nature neuroscience

Neural correlates of categorical perception in learned vocal communication

Jonathan F Prather[1], Stephen Nowicki[1,2], Rindy C Anderson[2], Susan Peters[2] & Richard Mooney[1]

The division of continuously variable acoustic signals into discrete perceptual categories is a fundamental feature of vocal communication, including human speech. Despite the importance of categorical perception to learned vocal communication, the neural correlates underlying this phenomenon await identification. We found that individual sensorimotor neurons in freely behaving swamp sparrows expressed categorical auditory responses to changes in note duration, a learned feature of their songs, and that the neural response boundary accurately predicted the categorical perceptual boundary measured in field studies of the same sparrow population. Furthermore, swamp sparrow populations that learned different song dialects showed different categorical perceptual boundaries that were consistent with the boundary being learned. Our results extend the analysis of the neural basis of perceptual categorization into the realm of vocal communication and advance the learned vocalizations of songbirds as a model for investigating how experience shapes categorical perception and the activity of categorically responsive neurons.

One way that perception enables adaptive behavior is by grouping variable stimuli into classes. An impressive example of this synthetic capacity of the nervous system occurs when continuous changes in a stimulus parameter are perceived as discrete perceptual categories[1]. Categorical perception has a prominent role in processing acoustic signals that are important to social communication[2], including vocalizations made by humans[3,4], monkeys[5], rodents[6], birds[7] and frogs[8,9]. In spoken English, for example, the distinction between the speech sounds that elicit the phonemes /ba/ and /pa/ stems from a difference in delay time between the initial sound made by opening the lips and the onset of vocal fold vibration (voice onset time). Voice onset time varies considerably from individual to individual, but this variation is perceived categorically[3,4], with native English speakers recognizing a categorical boundary between /ba/ and /pa/ at ∼12 ms[3,4]. Presumably, categorical processing of speech more generally facilitates comprehension by generating perceptual constancy in the face of individual variation in many dimensions, including vocal pitch, timbre and tempo.

Despite the prominent role of categorical perception in the processing of speech and other natural signals, the neural mechanisms underlying this cognitive ability remain poorly studied. The fact that some animals have been shown to perceive human speech sounds categorically, even when they do not learn their own species-typical vocalizations[6], suggests that the capacity for categorical perception reflects some fundamental and innate brain mechanism. At the same time, however, the underlying mechanism must also be sensitive to experience because the perceptual boundaries that define speech sounds vary across native speakers of different languages[3,4], indicating that the boundaries, as with other features of speech, are learned. What

is lacking is a model system in which the neural underpinnings of categorical perception can be studied for a natural, learned communication system. Here we asked whether sensorimotor neurons that are known to be important for learned vocal communication in songbirds could also mediate categorical perception of learned vocalizations.

Swamp sparrows (*Melospiza georgiana*) are especially appropriate for examining the neural events underlying categorical perception of learned vocalizations for several reasons. First, as with all songbirds studied to date, swamp sparrows learn their song notes by imitation[10], a feature of human speech that is otherwise rare among animals[11]. Second, behavioral experiments have shown that male swamp sparrows use categorical perception to distinguish fundamental acoustic elements in their species-typical vocal repertoire[7]. Swamp sparrow songs comprise repeated groups of 2–5 'notes' (**Fig. 1a**), which are composed of short pure-tonal frequency sweeps, with note categories differing primarily in duration, bandwidth and rate of change in frequency[12]. Previous work has shown that males use categorical perception to distinguish between note types that, similar to the phones produced in speech, are produced with considerable variation by different individuals but are grouped into natural categories. Specifically, note types in categories I and VI share most spectral features, but differ along a duration continuum[7,12,13], and individual male swamp sparrows in a wild population in New York perceived these differences categorically, with an estimated perceptual boundary at a duration of 13 ms (**Fig. 1b,c**)[7].

A third feature that advances swamp sparrows as a model for investigating neural activity underlying categorical perception is that their brains, as in other songbirds, contain a discrete set of specialized

