Published in final edited form as:
Neuron. 2013 July 10; 79(1): 141–152. doi:10.1016/j.neuron.2013.04.038.

Sparse and background-invariant coding of vocalizations in auditory scenes

David M. Schneider[1,3] and **Sarah M. N. Woolley**[1,2]
[1]Program in Neurobiology and Behavior, Columbia University, New York, NY 10032, USA

[2]Department of Psychology, Columbia University, New York, NY 10027, USA

Summary

Vocal communicators such as humans and songbirds readily recognize individual vocalizations, even in distracting auditory environments. This perceptual ability is likely subserved by auditory neurons whose spiking responses to individual vocalizations are minimally affected by background sounds. However, auditory neurons that produce background-invariant responses to vocalizations in auditory scenes have not been found. Here, we describe a population of neurons in the zebra finch auditory cortex that represent vocalizations with a sparse code and that maintain their vocalization-like firing patterns in levels of background sound that permit behavioral recognition. These same neurons decrease or stop spiking in levels of background sound that preclude behavioral recognition. In contrast, upstream neurons represent vocalizations with dense and background-corrupted responses. We provide experimental evidence suggesting that sparse coding is mediated by feedforward suppression. Finally, we show through simulations that feedforward inhibition can transform a dense representation of vocalizations into a sparse and background-invariant representation.

Introduction

In natural environments, important sensory stimuli are accompanied by competing and often irrelevant sensory events. Although simultaneous sensory signals can obscure one another, animals are adept at extracting important signals from noisy environments using a variety of sensory modalities (Born et al., 2000; Jinks and Laing, 1999; Raposo et al., 2012; Wilson and Mainen, 2006). As a striking yet common example of this perceptual ability, humans and other vocally communicating animals can recognize and track individual vocalizations in backgrounds of conspecific chatter (Cherry, 1953; Gerhardt and Klump, 1988; Hulse et al., 1997).

The ability to extract an individual vocalization from an auditory scene is thought to depend critically on the auditory cortex (Naatanen et al., 2001). In the human auditory cortex, population brain activity selectively reflects attended vocalizations within a multi-speaker environment (Mesgarani and Chang, 2012), and in humans and birds, population activity is stronger for vocalizations presented in levels of background sound that permit their behavioral discrimination compared to levels of background sound that do not (Binder et al.,

Correspondence author: Sarah M. N. Woolley, sw2277@columbia.edu.
[3]Present address: Department of Neurobiology, Duke University, Durham, NC 27710, USA

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.

2004; Boumans et al., 2008). Individual auditory cortical neurons appear well suited to encode vocalizations presented in a distracting background, in part because the acoustic features to which individual cortical neurons respond are more prevalent in vocalizations than in other sound classes (deCharms et al., 1998; Woolley et al., 2005). Further, in response to vocalizations, auditory cortical neurons often produce sparse and selective trains of action potentials (Gentner and Margoliash, 2003; Hromadka et al., 2008) that are theoretically well suited to extract and encode individual vocalizations in complex auditory scenes (Asari et al., 2006; Smith and Lewicki, 2006). However, electrophysiology studies have found that single neuron responses to individual vocalizations are strongly influenced by background sound (Bar-Yosef et al., 2002; Keller and Hahnloser, 2009; Narayan et al., 2007). Discovering single cortical neurons that produce background-invariant spike trains and neural mechanisms for achieving these responses would bridge critical gaps among human and animal psychophysics, population neural activity, and single-neuron coding.

Here, we identify a population of auditory neurons that encode individual vocalizations in levels of background sound that permit their behavioral recognition, and we propose and test a simple cortical circuit that transforms a background-sensitive neural representation into a background-invariant representation, using the zebra finch (*Taeniopygia guttata*) as a model system. Zebra finches are highly social songbirds that, like humans, communicate using complex, learned vocalizations, often in the presence of conspecific chatter.

Results

Behavioral recognition of songs in auditory scenes

We first measured the abilities of zebra finches to behaviorally recognize individual vocalizations (songs) presented in a complex background, a chorus of multiple zebra finch songs. We trained 8 zebra finches to recognize a set of previously unfamiliar songs using a Go/NoGo task (Gess et al., 2011) (Figure 1a) and we tested their recognition abilities when songs were presented in auditory scenes composed of one target song and the chorus (Figure 1b). We randomly varied the signal-to-noise ratio (SNR) of auditory scenes across trials by changing the volume of the song (48 to 78 dB SPL, in steps of 5 dB) while keeping the chorus volume constant (63 dB; Figure 1b). Birds performed well on high-SNR auditory scenes immediately after transfer from songs to auditory scenes (Figure S1), indicating that they recognized the training songs embedded in the scene. At high SNRs, birds performed as well as when the background was absent and their performance decreased sharply around 0 dB SNR (Figure 1c), in close agreement with the abilities of human subjects to recognize speech in noise (Bishop and Miller, 2009).

Transformation from dense to sparse coding of song

We next recorded the activity of single neurons at multiple stages of the auditory pathway while birds heard the songs that they had learned during behavioral training, the chorus alone and the auditory scenes used in behavioral testing. From each bird, we recorded single unit responses in the auditory midbrain (MLd, homolog of mammalian inferior colliculus, n = 100), the primary auditory cortex (Field L, thalamo-recipient and immediately adjacent regions, n = 99) and a higher-level auditory cortical region (NCM, n = 170; Figure 2a) that receives synaptic input from the primary auditory cortex (Table S1). Most primary auditory cortex (AC) neurons were recorded in the subregion L3, which provides the majority of input to the higher-level cortical region NCM (Figure S2). All electrophysiology experiments were performed with awake, restrained animals.

Action potential widths of higher-level AC neurons formed a continuous distribution with two clear peaks (p = 0.0001, Hartigan's dip test), suggesting two largely independent populations (Figure 2b). Higher-level AC neurons with narrow action potentials (0.1 to 0.4 ms) were classified as narrow spiking (NS, n = 35; 0.254+/−0.047 ms, mean+/−sd), while neurons with broad action potentials (>0.4 ms) were classified as broad spiking (BS, n = 135; 0.547+ −0.102 ms, mean+/−sd). BS and NS neurons were also largely segregated based on song-driven firing rate, with 90% of BS neurons firing fewer than 3.5 sp/s and 89% of NS neurons firing greater than 3.5 sp/s. In contrast to this bimodal distribution, widths of midbrain and primary AC action potentials formed unimodal distributions with peaks in the NS range and tails extending into the BS range that included only a small fraction of neurons (7%, midbrain; 11%, primary AC). None of the BS-like midbrain neurons had driven firing rates less than 3.5 sp/s and only 2% of primary AC neurons fired fewer than 3.5 sp/s. These analyses suggest that the higher-level AC contains a largely unique population of neurons with very broad action potentials and low firing rates.

Although individual neurons in each brain area responded to song playback with increased firing rates relative to spontaneous firing (mean z-scores of 4.17, 4.40, 3.31 and 1.36 in midbrain, primary AC, higher-level AC NS and BS populations, respectively), individual BS neurons in the higher-level AC fired fewer spikes, produced more precise spike trains, and were highly selective for individual songs. Song-driven firing rates of BS neurons were significantly lower than those of neurons in the midbrain, primary AC or NS neurons in the higher-level AC (2.4+/−2.7, 39.8+/−25.4, 32.4+/−20.1, 19.0+/−11.7 Hz, respectively; Figure 2d). Despite the low firing rates of BS neurons, the spikes that individual BS neurons produced were highly reliable across multiple presentations of the same stimulus. To quantify the precision of individual neurons, we computed the shuffled-autocorrelogram (SAC) from the spiking responses to individual songs. The value of the SAC at zero millisecond lag is termed the correlation index and it describes the propensity for a neuron to spike with submillisecond precision across multiple presentations of the same song, with a value of 1 indicating chance and larger values indicating greater degrees of trial-to-trial precision. BS neurons in the higher-level AC had significantly higher correlation index values (10.3+/13.0) than did midbrain, primary AC or higher-level AC NS neurons (correlation index of 2.8+/−3.1; 2.5+/−1.9; 2.1+/−0.7, respectively; Figure 2e). Also in contrast to other populations, BS neurons were typically driven by a subset of songs (6.9+/−5.2 out of 15), while midbrain, primary AC, and higher-level AC NS neurons responded to nearly every song (14.4+/−2.5; 14.7+/−1.9; 14.96+/−0.21 out of 15, respectively). We quantified response selectivity as 1 – (n/15), where n was the number of songs to which an individual neuron reliably responded. BS neurons in the higher-level AC were significantly more selective than were neurons in other populations (Figure 2f). Broad and narrow populations of neurons in the midbrain and primary AC did not differ in the neural coding of song (Figure S4). Further, we found no systematic relationship between response properties of primary AC neurons and anatomical location along the dorsal-ventral or anterior-posterior axes, each of which correlates with the location of subregions (Figure S5). Together, these results show that the neural coding of song changes minimally between the midbrain and primary AC, but a stark transformation in song coding occurs between the primary AC and BS neurons in the higher-level AC.

As a population, BS neurons represented songs with a sparse and distributed population code, in contrast to neurons in upstream areas. The BS neurons driven by a particular song each produced discrete spiking events at different times in the song (Figure 3a), resulting in a sparse neural representation that was distributed across the population. We quantified population sparseness by measuring the fraction of neurons in each population that were active during a sliding window of 63 ms, which is the average duration of a zebra finch song note (the basic acoustic unit of song; see spectrogram in Figure 3a). While more than 70% of neurons in upstream auditory areas fired during an average 63 ms window, fewer than 5% of BS neurons were active during the same epoch (Figure 3b).

Despite the markedly different population coding of song in the BS population compared to the NS and upstream populations (Figure S3), the temporal pattern produced by the BS population was similar to the temporal patterns produced by the dense coding populations (Figure 3c). Each population fired throughout the duration of a song, followed the temporal envelope of the song and was most strongly driven by syllable onsets. These findings show that the neural representation of individual songs transforms from a dense and redundant code in the midbrain and primary AC to a sparse and distributed code in a subpopulation of neurons in the higher-level AC.

