Acoustic and neural bases for innate recognition of song

C. S. Whaling*, M. M. Solis†, A. J. Doupe†, J. A. Soha*, and P. Marler*‡

*Animal Communication Laboratory, Section of Neurobiology, Physiology, and Behavior, University of California, Davis, CA 95616; and †Keck Center for Integrative Neuroscience, Departments of Physiology and Psychiatry, University of California, San Francisco, CA 94143

ABSTRACT In behavior reminiscent of the responsiveness of human infants to speech, young songbirds innately recognize and prefer to learn the songs of their own species. The acoustic and physiological bases for innate recognition were investigated in fledgling white-crowned sparrows lacking song experience. A behavioral test revealed that the complete conspecific song was not essential for innate recognition: songs composed of single white-crowned sparrow phrases and songs played in reverse elicited vocal responses as strongly as did normal song. In all cases, these responses surpassed those to other species' songs. Although auditory neurons in the song nucleus HVc and the underlying neostriatum of fledglings did not prefer conspecific song over foreign song, some neurons responded strongly to particular phrase types characteristic of white-crowned sparrows and, thus, could contribute to innate song recognition.

The brain is clearly not a "blank slate" prior to auditory experience and vocal learning. For instance, human infants display innate and categorical recognition of phonemes of human languages (1–4). Likewise, young sparrows prevented from hearing adult vocalizations readily discriminate between conspecific (i.e., own species) and foreign songs (5–9). Thus, newborn songbirds and humans can focus attention on species-specific vocalizations in preparation for learning to sing or to speak. To define the mechanisms enabling a young bird to identify appropriate song models, it is important to study the acoustic features of conspecific song that are critical for its recognition, as well as the neural responses to conspecific song. We therefore used vocal responses to normal and modified song stimuli, in parallel with electrophysiological recordings, to investigate the acoustic and neural bases of innate song recognition in very young, naive white-crowned sparrows (*Zonotrichia leucophrys nuttalli*, Nuttall's subspecies).

METHODS

Animals. White-crowned sparrow nestlings ($n = 70$; 4–7 days of age) were collected from Bodega Bay, CA, during spring of 1994 and 1995. Once in captivity, nestlings were hand-reared and prevented from hearing adult song. Songs heard while in the egg or during the first week of life have no influence on learning in songbirds (10); as a precaution, however, songs of our subjects' natal dialect (i.e., Bodega Bay) were not used in any behavioral or neurophysiological studies. Sex was determined by laparotomy under Metofane anesthesia (Pitman–Moore, Mundelein, IL). Males and females were used in both behavioral and neurophysiological studies.

Playback Experiments. Fledglings were housed individually in sound isolation chambers until behavioral tests were performed at 13–23 days of age (mean, 18 days). We recorded the vocal responses of individual fledglings to playback of song

stimuli according to the methods of Nelson and Marler (9). Each stimulus was presented 10 times, at 10-s intervals. Each bird's vocalizations were recorded during this test period and also during a pre-trial period of equal duration. The number of calls produced during the pre-trial period was subtracted from the number of calls produced during the playback test to give the vocal response to song. Vocal responses to different stimuli were compared using Wilcoxon paired signed ranks tests (two-tailed). Analyses were performed with Systat (SPSS, Evanston, IL).

In the first experiment (1994), birds ($n = 31$) were presented with seven song types: songs (*i*) of the same subspecies as the subjects, (*ii*) of another subspecies (*Zonotrichia leucophrys oriantha*), (*iii*) of both subspecies raised in acoustic isolation ("isolate" songs), and (*iv*) of three foreign species, the song sparrow (*Melospiza melodia*), the savannah sparrow (*Passerculus sandwichensis*), and the zebra finch (*Taeniopygia guttata*). Five different examples of each song type were used, one of which was heard by each bird. A bird was presented with the seven song stimuli in random order, one each hour, on a single day. In the second experiment (1995), birds ($n = 28$) were presented with six song types: (*i*) songs of their subspecies, (*ii*) reversed song of their subspecies, (*iii*) three "repeated phrase" songs constructed from repetitions of fragments or "phrases" of white-crowned sparrow song (whistles, buzzes or trills; see Fig. 1) while preserving normal song length and internote interval, and (*iv*) foreign song (song sparrow song). Nine different versions of each of the six song types were used, three of which were presented to each bird, one each day, over the course of 3 days. Each day, the six song stimuli were tested in random order, one each hour, over the course of 6 h. No song was tested twice in a single bird. To test for sex differences in recognition of conspecific song, we subtracted each bird's mean response to foreign song from its mean response to normal white-crowned sparrow song. The resulting differences were then compared using a Mann–Whitney test. Statistical analyses were performed using Systat.

