20(14):5420–36. Erratum in: *J Neurosci* 20(15). [PubMed: 10884326]
- Nagel KI, Doupe AJ. Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron.* 2008 Jun 26; 58(6):938–55. [PubMed: 18579083]
- Narayan R, Graña G, Sen K. Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol.* 2006; 96(1):252–8. [PubMed: 16571738]
- Nelson DA. Song frequency as a cue for recognition of species and individuals in the field sparrow (*Spizella pusilla*). *J Comp Psychology.* 1989; 103(2):171–6.
- Nelson DA, Marler P. Categorical perception of a natural stimulus continuum: birdsong. *Science.* 1989; 244:976–978. [PubMed: 2727689]
- Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, *Serinus canarius*. *J Comp Neurol.* 1976 Feb 15; 165(4):457–86. [PubMed: 1262540]

- O'Shaughnessy D. Speaker recognition. *IEEE Acoustics, Speech, and Signal Processing*. 1986; 3(4):4–17.
- Ringach DL, Hawken MJ, Shapley R. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J Vis*. 2002; 2(1):12–24. [PubMed: 12678594]
- Ringach D, Shapley R. Reverse correlation in neurophysiology *Cognitive Science: A Multidisciplinary Journal*. 28. 2004; 2:147–166.
- Rosen MJ, Mooney R. Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron*. 2003; 39(1):177–94. [PubMed: 12848941]
- Scheich H, Bonke BA, Bonke D, Langner G. Functional organization of some auditory nuclei in the guinea fowl demonstrated by the 2-deoxyglucose technique. *Cell Tissue Res*. 1979; 204(1):17–27. [PubMed: 527021]
- Searcy, WA.; Nowicki, S. Functions of song variation in sparrows. In: Hauser, MD.; Konishi, M., editors. *The Design of Animal Communication*. MIT Press; Cambridge, MA: 1999. p. 577-595.
- Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol*. 2001; 86(3):1445–58. [PubMed: 11535690]
- Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD. Adaptive filtering enhances information transmission in visual cortex. *Nature*. 2006; 439(7079):936–42. [PubMed: 16495990]
- Sheft S, Ardoint M, Lorenzi C. Speech identification based on temporal fine structure cues. *JASA*. 2008; 124:562–575.
- Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA. Temporal symmetry in primary auditory cortex: implications for cortical connectivity. *Neural Comput*. 2007; 19:583–638. [PubMed: 17298227]
- Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am*. 2003; 114(6 Pt 1):3394–411. [PubMed: 14714819]
- Stoddard PK, Beecher MD, Horning CL, Campbell SE. Recognition of individual neighbors by song in the song sparrow, a species with song repertoires. *Behavioral Ecology and Sociobiology*. 1991; 29(3):211–215.
- Stripling R, Volman SF, Clayton DF. Response modulation in the zebra finch neostriatum: relationship to nuclear gene regulation. *J Neurosci*. 1997; 17(10):3883–93. [PubMed: 9133406]
- Terleph TA, Mello CV, Vicario DS. Auditory topography and temporal response dynamics of canary caudal telencephalon. *J Neurobiol*. 2006; 66(3):281–92. [PubMed: 16329130]
- Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci*. 2000; 20:2315–31. [PubMed: 10704507]
- Theunissen FE, Shaevitz SS. Auditory processing of vocal sounds in birds. *Curr Opin Neurobiol*. 2006; 16(4):400–7. [PubMed: 16842993]
- Vates GE, Broome BM, Mello CV, Nottebohm F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J Comp Neurol*. 1996; 366:613–642. [PubMed: 8833113]
- Vignal C, Mathevon N, Mottin S. Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behavioural Processes*. 2008; 77(2): 191–198. [PubMed: 17980974]
- Walden BE, Montgomery AA, Gibeily GJ, Prosek RA, Schwartz DM. Correlates of psychological dimensions in talker similarity. *Journal of Speech and Hearing Research*. 1978; 21(2):265–275. [PubMed: 703276]
- Wang L, Narayan R, Graña G, Shamir M, Sen K. Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci*. 2007; 27(3):582–9. [PubMed: 17234590]
- Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature*. 2005; 435:341–346. [PubMed: 15902257]
- Wild JM, Karten HJ, Forst BJ. Connections of the auditory forebrain in the pigeon (*Columba livia*). *J Comp Neurol*. 1993; 337(1):32–62. [PubMed: 8276991]

