

Neural Systems for Individual Song Recognition in Adult Birds

T.Q. GENTNER

Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois, USA

ABSTRACT: The songbird auditory system is an excellent model for neuroethological studies of the mechanisms that govern the perception and cognition of natural stimuli (i.e., song), and the translation of corresponding representations into natural behaviors. One common songbird behavior is the learned recognition of individual conspecific songs. This chapter summarizes the research effort to identify the brain regions and mechanisms mediating individual song recognition in European starlings, a species of songbird. The results of laboratory behavioral studies are reviewed, which show that when adult starlings learn to recognize other individual's songs, they do so by memorizing large sets of song elements, called motifs. Recent data from single neurons in the caudal medial portion of the mesopallium are then reviewed, showing that song recognition learning leads to explicit representation of acoustic features that correspond closely to specific motifs, but only to motifs in the songs that birds have learned to recognize. This suggests that the strength and tuning of high-level auditory object representations, of the sort that presumably underlie many forms of vocal communication, are shaped by each animal's unique experience.

KEYWORDS: audition; perception; representational plasticity; animal communication; neural coding; object recognition; adult learning

Recent research on the neurobiology of birdsong has focused so consistently and successfully on song acquisition and production that the communication function of song is sometimes overlooked. Like other communication signals, the adaptive function of birdsong is to control or influence the behavior of others, usually conspecifics.¹ Implicit in this function is the notion that communication signals transmit information between the sender of the signal and the receivers. The success of this transmission rests on predictability. When a singer produces a specific pattern of acoustic energy, it does so under the expectation that it will be perceived in a predictable way—that is, that it will elicit the intended behavior in the receiver. Without the predictable correspondence between production and perception, signals would lose their functionality. Conversely, a functional signal implies a reliable correspondence between production and perception.

Address for correspondence: T.Q. Gentner, Department of Organismal Biology and Anatomy, University of Chicago, 1027 E. 57th Street, Chicago, IL 60637. Voice: 773-702-8090. t-gentner@uchicago.edu

Ann. N.Y. Acad. Sci. 1016: 282–302 (2004). © 2004 New York Academy of Sciences.
doi: 10.1196/annals.1298.008

In the case of birdsong, functions differ with the sex of the recipient. To other males, a bird's song may make a statement about the singer's species, sex, and territoriality. To females, male song may announce the singer's species, its readiness to mate, and/or convey important information about the singer's fitness. These functions, or rather the production/perception correspondences they require, constrain the acoustic structure of the signal and thereby limit the universe of functional signals that a sender may produce. For the receiver, these functions imply corresponding constraints in the form of (1) a sensory apparatus selectively tuned to signal properties, (2) perceptual mechanisms for the classification of the signals, and (3) cognitive (i.e., sensorimotor) mechanisms for the translation of perception into action.² These constraints provide starting points for neuroethological studies of the mechanisms governing the processing of natural communication signals.

Studies of the song recognition system require clarity both with respect to evolutionary constraints on the system and to the perceptual/cognitive operations that the system governs directly. One such operation, common to many species of songbirds, is the recognition of individuals based on their vocalizations.^{3,4} Indeed, song recognition, i.e., the ability to associate specific songs with specific singers or other referents, is a prerequisite for many of the decisions involved in such elaborate social behaviors as female choice, female preferences, kin recognition, and territoriality (see ref. 5). Our ultimate goal is the identification of the brain regions and mechanisms mediating song recognition; namely, those structures where learned representations of songs are localized and whose differential activation is correlated with a behaviorally defined recognition process. This chapter summarizes my research strategy, describes some of the recent findings, and suggests directions for future research.

DECONSTRUCTING SONG RECOGNITION BEHAVIOR: THE "MEANING" OF SOUNDS

Learned (in contrast to innate) song recognition may be conceptualized as a sequence of subprocesses, beginning with the discrimination of a to-be-recognized "target" sound from noise and other potentially relevant sounds. The target sound must then be classified or associated with some specific referent (e.g., a mate or an intruder), and the association retained until the sound's next occurrence. Recognition of this form can thus be seen as a classification problem in associative learning and memory. This is true whether one considers the elementary "recognition" of a sound as novel vs. familiar, or the more complex associations that form between a specific sound and a specific individual, motivational state, location in space, etc. Importantly, these associations can vary within and between species depending on the behavioral context in which song recognition occurs, as, for example, when songs are used to recognize different individuals, kin, dialects, or species. The functional plurality of song probably reflects the coding of information at multiple levels within song and, as noted above, is also likely to evoke selection pressures on different perceptual and cognitive mechanisms. For my present purposes, I focus on mechanisms of "individual vocal recognition" of same-species members—so-called conspecifics—in which songs are associated with individuals or groups of individual conspecifics.

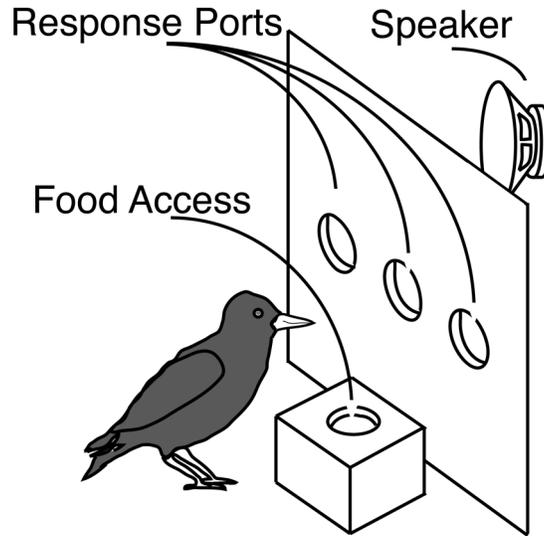
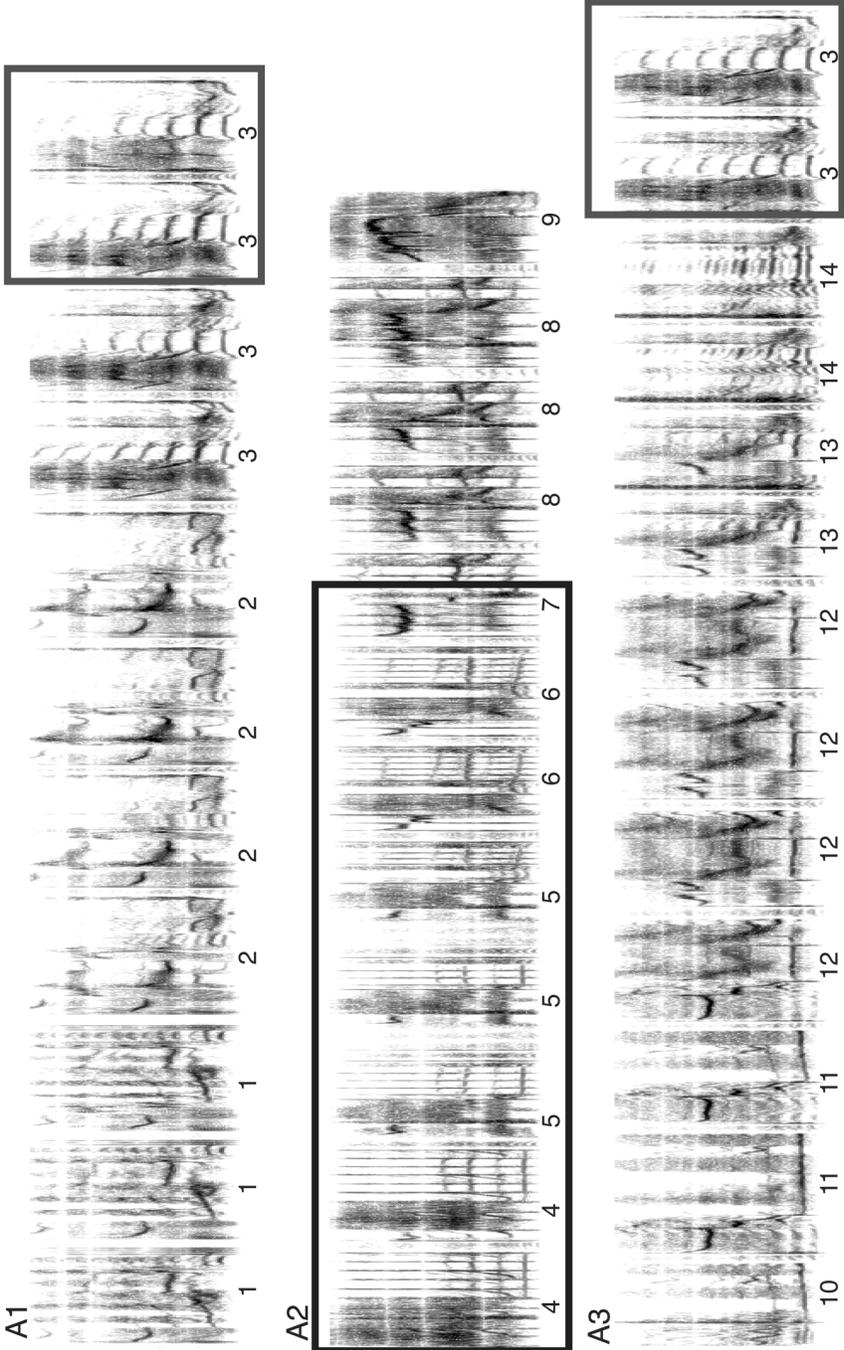