Electrophysiology. Birds ($n = 11$) collected as for the behavioral studies were kept in acoustic isolation until extracellular recordings were made in the neostriatum at 23–42 days of age. Birds were anesthetized with urethane (7.5 ml/kg) and valium (1.5 ml/kg). Surgery and extracellular recording were performed as described by Doupe (11). Recording sites included HVc (acronym as specified in ref. 12) and the neostriatum immediately below and medial to HVc, to a maximum depth of 700 μm below the center of the ventral border of HVc (the auditory "shelf" of HVc; refs. 12 and 13). Recordings did not include the caudomedial neostriatum (13, 22). Test stimuli consisted of broad band noise bursts, tone bursts from 500 Hz to 8 kHz (usually presented in 500 Hz or 1 kHz increments), normal, reversed, and isolate white-crowned sparrow songs, and foreign songs. Their peak sound pressure level was 70 dB. Multiple exemplars of songs from each song type were presented to each cell (10–20 times each, interleaved). Neurons were included in detailed data analysis if the spike rate during

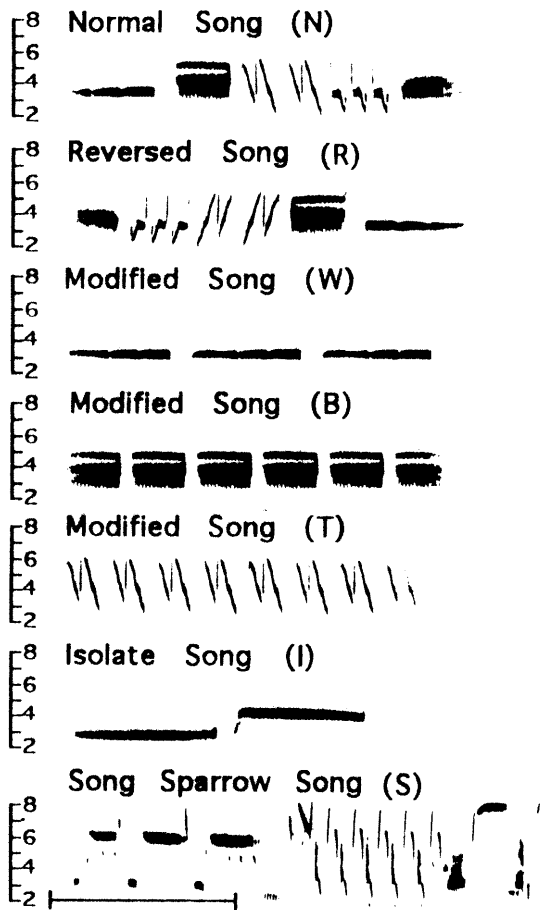‡To whom reprint requests should be addressed. e-mail: prmarler@ucdavis.edu.

FIG. 1.   Sound spectrograms of some of the test stimuli used in playback experiments. Normal (N), isolate (I), and savannah sparrow song (S) were presented in the first experiment. Reversed song (R), whistle songs (W), buzz songs (B), and trill songs (T), as well as N and S, were tested in the second experiment. Buzzes have a larger bandwidth and more amplitude modulation than whistles. Trills are repeated series of discrete frequency-modulated (FM) sweeps. Note that the song sparrow song contains short whistles and a trill. (Scale bar represents 1 second.)

at least one song stimulus was significantly greater than the spontaneous rate (paired *t* test, $P < 0.05$). Thirteen single units and 14 small clusters (maximum 3 neurons/cluster) met this criterion for a significant song response; 6 units were from HVc and the rest were from the shelf. Twenty-four additional single units or clusters were excluded from further analysis; many of these either had significant responses to tone bursts

but not to any song stimuli or had inhibitory responses to most songs.