- Woolley SM, Gill PR, Theunissen FE. Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *J Neurosci*. 2006; 26:2499–512. [PubMed: 16510728]
- Woolley SM, Gill PR, Fremouw T, Theunissen FE. Functional groups in the avian auditory system. *J Neurosci*. 2009; 29(9):2780–93. [PubMed: 19261874]
- Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci*. 2005; 8:1371–9. [PubMed: 16136039]
- Zahn, RA. *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. Oxford University Press; 1996.

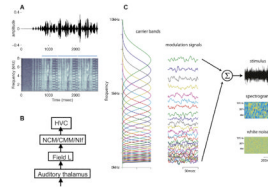


Fig. 1.

A) Typical song of a zebra finch, shown both as an oscillogram (sound pressure vs time, top panel) and spectrograms (frequency vs time, with amplitude indicated by relative lightness, lower panel). Each song is composed of several repeated sequences of syllables known as ‘motifs’, indicated by the blue bars over the spectrograms. This song is also song A of the behavioral experiments described later. B) Simplified schematic of the songbird central auditory hierarchy. The sensorimotor song control nucleus HVC is shown here as the highest level, and receives input directly or indirectly from a less selective sensorimotor nucleus known as Nif (nucleus interface) and the high-level auditory nuclei CM (caudal mesopallium) and NCM (caudomedial nidopallium). CM and NCM receive input from the primary auditory cortex equivalent of birds, known as field L. C) Schematic of naturalistic stimulus construction. Thirty-two overlapping frequency bands (left-hand column) were modulated by independent amplitude envelopes (middle column). The stimulus was the sum of these bands, shown as both an oscillogram (top panel) and as a spectrogram (second panel) in the right-hand column. The naturalistic stimulus was smoother in both time and frequency than a pure white noise stimulus (typical noise segment in bottom panel).