FIGURE 1. Schematic of operant training apparatus. Birds were given access to a metal panel housing three response buttons, and a feeder. Pressing the center button triggered playback of a training song. For the two-alternative choice procedure (see text), subjects had to peck the left button after hearing one-half of the training songs, and the right button after hearing the other half to receive a food reward. For the go/nogo procedure (see text), subjects had to peck the center button after hearing one-half of the training songs (“go” response, S+stimulus), and had to withhold responses altogether after hearing the other half (“nogo” response, S–stimulus). Only “go” responses to the S+stimuli were reinforced with food.

To study individual vocal recognition, one first needs to understand the behavioral schemes that birds use to classify and organize the songs of conspecific singers. Because the ultimate goal is to study the neural mechanism, I sought an operational definition of song recognition that permitted direct laboratory study, such that the “meaning” of a sound could be experimentally manipulated and the behavioral consequences of its presentation assessed. To do this, we used a set of operant conditioning techniques that required subjects to make one response to the songs of a specific bird and a different response to the songs of one or more other birds.^{6,7} Typically, the birds are first trained to obtain food by pecking buttons on a panel mounted on the side of their cage (FIG. 1). They are then reinforced with food for pecking one button, say the left, every time they hear a song from male “A”, and for pecking another button, in this case the right, every time they hear a song from male “B”. Tasks such as this, in which two sets of stimuli (songs) are associated with similar operantly reinforced behaviors (peck right/ peck left) are called two-alternative choice tasks (2AC). In a close variant, the so-called go/nogo procedure, behavioral responses to only one set of stimuli are reinforced, leading the subject to cease responding to the nonreinforced stimuli. With both training procedures, subjects become proficient at recognizing stimuli in each class, and there are innumerable ways that both the task and the stimuli can be varied to ask specific questions about the



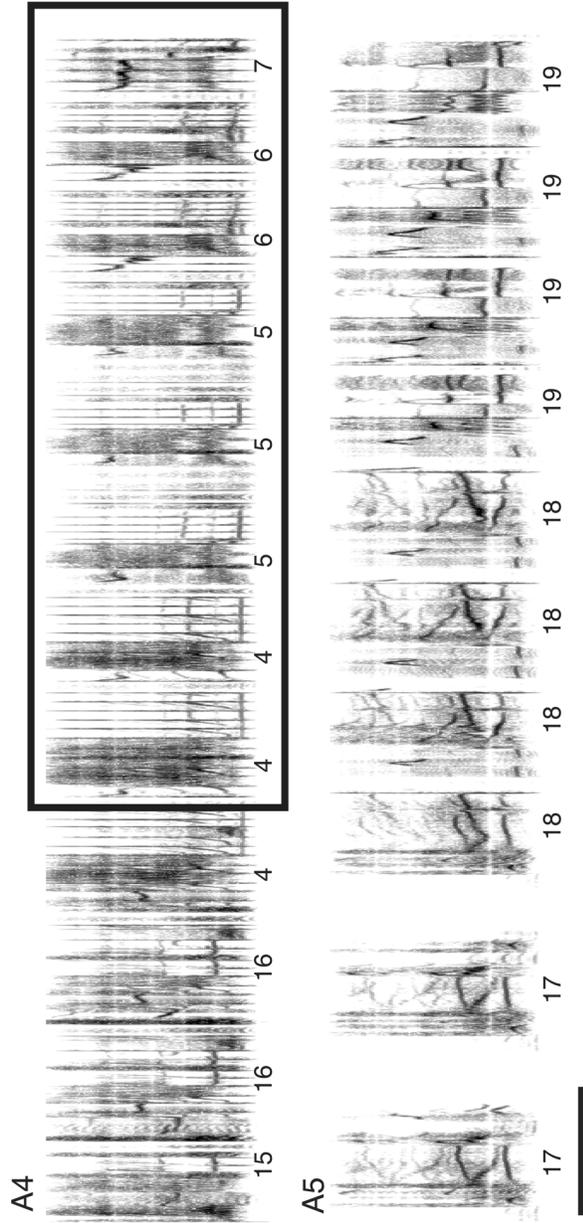


FIGURE 2. Starling song stimuli. Sonograms of typical stimulus exemplars used for operant training of song recognition. Each row shows a different (~10 s) sample of song. All samples shown were drawn from the songs of a single male starling. Unique motifs are labeled with different numbers along the bottom of each sonogram. The boxes outline similar motif sequences in different exemplars (i.e., songs). Note the high motif variability across song exemplars, the repetitive motif sequencing with songs, and the stereotyped note structure within similar motifs. Scale bars show 1 s. Frequency range 0–10 kHz.

acoustic features, as well as associative processes such as generalization and categorization, that guide song recognition.

Subjects in these experiments were European **starlings** (*Sturnus vulgaris*). Male starlings tend to sing in long continuous episodes (i.e., *bouts*). Song bouts, in turn, are composed of much smaller acoustic units referred to as *motifs*^{8,9} (FIG. 2), and these, in turn, are composed of still smaller units called *notes*. Notes can be broadly classified by the presence of continuous energy in their spectrotemporal representations. Although a motif may consist of several notes, the note pattern within a motif is usually relatively stereotyped between successive renditions of that motif. Commonly, each motif is repeated two or more times before the next one is sung. Thus, starling song appears (acoustically) as a sequence of changing motifs, where each motif is an acoustically complex event (FIG. 2). Different motifs can vary in duration from roughly 200 to 1,000 ms, and the number of unique motifs that a male starling can sing (i.e., his repertoire size) can be very large. Consequently, different song bouts from the same male are not necessarily composed of the same set of motifs. Over time, however, the songs of a specific male can be characterized by a set of motifs typical of that male. Although some sharing of motifs does occur among captive males,^{10,11} the motif repertoires of different males living in the wild are generally unique.^{7-9,12,13} Thus, learning which males sing which motifs can provide a diagnostic cue for song recognition.

The results of our behavioral studies support the idea that individual song recognition operates at the level of the motif. Starlings trained to recognize sets of songs from different conspecific individuals can correctly recognize sets of novel song bouts from the same singers^{7,14} (FIG. 3). One class of acoustic cues that could permit the recognition of novel song bouts are those resulting from idiosyncratic source and/or filter properties of each individual's vocal apparatus (i.e., "voice" characteristics). The use of voice characteristics (e.g., vocal timbre, the frequency of glottal pulsation, and spectral contours imparted by laryngeal morphology) is well documented for individual talker recognition in humans (e.g., ref. 15). The hypothesis that recognition is based at the level of the motif was tested by asking the birds to recognize novel song bouts that have *no* motifs in common with the training songs (FIG. 3). The hypothetical role of voice characteristics was tested by training birds to recognize isolated motifs shared by two different males (e.g., bird A singing motif 1 and bird B singing motif 2), and then watching recognition after *switching the motif but not the singer* (i.e., bird A singing motif 2 and bird B singing motif 1). In both cases, recognition falls to chance levels, suggesting the importance of motifs and eliminating a critical role for voice characteristics.