The mean response strength (stimulus-evoked firing with spontaneous rate subtracted) to each song category (normal conspecific, isolate, reversed, foreign) was calculated for each neuron by averaging its response strengths to all stimuli belonging to a specific category. The mean response strength of a neuron to each phrase type was calculated from all songs that elicited a significant response. Phrase types were identified by visual inspection of spectrograms (Kay Elemetrics 7800, Lincoln Park, NJ). If phrases from foreign songs could not be classified as whistles, buzzes or trills, they were not used in the analysis (43% of foreign song phrases).

RESULTS

Playback Experiments. Isolate songs, produced by birds without tutoring experience, presumably reflect features of song that are innately specified (14, 15). In white-crowned sparrows, isolate songs include whistles but lack the trills, complex syllables, and buzzes characteristic of normal song (Fig. 1). Despite this simplicity, the overall duration of isolate song is approximately normal (7). It is possible that this innate song information ("innate template") is used for song recognition; to test this, we compared responses of naive birds to isolate song with those to normal song.

The two subspecies of white-crowned sparrows, *Z. l. nuttalli* and *Z. l. oriantha*, differ in genetic makeup, natural history, vocal learning, and details of song structure. We also compared responses to songs of these different subspecies, to determine if the conspecific recognition capability of naive sparrows could distinguish the subtle differences between subspecies, in addition to the greater differences between conspecific and foreign song. Finally, we tested a variety of foreign songs. Two types of foreign song came from the song sparrow and the savannah sparrow, species that live in the same area as Nuttall's white-crowned sparrow (Fig. 1). These foreign songs are heard but not normally learned by young white-crowned sparrows; therefore, they must be identified as foreign and rejected during song learning (7). This discrimination performed in nature is tested when responses elicited by these foreign songs are compared with those elicited by conspecific song. The third set of foreign songs came from the zebra finch, an Australian species whose songs are never normally heard by white-crowned sparrows and are acoustically very different.

Strikingly, the vocal response to isolate song matched that to normal song, consistent with a role for the innate template in conspecific song recognition (Fig. 2*A*). Responses to *nuttalli* subspecies songs tended to be stronger than to *oriantha* subspecies songs, but not significantly so ($Z = 1.60$, $P = 0.11$). Responses to all foreign songs were significantly less than those
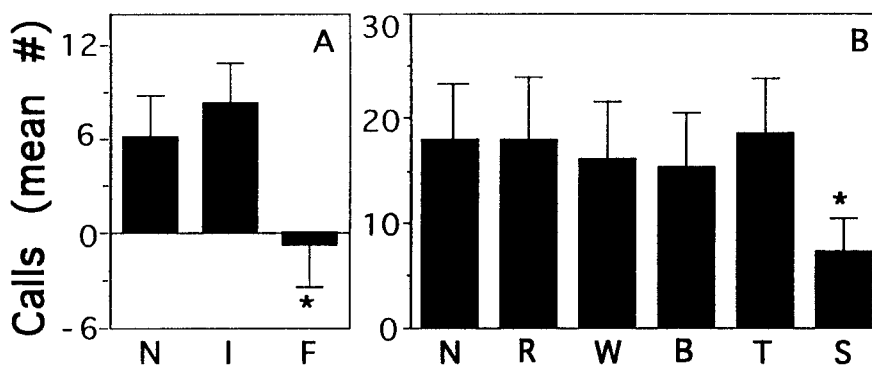

FIG. 2.   Mean (±SEM) number of calls given during playback experiments after subtraction of pre-trial responses in the first (*A*) and second (*B*) experiment. Abbreviations are explained in Fig. 1; "F" depicts the mean response to the "foreign" song of song sparrows, savannah sparrows, and zebra finches. *, $P < 0.05$ compared with normal song.

to conspecific song ($Z = 2.38$, $P < 0.05$), as previously demonstrated (9). Despite the strong acoustic differences between zebra finch songs and song sparrow and savannah sparrow song, responses to finch song were not significantly different from those to these sparrows.

The potency of isolate song suggests that isolate and normal song share acoustic features that act as conspecific markers. Both song types contain one or more whistles. Whistles are produced by white-crowned sparrows without tutoring, are universal across dialects, and are distinct from song phrases produced by foreign species in our study area; thus, they alone could suffice as conspecific markers. Alternatively, phrase order (i.e., whistles first) or all phrase types might contribute to conspecific identification.