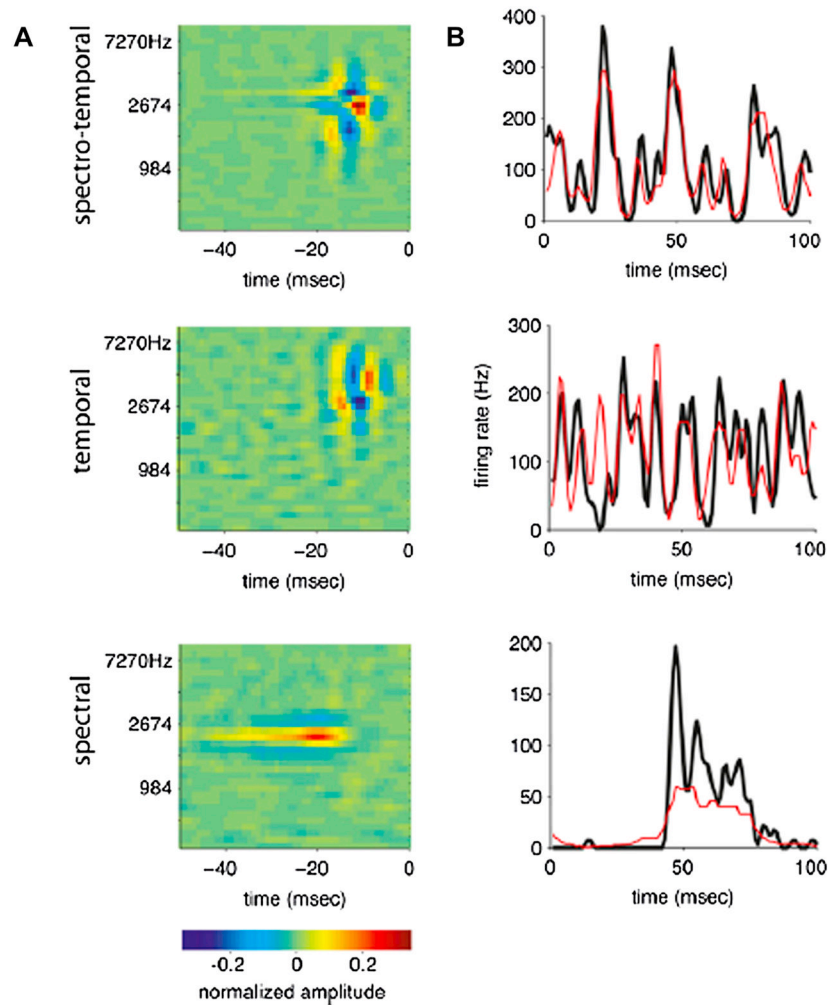


Fig. 2. Examples of three common types of spectro-temporal receptive fields. A) Spectro-temporal receptive field (STRF) and real and predicted peri-stimulus time histograms (PSTHs) of a cell sensitive to both spectral and temporal modulations. The central positive peak of the STRF is flanked by negative regions in both spectrum and time. The PSTH shows the cell's average response to 50 repeated stimulus segments (right-hand panel, black line). The prediction (red line) is generated by convolving the STRF with this stimulus and passing the result through a nonlinearity derived from the data (see Nagel and Doupe, 2008). The correlation coefficient between prediction and PSTH for this cell was 0.70, making it one of the better fits in our population. B) A cell sensitive to temporal modulations. Positive and negative subfields of the STRF are arranged sequentially in time. The correlation coefficient between PSTH and prediction for this cell was 0.48, just below the population mean of 0.52 ± 0.14 (standard deviation). C) A cell sensitive to spectral modulations. The positive region of the STRF is extended in time and flanked by negative spectral sidebands. The correlation coefficient between prediction and PSTH for this cell was 0.58. The color scale below the spectrograms shows the normalized amplitude of the neural response.

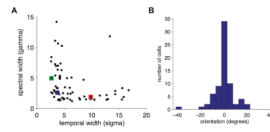


Fig. 3.

A) Spectral width versus temporal width for all STRFs ($n = 71$). Width parameters were obtained by fitting the bivariate Mexican hat model to each STRF. They show an L-shaped distribution, with most cells narrowly tuned in spectrum or time, or both. Example cells from Figure 2 are indicated by blue (spectro-temporal, example from Figure 2A), green (temporal, example from Figure 2B), and red (spectral, example from Figure 2C) squares. B) Distribution of orientation parameters obtained by fitting a version of the bivariate Mexican hat model including an orientation term. Most cells show orientations near 0, indicating that they are aligned largely along the temporal and/or spectral axes.

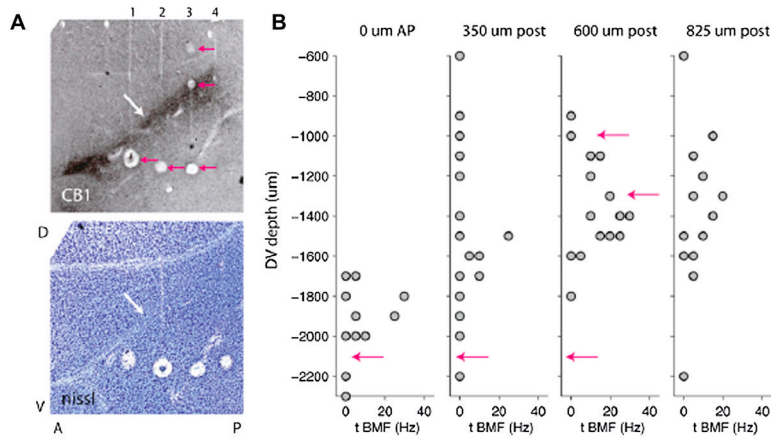


Fig. 4.