[1]Department of Neurobiology, 101 Research Drive, Duke University Medical Center, Durham, North Carolina 27710, USA. [2]Department of Biology, 125 Science Drive, Duke University, Durham, North Carolina 27710, USA. Correspondence should be addressed to R.M. (mooney@neuro.duke.edu).
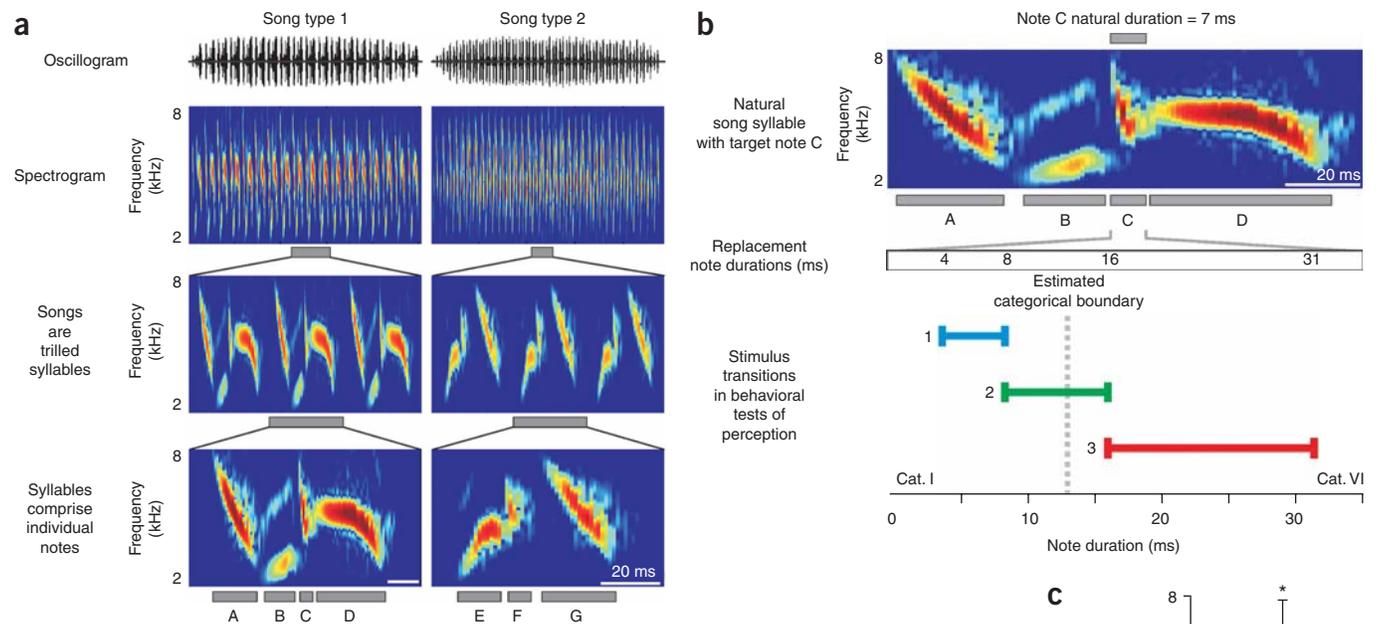
Figure 1 Earlier behavioral tests with swamp sparrows reveal categorical perception of changes in individual song notes. (**a**) Male swamp sparrows sing a small repertoire of distinct song types composed of trilled syllables that comprise multiple notes (2–5 song types, each ~2-s duration). (**b**) Categorical perception in swamp sparrows was demonstrated previously[7] by splicing out one note in the syllable (note C) and replacing it with a note with similar spectral characteristics, but different duration. Replacement note durations differed by equal increments on a logarithmic scale. (**c**) Behavioral testing revealed that swamp sparrows perceived strong differences, evident as robust expression of aggressive displays, when note transitions spanned an estimated boundary of 13 ms (stimulus set 2), but perceived little or no difference when transitions did not cross that boundary (stimulus sets 1 and 3, figure adapted from ref. 7), indicating that swamp sparrows employ categorical perception in their song discrimination. Error bars denote s.e.

sensorimotor structures devoted to learning, producing and perceiving song[14–19]. Previously, we found that one of these structures, the nucleus HVC (**Fig. 2a**), contained a certain class of striatum-projecting neurons (HVC_X cells; **Fig. 2b,c**) that respond to only one song type in the bird's repertoire[20,21]. Three observations implicate song type–selective HVC_X neurons in song perception. First, song perception is impaired by lesions to HVC[15,16]. Second, song perception is also impaired by lesions to the striatal portion of an anterior forebrain pathway into which HVC_X cells project their axons[22]. Finally, some HVC_X cells show a sensorimotor correspondence that is reminiscent of mirror neurons in the monkey cortex, which are hypothesized to be important in perception[20,23].

Although HVC_X cells appear to be well-suited to a role in perception, the extent to which their auditory properties are correlated with the bird's perception of categorical differences among song features is unknown. Therefore, we used a lightweight chronic recording device[20,24] to assess whether HVC_X cells in freely behaving adult male swamp sparrows responded categorically to changes in note duration, known from behavioral experiments to be an important feature that distinguishes between note-type categories I and VI (ref. 7).

RESULTS

In a set of initial experiments, we recorded from antidromically identified HVC neurons in freely behaving adult male swamp sparrows and presented each bird's song types through a speaker located near its perch. These recordings confirmed that HVC_X neurons responded to a single song type in the bird's repertoire (the primary song type) and revealed that robust auditory responses could be evoked in interneurons (HVC_{INT} cells). In contrast, all of the cells of

the other projection neuron type in HVC that we sampled, which innervates the song premotor nucleus RA (HVC_{RA} cells), were unresponsive to auditory stimuli.