Innate song recognition may depend on the presence of the introductory whistle or on other song features as well. To test these possibilities, fledglings were tested with "repeated phrase" songs. These songs reiterated single characteristic phrases of conspecific song (whistles, buzzes, or trills; Fig. 1) while preserving normal song length and internote interval. To assess the role of temporal order in song recognition, we also tested reversed white-crowned sparrow song (Fig. 1). Vocal responses to these modified songs were compared with normal conspecific song and foreign song (song sparrow song).

All "repeated phrase" songs were as potent as normal white-crowned sparrow song in eliciting vocal responses (Fig. 2B; $Z = 1.88$, $P > 0.2$ for whistle songs; $Z = 1.83$, $P > 0.4$ for buzz songs; $Z = 0.09$, $P > 0.9$ for trill songs; Bonferroni inequality correction for multiple *post hoc* tests). Thus, innate recognition did not depend exclusively on whistles. Moreover, these naive birds vocalized strongly to conspecific phrases even when presented out of the context of the entire song. Likewise, responses to reversed and normal songs were equivalent ($Z = 0.42$, $P > 0.9$). Either naive young birds are insensitive to phrase order or they recognize those song components that are only minimally changed when the song is reversed, like whistles and buzzes. As expected, responses to foreign song were significantly weaker than those to normal song ($Z = 3.56$, $P < 0.005$).

Vocal responses from the 1995 cohort of birds were greater to all stimuli than those from the 1994 cohort (Fig. 2). The 1995 group contained five highly vocal fledglings from four different broods. Although these subjects chirped at a high rate, their responses to normal and modified conspecific songs were still greater than those to foreign songs. In addition, male and female birds did not differ in their vocal responses to conspecific song (Mann–Whitney $U = 110$, $P = 0.4$). This finding suggests that females are also equipped to recognize conspecific song, even though singing is a predominantly male behavior in this species.

Electrophysiology. To investigate the neural basis of behavioral conspecific recognition, we recorded auditory responses of neurons in brain regions involved in auditory processing and song learning and production (17–22). The song nucleus HVc and the surrounding neostriatum are critical for species recognition in adult female canaries (16) and have strong responses to conspecific song or learned songs in birds with song experience (12, 18–22).

In HVc and underlying auditory neostriatum of naive birds, we recorded neural responses to the same song stimuli presented in the playback experiments. As was seen with the behavioral results, neurons responded strongly and equivalently to conspecific songs, isolate song, and reversed song. In contrast to the behavioral results, however, neostriatal units were also equally responsive to foreign and conspecific song (Fig. 3 A and B); thus, there were no significant differences between the mean response strengths to normal conspecific, isolate, reversed songs and foreign song (one-way ANOVA, $F_{3,92} = 0.251$; $P < 0.87$). This was true even for comparisons between pairs of song categories (paired *t* tests), and for
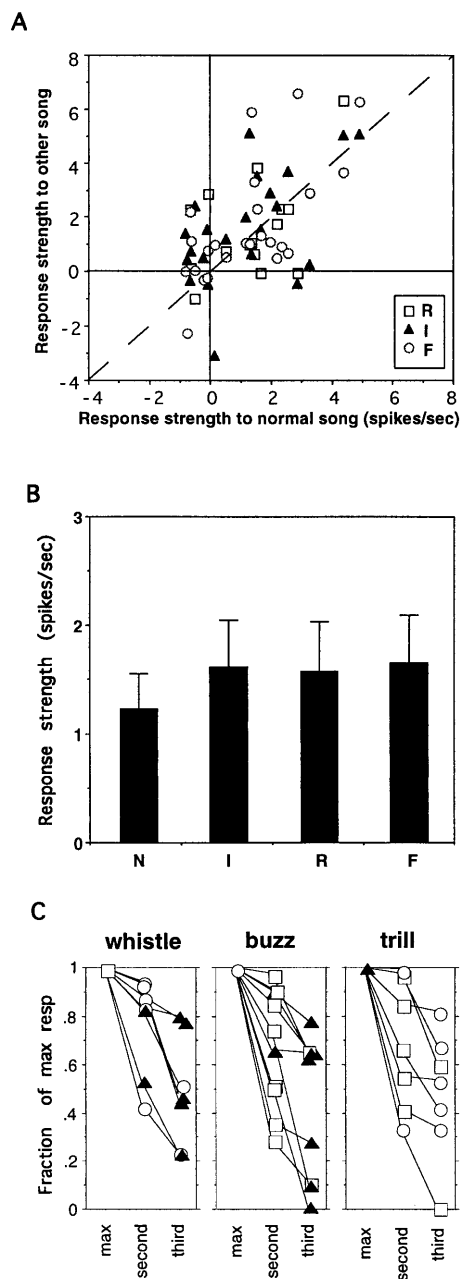