Sparse coding neurons extract songs from scenes

We next examined the coding of individual songs in auditory scenes. Figure 4a shows responses of representative neurons to a song presented at multiple sound levels, chorus and auditory scenes presented at multiple SNRs. BS neurons in the higher-level AC responded reliably to songs in levels of chorus that permitted behavioral recognition, but largely stopped firing in levels of chorus that precluded behavioral recognition (see Figure 1c). In response to auditory scenes at SNRs below 5 dB, BS neurons fired fewer spikes than to the songs presented alone, indicating that the background chorus suppressed BS neurons' responses to songs (Fig. 4b). In contrast, midbrain, primary AC and higher-level AC NS neurons fired more in response to auditory scenes than to songs presented alone, consistent with the higher acoustic energy of auditory scenes compared to the song or chorus comprising them.

Higher-level AC BS neurons produced highly song-like spike trains in response to auditory scenes at SNRs that permitted behavioral recognition (Figure 5a). In contrast, neurons in upstream auditory areas and higher-level AC NS neurons produced spike trains that were significantly corrupted by the background chorus, including at SNRs that permitted reliable behavioral recognition. We quantified the degree to which each neuron produced background-invariant spike trains by computing the correlation between responses to auditory scenes and responses to the song component (R_{song}) and chorus component (R_{chor}) when presented alone. From these correlations we calculated an extraction index, $(R_{song} - R_{chor})/(R_{song} + R_{chor})$, which was positive when a neuron produced song-like responses and was negative when the neuron produced chorus-like responses.

The extraction indexes of BS neurons were significantly greater than the extraction indexes of upstream neurons and NS neurons, particularly at SNRs that permitted reliable behavioral recognition (Figure 5b). On average, BS neurons produced song-like spike trains at SNRs greater than 0 dB, whereas midbrain, primary AC and higher-order AC NS neurons produced song-like spike trains only at SNRs greater than 5 dB. The extraction index curves of BS neurons decreased precipitously between +5 and −5 dB SNR, in close agreement with psychometric functions (see Fig. 1c), whereas the extraction index curves of midbrain, primary AC and higher-level AC NS neurons decreased linearly. To quantify the rate at which the neural and behavioral detection of songs in auditory scenes changed as a function of SNR, we fit each extraction index curve and each psychometric curve with a logistic function, from which we measured the slope of the logistic fit. We found that extraction index curves of BS neurons and psychometric functions of behaving birds had similarly step-like shapes, and that they were both significantly more steep than extraction index curves of midbrain and primary AC neurons (Figure 5c). Interestingly, neurons in each brain area were equally good at extracting trained and unfamiliar songs (data not shown), indicating that training and behavioral relevance were not critical for the neural extraction of songs from auditory scenes. Further, segregating neurons in the midbrain and primary AC into broad and narrow populations revealed no significant differences in the extraction of songs from auditory scenes (Figure S4). These findings show that BS neurons represent individual songs in auditory scenes at SNRs that match birds' perceptual abilities to recognize songs in auditory scenes, in contrast to NS and upstream neurons.

Feedforward suppression sparsifies neural responses

The BS population represented individual songs with a sparse population code, in contrast to the representation of songs in upstream populations, and we next aimed to understand how a sparse sensory representation arises in the BS population. One neural mechanism for producing sparse sensory responses is with neurons that are only sensitive to very specific stimulus features. To determine whether BS neurons were sensitive to particular acoustic features, we computed a percentage similarity score (Sound Analysis Pro, Tchernichovski et al., 2000) for every pair of notes to which an individual BS neuron responded. Percentage similarity score describes the acoustic similarity of a pair of notes based on measures of pitch, amplitude modulation, frequency modulation, Weiner entropy and goodness of pitch. Like neurons in other auditory populations, pairs of notes to which a BS neuron responded were spectrotemporally more similar to one another (percentage similarity score, 69.2+/−28.3) than were notes selected at random (percentage similarity score, 45.8+/−27.2, mean +/− sd; $p<0.0001$; Figure S6a–b). However, unlike other recorded neurons, BS neurons often failed to respond to every iteration of a note that was repeated multiple times in a song (Figure S6c and see Figure 7a), and notes that were spectrotemporally similar to a response-evoking note often failed to evoke a response (see Figure S6b). These observations indicate that although individual BS neurons were sensitive to particular acoustic features, acoustic features alone may be inadequate for predicting their responses.

To quantitatively assess the acoustic features to which BS neurons were tuned, we next computed spectrotemporal receptive fields (STRFs). STRFs provide an estimate of the acoustic features to which a neuron is sensitive, and the complexity of a receptive field can indicate a neuron’s selectivity for complex or rarely occurring acoustic features. Such highly selective feature detectors could lead to sparse firing patterns and could potentially differentiate between subtle variations of a repeated note, as we observed in the BS population. For each neruron, we computed a STRF based on the spiking responses to all but one of 15 songs, and we validated each STRF by using it to predict the response to the song not used during STRF estimation. The STRFs of midbrain, primary AC and higher-level AC NS neurons showed clear tuning for particular acoustic features (Figure S6d) and could be used to accurately predict neural responses to novel stimuli (Figure S6e). In contrast, the acoustic features to which BS neurons in the higher-level AC were sensitive were poorly characterized by STRFs, and STRFs of BS neurons were poor predictors of neural responses to novel stimuli. These results suggest that the responses of BS neurons may be modulated by more than the short time-scale acoustic features that are typically coded by upstream populations.

To determine whether BS neurons were sensitive to long time-scale acoustic information (10s to 100s of ms), we presented individual notes independent of their acoustic context in songs. We reasoned that if BS neurons are highly selective feature detectors that were only sensitive to short time-scale information, they should respond to the same subset of notes when presented independently or in the context of a song. We further predicted that BS neurons should retain their selectivity for some iterations of a repeated note but not for others. Contrary to these predictions, BS neurons responded to 8 times more notes when they were presented independently (in the absence of acoustic context) than in the context of the song ($p<0.05$, Wilcoxon; Figure 6 a–b). Further, when notes were presented independently, BS neurons tended to respond to more iterations of a repeated note than when they were presented in the context of song (see Figure 6a). The finding that BS neurons can respond to notes that do not drive a response during song indicates that preceding notes within a song suppress a neuron’s response to subsequent notes.

To measure the time course of contextual suppression during the playback of song, we systematically increased or decreased the interval between notes that evoked responses and the notes immediately preceding them (Figure 6c). We found that acoustic context influenced BS neuron responses to subsequent notes with interactions lasting at least 100 ms (Figure d). The suppression induced by preceding notes did not require that the neuron respond to the preceding notes (e.g. Figure 6c), suggesting that contextual suppression is synaptic rather than due to intrinsic hyperpolarizing currents, which are typically activated after spiking (Cordoba-Rodriguez et al., 1999). Removing the acoustic context had no effect on the number of notes to which NS or primary AC neurons responded (data not shown). Therefore, presenting notes in the context of song prevented BS neurons, but not other neurons, from responding to notes that were capable of driving spiking responses and this suppression was not spiking dependent.

The context dependence of responses to songs suggests a role for synaptic inhibition in contextual suppression. We next explicitly tested the role of GABA in the contextual suppression of song responses by presenting songs while locally blocking inhibitory synaptic transmission within the higher-level AC using the selective GABA-A receptor antagonist gabazine (Thompson et al., 2013). We found that BS neurons responded to 9 times as many notes with inhibition blocked than without ($p<0.05$, Wilcoxon; Figure 7a–b), in agreement with the increase in responsive notes found by removing the acoustic context. Further, the additional notes to which neurons responded under gabazine were spectrotemporally similar to the notes that evoked a response under non-gabazine conditions (percentage similarity score of non-gabazine responsive vs. gabazine responsive notes, 64.2+/−31.1; percentage similarity score of randomly selected notes, 45.8+/−27.2, mean +/− sd; $p<0.0001$). Blocking inhibition had no effect on the number of notes to which NS neurons responded ($p>0.05$, Wilcoxon; Figure 7c) and blocking inhibition in the primary AC had no effect on the number of notes to which primary AC neurons responded ($p>0.05$, Wilcoxon, data not shown).

A functional circuit for sparse and background-invariant neural representations

Presenting notes independently or blocking inhibition in the higher-level AC both increased the number of notes to which BS neurons were responsive. Under both experimental conditions, the additional notes to which a BS neuron responded were spectrotemporally similar to notes to which the neuron responded without experimental manipulation (data not shown), suggesting that BS neurons received spectrotemporally tuned input that was suppressed under normal song conditions. Song manipulation experiments showed that preceding song notes provided feed-forward suppression and gabazine experiments suggested that this suppression was mediated by synaptic inhibition. Taken together, these findings are suggestive of a cortical architecture of feedforward inhibition, similar to that described in the mammalian auditory cortex (Tan et al., 2004; Wehr and Zador, 2003).

We next designed and simulated a putative circuit of feedforward inhibition that is based in part on the assumptions that NS neurons are inhibitory whereas BS neurons are excitatory, and that excitatory and inhibitory inputs to BS neurons are matched in spectral tuning. Although these assumptions are supported by anatomical, pharmacological and physiological studies (Vates et al., 1996; Atencio and Schreiner, 2008; Mooney and Prather, 2005; see Discussion), they have not been explicitly tested. Rather than to propose an exact wiring diagram, the purpose of the model is to test the hypothesis that a simple circuit of feedforward inhibition can reproduce the sparse and background-invariant song representations that we observed in BS neurons.