To test whether individual HVC_X neurons or interneurons express categorical responses to changes in the duration of notes in the primary song type, we first identified the primary song type to which a cell responded and then identified a note in the repeated syllable comprising that song type that was in either a category I (<13 ms) or a category VI (>13 ms) note type, as defined in previous studies[7,12,13]. We then replaced the category I or category VI note in the primary song type with a note that had similar spectral features, but had a different duration, following the same procedures used in prior behavioral studies (**Fig. 3a**), in which replacement notes were chosen that differed in their durations by equal increments on a logarithmic scale[7] (see Methods for additional details). The resulting synthetic syllable was assembled into a trill to form a stimulus song, with all other notes, internote intervals and trill rate being identical to the natural song. Repeating this procedure using different replacement notes to create each stimulus, we presented to each cell a set of song stimuli comprising 5–11 variants of the primary song type, each differing only in the duration of a single replacement note in each trilled syllable, with durations spanning a range from notes that would be classified as category I to those that would be classified as category VI.

HVC_X neurons display categorical auditory responses

Acoustic presentation of variants of the primary song type revealed that the auditory responses of HVC_X neurons, but not interneurons, are highly sensitive to changes in note duration (**Fig. 3**). Most HVC_X neurons responded robustly to song stimuli containing replacement
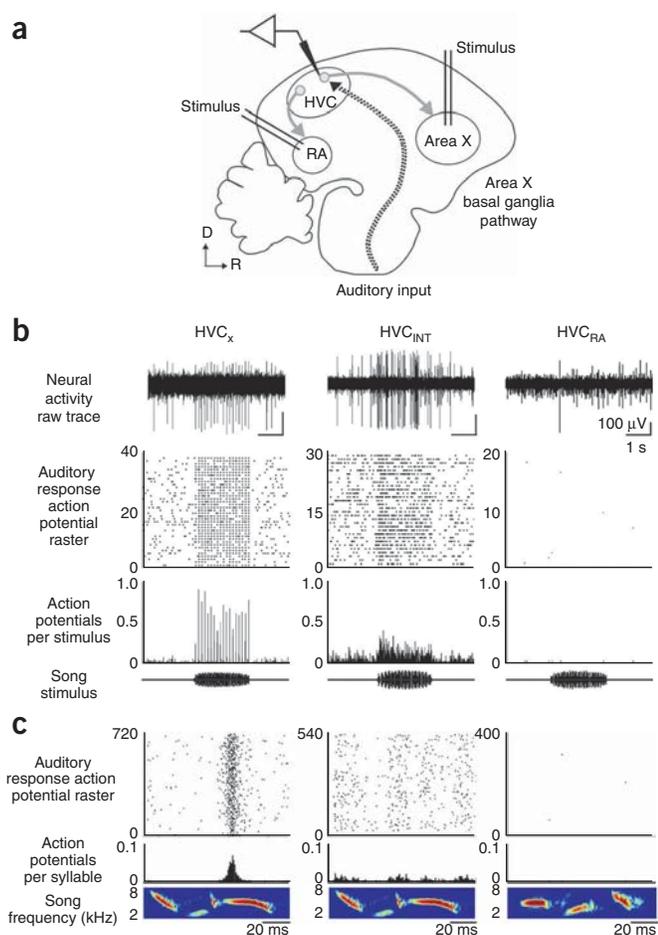
Figure 2 Extracellular recordings reveal differences in the auditory response of identified single units in nucleus HVC. (**a**) A parasagittal schematic of the swamp sparrow brain at the level of HVC. Antidromic stimulation was used to identify HVC neurons[19] as projecting to striatal areas that are implicated in song learning and perception (Area X), as projecting to premotor structures (robust nucleus of the arcopallium, RA) and of making local connections in HVC (HVC interneurons, omitted for clarity). Each cell type in HVC receives auditory input[45] (dotted line; D = dorsal, R = rostral). (**b**) Auditory responses of HVC_X neurons ($n = 24$ of 29 cells, 5 birds) were phasic and typically evoked by only one song type in the bird's repertoire (primary song type[20], $n = 23$ of 24 responsive cells, identity of primary song type varied across cells). Auditory responses of HVC_{INT} cells ($n = 18$ of 18 cells, 5 birds) consisted of tonic increases in activity in response to many or all song types in the repertoire. In contrast, HVC_{RA} neurons were unresponsive to any stimulus that was presented (top, raw data traces; middle, auditory response raster; bottom, auditory response PSTH, 10-ms bin size; $n = 16$ cells, 5 birds). (**c**) HVC_X phasic responses occurred at a restricted phase in the syllable of the primary song type (top, auditory response raster; bottom, auditory response PSTH, 1-ms bin size), whereas HVC_{INT} cells responded throughout the syllable duration.

Fig. 3b,c). These results indicate that categorical responses to changes in note duration are shown by only one subset of auditory responsive HVC neurons, namely those that project to a striatal pathway that is important in song perception[15,22].