The most reasonable alternative hypothesis is that starlings learn to recognize the songs of individual conspecifics by memorizing sets of motifs that are associated with individual singers. If this is true, then once recognition is learned, it should be possible to control it systematically by varying the proportions of motifs in a "target" bout that come from two vocally familiar males. If subjects memorize a large set of motifs from each singer, recognition behavior should be correlated with the relative proportions of familiar motifs from different males independent of the specific motifs comprising a given song. If they attend to the presence (or absence) of a single motif or a small set of motifs, recognition should not follow relative motif proportions, and should not generalize between songs in which different motifs from the same singer make up similar proportions. To test these ideas, I again trained starlings



to recognize sets of songs from different individual males, and then watched as subjects classified novel song bouts in which motifs from the training songs were combined in several different ways.¹⁶ Consistent with the motif memorization hypothesis, we observed an approximately linear relationship between song classification and the relative proportions of familiar motifs from different singers composing each bout (FIG. 3). This suggests that when starlings are compelled to classify individual conspecific songs, they do so by memorizing large numbers of unique

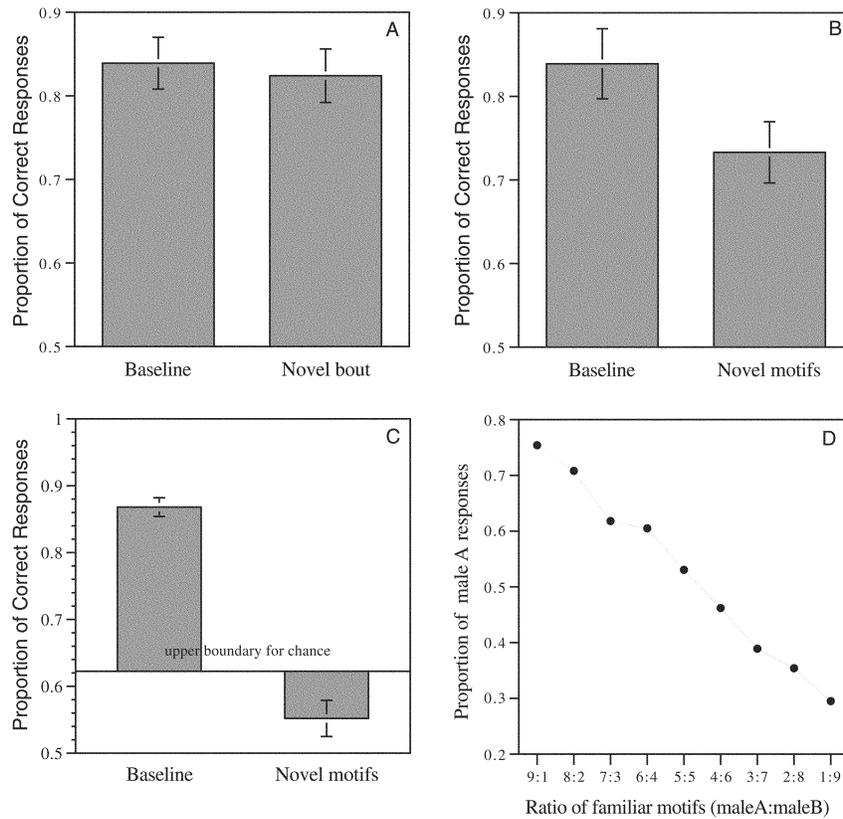


FIGURE 3. Individual song recognition behavior in European starlings. (A) Mean (\pm SEM) proportion of correct responses given during asymptotic performance on an operant recognition task (Baseline) and during initial transfer to novel songs containing familiar motifs (Novel bout). (B and C) Mean (\pm SEM) proportion of correct responses during transfer from the baseline training to novel songs from the same singers composed of novel motifs. Data in B show the transfer when the subjects were exposed to the training and test songs outside of the operant apparatus. Data in C show the results for the same transfer after controlling for this experience. Note that in C recognition of the novel motifs falls below chance. (D) Data showing the responses to chimeric songs composed of familiar motifs from two different singers. Note the close and approximately linear relationship between the ratio of familiar motifs from each singer and the subjects' recognition response.

song components (i.e., motifs) and then organizing subsets of these motifs into separate classes. From a human perspective, this might seem a suboptimal strategy to solve this vocal recognition problem. However, classifying songs according to their component structure represents a functionally parsimonious method of organizing these complex, but phonologically disjoint, sets of acoustic signals. Since individual starlings tend to possess unique motif repertoires, it is likely that, under natural conditions, these disjoint sets and the associated perceptual classes will correspond to individual identity.

NEURAL SUBSTRATES OF SONG RECOGNITION

From an operational standpoint, we consider individual recognition as a function that matches two representations—that of the incoming “to-be-recognized” (“target”) sensory signal, and that of the memory (or memories) of similar, recognizable signals. The behavioral data suggest strongly that in starlings these representations should correspond to the functional components of songs, i.e., motifs. Moreover, because song recognition is an associative process, mnemonic representations may reflect the acoustics of the signal *and* the reward contingencies associated with specific signals. As a first step in understanding the neural substrates of song recognition, we attempted to localize these representations in the starling brain, and describe their form with respect to behaviorally relevant variation in conspecific songs.

Several lines of earlier work suggested auditory regions in the forebrain as likely locations in which one might observe neural correlates to individual vocal recognition in starlings. The large-scale architecture and general pattern of connectivity within the starling auditory forebrain appears to be similar to that observed for other songbird species¹⁷ (and see Theunissen *et al.*, Song Selectivity..., this volume). The field L complex is the primary telencephalic target for auditory information arriving via parallel pathways from the thalamus; interconnected subregions of field L project to the caudal medial nidopallium (NCM) and reciprocally to the lateral portion of the caudal mesopallium (clM). The NCM and clM share reciprocal connections with the caudal medial portion of the mesopallium (cmM), and projections from clM to Nif and the HVC shelf provide the likely source of auditory input to the classic “song control” system (Theunissen *et al.*, Song Selectivity..., this volume). Thus, within the sensory hierarchy, NCM and the caudal mesopallial structures (clM, cmM) sit in positions analogous to secondary auditory cortices in mammals and, by extension, then are likely to be involved in the processing of behaviorally relevant complex stimuli such as conspecific song. This reasoning is supported by a number of studies.

In European starlings, neurons throughout the auditory telencephalon show complex patterns of tonotopic organization^{18,19} that can be differentiated on the basis of the direction of the tonotopic gradient and tuning curve bandwidth,^{20,21} and these patterns appear to respect the anatomical boundaries of the field L complex. Neurons in L1 and L3 subregions of field L have lower response rates to tone bursts than those in L2 and show greater selectivity to species-specific vocalizations.^{22–25} This selectivity is borne out by the complexity of the spectrotemporal receptive fields (STRFs) for many neurons within field L. Indeed, more reliable estimates of the STRF are derived from responses to conspecific vocalizations than from tone pips²⁶ (cf. ref. 27).