In the circuit shown in Figure 8a, both BS and NS higher-level AC neurons receive direct excitatory input from the primary AC, and NS neurons provide delayed and sustained inhibition onto BS neurons. In response to a brief input from the primary AC, a simulated BS neuron receives a burst of excitation followed by delayed and prolonged inhibition (Figure 8a inset). Based on this temporal filter, we simulated the spiking activity of BS neurons (n = 70), each of which received as input the responses of an individual primary AC neuron (n = 70) to songs, chorus and auditory scenes. Primary AC activity was simulated using receptive fields estimated from responses to songs (Calabrese et al., 2011). Simulations of this circuit transformed dense and continuous primary AC responses to song into sparse responses that were selective for a subset of songs, firing reliably in response to specific notes (Figure 8b). The firing rate, selectivity and sparseness of simulated BS neurons were similar to those observed in experimentally recorded BS neurons (Figure S7). In response to auditory scenes at SNRs above 0 dB, simulated BS neurons produced precise spike trains similar to those produced in response to the song presented alone, and at low SNRs, most simulated BS neurons stopped firing (Figure 8c). As in recorded responses, simulated BS neurons extracted individual songs from auditory scenes better than simulated primary AC neurons at high and intermediate SNRs (Figure 8d). Using raw PSTHs from primary AC neurons as inputs to the model rather than simulated PSTHs produced similar results (data not shown). Together, these simulations show that a cortical circuit of feedforward inhibition can accurately reproduce the emergence of sparse and background-invariant song representations.

Discussion

These findings are the first to report a population of auditory neurons that produce background-invariant responses to vocalizations at SNRs that match behavioral recognition thresholds. Individual broad spiking (BS) neurons in the higher-level AC respond sparsely and selectively to a subset of songs, in contrast to narrow spiking (NS) neurons and upstream populations. BS neurons largely retain their song-specific firing patterns in levels of background sound that permit behavioral recognition, and stop firing at SNRs that preclude behavioral recognition. These results suggest that the activity of BS neurons in the higher-level AC may serve as a neural mechanism for the perceptual extraction of target vocalizations from complex auditory scenes that include the temporally overlapping vocalizations of multiple individuals.

To measure behavioral recognition, we trained birds to report the identity of an individual song presented simultaneously with a distracting chorus using a Go/NoGo task. Although Go/NoGo behaviors are typically described as discrimination tasks, a variety of strategies could be used to perform the task, all of which require subjects to detect target sounds but not necessarily to discriminate among them. In our physiological experiments, neural responses were recorded during passive listening and reflect the abilities of neurons to detect, but not necessarily discriminate among, songs within auditory scenes. Our physiology results show that BS neurons in the higher-level AC provide a signal that could be used for accurate detection of target vocalizations in auditory scenes at SNRs that match behavioral thresholds, regardless of the strategy birds used during behavioral testing. It is still unclear how or where these neural signals are integrated with decision-making and motor-planning circuits to produce the appropriate behavioral response during the recognition task.

By analyzing the action potential shape of individual cortical neurons, we identified largely independent narrow and broad spiking populations in the higher-level AC and found that these populations could play unique functional roles in the processing of songs and auditory scenes. A small fraction of midbrain and primary AC neurons have action potential widths that we call broad (>0.4 ms), but action potential widths in these regions did not form bimodal distributions and BS and NS neurons in these regions did not show significant differences in responses to songs or auditory scenes. Categorizing intermingled neurons based on action potential width has been critical for understanding neural coding in the songbird vocal production system and in the mammalian cortex (Dutar et al., 1998), in large part because broad and narrow spiking neurons in these systems tend to form distinct excitatory and inhibitory populations. Whether NS and BS neurons in the higher-level AC comprise distinct inhibitory and excitatory populations remains to be tested.

In agreement with many previous reports, we find that the neural representation of communication sounds transforms at subsequent stages of auditory processing (e.g. Chechik et al., 2006; Meliza and Margoliash, 2012). Our findings provide strong evidence that the representation of songs and auditory scenes is transformed dramatically between the primary and higher-level AC. However, we cannot rule out the possibility that significant transformations in the neural coding of songs and auditory scenes occur within the primary AC, and that these transformations are inherited by the higher-level AC. Further studies are necessary to fully describe the representation of auditory scenes at multiple stages in the primary AC (see Meliza and Margoliash, 2012) and to look at monosynaptic transformations between projection neurons in the primary AC and neurons in the higher-level AC.

Our results differ in two important ways from recent findings in another songbird species, the European Starling (Meliza and Margoliash, 2012). First, we see a large increase in selectivity between the primary AC and the higher-level AC, but only in BS neurons. In contrast, in the auditory cortex of the European Starling there is a smaller (but significant) increase in selectivity between the two stages of processing and only small differences in selectivity between BS and NS populations. These differences could be attributed to multiple factors including anesthetic state (awake vs. urethane anesthetized), method for computing selectivity (song vs. motif) or boundary criteria between action potentials of NS and BS populations. Second, we found no effect of learning on song coding or auditory scene processing in the higher-level AC, in contrast with previous reports from the European Starling (e.g. Gentner and Margoliash, 2003; Meliza and Margoliash, 2012), which may suggest differences in cortical plasticity between species with open-ended (European Starling) and close-ended (zebra finch) learning periods.

We propose and model a cortical circuit based on feedforward inhibition that recapitulates salient aspects of the neural coding transformations observed between the primary and higher-level AC. Although the results of the simulation are in close agreement with our physiological and pharmacological findings, the model makes assumptions regarding the identity and connectivity of excitatory and inhibitory neurons, and the relative timing of excitatory and inhibitory inputs. The model also assumes that excitatory and inhibitory inputs to BS neurons are perfectly co-tuned in frequency, since in the model excitation is directly supplied and inhibition is indirectly supplied by the same neuron in the primary AC. Although we do not explicitly verify these assumptions, they are supported by previous studies showing that the higher-level AC receives direct synaptic input from the primary AC and is richly interconnected by local interneurons (Vates et al., 1996), and that neurons in the songbird (Mooney and Prather, 2005) and mammalian (Atencio and Schreiner, 2008) cortex can be segregated based on action potential width into excitatory (broad) and inhibitory (narrow) populations. Our data show that primary AC and NS neurons in the higher-level AC have similar spike train patterns, firing rates, selectivity and STRFs, in support of NS neurons receiving direct excitatory input from the primary AC. Lastly, spectrally co-tuned but temporally offset excitation and inhibition have been demonstrated in the mammalian auditory cortex (Wehr and Zador, 2003). Our proposed model captures our experimental findings and makes testable hypotheses about how the auditory cortex is organized to transform behaviorally relevant information.

Across organisms and sensory modalities, examples of sparse coding (Crochet et al., 2011; DeWeese et al., 2003; Stopfer et al., 2003; Weliky et al., 2003), contextual sparsification (Haider et al., 2010; Vinje and Gallant, 2000) and feedforward inhibition (Tiesinga et al., 2008; Vogels et al., 2011; Wehr and Zador, 2003) are common. The ubiquity of these neural phenomena and the necessity of social animals to extract communication signals from noisy backgrounds suggest that our results may demonstrate a basic mechanism for generating sparse codes from dense codes and for the neural extraction of important sensory signals from complex environments.

Experimental Procedures

Behavioral Training and Testing

Eight male zebra finches were trained to recognize the songs of other zebra finches using a Go/NoGo operant conditioning paradigm (Gess et al., 2011). For each bird, two songs were selected from a group of 15 as Go stimuli and two songs were selected as NoGo stimuli. Sounds were presented through a free field speaker located directly above the bird. Each bird was trained on a different set of four songs. Birds reached a performance level of 80% correct after 1500 to 10,000 trials, after which we tested their abilities to recognize the Go and NoGo songs when they were part of auditory scenes. Auditory scenes were interleaved with trials containing only the song or only the chorus. Positive and negative outcomes for hits and false alarms were the same during testing with auditory scenes as they were during training with songs, and chorus-alone trials were reinforced randomly. Each bird performed at least 3300 trials during behavioral testing (100 per unique stimulus), and all testing trials were included for computing psychometric functions.

Behavior and physiology were performed sequentially rather than simultaneously because (i) the low yield of simultaneous physiology and behavior would have limited the surveying of neurons in multiple auditory areas and sampling of neurons throughout the volume of each area; (ii) higher-level AC BS neurons were sparse firing and difficult to isolate, further decreasing the yield of simultaneous physiology and behavior experiments; (iii) higher-level AC BS neurons were responsive to only a subset of songs, and not necessarily those that birds were trained to discriminate; and (iv) in the time during which BS neurons were isolated, birds were unlikely to perform a sufficient number of trials to obtain meaningful results. Sequential behavior and physiology allowed for accurate characterization of psychometric functions and high yields of well-isolated neurons at multiple stages of the auditory pathway.

Stimuli

Behavioral and electrophysiological experiments were performed with the same set of song, chorus and auditory scene stimuli. The songs were from 15 unfamiliar zebra finches. The zebra finch chorus was created by superimposing the songs of 7 unfamiliar zebra finches that were not included in the library of individual songs. To remove energy troughs from the chorus, we applied a time-varying scaling function that was inversely proportional to the RMS energy, averaged over a sliding 50 ms window. This was done so that chorus amplitude troughs did not influence the detection of each song differently by allowing “dip listening” (Howard-Jones and Rosen, 1993). Each song was 2.0 seconds in duration. For both behavioral training and electrophysiology, each individual song was flanked by 0.25 seconds of zebra finch chorus, resulting in total durations of 2.5 seconds. We used flanking chorus to eliminate onset and offset cues that could signal the song identity during behavioral recognition and because variations in the strength and timing of the onset response across stimuli could provide potent cues for neural discrimination. Auditory scenes were composed of an individual song presented simultaneously with the chorus. We varied the SNR of the auditory scene by varying the song level (48 to 78 dB, in steps of 5 dB) while keeping the chorus level constant (63 dB). All neural analyses were constrained to the central 2 seconds that were unique to each stimulus.

Songs were separated into notes based on changes in overall energy and transitions in spectrotemporal features. When two contiguous notes morphed into one another without any obvious transition point, the note sequence was left intact and presented as a single "note". To determine the acoustic similarity between pairs of notes we compared their spectrotemporal features using Sound Analysis Pro (Tchernichovski et al., 2000). For every pair of notes we computed a percentage similarity score that quantified their overall acoustic similarity based on measures of pitch, amplitude modulation, frequency modulation, Weiner entropy and goodness of pitch. Notes that were spectrotemporally similar to one another had percentage similarity scores near 100%, whereas notes that were spectrotemporally different from one another had percentage similarity scores near 0%. We also computed individual acoustic features for each note. To determine whether a BS neuron in the higher-level AC was responsive to particular spectrotemporal features, we computed the percentage similarity between notes that evoked responses and we compared these values to the percentage similarity between notes selected at random.

