2pAB2. Mechanisms of perceiving communication sounds in scenes
Sarah M. Woolley*

*Corresponding author's address: Psychology, Columbia University, 406 Schermerhorn Hall, New York, NY 10027, sw2277@columbia.edu

Vocal communicators must perceive the vocal signals of social partners in complex auditory scenes that include distracting background sounds. The auditory system must therefore parse auditory scenes into multiple information streams and/or accurately encode individual vocalizations despite the presence of competing sounds. Explaining mechanisms whereby neural representations of vocalizations are extracted from neural representations of scenes is an important part of understanding how auditory processing leads to perception of communication signals in complex scenes. We study how songbirds recognize individual vocalizations (songs) in scenes of conspecific choruses. We combine behavioral studies with neurophysiological studies of song and scene coding in midbrain and cortex. We find dramatic transformations in the neural coding of songs and scenes between different regions of auditory cortex. Neural representations of individual songs are dense and non-selective in the midbrain and primary cortex, but are sparse and highly selective in higher cortex. Sparse coding neurons produce background-invariant responses to individual songs in scenes, providing a potential neural mechanism for the perception of individual communication vocalizations in complex auditory scenes. Acoustic manipulations of song and pharmacological manipulations of neural coding suggest that sparse and background-invariant representations of songs in higher cortex are due to context-dependent inhibition.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

The songbird auditory system offers a unique opportunity to investigate the role of auditory processing in the perception of complex, learned vocal communication sounds. While all vertebrates have auditory systems and many communicate vocally, few communicate using learned sounds. Only humans, some cetaceans, and three clades of birds (parrots, hummingbirds and songbirds) are known to learn their vocalizations (Brenowitz and Beecher, 2005). Other animals produce and perceive unlearned vocalizations. Currently, the songbird is the only laboratory animal model of auditory processing and perception in vocal learners. Studies of songbird vocal behavior, including its dependence on learning, reveal numerous parallels between speech and birdsong. Examples include: 1) a basic organization of spectrotemporally complex syllables produced in temporal sequences; 2) critical periods for vocal learning; 3) learning through imitation of adult models; 4) the dependence of vocal behavior on auditory feedback; and 5) the use of unique vocal signals to recognize individuals (see Doupe and Kuhl, 1999 for review). These similarities in human and songbird vocal communication and the evolutionary conservation of auditory circuitry (Butler et al., 2011; Wang et al., 2010) suggest that the songbird is an ideal model system in which to study how the auditory system subserves perception of vocal communication sounds.

Complex auditory perceptual abilities that are classically described in human psychoacoustics can be studied in songbirds in which neural mechanisms can be worked out experimentally. One example is the ability to understand communication vocalizations in acoustically cluttered and distracting environments. Zebra finches (Taeniopygia guttata) are highly social songbirds that, like humans, communicate in complex acoustic environments that are filled with the vocalizations of many individuals. Zebra finches learn to recognize and discriminate among the unique songs of other individual birds (Fig. 1; Cynx and Nottebohm, 1992; Cynx 1993; Gess et al., 2011), similar to our recognition of unique human voices producing speech segments. Birds can also identify individual songs in complex scenes (Hulse et al., 1997; Schneider and Woolley, 2012). This ability in songbirds allows neuroscientists to study parallels between the perception and the neural processing of vocal communication sounds in auditory scenes.

Our understanding of how neural representations of songs transform along the auditory pathway can suggest auditory coding principles that are applicable to speech processing and perception. Studies on the auditory coding of song and other sounds in the zebra finch brain have led to insights into the coding properties of neurons that underlie perception of communication sounds (Woolley et al., 2005, 2009; Wang et al., 2007; Schneider and Woolley, 2011). We now have a general understanding of the spectral and temporal tuning properties of neurons from the midbrain to the cortex (Theunissen et al., 2000; Sen et al., 2001; Woolley et al., 2005, 2006, 2009; Amin et al., 2010; Schneider and Woolley, 2010, 2011). Additionally, auditory tuning mechanisms that facilitate neural coding of vocalizations over other sounds are well described in the zebra finch (Woolley et al., 2005, 2006; Schneider and Woolley, 2011). Knowledge of auditory coding properties in the songbird allows us to generate hypotheses about how the response properties of human auditory neurons may subserve speech processing and perception. This paper focuses on the perceptual abilities of songbirds to recognize individual songs, including in complex auditory scenes, and neural coding properties that may facilitate perception of communication sounds in complex environments.

Song Recognition in Birds

Auditory recognition training and testing using Go No-Go operant tasks show that zebra finches successfully learn to discriminate among songs in experimental settings (Cynx and Nottebohm, 1992; Cynx, 1993; Gess et al., 2011). Figure 1 shows the song discrimination learning curves of four adult male zebra finches trained to discriminate among four unique, unfamiliar zebra finch songs, with three exemplars of each song presented over a loudspeaker during training. Birds were trained to peck a sensor to initiate a song presentation and then peck the sensor a second time when they heard one of two songs (Go songs) and to refrain from pecking the sensor when they heard one of the other two songs (No-Go songs). Go and No-Go songs were randomized across birds. Because birds learn to recognize and respond correctly to different songs, birds’ abilities to accurately recognize songs in complex scenes can be tested by presenting target songs that birds have been trained to recognize and competing background sounds simultaneously. For example, studies in which target songs are buried in background scenes at varying intensities can be used to determine the role of signal to noise ratio and acoustic features of target sounds versus background sounds in song perception. Our experiments measure the accuracy of song discrimination when target songs are presented in combination with background choruses of multiple birds’ songs (Schneider and Woolley, 2012). These experiments show that zebra finches recognize target songs in scenes at signal to noise ratios that are similar to the signal to noise ratios at which humans recognize target sounds in auditory scenes (Bishop and Miller,
2009), providing an animal model for examining parallels between the auditory perception of vocalizations and the neural coding of those same vocalizations.