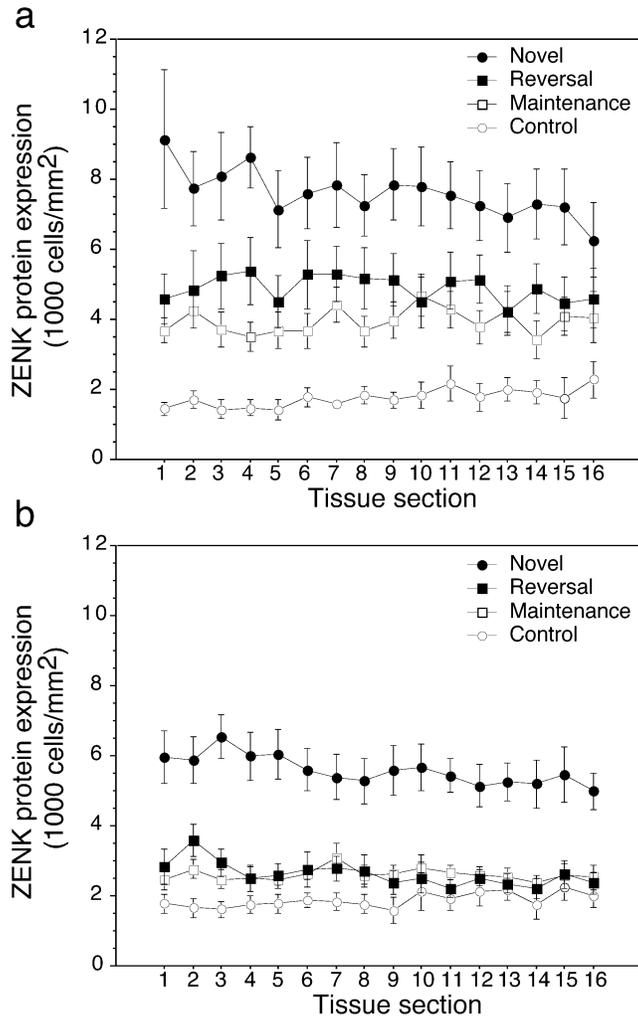


FIGURE 4. IEG *zenk* expression during vocal recognition. Zenk protein expression density in (a) NCM and (b) cmM following exposure to conspecific song under three different (see text) operant training regimes. In NCM, the response to novel songs is significantly elevated above that for all other groups. In cmM, expression was elevated in all three groups of operantly trained, and was higher still in the birds that heard novel songs. The 16 most medial tissue sections (40 μ m) are shown (1, most medial; 16, most lateral) with the data collapsed across hemispheres.

The general pattern of increasing response selectivity along the sensory hierarchy areas continues into NCM and cM,²⁴ suggesting that these regions are involved in the extraction of complex features. Early data from white-crowned sparrows are consistent with this in identifying a small subset of NCM neurons that are selective for the direction of frequency modulation in particular elements of conspecific song.²⁸ More extensive studies using complex stimuli demonstrate that both cM and NCM are broadly tuned, i.e., selectively responsive, to conspecific songs,²⁹ with a hierarchy of increasingly nonlinear components driving neural responses from field L to cM respectively.³⁰ A similar sensitivity to complex acoustic features characteristic of conspecific songs is apparent in NCM. Repeated presentation of the same conspecific song gives rise to song-specific habituation of the NCM response,^{31,32} suggesting that NCM plays an important role in vocal recognition.³³

The second line of research supporting the role of NCM and cM in the processing of conspecific song comes from studies of stimulus-driven expression of the immediate-early-gene (IEG) *zenk*, a putative marker for song-induced experience-dependent plasticity.³⁴ Both NCM and cM show a rapid upregulation of the IEG *zenk* in response to the presentation of conspecific songs compared to a variety of other acoustic stimuli³⁵ (and see Mello, this volume). The *zenk* response is tuned to the acoustics of particular conspecific song syllables,³⁶ and habituates to the repeated presentation of the same conspecific song³⁷ on a time scale that mirrors the song-specific modulation of NCM neuron firing rates. In addition, pairing the presentation of song with an aversive stimulus leads to increased expression relative to controls in which song is unpaired, suggesting that associative mechanisms may mediate at least a portion of the NCM IEG response.³⁸

To explore this idea in the context of individual vocal recognition, I trained starlings to recognize sets of conspecific songs using the operant two-alternative choice procedure. Once they had learned to recognize the baseline set of songs I transferred one-third of them to novel songs from different singers, kept one-third on the training songs, and reversed the response contingencies of the training songs for the final third. Following these transfer sessions *Zenk* expression was visualized in NCM and cM. The IEG response in NCM was consistent with the habituation and novelty responses described above. The density of *Zenk* protein expression in NCM was elevated *only* in the birds required to classify the novel songs; expression in the other two groups was no different than in control birds that heard no song but worked the apparatus for food. The pattern of expression in cmM, however, was qualitatively different, with all three groups showing significantly elevated levels of *Zenk* protein expression relative to the silent controls and birds in the novel song condition showing still higher levels relative to the other two groups of birds (FIG. 4). Thus, in addition to the effect of song novelty, *Zenk* protein expression in cmM also appeared to correlate with the ongoing recognition of familiar songs.

LEARNED REPRESENTATIONS OF SONGS IN CM

To examine the role of cmM in the representation of learned conspecific song in adult birds, I again trained starlings to recognize two sets of conspecific songs, using both two-alternative choice (2AC) and go/nogo (GNG) operant procedures (see FIG. 1). Both procedures teach the subject to recognize all the training songs, but for

the 2AC procedure the songs in both sets are associated with similar positively reinforced behaviors (pecks to two different buttons), and for the GNG procedure only one set of songs is associated with a positively reinforced behavior. Regardless, in both training regimens, subjects learned very accurate recognition of the training songs. Once asymptotic behavior was attained, I anesthetized each subject with urethane and recorded the extracellular responses of single neurons in the cmM to an ensemble of acoustic stimuli. The stimulus ensemble comprised the same song stimuli used during operant recognition training (termed “familiar” songs), a balanced number of novel conspecific song stimuli (termed “unfamiliar” songs), and two synthetic stimuli. I used three sets of songs recorded from three different males. Each of the different song stimulus exemplars was a 10-s sample of continuous singing, taken from a single bout of a given male’s song. Two sets of songs served as training stimuli for each subject, and the third set was used as the novel song stimuli. Thus, the stimulus ensemble was similar for each subject, except that the familiarity or novelty of any given song varied systematically. Neurons were tested with 33 to 77 unique song motifs (73 to 178 total motifs), depending on the exact set of familiar songs used to train each animal and the set of unfamiliar songs used during testing.³⁹

As a *population*, cmM neurons responded selectively to the class of familiar songs. The mean response strength, which reflects the cell’s mean spike rate and variance,³⁹ was strongly and significantly biased toward familiar (i.e., training) songs compared with unfamiliar songs. The strong response bias for familiar songs was consistent in animals trained under both the two-alternative choice and go/nogo operant regimes (FIG. 5). Thus, in this paradigm song recognition learning shapes the responses of cmM neurons.

The observed plasticity in cmM could result from either “bottom-up” or “top-down” processes, or both. By bottom-up, I mean that the response is driven in a predictable way by patterns of acoustic variation in the stimulus. In the extreme case, all the information represented by the cell’s response is present in the acoustic variation of the signal. Thus, plasticity may result simply from exposure to a given set of songs. In contrast, top-down processes refer to the attentional, motivational, and/or reward mechanisms that might also shape a cell’s response, but which are largely independent of any particular signal acoustics. Our associative model of individual song recognition requires both top-down and bottom processes.