Electrophysiology

Using electrophysiology techniques that have been previously described (Schumacher et al., 2011), we recorded the spiking activity of individual auditory neurons along three stages of the ascending auditory pathway in 8 awake birds; neurons were recorded from the mesencephalicus lateralis dorsalis (MLd, midbrain), Field L (used as a proper name, primary auditory cortex) and caudomedial nidopalliam (NCM, higher-level AC). Birds were not anesthetized during physiology but were restrained with a metal post affixed to the skull and a jacket around their bodies. Booth lights were on throughout the recording session. Craniotomies were made bilaterally at stereotaxic coordinates measured relative to the bifurcation of the sagittal sinus and centered over each of the three areas: MLd, 2.7 mm medial, 2.0 mm rostral; Field L, 1.3 mm medial, 1.3 mm rostral; NCM, 0.6 mm medial, 0.6 mm rostral. Glass pipettes (3 to 12 MOhm impedance) were used to record extracellular signals in each brain area. On each recording day, neurons were recorded in the midbrain of one hemisphere and the primary or higher-level AC of the other hemisphere, and locations were changed on subsequent days. Physiological recordings were made for up to 14 days after the last day of behavioral testing. On the last day of physiology, BDA injections were made along the recording paths to estimate recording sites.

From the responses of individual neurons we measured average firing rates, z-scores, precision and selectivity. Z-scores were measured as (driven firing rate – baseline firing rate)/(standard deviation of baseline firing rate). We quantified trial-to-trial precision by first computing the shuffled autocorrelogram using the spiking responses to individual songs (Joris et al., 2006). The shuffled autocorrelogram quantifies the propensity of neurons to fire spikes across multiple presentations of the same stimulus at varying lags. The correlation index is the shuffled autocorrelogram value at a lag of 0 ms, and it indicates the propensity to fire spikes at the same time (+/− 0.5 ms) each time the stimulus is presented. To quantify selectivity, we first determined the number of songs that drove at least one significant spiking event. Significant spiking events were defined by two criteria: (i) the smoothed PSTH (binned at 1 ms and smoothed with a 20 ms Hanning window) had to exceed baseline activity ($p<0.05$), and (ii) during this duration, spiking activity had to occur on > 50% of trials. Selectivity was then quantified as $1 - (n/15)$, where n was the number of songs (out of 15) that drove at least one significant spiking event.

To quantify population sparseness, we computed the fraction of neurons that produced significant spiking events during every 63 ms epoch, using a sliding window. We then quantified the fraction of neurons active during each window, with low values indicating higher levels of sparseness. To create population PSTHs, we first computed the PSTH of each individual neuron within a population in response to a single song, smoothed with a 5 ms Hanning window. We then averaged the PSTHs of every neuron in a population, without normalizing.

To quantify the degree to which neural responses to auditory scenes reflected the individual song within the scene, we computed an extraction index using the PSTHs to a scene at a particular SNR, as well as the PSTHs to the song and chorus components of that scene. From these PSTHs we computed two correlation coefficients: R_{song} was the correlation between the song and scene PSTHs and R_{chor} was the correlation coefficient between the scene and the chorus PSTHs. The extraction index was defined as $(R_{song} – R_{chor})/(R_{song} + R_{chor})$. Other methods for quantifying the extraction index from the PSTHs or from single spike trains produced qualitatively and quantitatively similar results.

Spectrotemporal receptive fields (STRFs) were calculated from the spiking responses to individual vocalization and the corresponding spectrograms using a generalized linear model, as has been previously described (Calabrese et al., 2011). We validated the predictive quality of each STRF by predicting the response to a song not used during estimation. We then calculated the correlation coefficient between the predicted and actual PSTHs.

We performed an ANOVA to determine the impact of bird ID, recording day, training performance, and recording hemisphere on neural selectivity and neural extraction from auditory scenes, and found that none of these variables were correlated with neural results (p>0.1).

Pharmacology

Local and temporary administration of the GABA-A receptor antagonist gabazine was performed simultaneously with electrophysiology using a carbon electrode coupled to a three-barrel pipette (Carbostar). Two pipettes were filled with 0.9% saline and one pipette was filled with 2.7 mM gabazine diluted in 0.9% saline. An Injection current of 30 nA was used to deliver both drug and vehicle, and a retention current of −30nA was used at all other times. A variable current was passed through the second saline barrel to balance the net current at the tip of the electrode. Physiology experiments during gabazine administration were started 2–5 minutes after beginning iontophoresis, which was continued throughout the drug phase. Immediately following gabazine administration, saline was administered for 5 minutes before and continuously throughout the wash-out phase.

Simulations

To simulate the activity of a primary AC neuron, we convolved the STRF of a primary AC neuron with the spectrograms of songs, chorus and auditory scenes. By rectifying the resultant with an exponential we generated a simulated PSTH that was highly similar to the PSTH recorded *in vivo* (r>0.60). We generated spike trains by sampling each PSTH with a Poisson spike generator and we simulated 10 trials of every stimulus.

The kernel defining the BS temporal filter was a mixture of excitatory and inhibitory Gaussians with different delays and variances, representing excitation from the primary AC and delayed inhibition from NS neurons, and was constant for every simulated BS neuron. We simulated multiple BS neurons, each of which had the same temporal filter but received input from a different primary AC neuron. In this way, each BS neuron inherited a spectrotemporal filter from the primary AC, onto which was applied a temporal kernel. The width of the excitatory Gaussian corresponded to the duration of a typical BS spiking event (~15 ms) and the width of the inhibitory Gaussian corresponded to the duration over which contextual suppression was observed *in vivo* (~100 ms). Because a single primary AC neuron provided input to the BS and NS neuron, the excitation and inhibition that each BS neuron received were co-tuned. To simulate BS spiking activity, we convolved a primary AC PSTH with the BS temporal kernel shown in Figure 5a. To the resultant of this convolution we added an offset, rectified the outcome with an exponential filter, and generated spiking activity with a Poisson spike generator. We quantified simulated primary AC and BS spike trains with the same methods described above for recorded spike trains.

Statistical Analysis

For statistical analysis, the non-parametric Kruskal-Wallis and Wilcoxon rank-sum tests were used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Moore, J. Schumacher and A. Calabrese for critical comments on the manuscript and for their input on experimental design and data analysis; G. Pollak generously donated equipment used for pharmacology experiments. D. M. S. was supported by the NIH (F31-DC010301) and S. M. N. W. was supported by the NIH (R01-DC009810) and the NSF (IOS-0920081).

References

Asari H, Pearlmutter BA, Zador AM. Sparse representations for the cocktail party problem. J Neurosci. 2006; 26:7477–7490. [PubMed: 16837596]

Atencio CA, Schreiner CE. Spectrotemporal processing differences between auditory cortical fast-spiking and regular-spiking neurons. J Neurosci. 2008; 28:3897–3910. [PubMed: 18400888]

Bar-Yosef O, Rotman Y, Nelken I. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J Neurosci. 2002; 22:8619–8632. [PubMed: 12351736]

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci. 2004; 7:295–301. [PubMed: 14966525]

Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. J Cogn Neurosci. 2009; 21:1790–1805. [PubMed: 18823249]

Born RT, Groh JM, Zhao R, Lukasewycz SJ. Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. Neuron. 2000; 26:725–734. [PubMed: 10896167]

Boumans T, Vignal C, Smolders A, Sijbers J, Verhoye M, Van Audekerke J, Mathevon N, Van der Linden A. Functional magnetic resonance imaging in zebra finch discerns the neural substrate involved in segregation of conspecific song from background noise. J Neurophysiol. 2008; 99:931–938. [PubMed: 17881485]

Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PLoS One. 2011; 6:e16104. [PubMed: 21264310]

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

Cherry EC. Some Experiments on the Recognition of Speech, with One and with 2 Ears. J Acoust Soc Am. 1953; 25:975–979.

Cordoba-Rodriguez R, Moore KA, Kao JP, Weinreich D. Calcium regulation of a slow post-spike hyperpolarization in vagal afferent neurons. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96:7650–7657. [PubMed: 10393875]

Crochet S, Poulet JF, Kremer Y, Petersen CC. Synaptic mechanisms underlying sparse coding of active touch. Neuron. 2011; 69:1160–1175. [PubMed: 21435560]

deCharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons. Science. 1998; 280:1439–1443. [PubMed: 9603734]

DeWeese MR, Wehr M, Zador AM. Binary spiking in auditory cortex. J Neurosci. 2003; 23:7940–7949. [PubMed: 12944525]

Dutar P, Vu HM, Perkel DJ. Multiple cell types distinguished by physiological, pharmacological, and anatomic properties in nucleus HVc of the adult zebra finch. J Neurophysiol. 1998; 80:1828–1838. [PubMed: 9772242]

Gentner TQ, Margoliash D. Neuronal populations and single cells representing learned auditory objects. Nature. 2003; 424:669–674. [PubMed: 12904792]

Gerhardt HC, Klump GM. Masking of Acoustic-Signals by the Chorus Background-Noise in the Green Tree Frog - a Limitation on Mate Choice. Anim Behav. 1988; 36:1247–1249.