FIGURE 1. Song recognition learning curves for four male zebra finches show that birds learn to discriminate among the unique songs produced by individual birds. Birds were trained using a Go No-Go operant condition procedure. Each line shows the cumulative performance of a single bird over 100-trial training blocks to criterion. Birds reached criterion performance (80% correct on two consecutive trial blocks) in between 10 and 24 blocks. The solid black line represents 50% correct responses or chance performance. The dotted line represents criterion correct performance.

Neural coding of songs in the auditory cortex and its relationship to perception

The neural coding of songs in the zebra finch auditory system dramatically transforms between the midbrain and higher cortex. In the midbrain, single neurons respond to an individual song with spikes that reliably occur at specific points in time as the bird hears the song (Schneider and Woolley, 2010). The same neuron produces a different response to another song because the acoustics of the two songs differ and the spectrotemporal features to which the neuron is tuned occur at different times in different songs. Other neurons produce responses that differ from the responses of the first neuron because each neuron’s tuning properties are unique. Despite acoustic differences among songs and tuning differences among neurons, most midbrain neurons respond robustly to most songs. This coding scheme results in a dense and redundant neural representation of a song in the auditory midbrain. Between the input (thalamo-recipient) layer of auditory cortex and higher auditory cortical regions, the coding of songs transforms. Early in the cortical processing pathway, responses are non-selective (i.e. each neuron responds to a high proportion of presented songs) and include many spikes in the response to a single song, similar to the responses observed in the midbrain. At the highest levels of auditory cortex, single neuron responses are selective (i.e. one neuron responds to a small subset of songs) and are characterized by few spikes in the response to a song (Fig. 2). The few spikes that do occur during song presentation are reliable, meaning that they occur at similar times in the song presentation each time a song is presented. Because responses are so “sparse”, each neuron produces a highly distinct response pattern to each song, if it responds to the song. Higher cortical regions therefore represent songs in a sparse and distributed spiking code. The hierarchical transformation of song coding in the songbird auditory cortex is similar to transformations in sensory representations in other systems (Graham and Field, 2007).

Selective and sparse neural coding may facilitate the coding of target sounds such as individual songs in complex scenes such as song choruses. Sparse coding auditory neurons produce song responses that are more precise than song responses in upstream auditory regions; the temporal patterns of their responses are highly similar over multiple presentations of the same song (Schneider and Woolley, 2012). Such selective, sparse and precise song coding may facilitate the perceptual learning and recognition of complex sensory signals such as songs. Response selectivity is inversely correlated with the strength of responses to continuous sounds such as background noise that don’t contain silent epochs. For this reason, we have hypothesized that sparse coding neurons in the songbird auditory cortex subserves the perception of individual songs in scenes that include background choruses of temporally overlapping songs.
FIGURE 2. Responses of single neurons to songs transform from dense to sparse between the input layer and higher zebra finch auditory cortex. Raster plots show timing of spikes in response to the song shown in the spectrogram above. Left, spike rasters of four neurons in the thalamo-recipient layer show that responses are non-selective and dense. Middle, responses of four neurons in the cortex downstream of the thalamo-recipient region are also non-selective and dense, but have lower firing rates than in the thalamo-recipient layer. Right, responses in the higher cortex are sparse, with few, precise bursts of spikes. Small rows are responses of one neuron to one song presentation. Large rows are responses of one neuron to ten presentations (trials) of the song.

Because laboratory studies allow the careful control and systematic variation of signal to noise ratios of target sounds and scenes, songbirds can be used to correlate the presence and absence of perceptual recognition with neural coding of target sounds and complex scenes at varying signal to noise ratios. Our experiments compare the song (signal) to background chorus (noise) ratios at which birds accurately discriminate among individual songs buried in choruses and the signal to noise ratios at which auditory neurons respond to individual songs in choruses. The strength and temporal patterns of auditory responses to songs presented alone and songs presented in choruses are compared to determine whether song responses are resistant to or corrupted by background choruses. Results suggest that sparse coding neurons produce background-invariant responses to songs in choruses at lower signal to noise ratios than do dense coding, non-selective neurons. In contrast, responses of non-selective, dense coding neurons reflect the combined acoustics of target songs and choruses. We conclude that the coding behavior of sparse firing, higher cortical neurons parallels recognition behavior in perceptual tests of song discrimination in complex scenes, whereas the coding behavior of upstream auditory neurons does not.

ACKNOWLEDGMENTS

I thank Lana Rosis and Jordan Moore for providing the data shown in Figures 1 and 2, and David Schneider for conducting the studies on song recognition in scenes and neural coding of songs and scenes discussed here. This work is supported by the NIH (F31-DC010301 and R01-DC009810) and the NSF (IOS-0920081).

REFERENCES