To probe the link between these categorical neuronal responses and song perception, we first estimated the categorical boundary evident in the auditory responses of HVC_X neurons. Interpolation at the transition from strong to weak responses yielded an estimated categorical response boundary at 21 ± 4 ms (mean ± s.d., $n = 19$ cells, 5 birds). We further probed the location of the categorical response boundary at a higher resolution in a subset of HVC_X cells using stimuli containing replacement notes with durations near that of the estimated boundary (19, 22 and 25 ms, $n = 10$ cells, 2 birds; **Fig. 3d** and **Supplementary Figure 3** online). This higher-resolution investigation of the categorical transition yielded an estimated boundary of 20 ± 4 ms, very similar to that obtained using the full dataset (21 ± 4 ms, $P = 0.47$, unpaired t test). Such a sharp response transition is a hallmark of categorical activity[3,4,28] (**Supplementary Methods** online), further establishing the auditory responses of HVC_X neurons as categorical.

The HVC_X response boundary marks a perceptual boundary

Although the categorical responses of HVC_X cells are consistent with their having some role in the categorical perception of note type on the basis of duration, we found it curious that the categorical boundary estimated from our neural data (~21 ms) was different from the perceptual boundary that was previously reported for swamp sparrows on the basis of behavioral testing (~13 ms)[7]. This discrepancy suggests that the neural responses that we measured are not linked to perception of note duration or, alternatively, that the perceptual boundary in the birds that we studied differed from the previously reported values. All of the birds from which we collected neural data were obtained from a population in northwestern Pennsylvania, whereas the previous behavioral study examined a population from upstate New York, some 540 km distant[7]. As with most songbirds, geographically distinct populations of swamp sparrows learn different song dialects[29], raising the possibility that the perceptual boundary separating category I and category VI notes may be influenced by learning and thus may vary across populations.

To test whether the discrepancy between our neural data and previous behavioral measures of a categorical boundary reflects a learned dialect difference, we first compared the acoustic characteristics of category I and category VI notes obtained from the Pennsylvania and

notes with durations in the range of the target note category, but weakly or not at all to variants of the primary song type with replacement notes having durations outside of that category ($n = 19$ of 22 HVC_X cells, 5 birds). Robust responses were invariably evoked by stimuli containing replacement notes with durations that would fall unambiguously into the same category as the target note in the natural song (for example, note durations of 4 or 8 ms, target note duration of 7 ms, replacement notes near the categorical boundary discussed below; **Fig. 3a**). Following the criteria established in previous studies of categorical perception[7,25–27], we termed these cells as being categorically responsive (see Methods for additional details). Categorical responsiveness was evident in HVC_X cells both within and across birds (within bird: $n = 4$ of 4 cells, 1 bird, $P < 0.001$; across birds: $n = 19$ cells, 5 birds, $P < 0.001$; Mann-Whitney U test; **Fig. 3b,c**), regardless of whether the original target note was of the shorter category I type or of the longer category VI type (**Fig. 3d**).