FIG. 3. Responses of HVc and neostriatal auditory neurons in the naive sparrow. (*A*) Each neuron's response strength to normal song (N) is plotted against its response strength to reversed (R), isolate (I), or foreign song (F). Although some neurons responded less to other songs than to normal song, the majority of units lie along the dotted line, indicating their equal responses to the stimuli compared. (*B*) Histogram of mean response strengths of all auditory neurons to each song category. Error bars are SEMs. (*C*) Response profiles of neurons with maximum responses to whistles (*Left*), buzzes (*Center*), or trills (*Right*). The *x* axis indicates the rank of the response to each phrase type; mean response strength to whistles are marked as squares, buzzes as circles, and trills as triangles. The *y* axis depicts response strength to phrases, normalized by the maximum response. Each connected line represents the mean response of a single neuron to whistle, buzz, and trill phrases. Neurons with strong phrase preferences have steep slopes.

individual neurons (Fig. 3A). No differences were observed between male and female neural responses.

Neurons responsive to both foreign and conspecific song may have been responding to acoustic components shared by both types of song. For example, some neurons fired strongly
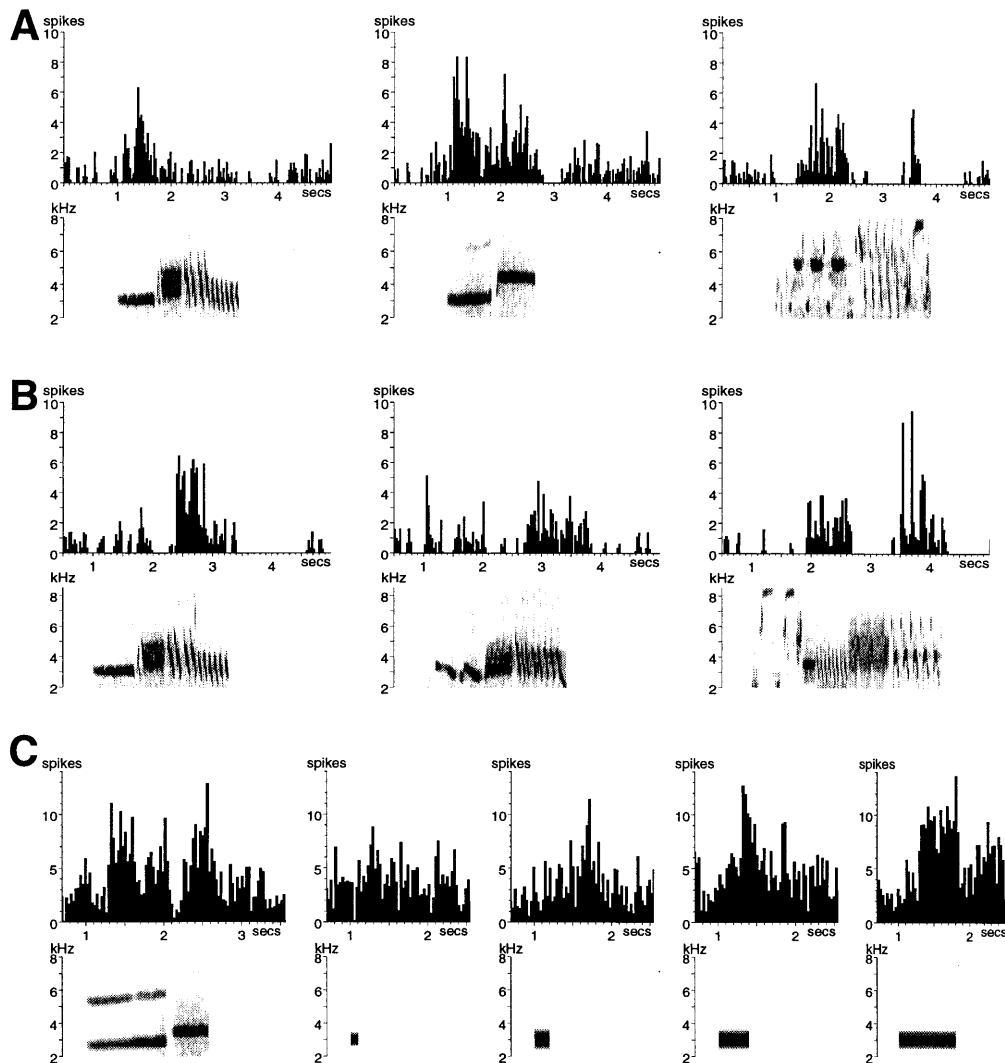
FIG. 4. Responses of neostriatal neurons to songs and their component phrases. (*A*) Peristimulus time histograms of the response of a single unit 200 μm below HVc shows its strong response to the whistles present in normal white-crowned sparrow song (*Left*), isolate song (*Center*), and song sparrow song (*Right*). (*B*) This small cluster of units 300 μm below HVc had strong responses to the trills present in both conspecific (*Left* and *Center*) and foreign (*Right*) songs. (*C*) A single unit in HVc responded strongly to isolate song (*Left*). This cell also responded with increasing strength as the duration of a tone burst increased (four panels on the right); this tone burst shared the frequency range of the whistle in the isolate song.