Consistent with the associative model, and thus the important role of associative learning in individual song recognition, a significant portion of the response tuning in cmM appears to be under the control of different reinforcement signals. Whereas the subjects trained using the two-alternative choice procedure showed no reliable difference between response strengths associated with the two sets of training songs, those trained with the go/nogo procedure did. That is, songs associated with positive reinforcement (S+stimuli) elicited significantly stronger responses than those associated with no reinforcement (S–stimuli; FIG. 5). Importantly, this difference was not the result of overall failure to respond to the S–stimulus, because the S–stimuli elicited stronger responses than unfamiliar songs. Thus, although animals in both operant regimes learned to discriminate equally well between the sets of training songs, the stimulus reinforcement contingencies specific to the training regime (see FIG. 1) had differential effects on the distributions of neuronal responses. These task-specific effects may reflect the use of different behavioral strategies to solve each task, or perhaps the differential cost of incorrect responses under each regime. In the go/

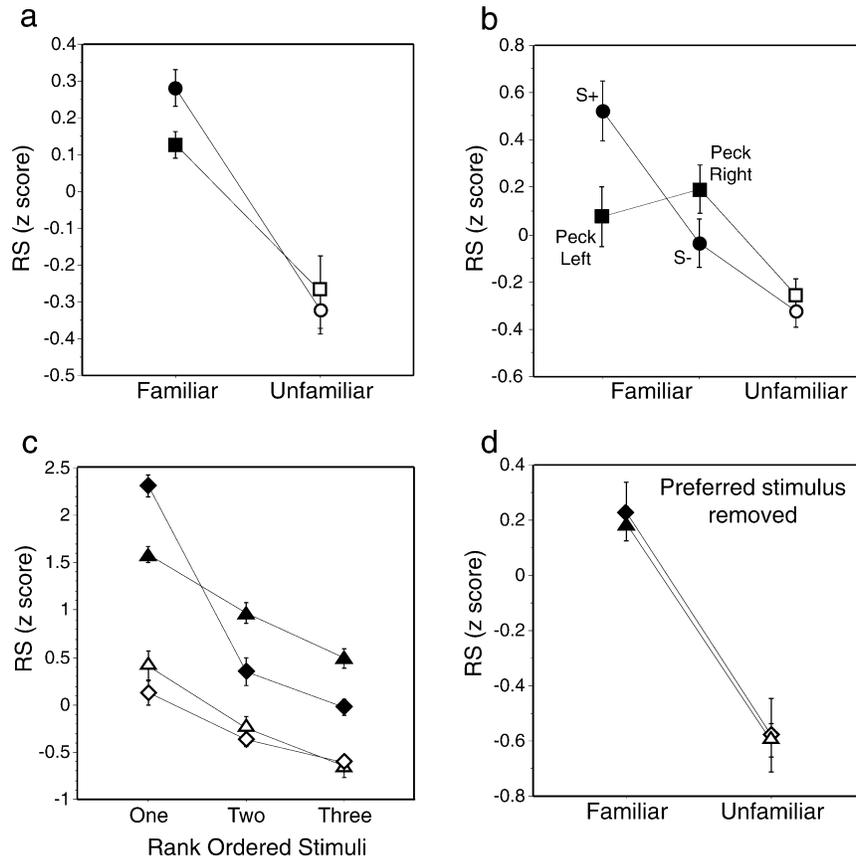


FIGURE 5. cmM neural responses. (a) Normalized response strength (z-scores) for familiar (*filled symbols*) and unfamiliar (*open symbols*) songs, split by training regime (two-alternative choice: ■, □; go/nogo: ●, ○). (b) Response strength as in a, but with the two sets of training stimuli and their accompanying responses labeled separately. (c) Rank-ordered response strength (z-scores) for the three most potent familiar (*filled symbols*) and unfamiliar (*open symbols*) songs for song-selective (◆, ◇) and nonselective neurons (▲, △). The interaction between stimulus rank-order and response selectivity among the familiar songs is significant, showing the strong bias in the song-selective cells for a single stimulus. The difference between the curves for the *filled* and *open symbols* shows the population-level bias for familiar songs. (d) Response strength z-scores for song-selective (◆, ◇) and nonselective neurons (▲, △) neurons, with the response to the preferred stimulus (the stimulus that elicited the strongest response) removed from the analysis. The differences are significant. All values are reported as means \pm SEM.

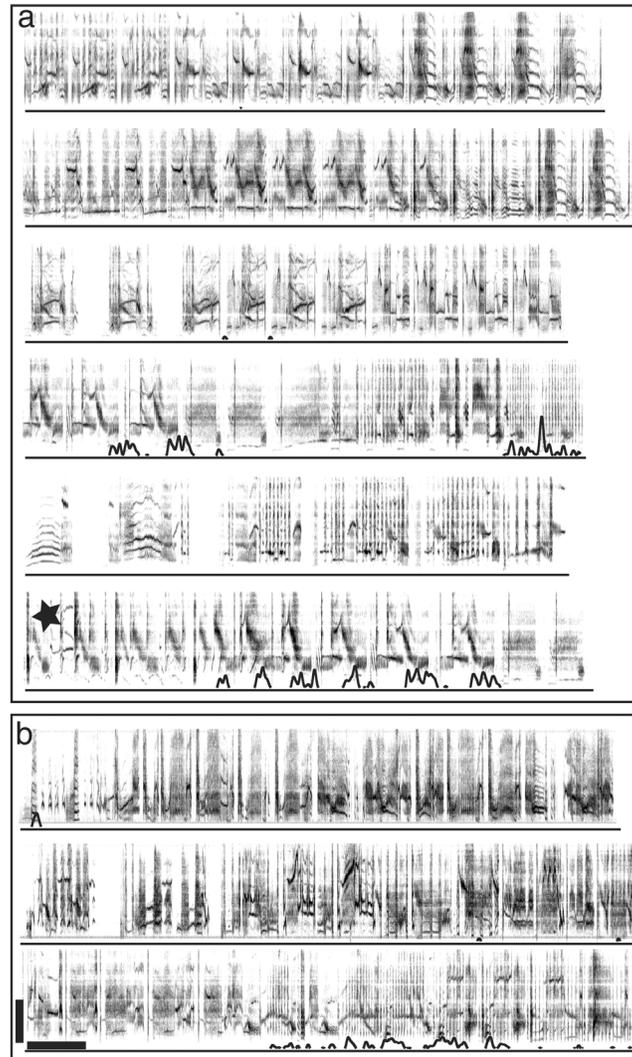


FIGURE 6. Typical single neuron responses in the cmM. **(a,b)** Response of a selective cmM unit to nine different song stimuli. Familiar songs are outlined in panel **a**, unfamiliar in panel **b**. The peristimulus time histogram (PSTH) of the response is superimposed over the sonogram of each stimulus. **(c)** Detailed view of the response of the unit in **(a)** and **(b)** to the song stimulus denoted by the black star. Traces from top to bottom show the raw spike waveform for a single stimulus presentation showing excellent single unit isolation, the

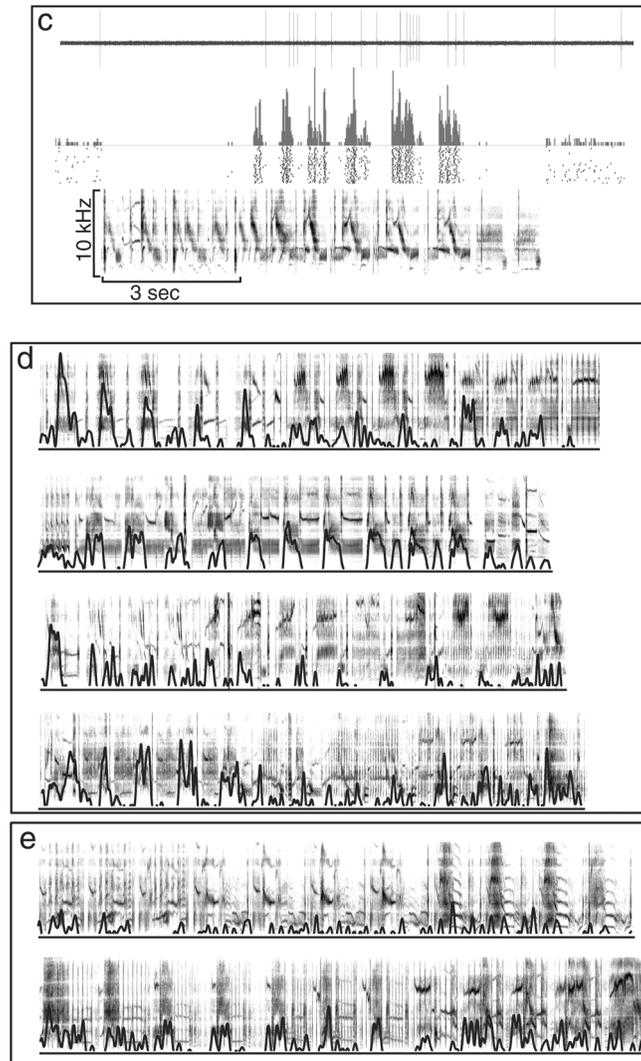


FIGURE 6. *Continued.*

PSTH, spike raster plots for several stimulus repetitions showing the reliability of the response, and the stimulus sonogram. (d,e) Example of another, very marginally selective, cmM neuron responding to four familiar songs (d) and two unfamiliar songs (e). Horizontal and vertical scale bars show 5 kHz and 1 s, respectively.

nogo condition, the failure to recognize an S+stimulus meant a missed opportunity for food, whereas the failure to recognize an S−stimulus cost only a slight increase in the already present delay to reinforcement following the next S+stimulus. Regardless of the specific behavioral strategies underlying differences between the two-alternative choice and go/nogo procedures, it is clear that bottom-up (i.e., stimulus-driven) mechanisms alone cannot account for the shaping of song representations in cmM. If the absolute or relative levels of stimulus activation were the only forces shaping cmM responses, then the response strength associated with the S+ and S−stimulus would not have been significantly different. Additional top-down mechanisms such as those governing attention, motivation or associative processing, etc., must also play a role.

