Evidence for high-level representational plasticity underlying the recognition of acoustically complex communication signals

T.Q. Gentner, D. Margoliash
Department of Organismal Biology and Anatomy
University of Chicago

Data suggest that the caudo-medial ventral hyperstriatum (cmHV) in songbirds plays a role in the recognition of conspecific vocal communication signals (birdsongs). We recorded extracellularly from neurons in the cmHV of European starlings after they were trained to recognize multiple conspecific songs. Most neurons were strongly selective for training songs over novel songs and synthetic stimuli. Responses were usually tuned to specific features of one training song. Linear models of each neuron's spectro-temporal receptive field poorly predicted the actual response. The results suggest dramatic reorganization of high-level feature representations, reflecting the behavioral relevance and higher-order statistics of specific songs.

The competitive reproductive environments in which birdsongs function require adult songbirds to discriminate and classify conspecific vocalizations based on intra-species acoustic variation. In fact, one of the most widespread vocal behaviors in songbirds is song recognition. Nearly every species of songbird investigated to date shows some form of discrimination and classification of different (usually conspecific) songs in an individual's immediate environment. The variety and complexity of these vocal communication signals, along with the wealth of knowledge regarding the behavioral functions of song, make the songbird auditory system an excellent model for the perceptual and cognitive processing of natural acoustic signals. Little is known about the neural basis of song recognition behavior. However, data from a variety of electrophysiological and stimulus driven gene expression studies suggest that several regions in the songbird auditory forebrain contribute to the perceptual processing of conspecific song, and thus are likely to be involved in the recognition of these vocal communication signals.

As with many vocal communication signals, birdsongs are acoustically complex. Their spectro-temporal structure varies over a large number of dimensions, often simultaneously. The representation of such signals is potentially problematic. If features or feature conjunctions are represented by single cells, then the number of percepts derived from complex stimuli can quickly overwhelm the capacity of the representational system. This problem can be ameliorated by appealing to distributed coding schemes, in which large ensembles of cells code for features or conjunctions. However, fully distributed coding schemes violate the topographical and stimulus specificity of neural responses observed throughout most sensory systems. The short comings of both coding schemes can be avoided if the representational system is biased dynamically toward non-arbitrary signals (see Genter and Margoliash , in press). Under such conditions, an approximately hierarchical organization is preserved for signals encountered under predictable environmental conditions, and more distributed representations are employed for unpredictable (i.e. arbitrary) signals. While this synthetic approach is biologically plausible in many respects, it requires substantial representational plasticity in the higher-order responses of the sensory system in order to accommodate changes in the set of relevant signals. In mammals there is evidence for representational plasticity in the frequency tuning properties in primary auditory cortex (AI), and in the topographical maps in other sensory cortices. Here we report that single unit responses in a region of the avian auditory forebrain are closely tuned to the acoustic features in familiar songs, but not to those in unfamiliar songs. We argue that this pattern suggests a dramatic reorganization of high-order feature selectivity that reflects the behavioral relevance of particular songs. This provides an important component of the neural basis for vocal recognition, and support for a synthetic coding scheme.

We trained European starlings (Sturnus vulgaris) to discriminate accurately among multiple songs of conspecific males using an operant procedure, then recorded from single units in cmHV while the animal was under urethane anesthesia. The operant training stimuli were generated from recordings of two adult male starlings. Ten continuous samples (~10 s each) were taken from within each song bout, then divided into two sets. Each set contained 2-5 exemplars from a single male. Subjects were naive to all training stimuli at the start of the experiment. We trained each subject (N=6) to discriminate among the training songs, using either a 2-alternative choice or go/nogo operant procedure. All subjects learned to discriminate accurately among the different exemplars regardless of training regime. Once a subject accurately recognized the training songs, we anesthetized the animal (20% Urethane) and began the recording session, employing well-described extracellular physiological techniques. We report here on 39 cells, confirmed via post-mortem histology, to be within cmHV.

Along with the training stimuli, we presented two to five novel conspecific songs. Each novel exemplar was ~10 s long, and was recorded under similar conditions to the training stimuli, but from different starlings. Subjects also heard two synthetic stimuli: bandpass white noise bursts, and FM sweeps. All test stimuli were used to search for auditory units. Once an auditory unit was isolated, the test stimuli were presented in random order in epics comprising five to ten reps of each stimulus. The delay between successive stimulus events ranged from 5 - 20 s.

A large majority (~75%) of the auditory units in cmHV showed a selective response to the training songs over the novel songs and the synthetic stimuli. In many cases, the selective response was tuned to a specific acoustic feature within a single training song, resulting in phasic PSTH response patterns composed of spike trains that were highly reliable across trials. In a smaller number of cases, selectivity was apparent by an significantly higher mean response rate for one or more of the training songs than for any of the novel stimuli. These phasic and more tonic response classes may represent separate physiological classes of cells. Approximately 14% of isolated units showed auditory responses that were non-selective for any of the stimuli presented, while only ~3% showed selective responses to the novel song stimuli.

To quantify these responses, we constructed the spectro-temporal receptive field (STRF) of each cell using a forward regression technique which normalized the spike triggered average by the auto-correlation matrix of the stimulus in the frequency domain (Theunissen et al., Network: Comput. Neural Syst, 12: 289). This generates an impulse response filter that we use in a jackknife algorithm to predict a cell's response to the test stimuli. Overall, the linear filter was a poor predictor of the cell's firing rate, suggesting that the selectivity apparent in each response profile derives largely from the higher-order statistics of specific spectro-temporal patterns in the training songs.

These data strongly suggest that the cmHV responds to changes in the set of relevant signals by dramatically altering its representational structure to efficiently code the higher-order acoustic features of these signals. That is, the representational structure of the system appears to emerge as the behavioral relevance and statistics of a natural signal are learned.