to specific phrase types (whistles, buzzes, or trills) found in both foreign and conspecific songs (Fig. 4*A* and *B*). The phrase preferences of individual neurons are shown by ranking their mean firing rates to whistle, buzz and trill phrases from all songs (Fig. 3*C*). Nine neurons responded to one phrase type at least twice as much as to other phrase types; their preferences included all phrase types ($n = 3$ for trills, $n = 2$ for whistles, $n = 4$ for buzzes). Such neurons might serve as phrase detectors and could underlie the strong behavioral responses to synthetic songs composed of only one repeated white-crowned sparrow phrase type. Neurons responding well to all phrase types may be sensitive to simpler acoustic features, such as frequency content, that are shared by different phrase types.

Although only a small number of neurons preferred whistles, many had strong responses to whistles, which are universally present in normal white-crowned sparrow and isolate songs ($n = 14$). In six of these whistle-responsive neurons, however, short tone bursts (100–300 ms) in the same frequency range as the whistle failed to mimic the response to whistles. Moreover, some neurons with weak or no responses to short tone bursts markedly enhanced their responses when tone burst duration was increased to approximate the length of normal song

whistles (500–1,000 ms; $n = 4/8$ neurons tested; Fig. 4*C*). This duration sensitivity raises the possibility that some neurons are innately tuned to temporal features of normal white-crowned sparrow song.

DISCUSSION

In behavioral testing, young naive birds vocalized selectively to conspecific songs and song phrases, including isolate songs produced by birds lacking tutor song experience. Songs made up of single, repeated phrases and reversed songs were as effective as normal song in eliciting selective vocal responses, indicating that innate recognition of song is not dependent on normal phrase order or song complexity. All phrase types were as potent as normal songs, suggesting innate recognition can rely on any of the three phrase types. Alternatively, these phrases could share acoustic features (e.g., timbre, pitch) that are detected by the recognition mechanism. The strong responses elicited by all repeated phrase songs suggest that the innate template that guides song production in isolate sparrows cannot be solely responsible for conspecific song recognition because this template only specifies whistles. Although whistles

were not the only cue for innate recognition, they appear to influence the selection of a tutor model during song acquisition (J.A.S. and P.M., unpublished data).