Response Selectivity

The strong population response bias for familiar songs obviously indicates that cmM neurons did not respond equally to all songs in the test ensemble. To examine this bias in more detail, I developed a selectivity index (SI) score, in which a neuron's strongest response to a test stimulus is compared to that cell's response to all test stimuli.³⁹ An SI score close to one indicates that the cell responded to only one stimulus, while smaller SI scores indicate responses more evenly distributed across all stimuli. If the SI score for a cell was significantly different from a random distribution of SI scores, the cell was termed "song-selective" for the song that elicited the maximum response. The remaining cells were termed "nonselective." By this definition, roughly 64% of the cells in the cmM gave a selective response to one of the test stimuli; of these cells, a significant proportion (93%) preferred one of the training songs (FIG. 5). For nonselective cells, the song that elicited the strongest response was not significantly more likely to be either familiar or unfamiliar.

To our surprise, the population-level difference between the mean responses to the familiar and unfamiliar songs was not wholly dependent on the response to the preferred song. Even when the preferred song was removed from the comparison, the mean response strength associated with the familiar songs remained significantly greater than that for the unfamiliar songs (FIG. 5). This was observed for nonselective as well as selective cells (FIG. 5), demonstrating that as a population cmM neurons evince selectivity for the class (or a non-singleton subset) of familiar songs. The strong bias in the selectivity of single neurons for familiar songs resembles biases seen in the song control system for a bird's own song.^{40,41} Here, however, I am describing neuronal selectivity for conspecific songs that adults have learned to recognize, in neurons outside of the song control system.

Motif Selectivity

Both song-selective and nonselective cells could respond with phasic or tonic patterns of activity, and with high or low spontaneous rates, but on average the song-selective cells had more phasic responses and lower spontaneous firing rates than nonselective cells. Thus, most cells with near-zero spontaneous firing rates and highly phasic responses were selective (FIGS. 6 and 7). For many of the song-selective cells, responses were restricted to one or a small number of repeated motifs within one or a few songs, typically with suppression of background activity for all other motifs. Be-

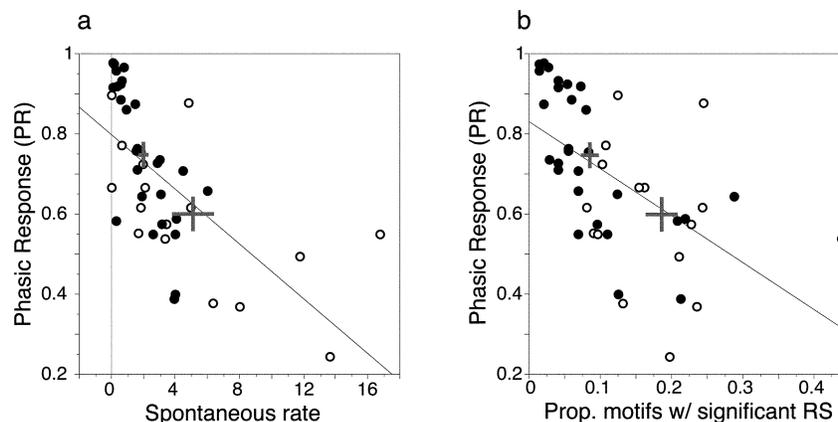


FIGURE 7. Response scattergrams. **(a)** Scatter plot showing the distribution of phasic responses (PR³⁹) across the population of selective (*filled symbols*) and nonselective (*open symbols*) cmM neurons, as a function of spontaneous firing rate (spikes/s). The means (\pm SEM) for each class are shown as the gray crosses. The line shows the significant linear regression. **(b)** As in **a**, but with PR plotted as a function of the proportion of motifs that elicited significant responses. The line shows the significant linear regression.

cause these responses appeared to be driven by acoustic variation at the level of the motif, we simply counted the number of motifs that elicited a significant response (i.e., an increase in firing rate relative to spontaneous activity) from each cell. Consistent with their tendency toward phasic response patterns, the song-selective cells responded to on average about 8 motifs, whereas nonselective cells were driven by significantly more motifs, roughly 20, on average.

To quantify the relationship between motif features and each neuron's spike response, we used a multiple linear regression between the response strength associated with each motif in a given cell's test ensemble and the coefficients from a two-dimensional wavelet decomposition of the sonogram for each motif. (Wavelet decomposition is a very efficient data compression technique that allows us to characterize the high-dimension acoustic space of each motif by a relatively small list of weighted features, i.e., coefficients. To the extent that variation among the wavelet coefficients is correlated with variation in the spike rate of a single neuron, one can say that the neuron codes for those features represented by the significant regression coefficients. Thus, we are concerned primarily with two measures: the mean R^2 values across cells, which provide us with an estimate of the overall quality of the regressions, and the proportions of cells within different classes that show a significant regression with the wavelets). The mean R^2 value for regressions with the song-selective cells was significantly greater than that for the nonselective cells, indicating that acoustic features in individual motifs predicted the responses of the song-selective cells better than those of nonselective cells. In addition, the regressions were significant for roughly 60% of the song-selective cells, but only about 18% of the nonselective cells. This suggests strongly that the "song" selectiv-

ity we have described in this population of cmM neurons is derived from selective tuning for spectrotemporal features centered at the level of the motif.

Characterizing these critical spectrotemporal features in any detail is a complex task. Because of its emphasis on data compression, the wavelet decomposition technique tends to yield only arbitrary characterizations of the features driving each response. It was very difficult to reconstruct a spectrotemporal image of the relevant features that drove a cell using only the significant wavelets. What can be said is that the cells do not appear to be tuned in a linear way to relatively “simple” acoustic features, such as a sound at a given frequency or repetition rate. Consistent with the observation of others,³⁰ the cmM neurons we describe here are poorly modeled by spectrotemporal receptive fields (STRFs). Since STRF techniques provide an estimate of the optimal *linear* combination of spectrotemporal features that maximally drive a given cell (see Theunissen *et al.*, Methods..., this volume), the observation that STRFs provide poor predictions of a cell’s response to a stimulus indicate a predominantly nonlinear response. Thus, neurons in cmM are driven by relatively complex acoustic features corresponding to specific motifs. We cannot yet say what those features are precisely. Nonetheless, I have demonstrated the selective neural representations for acoustic features diagnostic of individual or small sets of motifs. Such representation could obviously contribute to assessing the proportion of familiar motifs within a given song, which our earlier behavioral work has shown is important for learned recognition of individual conspecific songs in starlings.¹⁷

CONCLUSIONS

I have described a population of neurons selective for individually distinct conspecific songs that starlings that have been operantly trained to recognize—neurons that are, in other words, tuned to learned communication signals in adult animals. The neuronal population is found in a nonprimary forebrain auditory region, exhibits increased responses to the set of learned songs compared to novel songs, and exhibits differential responses to categories of learned songs based on recognition training contingencies. Within the population, many cells respond highly selectively to a subset of specific motifs present only in one individual’s songs. Since behavioral recognition of individual songs is also driven by motif acoustics, one may infer that the experience-dependent neuronal selectivity shown here contributes substantially to individual song recognition behavior in this species. Data from lesion studies and preliminary recordings in awake, behaving starlings (Gentner and Margoliash, unpublished data) are consistent with this inference.