A) Histological sections showing the regions of field L and the locations of electrode penetrations in a mapping experiment. (Top panel) Slide stained with an antibody to CB1, the cannabinoid receptor, which selectively labels the input area L2 (white arrow). Above the stained area is area L1, and below it is area L3. Tracks of four electrodes can be seen crossing the three layers of field L. Pink arrows indicate marker lesions. (Bottom panel) Nissl-stained slide adjacent to above showing the laminae that define the borders of the field L complex, as well as the diagonal fiber tract immediately adjacent to field L2 (white arrow). Lesions from all four electrodes are visible. B) Temporal BMF of raw multiunit STAs as a function of recording depth on each of four electrodes. Pink arrows mark the depths of the lesions shown in the top panel of (A). Sites with higher best modulation frequencies are found within a restricted range of depths that is deeper and narrower for the anterior electrodes and shallower and wider for the posterior electrodes. The location of these sites containing cells with faster preferred modulation rates corresponds well to the location of the darkly-stained area in the CB1 slide, suggesting that these “faster sites” are localized to area L2.

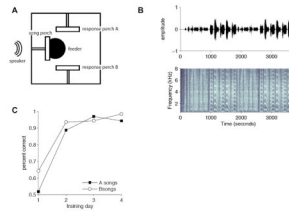


Fig. 5.

A) Diagram of the operant cage seen from above. Three perches are located on three sides of the cage. After a hop on the central perch, the computer plays a song from the speaker located directly in front of it. The bird can respond by hopping on either of the two response perches. After a correct answer, a feeder beneath the cage is raised for 2–5 seconds, allowing the bird access to seed. After an incorrect answer, all perches cease to function for a 20–30 second time-out. B) Oscillogram and spectrogram of song B as used for training (song A is in Figure 1A). The goal of our study was to train birds to associate each response perch with the songs of one individual: the two individuals differed in the temporal pattern and frequency content of their syllables. 50 songs from each individual were used in training, to avoid over-training on a single song rendition. C) Learning curve for one representative individual showing percent of correct responses on A (black) and B (white) songs as a function of the number of days of discrimination training. By the second day of training, this bird performed well above chance on both stimulus types. This bird had 9 days of pre-training (song, food, and sequence modes, as described in text) before beginning discrimination trials. All birds showed greater-than-chance performance by the second day of training.

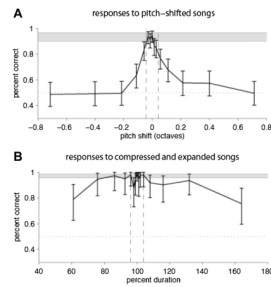


Fig. 6.

A) Performance of a typical bird on pitch-shifted probe stimuli. Percent correct as a function of pitch shift (black line). Error bars represent 95% confidence intervals on percent correct obtained by fitting data to a binomial distribution. The blue bar at the top represents the mean \pm one standard deviation of the bird's performance on control training trials, averaged across days ($n=16$). Dashed lines at 97% and 103% represent \pm one standard deviation of the range of natural pitch variation. The bird's performance overlapped with control performance within this range and fell off rapidly outside it. All birds showed a decrease in performance for songs outside the range of natural variability and chance behavior for the largest pitch shifts. B) Performance of a typical bird for duration probe stimuli. Percent correct as a function of song duration (black line). Error bars and blue bar as described in the legend for Figure 6A. Dashed lines representing the range of natural variability are at 96% and 104%. From 76 to 132% durations, the bird's performance on probes overlapped with her performance on control trials, and she performed significantly above chance on all probes. All birds performed significantly above chance for all duration probe stimuli, although like the example bird shown, their performance decreased below control levels for the largest changes, i.e., 61% and 164% duration.