Beyond the sheer abundance of categorically responsive HVC_X neurons, two observations indicated that the HVC_X neuronal population encodes note duration in a categorical manner. First, the three HVC_X cells that did not respond categorically also did not show bell-shaped tuning curves to intermediate note durations, as might be expected if note duration is represented continuously (**Supplementary Figure 1** online). Second, our dataset differed substantially from a model in which note duration is encoded linearly by HVC_X cells (**Supplementary Figure 2** online). In contrast to HVC_X neurons, HVC_{INT} cells responded similarly regardless of whether the replacement note was of the same category or a different category as the target note ($n = 16$ of 18 cells, 5 birds, $P = 0.24$, Mann-Whitney U test;
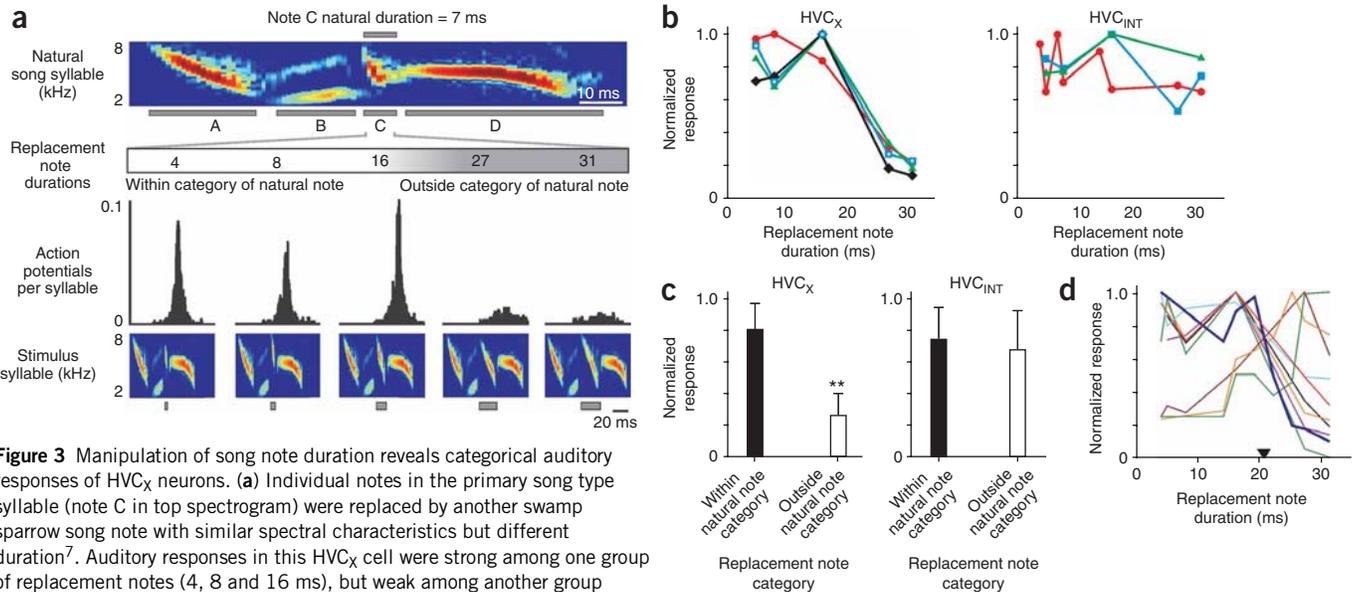
Figure 3 Manipulation of song note duration reveals categorical auditory responses of HVC_X neurons. (**a**) Individual notes in the primary song type syllable (note C in top spectrogram) were replaced by another swamp sparrow song note with similar spectral characteristics but different duration[7]. Auditory responses in this HVC_X cell were strong among one group of replacement notes (4, 8 and 16 ms), but weak among another group (27 and 31 ms; third row, auditory response PSTH, 1-ms bin size; bottom row, stimulus syllables, gray boxes indicate replacement notes). (**b**) Categorical responses such as those in **a** were evident in all HVC_X cells in this bird that responded to the same song type (*n* = 4, left, open blue squares indicate the cell in **a**). In contrast, categorical activity was not evident in any of the HVC_INT neurons that responded to the same song type in the same bird (*n* = 3, right). In behavioral testing, this same bird responded to differences between stimuli that crossed the putative neurophysiological boundary but not to differences between stimuli that did not cross the boundary (**Supplementary Fig. 4**). (**c**) Among all cells tested in all birds (*n* = 22 of 24 auditory responsive HVC_X cells, 5 birds and 18 of 18 responsive HVC_INT cells, 5 birds), categorical activity was consistently evident for HVC_X neurons (*n* = 19 of 22 cells, 5 birds), with stronger response strengths being evoked by replacement notes in the same category as the note that was replaced. In contrast, categorical activity was very rarely evident in HVC_INT neurons (*n* = 2 of 18 cells, 2 birds). Error bars denote s.e. (**d**) Categorical responses were observed in HVC_X cells, regardless of whether the duration of note that was replaced was naturally short or long, and the location of the categorical boundary (black triangle) was consistent across HVC_X cells and birds (21 ± 4 ms, mean ± s.d.; ten cells shown for clarity). A subset of ten categorically responsive HVC_X cells (two birds) was tested further to probe the categorical boundary at a higher resolution (19, 22 and 25 ms; for example, thick blue line), revealing an estimated categorical boundary (20 ± 4 ms) that was very similar to that estimated using the full dataset (21 ± 4 ms; *P* = 0.47, unpaired *t* test).

New York populations. We found that the distributions of note durations differed between these two populations, as would be predicted by a dialect difference in this phonological feature (**Fig. 4a**). We next asked whether population-specific differences in note durations are paralleled by population differences in song perception. We tested the categorical perceptual boundary for note duration in the Pennsylvania population, replicating the behavioral experiment carried out on the New York population[7], in which habituation to repeated presentation of one song stimulus was followed by presentation of a test stimulus (that is, the stimulus transitions in **Fig. 4b,c**). This approach is a standard method for establishing categorical perception in human infants[30] and nonhuman animals[7,26].

Swamp sparrows are territorial and the introduction of novel song stimuli into their territory evokes robust aggressive displays such as wing waves[7,31]. We quantified this territorial behavior in a habituation/dishabituation procedure to determine whether the perceptual boundary of Pennsylvania birds more closely approximated the perceptual boundary measured behaviorally for New York birds[7] or the neural response boundary that we measured for Pennsylvania birds. We found that Pennsylvania birds responded as though they did not detect changes in note duration that spanned the perceptual boundary reported for New York birds (*P* > 0.05, stimulus transition 2 versus 1; Tukey's honestly significant difference (HSD) test; **Fig. 4b,c**). However, Pennsylvania birds clearly detected changes in note duration that spanned the neural response boundary detected in HVC_X neurons, indicating that the changes in duration also spanned a categorical perceptual boundary (*P* < 0.001, Kruskal-Wallis test; *P* < 0.05, stimulus transition 3 versus 1; *P* < 0.05, stimulus transition 3 versus 2;

Tukey's HSD). Thus, categorical perception of note duration in swamp sparrows, as measured in the field by behavioral testing, is accurately predicted by the auditory response properties of HVC_X neurons measured in the laboratory. Although it is extremely challenging to conduct behavioral tests that adequately capture the responsiveness of a wild territorial male songbird in a laboratory setting while simultaneously collecting neurophysiological data, we were able to obtain both behavioral and neurophysiological data at different times from one bird (**Supplementary Figure 4** online; see also **Figs. 2b,c** and **3a,b** and **Supplementary Figure 5** online). Although these findings are limited to a single individual, they indicate a direct within-individual correspondence between neurophysiological data and behavioral data related to those that we measured in a field setting.