The behavioral strategy for individual vocal recognition used by starlings suggests that the upper boundary for accurate recognition is set by the memory capacity of the system for specific motifs. To this end, both the acoustic complexity and the sheer numbers of to-be-remembered motifs are likely to pose significant challenges to the representational system’s capacity (cf. ref. 42). Although some coding efficiency is likely gained through adaptive sensory system specializations that lead to innate representational mappings, the idiosyncratic acoustic microenvironments encountered by any given individual are very hard to predict, and thus require representational plasticity mechanisms that operate on ontogenetic time scales. The observations of experience-dependent mechanisms that work to shape neural repre-

sentations of unique sets of behaviorally relevant auditory objects (motifs) demonstrate exactly this sort of coding efficiency, and suggest the hypothesis that long-term memory constraints derive directly from the perceptual mechanisms for coding complex stimuli.

While the explicit spectrotemporal parameters underlying the selectivity of cmM neurons have not been described, the acoustic features of recognized motifs (as represented by the wavelet coefficients) were correlated with the response strength of many cmM neurons. Yet, a purely acoustic account of cmM responses, involving putative “feature detectors” is unlikely to be convincing. The distribution of cmM response properties depends not only on the spectrotemporal patterns of motifs in familiar (and unfamiliar) stimuli, but also on the specifics of the conditioning paradigm. This sensitivity to conditioning context may reflect seemingly subtle differences in the reward contingencies in the two training regimes. Because song recognition mediates a variety of behaviors in both agonistic and antagonistic contexts, different forms of learning under more natural conditions may act on the recognition system to produce a variety of top-down effects on auditory response properties.^{43,44} The presence of these contextual influences on representational plasticity complicates the search for neural correlates in the absence of well-controlled behavior, and may account for the few reports of such neurons in sensory systems.

Even with the large stimulus repertoires that animals were trained to recognize, I estimate that less than 50% of the neurons responded to the test stimuli. The unresponsive neurons may not have been auditory, or may have been responding to other songs that these wild-caught birds had previously learned. Since we find almost no cells selectively tuned to the motifs in *unfamiliar* songs, the data argue that subpopulations of cmM neurons are selected from a pool composed primarily of neurons that have already been shaped by the animal’s prior experiences with conspecific songs, rather than from a large pool of unselective neurons. cmM neurons were more selective than those observed in the Field L of starlings, where cells commonly respond to numerous novel conspecific vocalizations.²² Response biases for species-specific vocalizations have been reported in Field L^{22,45} and cM of birds,²⁹ and in mammalian primary and secondary auditory cortex.^{46,47} Within this context, it is important to note that although cmM evinces a strong selectivity for familiar songs, the responses are not entirely exclusive. With the exception of the roughly 20% or so of cells that showed high phasic and selective responses, many of the cells were driven above spontaneous response levels by both familiar and unfamiliar songs. Thus, despite the prevalent bias for familiar songs, it does not appear that all information about unfamiliar songs is excluded from the system. Response to novel songs observed as modulations NCM neural firing rates,^{31,32} and IEG *zenk* responses,³⁷ may reflect the incorporation of novel information that under appropriate conditions is consolidated into longer-term representations in cmM. Obviously, for learning to take place, some novel information must be available.

Experience-dependent representational plasticity, reported in a variety of animals and sensory systems,⁴⁸ is typified by shifts in the topography of primary cortical receptive fields.^{49–52} In the auditory system, this plasticity leads to overrepresentation of the spectral and temporal properties of a learned stimulus.⁴³ Thus, in a hierarchical scheme of sensory processing, plasticity at primary levels should influence higher-order regions, such as cmM, so that the neuronal response properties and organization are expressly determined by an animal’s unique experience in behav-

iorally relevant tasks. These higher-order regions could contribute to recognition of individual conspecific vocalizations, and could influence motor tasks such as countersinging and vocal learning in juveniles.

Neural correlates to learned object recognition have also been reported in extrastriate visual cortex where cells are broadly tuned,^{53,54} and in prefrontal cortex⁵⁵ where selectivity for familiar objects is generally evidenced by a decrease in the numbers of neurons responding to a given object—an effect taken to indicate sharpening of the tuning for such objects.⁵⁶ In contrast, the proportion of cmM cells selective for familiar songs was much larger than that for unfamiliar songs, and many cells were sharply tuned. For vocal recognition, the predictability imparted by species-specific characteristics of vocalizations, and the constraints imposed by evolutionary history and experience, is likely to have produced a population of neurons predisposed to represent those vocalizations. I argue that the response properties of cmM neurons are elaborated continuously toward new functional representations depending upon the specific songs and behavioral contingencies an animal encounters. As in juvenile song learning,⁵⁷ the rules by which functional representations arise from biased populations may be complex, and are not likely to be well predicted by simple spectrotemporal similarities between the target songs and the initial representations of cmM neurons. The availability of “acquired neural representations” of auditory objects provides starlings, and perhaps other higher vertebrates, with an efficient mechanism for recognizing a wide and changing array of behaviorally important natural stimuli.

REFERENCES

1. KROODSMA, D.E. & E.H. MILLER. 1996. Ecology and Evolution of Acoustic Communication in Birds. Cornell University Press. Ithaca, NY.
2. MILNER, A.D. & M.A. GOODALE. 1995. The Visual Brain in Action. Oxford University Press. Oxford, UK.
3. FALLS, J.B. 1982. Individual recognition by sound in birds. *In* Acoustic Communication in Birds. D.E. Kroodsma & E.H. Miller, Eds.: 237–278. Academic Press. New York.
4. STODDARD, P.K. 1996. Vocal recognition of neighbors by territorial passerines. *In* Ecology and Evolution of Acoustic Communication in Birds. D.E. Kroodsma & E.H. Miller, Eds.: 356–374. Cornell University Press, Ithaca, NY.
5. GENTNER, T.Q. & D. MARGOLASH. 2002. Neuroethology of vocal communication: perception and cognition. *In* Acoustic Communication. A.M. Simmons, A.N. Popper & R.R. Fay, Eds.: 324–386. Springer-Verlag. New York.
6. HULSE, S.H. 1995. The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals. *In* Methods in Comparative Psychoacoustics, Vol. 10. G.M. Klump, R.J. Dooling, R.R. Fay & W.C. Stebbins, Eds.: 319–330. Birkhäuser Verlag. Basel.
7. GENTNER, T.Q. & S.H. HULSE. 1998. Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Anim. Behav.* **56**: 579–594.
8. ADRET-HAUSBERGER, M. & P.F. JENKINS. 1988. Complex organization of the warbling song in starlings. *Behav.* **107**: 138–156.
9. EENS, M., M. PINXTEN & R.F. VERHEYEN. 1991. Organization of song in the European starling: species-specificity and individual differences. *Belg. J. Zool.* **121**: 257–278.
10. HAUSBERGER, M. & H. COUSILLAS. 1995. Categorization in birdsong: from behavioural to neuronal responses. *Behav. Process.* **35**: 83–91.
11. HAUSBERGER, M. 1997. Social influences on song acquisition and sharing in the European starling (*Sturnus vulgaris*). *In* Social Influences on Vocal Development. C. Snowden & M. Hausberger, Eds.: 128–156. Cambridge University Press. Cambridge, UK.