Gess A, Schneider DM, Vyas A, Woolley SM. Automated auditory recognition training and testing. Anim Behav. 2011; 82:285–293. [PubMed: 21857717]

Haider B, Krause MR, Duque A, Yu Y, Touryan J, Mazer JA, McCormick DA. Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. Neuron. 2010; 65:107–121. [PubMed: 20152117]

Howard-Jones PA, Rosen S. Uncomodulated glimpsing in “checkerboard” noise. J Acoust Soc Am. 1993; 93:2915–2922. [PubMed: 8315155]

Hromadka T, Deweese MR, Zador AM. Sparse representation of sounds in the unanesthetized auditory cortex. PLoS Biol. 2008; 6:e16. [PubMed: 18232737]

Hulse SH, MacDougall-Shackleton SA, Wisniewski AB. Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (Sturnus vulgaris). J Comp Psychol. 1997; 111:3–13. [PubMed: 9090135]

Jinks A, Laing DG. A limit in the processing of components in odour mixtures. Perception. 1999; 28:395–404. [PubMed: 10615476]

Joris PX, Louage DH, Cardoen L, van der Heijden M. Correlation index: a new metric to quantify temporal coding. Hearing Research. 2006; 216–217:19–30.

Keller GB, Hahnloser RH. Neural processing of auditory feedback during vocal practice in a songbird. Nature. 2009; 457:187–190. [PubMed: 19005471]

Marler P. Song-learning behavior: the interface with neuroethology. Trends Neurosci. 1991; 14:199–206. [PubMed: 1713722]

Meliza CD, Margoliash D. Emergence of selectivity and tolerance in the avian auditory cortex. J Neurosci. 2012; 32:15158–15168. [PubMed: 23100437]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485:233–236. [PubMed: 22522927]

Mooney R, Prather JF. The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. J Neurosci. 2005; 25:1952–1964. [PubMed: 15728835]

Naatanen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. “Primitive intelligence” in the auditory cortex. Trends Neurosci. 2001; 24:283–288. [PubMed: 11311381]

Narayan R, Best V, Ozmeral E, McClaine E, Dent M, Shinn-Cunningham B, Sen K. Cortical interference effects in the cocktail party problem. Nat Neurosci. 2007; 10:1601–1607. [PubMed: 17994016]

Raposo D, Sheppard JP, Schrater PR, Churchland AK. Multisensory decision-making in rats and humans. J Neurosci. 2012; 32:3726–3735. [PubMed: 22423093]

Schumacher JW, Schneider DM, Woolley SM. Anesthetic state modulates excitability but not spectral tuning or neural discrimination in single auditory midbrain neurons. J Neurophysiol. 2011; 106:500–514. [PubMed: 21543752]

Smith EC, Lewicki MS. Efficient auditory coding. Nature. 2006; 439:978–982. [PubMed: 16495999]

Stopfer M, Jayaraman V, Laurent G. Intensity versus identity coding in an olfactory system. Neuron. 2003; 39:991–1004. [PubMed: 12971898]

Tan AY, Zhang LI, Merzenich MM, Schreiner CE. Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. J Neurophysiol. 2004; 92:630–643. [PubMed: 14999047]

Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. A procedure for an automated measurement of song similarity. Anim Behav. 2000; 59:1167–1176. [PubMed: 10877896]

Tiesinga P, Fellous JM, Sejnowski TJ. Regulation of spike timing in visual cortical circuits. Nat Rev Neurosci. 2008; 9:97–107. [PubMed: 18200026]

Thompson JV, Jeanne JM, Gentner TQ. Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. J Neurophys. 2013; 109:721–733.

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]

Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. Science. 2000; 287:1273–1276. [PubMed: 10678835]

Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science. 2011; 334:1569–1573. [PubMed: 22075724]

Wehr M, Zador AM. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature. 2003; 426:442–446. [PubMed: 14647382]

Weliky M, Fiser J, Hunt RH, Wagner DN. Coding of natural scenes in primary visual cortex. Neuron. 2003; 37:703–718. [PubMed: 12597866]

Wilson RI, Mainen ZF. Early events in olfactory processing. Annu Rev Neurosci. 2006; 29:163–201. [PubMed: 16776583]

Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci. 2005; 8:1371–1379. [PubMed: 16136039]

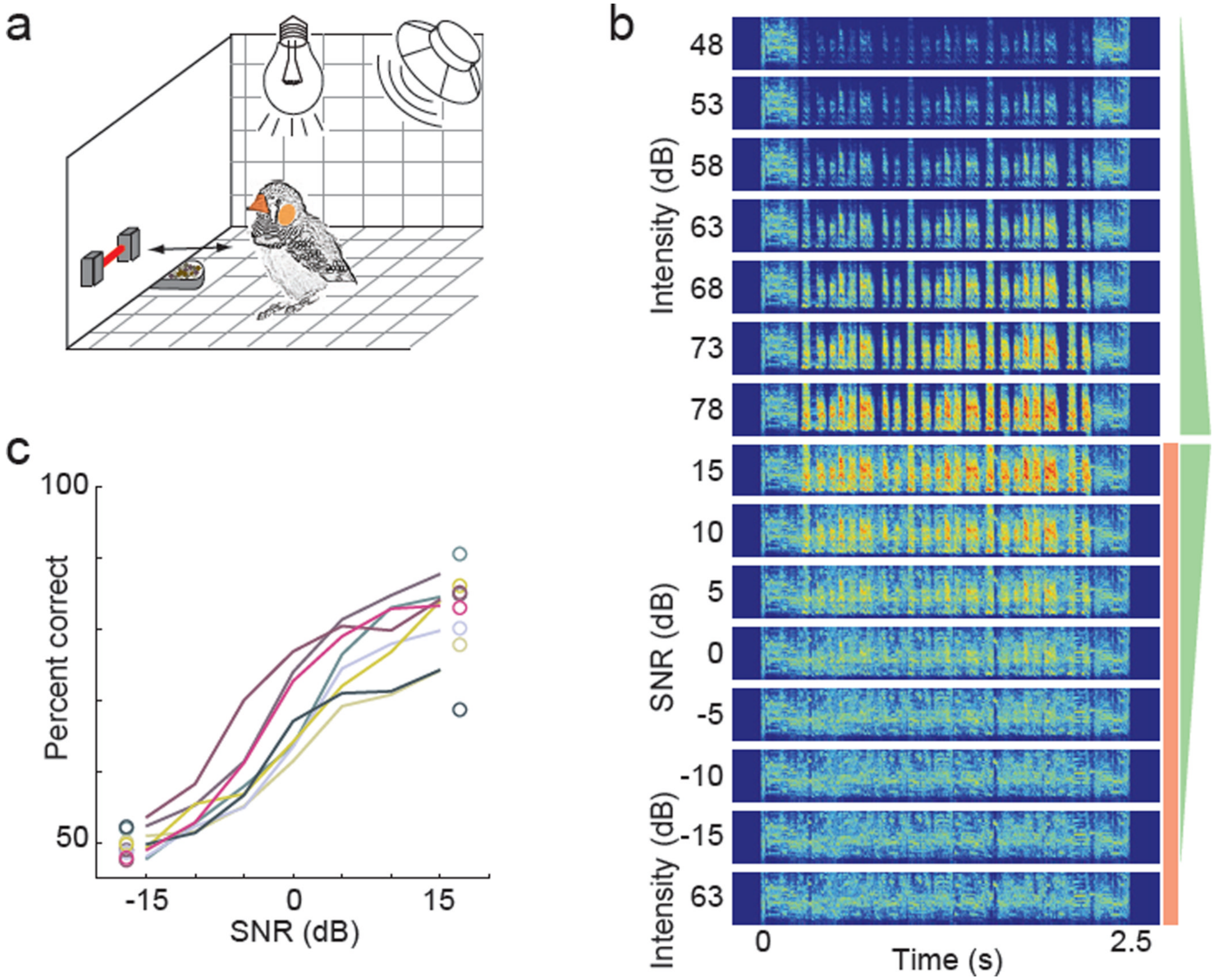

Figure 1. Behavioral recognition of songs in auditory scenes

(A) Birds were trained to recognize target songs in auditory scenes using a Go-NoGo task, in which birds initiated trials and responded to stimuli by breaking an infrared beam. Birds were rewarded with food for correct Go responses and punished with lights-out for incorrect NoGo responses. (B) Spectrograms showing frequency (ordinate: 0.25 to 8kHz) over time (abscissa) of a song presented at varying volumes and auditory scenes consisting of the song and a background chorus of conspecific songs presented at varying signal-to-noise ratios (SNRs). Chorus is shown at bottom. Green triangles and red rectangle on right schematize the volume of the song (green) and chorus (red) component comprising each sound. To minimize the facilitative effect that onset and offset cues have on behavioral and neural discrimination of vocalizations, each sound began and ended with the same 250 ms snippet of zebra finch chorus. (C) Birds' performance levels as a function of auditory scene SNR, and to songs alone (dots on right) and chorus alone (dots on left). Each colored line shows the data for one bird. See also Figure S1.