Note duration is the major determinant of HVC_X responses

Although prior behavioral work implicated note duration as the primary salient acoustic feature for categorical perception[7], the stimuli used in that study and the initial set of stimuli used here varied note duration without controlling for changes in other covariant features of the manipulated note, namely the rate of frequency modulation or frequency bandwidth. Thus it remains to be established whether note duration, rather than frequency modulation or bandwidth, is the salient feature underlying categorical perception of category I and category VI notes by swamp sparrows and the categorical responses of their HVC neurons. Although the larger stimulus set that was required to distinguish between these alternatives is impractical to employ in field studies of wild sparrows, such measurements could be made for the responses of single neurons.
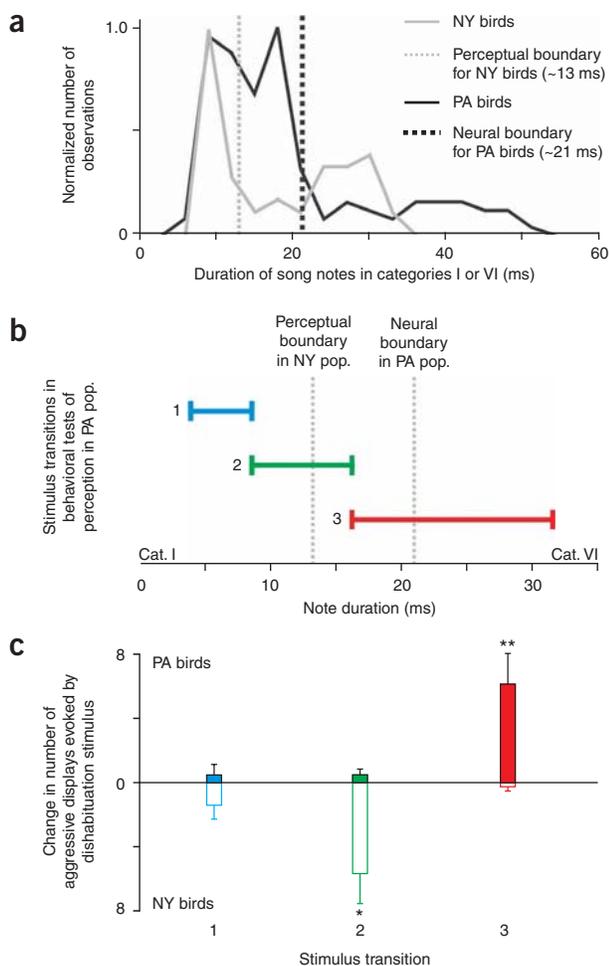
Figure 4 The perceptual boundary in swamp sparrows' categorization of note duration was predicted by the categorical boundary evident in auditory responses of HVC$_X$ neurons. (**a**) The distribution of category I and category VI[11] note durations in the songs of swamp sparrows from Pennsylvania (PA, thick black line) differed from the distribution in the songs of a previously studied population of swamp sparrows from New York[7] (NY, thin gray line; $P < 0.001$, Kolmogorov-Smirnov two-sample test, $n = 129$ notes, 62 songs from 35 Pennsylvania birds; $n = 52$ notes, 29 songs from 12 New York birds, 3-ms bin size), suggesting that the categorical boundary in each population (dotted lines) may be learned through population-specific auditory and vocal experience. (**b**) Stimuli in our behavioral tests contained note transitions that crossed no putative categorical boundary (4–8 ms), crossed the New York perceptual boundary but not the Pennsylvania neural boundary (8–16 ms), or crossed the Pennsylvania neural boundary but not the New York perceptual boundary (16–32 ms). (**c**) Behavioral testing revealed that Pennsylvania swamp sparrows (top, filled bars) perceived strong differences when the transition in note duration spanned the Pennsylvania neural boundary detected in HVC$_X$ neurons (stimulus set 3, $P < 0.001$, Kruskal-Wallis test; $P < 0.05$ for stimulus set 3 versus 1 and 3 versus 2, Tukey's HSD), but perceived little or no difference when the transition spanned the perceptual boundary detected in New York birds (stimulus set 2; $P > 0.05$, stimulus set 1 versus 2, Tukey's HSD) or spanned no putative boundary (stimulus set 1). Comparison of the behavioral data obtained from Pennsylvania birds (top, filled bars, mean ± s.e.) and from New York birds[7] (bottom, open bars, mean ± s.e.) made clear the population-specific differences in categorical perception of note duration (New York data adapted from ref. 7, * indicates $P < 0.05$, Kruskal-Wallis ANOVA, responses of individual Pennsylvania birds shown in **Supplementary Fig. 6** online).

categorical perception of these note types[7], is the salient song feature in driving categorical responses in HVC$_X$ neurons.