12. EENS, M., M. PINXTEN & R.F. VERHEYEN. 1989. Temporal and sequential organization of song bouts in the European starling. *Ardea* **77**: 75–86.
13. CHAIKEN, M., J. BÖHNER & P. MARLER. 1993. Song acquisition in European starlings, *Sturnus vulgaris*: a comparison of the songs of live-tutored, tape-tutored, untutored, and wild-caught males. *Anim. Behav.* **46**: 1079–1090.
14. GENTNER, T.Q., S.H. HULSE, G.E. BENTLEY & G.F. BALL. 2000. Individual vocal recognition and the effect of partial lesions to HVC on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.* **42**: 117–133.
15. BRICKER, P.D. & S. PRUZANSKY. 1976. Speaker recognition. *In* Contemporary Issues in Experimental Phonetics. N.J. Lass, Ed.: 295–326. Academic Press. New York.
16. GENTNER, T.Q. & S.H. HULSE. 2000. Perceptual classification based on the component structure of song in European starlings. *J. Acoust. Soc. Am.* **107**: 3369–3381.
17. VATES, G.E., B.M. BROOME, C.V. MELLO & F. NOTTEBOHM. 1996. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J. Comp. Neurol.* **366**: 613–642.
18. LEPPELSACK, H.J. & J. SCHWARTZKOPFF. 1972. Properties of acoustic neurons in the caudal neostriatum of birds. *J. Comp. Physiol. A* **80**: 137–140.
19. RUBSAMEN, R. & G.J. DORRSCHIEDT. 1986. Tonotopic organization of the auditory forebrain in a songbird, the European starling. *J. Comp. Physiol. A* **158**: 639–646.
20. HAÜSLER, U. 1996. Measurement of short-time spatial activity patterns during auditory stimulation in the starling. *In* Acoustical Signal Processing in the Central Auditory System. J. Syka, Ed.: 85–91. Plenum Press. New York.
21. CAPSIUS, B. & H.J. LEPPELSACK. 1999. Response patterns and their relationship to frequency analysis in auditory forebrain centers of a songbird. *Hear. Res.* **136**: 91–99.
22. LEPPELSACK, H.J. & M. VOGT. 1976. Responses of auditory neurons in forebrain of a songbird to stimulation with species-specific sounds. *J. Comp. Physiol. A* **107**: 263–274.
23. BONKE, D., H. SCHEICH & G. LANGNER. 1979. Responsiveness of units in the auditory neostriatum of the Guinea fowl (*Numida meleagris*) to species-specific calls and synthetic stimuli. I. Tonotopy and functional zones. *J. Comp. Physiol. A* **132**: 243–255.
24. MÜLLER, C.M. & H.J. LEPPELSACK. 1985. Feature extraction and tonotopic organization in the avian forebrain. *Exp. Brain Res.* **59**: 587–599.
25. THEUNISSEN, F.E. & A.J. DOUPE. 1998. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVC of male zebra finches. *J. Neurosci.* **18**: 3786–3802.
26. THEUNISSEN, F.E., K. SEN & A.J. DOUPE. 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* **20**: 2315–2331.
27. SCHÄFER, M., R. RUBSAMEN, G.J. DORRSCHIEDT & M. KNIPSCHILD. 1992. Setting complex tasks to single units in the avian auditory forebrain. II. Do we really need natural stimuli to describe neuronal response characteristics? *Hear. Res.* **57**: 231–244.
28. LEPPELSACK, H.J. 1983. Analysis of song in the auditory pathway of songbirds. *In* Advances in Vertebrate Neuroethology. J.P. Evert, B.R. Capranica & D.J. Ingle, Eds.: 783–799. Plenum Press. New York.
29. GRACE, J.A., N. AMIN, N.C. SINGH & F.E. THEUNISSEN. 2003. Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophys.* **89**: 472–487.
30. SEN, K., F.E. THEUNISSEN & A.J. DOUPE. 2001. Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* **86**: 1445–1458.
31. CHEW, S. J., C.V. MELLO, F. NOTTEBOHM, *et al.* 1995. Decrements in auditory responses to a repeated conspecific song are long lasting and require two periods of protein synthesis in the songbird forebrain. *Proc. Natl. Acad. Sci. USA* **92**: 3406–3410.
32. STRIPLING, R., S.F. VOLMAN & D.F. CLAYTON. 1997. Response modulation in the zebra finch neostriatum: relationship to nuclear gene expression. *J. Neurosci.* **17**: 3883–3893.
33. CHEW, S.J., D.S. VICARIO & F. NOTTEBOHM. 1996. A large-capacity memory system that recognizes the calls and songs of conspecifics. *Proc. Natl. Acad. Sci. USA* **93**: 1950–1955.
34. JONES, M.W., M.L. ERRINGTON, P.J. FRENCH, *et al.* 2001. A requirement for the immediate early gene Zif-268 in the expression of late LTP and long-term memories. *Nat. Neurosci.* **4**: 289–296.

35. MELLO, C.V., D.S. VICARIO & D.F. CLAYTON. 1992. Song presentation induces gene expression in the songbird forebrain. *Proc. Natl. Acad. Sci. USA* **89**: 6818–6822.
36. RIBEIRO, S., G.A. CECCHI, M.O. MAGNASCO & C.V. MELLO. 1998. Toward a song code: evidence for a syllabic representation in the canary brain. *Neuron* **21**: 359–371.
37. MELLO, C.V., F. NOTTEBOHM & D. CLAYTON. 1995. 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.* **15**: 6919–6925.
38. JARVIS, E.D., C.V. MELLO & F. NOTTEBOHM. 1995. Associative learning and stimulus novelty influence the song-induced expression of an immediate early gene in the canary forebrain. *Learn.Memory* **2**: 62–80.
39. GENTNER, T.Q. & D. MARGOLIASH. 2003. Neuronal populations and single cells representing learned auditory objects. *Nature* **424**: 669–674.
40. MARGOLIASH, D. 1983. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* **3**: 1039–1057.
41. MARGOLIASH, D. 1986. Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.* **6**: 1643–1661.
42. STODDARD, P.K., M.D. BEECHER, P. LOESCHE & S.E. CAMPBELL. 1992. Memory does not constrain individual recognition in a bird with song repertoires. *Behaviour* **122**: 274–287.
43. KILGARD, M.P. & M.M. MERZENICH. 1998. Plasticity of temporal information processing in the primary auditory cortex. *Nat. Neurosci.* **1**: 727–731.
44. BAO, S., V.T. CHAN & M.M. MERZENICH. 2001. Cortical remodeling induced by activity of ventral tegmental dopamine neurons. *Nature* **412**: 79–83.
45. SCHEICH, H., G. LANGNER & D. BONKE. 1979. Responsiveness of units in the auditory neostriatum of the guinea fowl (*numida-meleagris*) to species-specific calls and synthetic stimuli. 2. discrimination of iambus-like calls. *J. Comp. Physiol. A* **132**: 257–276.
46. WANG, X. & S.C. KADIA. 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J. Neurophysiol.* **86**: 2616–2620.
47. RAUSCHECKER, J.P., B. TIAN & M. HAUSER. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* **268**: 111–114.
48. GILBERT, C.D., M. SIGMAN & R.E. CRIST. 2001. The neural basis of perceptual learning. *Neuron* **31**: 681–697.
49. BAKIN, J.S. & N.M. WEINBERGER. 1990. Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Res.* **536**: 271–286.
50. RECANZONE, G.H., C.E. SCHREINER & M.M. MERZENICH. 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* **13**: 87–103.
51. KILGARD, M.P. & M.M. MERZENICH. 2002. Order-sensitive plasticity in adult primary auditory cortex. *Proc. Natl. Acad. Sci. USA* **99**: 3205–3209.
52. KAY, L.M. & G. LAURENT. 1999. Odor- and context-dependent modulation of mitral cell activity in behaving rats. *Nat. Neurosci.* **11**: 1003–1009.
53. LOGOTHETIS, N.K., J. PAULS & T. POGGIO. 1995. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**: 552–563.
54. KOBATAKE, E., G. WANG & K. TANAKA. 1998. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* **80**: 324–330.
55. RAINER, G., W.F. ASAAD & E.K. MILLER. 1998. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* **393**: 577–579.
56. RAINER, G. & E.K. MILLER. 2000. Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron* **27**: 179–189.
57. TCHERNICHOVSKI, O., P.P. MITRA, T. LINTS & F. NOTTEBOHM. 2001. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291**: 2564–2569.