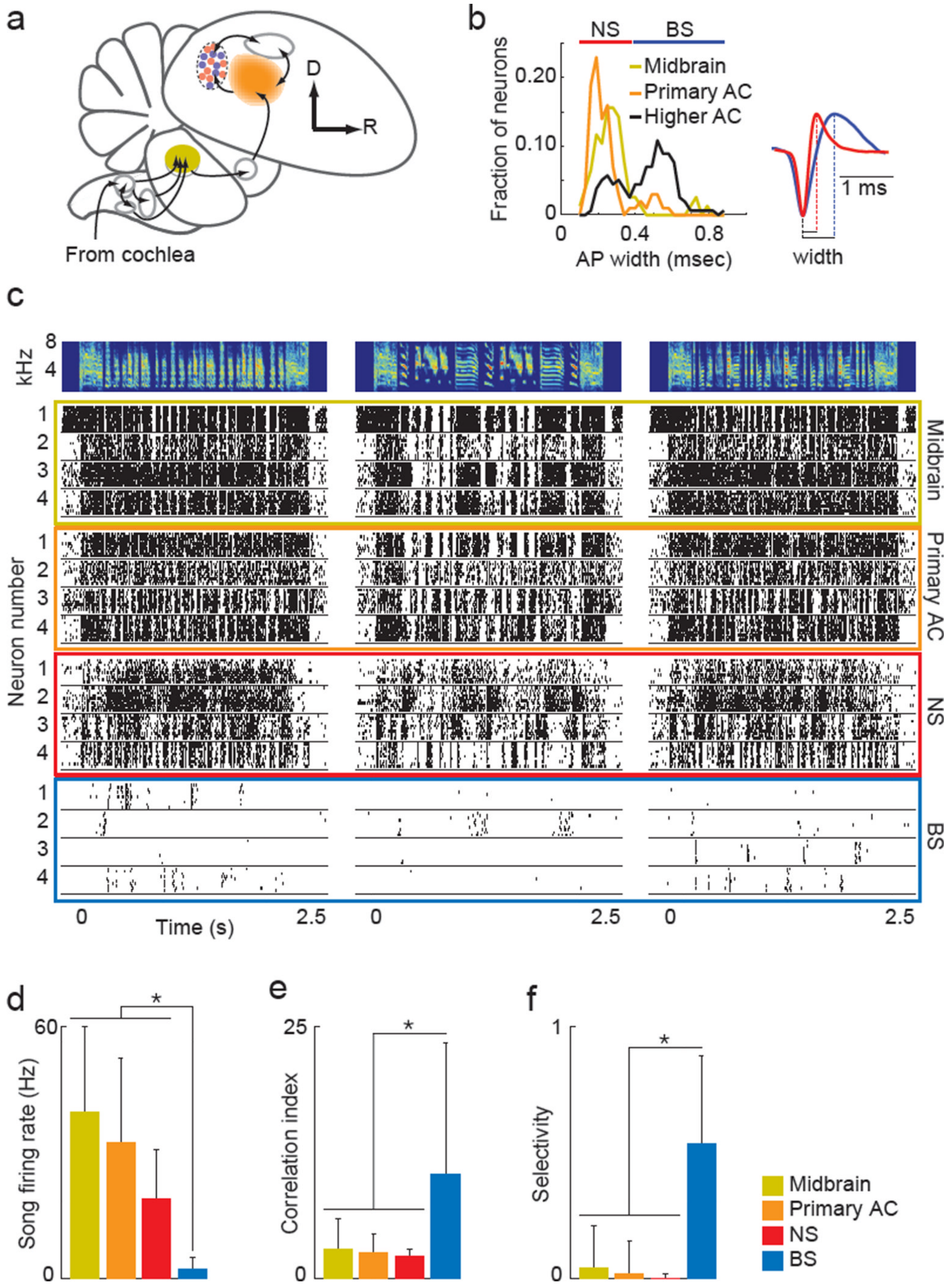

Figure 2. Neural transformations in the coding of vocalizations

(A) Schematic of the ascending auditory pathway. Neurons were recorded in the auditory midbrain (MLd, yellow), the primary auditory cortex (Field L, orange) and a higher-level region of the auditory cortex (NCM, red and blue). Other auditory areas are in gray. (B) Distributions of action potential widths in the three brain areas. Red and blue bars at top denote NS and BS ranges, respectively. Inset at right shows action potential widths of representative BS (blue) and NS (red) neurons in the higher-level AC. (C) Four example neurons from the midbrain (yellow) and primary AC (orange), and from each cell type in the higher-level AC (red, NS; blue BS) in response to three songs. Spectrograms of the three songs are on top. (D) Firing rates in response to songs. (E) Degree of millisecond precision in the spiking responses to repeated presentations of the same song, measured as the correlation index from the shuffled autocorrelogram. (F) Degree of selectivity for individual songs, measured as 1 – (n/15), where n is the number of vocalizations that drove a significant response. Legend corresponds to panels D–F. All bar graphs show mean +/− sd. Asterisks indicate $p < 0.05$, Kruskal-Wallis. See also Figure S2 and Table S1.

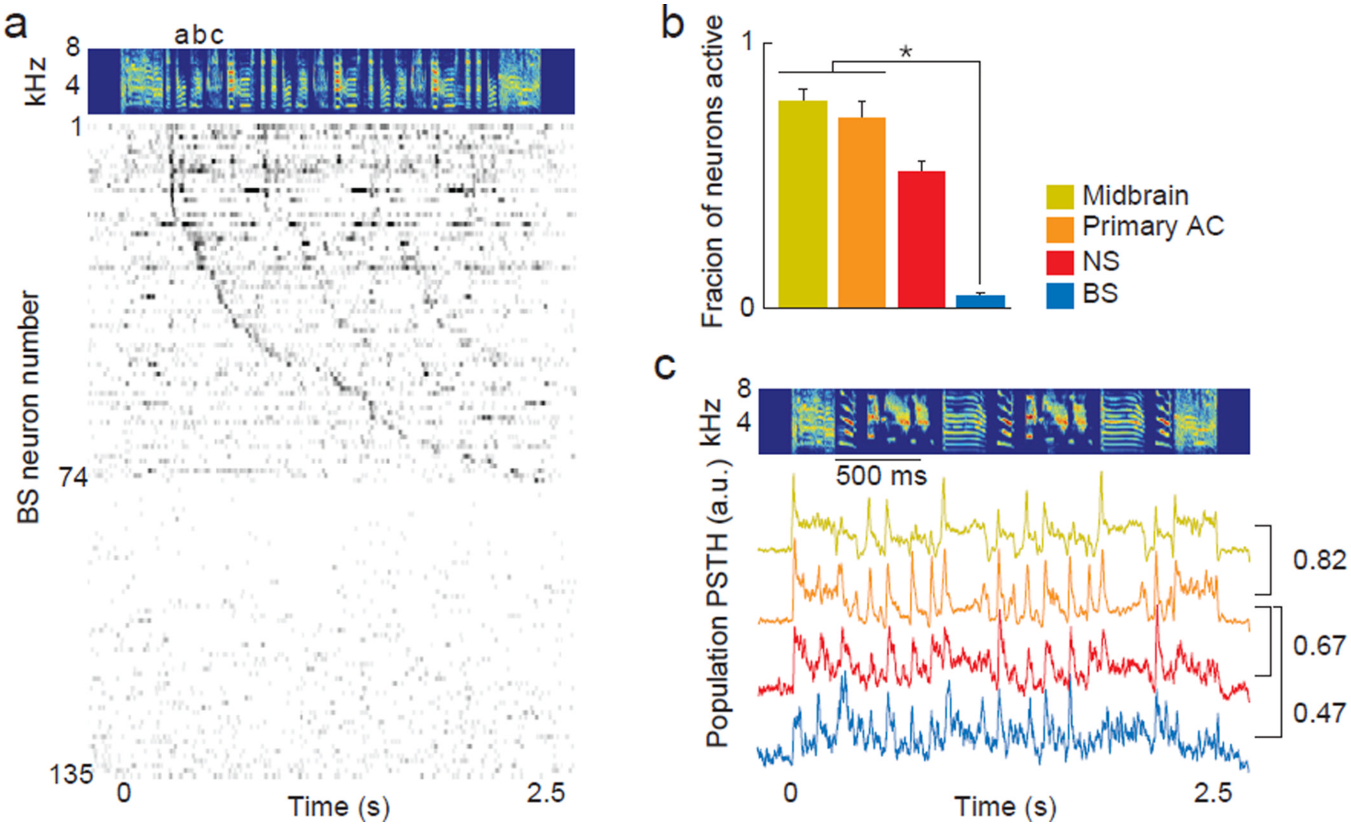

Figure 3. Population coding of vocalizations

(A) Neurogram of BS neurons in the higher-level AC in response to the song shown above. Each row shows the average firing rate over time for an individual neuron. Neurons were organized by the time of their first significant spiking event. Neurons 75 through 135 do not respond to this song. Gray scale is 0 (white) to 67 (black) spikes/sec. Letters above spectrogram indicate three unique notes. (B) Population sparseness measured as the fraction of all neurons active during each 63 ms epoch of song. Values near zero indicate a high degree of sparseness. NS neurons were not included in statistics because of small sample size. Bar graph shows mean +/− sd, asterisk indicates $p < 0.05$, Kruskal-Wallis. (C) Population peri-stimulus time histograms (PSTHs) showing the responses of all recorded neurons from each auditory area and each cell type in the higher-level AC to a song. Correlation coefficients between pairs of population PSTHs are shown at right. See also Figures S3, S4 and S5.