Categorical responses in sensorimotor HVC$_X$ neurons

In a prior study in which we recorded both auditory and singing-related activity in HVC$_X$ neurons of swamp sparrows, we established that individual cells show a precise sensorimotor correspondence[20]. In primates, neurons that display sensorimotor 'mirroring' are hypothesized to mediate perception of communication signals[23]. Therefore, we investigated whether categorical responsiveness was evident in the same cells for which this sensorimotor correspondence had been established.

Categorical auditory responsiveness was observed in all HVC$_X$ neurons for which we also recorded singing-related activity ($n = 5$

To begin to determine whether note duration was a primary determinant of HVC$_X$ neuronal responses, we tested a subset of HVC$_X$ neurons ($n = 10$ cells, 2 birds; different subset than the cells used in high-resolution testing of boundary location) with variants of the primary song type containing computer-generated replacement notes. The use of these synthetic notes allowed us to systematically change note duration while holding constant either the note's frequency modulation or frequency bandwidth (**Fig. 5a–c**). Syllables containing synthetic notes that closely replicated the features of the natural note that they replaced were just as efficacious as the natural syllable at eliciting responses from HVC$_X$ neurons ($P = 0.47$, Wilcoxon signed rank test; **Fig. 5a**). Furthermore, syllables containing synthetic notes with durations similar to those of the natural notes that they replaced, but with different frequency modulation or bandwidths, were also highly effective at eliciting responses from HVC$_X$ neurons (frequency modulation, $P = 0.13$; frequency bandwidth, $P = 0.43$; Wilcoxon signed rank test; **Fig. 5b,c**). Thus, these data indicate that neither frequency modulation nor frequency bandwidth has a primary role in influencing our results, and that note duration, the song feature previously implicated in affecting
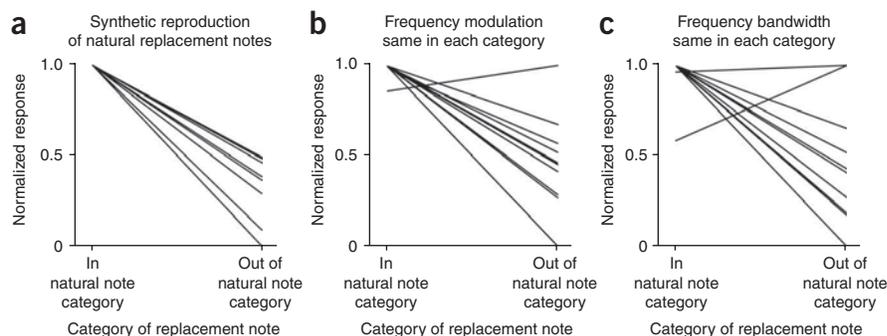
Figure 5 Categorical responses were evoked by changes in specific features of individual notes in the syllable. (**a–c**) A subset of ten categorically responsive HVC$_X$ neurons (two birds) was further tested to probe the acoustic stimulus features that directed categorical responsiveness. In all cells, computer-generated duplicates of natural replacement notes were sufficient to evoke a categorical response ($P = 0.47$, Wilcoxon signed rank test), establishing the validity of further testing using synthetic stimuli (**a**). Synthetic stimuli in which note duration was varied while either the rate of frequency modulation ($P = 0.13$; **b**) or the frequency bandwidth ($P = 0.43$, Wilcoxon signed rank test; **c**) was held constant evoked responses similar to those evoked by natural replacement notes, implicating note duration as the salient song feature driving categorical responses of HVC$_X$ neurons.

cells, 2 birds). Comparison of auditory and singing-related activity revealed that all of these neurons showed a precise sensorimotor correspondence[20] (for example, see **Supplementary Fig. 5**). These results suggest that neurons displaying a form of sensorimotor mirroring convey perceptual information about song.

DISCUSSION

This study provides, to the best of our knowledge, the first evidence for neurons that encode perceptual information about a phonological feature of learned vocal behavior, specifically information about a categorical perceptual boundary. Furthermore, our neural results accurately predicted a previously unknown geographic dialectical difference in this perceptual boundary, one that we subsequently confirmed using behavioral tests with wild birds. Variation in this perceptual boundary across swamp sparrow populations strongly suggests that both categorical perception and categorical neural responses in sparrows are affected by experience. Finally, by identifying a cell type that expresses categorical activity, we provide insights into the neural circuitry for categorical perception of note duration.

Categorical perception has been shown to have a prominent role in a wide variety of communication systems[3–9], including human speech[3,4]. Although categorical processing of both human and nonhuman vocalizations has been demonstrated in a variety of animals[3,4,6,8], neural correlates of this phenomenon await identification. Neurophysiological studies in other modalities, however, have identified that neurons can reliably encode information about perceptual categories. For example, neurons in the monkey inferotemporal cortex respond in a categorical manner to facial features[32], an important component of social signaling in primates, and brain imaging studies in humans have detected regions that show categorical sensitivity to differences in facial expression[33]. These studies strengthen the idea that categorical perception is mediated by categorically responsive neurons, and that the activity of those cells could be involved in communication. Our results identify neurons that respond categorically to learned vocal signals, extending the analysis of the neural basis of perceptual categorization to another major facet of communication with special relevance to human speech.