Although naive fledglings discriminated behaviorally between conspecific and foreign song, the population of neurons sampled electrophysiologically did not. Behavioral responses are the final product of many neuronal inputs and processing steps, and we may have recorded from areas relatively early in the auditory pathway, not yet selective for the entire conspecific song. Instead, these areas may only respond to components of songs, both conspecific and heterospecific. Consistent with this, we found neurons that responded preferentially to one of three basic phrase types found in white-crowned sparrow song, and others that responded best to whistles that were of the duration characteristic of this species. Such neurons may then project to other brain areas, not yet identified, where information from multiple neurons converges, eventually giving rise to single neurons selective for entire conspecific songs. Alternatively, behavioral conspecific discrimination may be mediated by the firing in concert of ensembles of whistle-, buzz-, or trill-responsive cells, together providing a recognition signal to the bird. Either way, responses to foreign song would be weaker than those to conspecific songs. Although some foreign songs also contain these phrase types, they are not completely composed of them, as are white-crowned sparrow songs.

The young birds' ability to recognize songs composed of single phrase types may underlie their capacity to identify and memorize the songs of any white-crowned sparrow that they encounter, whether or not phrase order varies or even whether or not all three phrase types are present (23, 24). Similarly, the responses of naive human infants to the phonemes of all human languages tested provide them with the capacity to learn any language (1–4). Moreover, the identification of neurons especially responsive to specific conspecific song phrases suggests that innately specified circuits in the naive brain might provide the initial substrate for vocal recognition and/or learning. Mechanisms by which speech recognition and production develop in humans may be elucidated by understanding how the songbird brain innately responds to a range of sounds, and then, as a result of experience, develops ever more specific responses (11, 20).

1. Eimas, P. D., Miller, J. L. & Jusczyk, P. W. (1987) in *Categorical Perception,* ed. Harnard, S. (Cambridge Univ. Press, New York), pp. 161–195.
2. Streeter, L. A. (1976) *Nature (London)* **259,** 39–41.
3. Kuhl, P. K. (1994) *Curr. Opin. Neurobiol.* **4,** 812–822.
4. Werker, J. F. & Polka, L. (1993) in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life,* ed. de Boysson-Bardies, B. (Kluwer, Dordrecht, The Netherlands), pp. 75–288.
5. Thorpe, W. H. (1958) *Ibis* **100,** 533–570.
6. Marler, P. & Peters, S. (1977) *Science* **198,** 519–521.
7. Marler, P. (1970) *J. Comp. Physiol. Psychol. (Monogr.)* **71,** 1–25.
8. Dooling, R. J. & Searcy, M. H. (1980) *Dev. Psychobiol.* **13,** 499–506.
9. Nelson, D. A. & Marler, P. (1993) *Anim. Behav.* **46,** 806–808.
10. Marler, P. & Peters, S. (1989) in *The Comparative Psychology of Audition: Perceiving Complex Sounds,* eds. Dooling, R. & Hulse, S. (Erlbaum, Hillsdale, NJ), pp. 243–273.
11. Doupe, A. J. (1997) *J. Neurosci.* **17,** 1147–1167.
12. Fortune, E. & Margoliash, D. (1995) *J. Comp. Neurol* **360,** 413–441.
13. Vates, G. E., Broome, B. M., Mello, C. V. & Nottebohm, F. (1996) *J. Comp. Neurol.* **366,** 613–642.
14. Konishi, M. (1965) *Z. Tierpsychol.* **22,** 770–783.
15. Marler, P. & Sherman, V. (1983) *J. Neurosci.* **3,** 517–531.
16. Brenowitz, E. (1991) *Science* **251,** 303–305.
17. Nottebohm, F., Stokes, T. M. & Leonard, C. M. (1976) *J. Comp. Neurol.* **165,** 457–486.
18. McCasland, J. S. & Konishi, M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7815–7819.
19. Margoliash, D. (1983) *J. Neurosci.* **3,** 1039–1057.
20. Volman, S. F. (1993) *J. Neurosci.* **13,** 4737–4747.
21. Leppelsack, H. J. & Vogt, M. (1976) *J. Comp. Neurol.* **107,** 263–274.
22. Mello, C. V., Vicario, D. S. & Clayton, D. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6818–6822.
23. Marler, P. (1984) in *The Biology of Learning*, eds. Marler, P. & Terrace, H. S. (Springer, Berlin), pp. 289–309.
24. Marler, P. & Nelson, D. A. (1995) in *Seminars in the Neurosciences,* ed. Marler, P. (Saunders, London), Vol. 4, pp. 415–423.