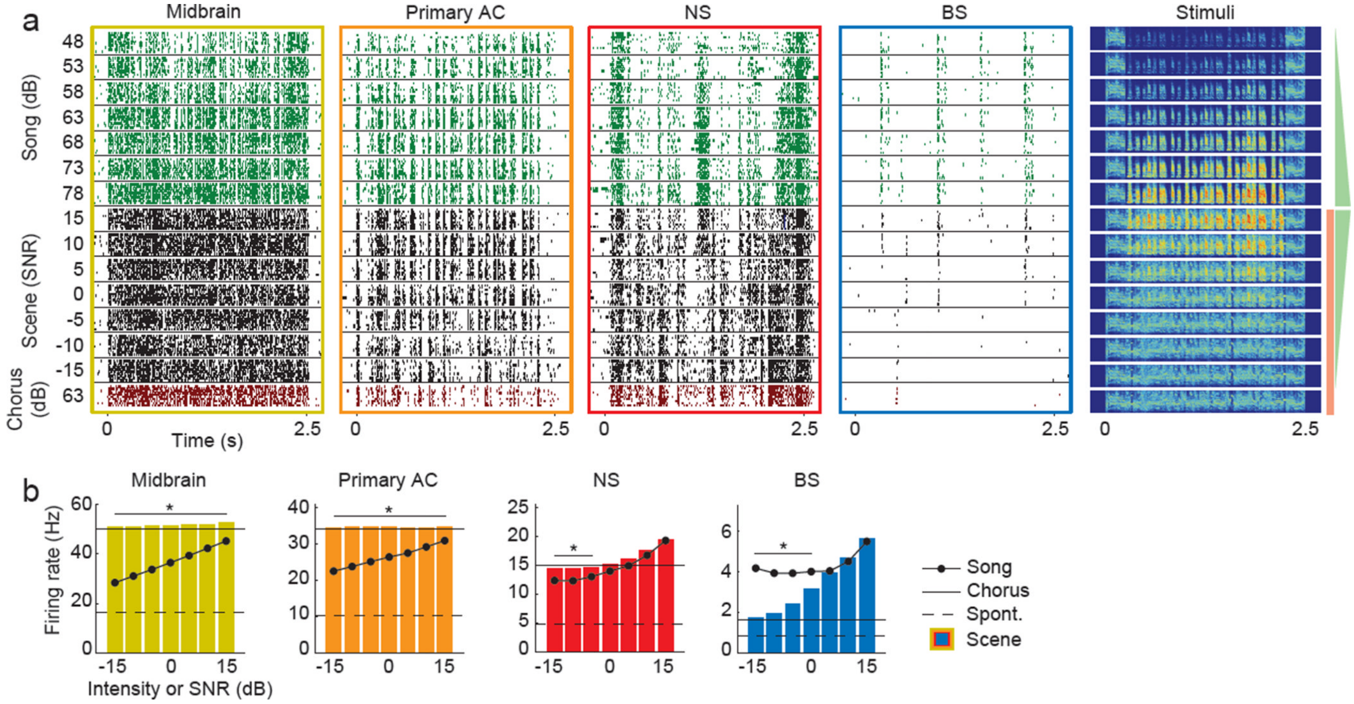

Figure 4. Neural encoding of songs in auditory scenes

(A) Examples of single neurons' responses to the songs, auditory scenes and chorus shown at the far right. Green spike trains are responses to songs, black to auditory scenes, and red to the chorus. Green triangles and red rectangle on right schematize the volume of the song (green) and chorus (red) components comprising each sound. (B) Average firing rates to songs at varying intensities (circles connected by solid line), auditory scenes at varying SNRs (bars), chorus (solid line) and silence (dashed line). Asterisks indicate signal to noise ratios for which the auditory scene and song firing rates are significantly different ($p < 0.05$, Wilcoxon). See also Figures S3 and S4.

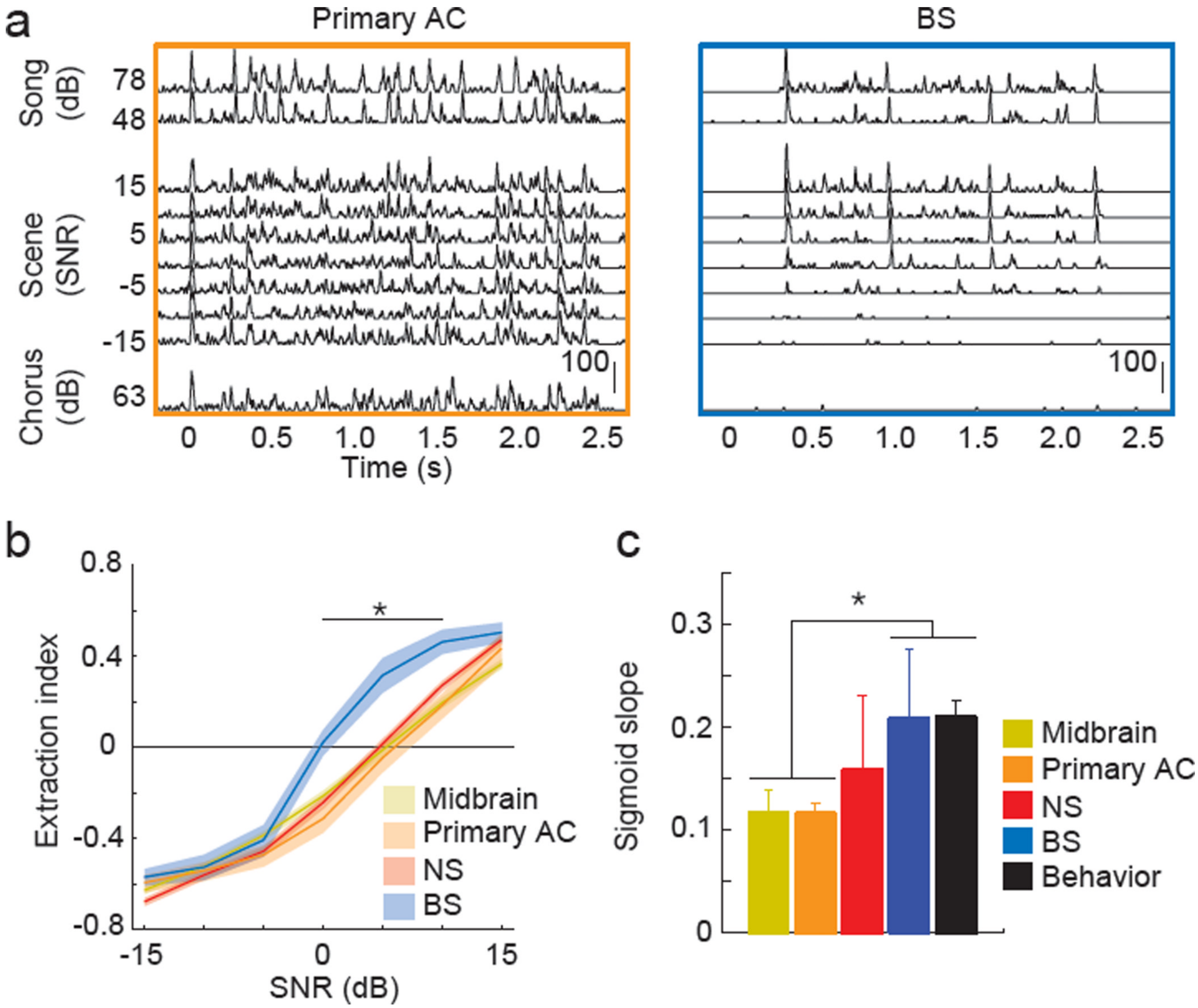

Figure 5. Background-invariant coding of auditory scenes

(A) Example PSTHs from an individual primary AC neuron (left) and BS higher-level AC neuron (right) to a song at highest and lowest intensity presented (top), to chorus (bottom) and to auditory scenes (middle). Scale bars show firing rate (Hz). (B) Extraction index shows the degree to which the response to auditory scenes was similar to the song response (positive numbers, +1 being identical) or the chorus response (negative numbers, −1 being identical). Solid lines show mean and shaded areas show +/−SEM. Asterisks indicate SNRs where BS neurons are significantly different than all other areas ($p < 0.05$, Kruskal-Wallis). (C) Slope of logistic fits of extraction index curves and psychometric functions.

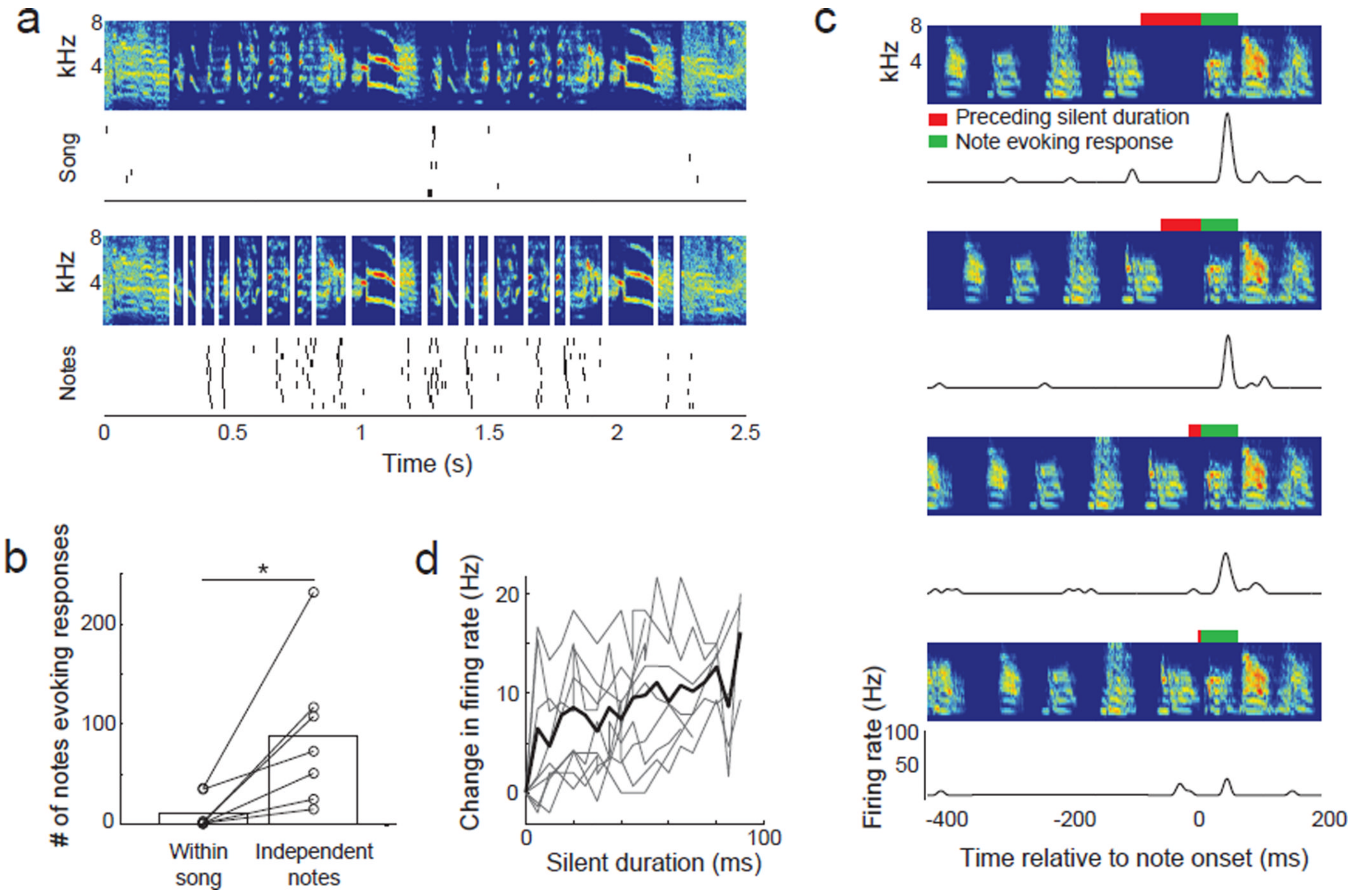

Figure 6. Sparsification due to local acoustic context

(A) Spectrogram and responses of a BS neuron to a song (top). Below are the spectrograms of individual notes (delineated by white vertical bars) and responses to notes, realigned to match the original spectrogram. (B) Number of notes to which BS neurons responded when presented within the song and when the notes were presented independently (n = 7). (C) PSTH responses of example neuron to songs with an extended or contracted silent gap (red bars) preceding a responsive note (green bars). The third spectrogram from the top shows the natural song. Contextual suppression does not depend on the neuron responding to preceding acoustic elements. (D) Change in firing rate as a function of the silent duration between a responsive note and preceding notes (n = 9). See also Figure S6.

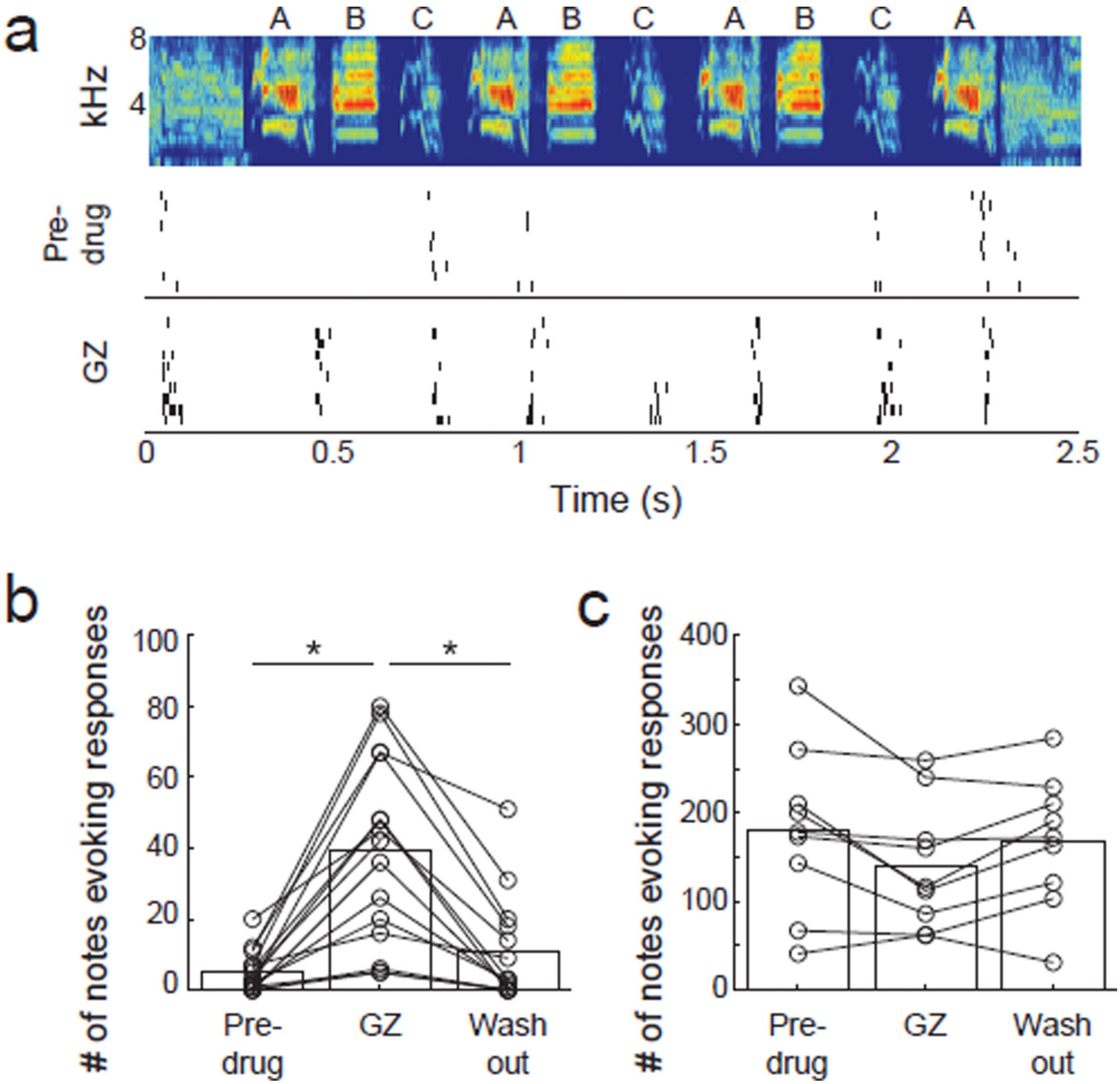

Figure 7. Sparsification due to synaptic inhibition

(A) Spectrogram (top) and responses of a BS neuron to a song without (middle) and with (bottom) local administration of gabazine. (B) Number of notes that BS neurons respond to before, during and after gabazine application (n = 14). (C) Number of notes that NS neurons respond to before, during and after gabazine application (n = 9). Asterisks indicate groups that are significantly different ($p < 0.05$, Wilcoxon).

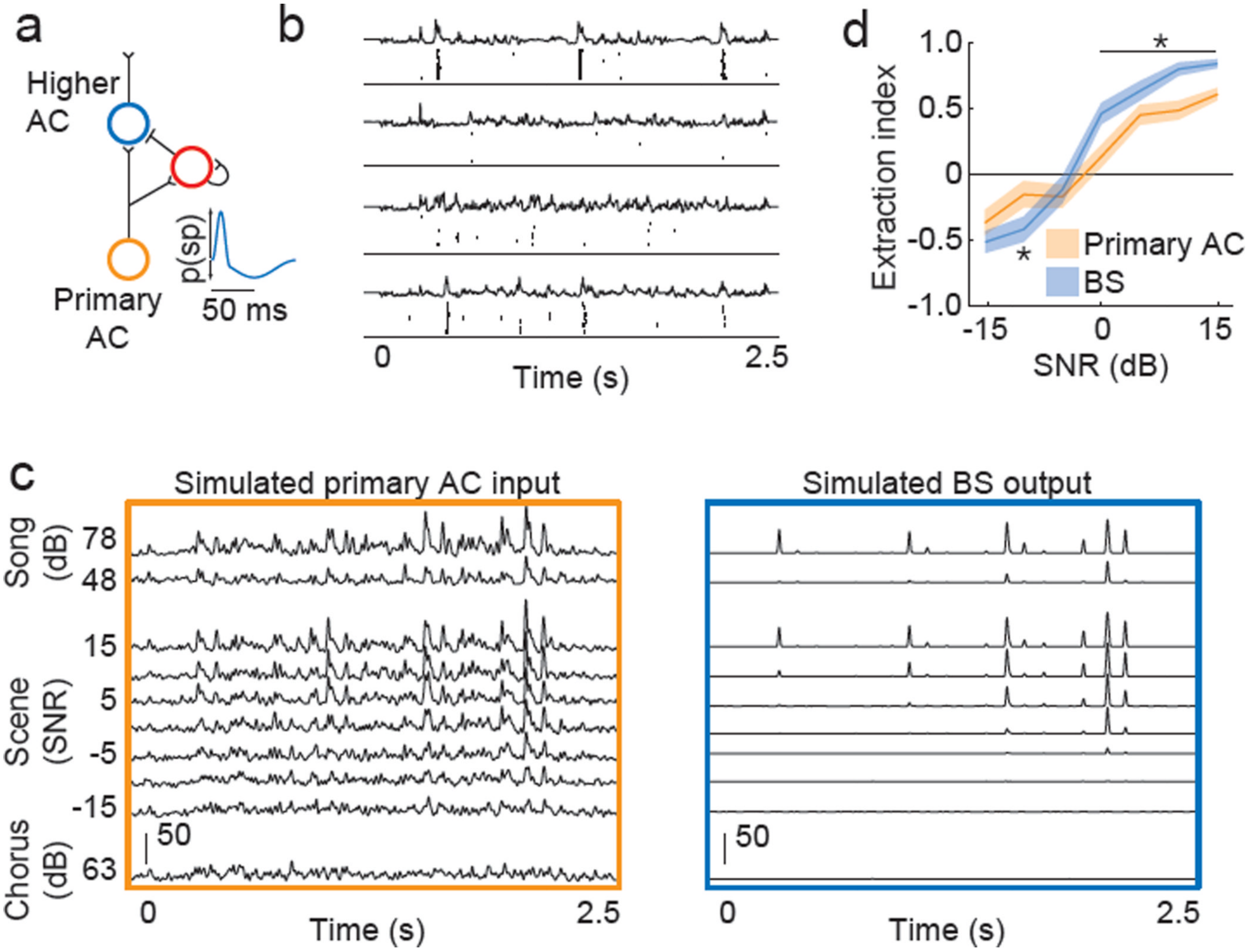

Figure 8. Simulating a functional circuit for sparse and background-invariant coding

(A) Functional circuit in which a primary AC neuron provides excitation to both BS and NS neurons in the higher-level AC. The NS neuron provides delayed and sustained inhibition onto the BS neuron. The auto-synapse onto the NS neuron represents any of a number of cellular or circuit mechanisms that could produce sustained firing that outlasts synaptic input to a neuron. Inset schematizes the change in spiking probability of BS neuron in response to a short burst of primary AC input. (B) Simulations of this circuit with primary AC responses to four different songs as input (continuous traces). Black ticks show spiking of a simulated BS neuron. (C) Simulations of this circuit with primary AC responses to auditory scenes as input (left). Average response of BS neuron to primary AC input is shown on right. Scale bars show firing rate (Hz). (D) Extraction index measured from the auditory scene responses of simulated primary AC (n = 70, orange) and higher-level AC BS neurons (n = 70, blue). Solid lines show mean and shaded areas show +/−SEM. Asterisks indicate SNRs at which the two populations are significantly different ($p < 0.05$, Wilcoxon). See also Figure S7.