To our initial surprise, we found that the categorical response boundary for note duration that we measured from neurons differed markedly from the previously reported categorical perceptual boundary for this system using behavioral methods[7]. We found, however, that the response boundary that we determined from physiological recordings of HVC$_X$ cells accurately predicted a previously unidentified population difference in this perceptual boundary, indicating that the perceptual boundary varies across geographically distinct swamp sparrow populations and consistent with the view that, as is the case for human speech[3,4], these boundaries can be influenced by early experience. The fact that bird songs vary geographically as a consequence of learning is widely documented[11,34], but our data are, to the best of our knowledge, the first to suggest an effect of learning on the categorical perception of a natural signal other than speech. The observations that human infants across cultures show an initially similar ability to categorize speech sounds[35] and that categorical boundaries of human speech can be detected by mammals that do not learn their vocalizations[6] suggest that categorization of vocal signals may exploit innate neural mechanisms. However, the role of social experience in shaping human speech perception, including the perception of categorical boundaries, is also well documented[3,4], indicating that the underlying neural mechanisms must also be pliable. Future studies in songbirds will be especially fruitful for addressing how acoustic and perhaps social experience shapes the development of categorical perception and categorically responsive neurons.

Our study localized categorical responses to HVC$_X$ projection neurons, which innervate a striatal pathway that lesion studies implicate in song perception and song learning[15,16,18,22] and may be regarded as being analogous to mammalian corticostriatal neurons. Our finding that the auditory responses of these projection neurons are closely linked to song perception provides a physiological basis for understanding how lesions to this pathway could interfere with song recognition. Furthermore, our findings indicate that categorical responses are shown by HVC$_X$ neurons that display precise sensorimotor mirroring[20], providing a link between mirror neurons and perception. In the songbird, categorically responsive sensorimotor neurons could enable auditory perception to directly guide subsequent behavior, as evident in countersinging in response to hearing a neighbor's song[20]. Notably, neurons that represent learned visual categories have been detected in regions of monkey parietal cortex that have been implicated in motor planning and decision-making[36]. The expression of categorical responses in sensorimotor neurons in both songbirds and primates could point to an efficient means by which the nervous system determines when variations in a sensory stimulus warrant similar or different behavioral responses.

Establishing that categorical responses are expressed by striatal-projecting HVC$_X$ neurons, but not by interneurons, may also provide a clue as to how and where encoding of perceptual attributes occurs. Prior studies have shown that although HVC$_X$ neurons and HVC interneurons receive shared excitatory auditory input[37,38], the auditory selectivity of interneurons more closely parallels the activity of auditory afferents to HVC[37] and the highly selective auditory responses of HVC$_X$ neurons require inhibitory sculpting through interneurons[39]. The presence of categorical auditory responses in HVC$_X$ neurons and the absence of such responses in HVC interneurons suggest that categorical activity emerges through the local circuitry of HVC[40] and the processes of synaptic integration in HVC$_X$ cells. Future studies using intracellular recordings could test this idea by determining whether HVC$_X$ neurons receive broadly responsive excitatory input from which categorical responses are sculpted by local inhibition.

Precise mapping of the neuronal response boundary revealed that the action potential discharge of HVC$_X$ cells is exquisitely sensitive to changes in note duration on the millisecond time scale. Sensitivity to fine temporal features of song has been described for HVC neurons in other songbirds[41], and intracellular and extracellular recordings from putative HVC$_X$ neurons reveal that their song-evoked auditory responses are strongly dependent on auditory context, such as sequences of notes or syllables[42–44]. This context sensitivity involves integration over hundreds of milliseconds[42] and is thought to depend at least in part on circuit mechanisms that are local to HVC[38,39]. Although our study cannot resolve whether categorical responsiveness to note duration arises in HVC, it does suggest that temporal aspects of song ranging from milliseconds to hundreds of milliseconds can be encoded by single neurons. The sensitivity of individual neurons to features over multiple timescales may constitute a strategy for optimizing encoding of complex stimuli that can vary in both local structure and global sequence.

METHODS

Subjects. The male swamp sparrows that were used in neural testing were collected as adults (age > 1 year) from their breeding grounds. All field tests of song perception were performed in Crawford County, Pennsylvania during May and June of 2007. All experiments were performed in accordance with protocols approved by the Duke University Animal Care and Use Committee.

Song stimulus preparation. Exemplars of each song in the bird's repertoire (2–5 song types) were recorded and digitized[20] to be used as stimuli and in the construction of additional stimuli to test categorical responsiveness. Song notes of the type implicated previously in swamp sparrow categorical perception[7] served as target notes for replacement (for example, note C in **Fig. 3a**). Each replacement note had similar spectral characteristics, but had a different duration than the target note (for example, **Fig. 3a**), and only one target note in the syllable was replaced in any stimulus. The possible durations of replacement notes were 4, 5, 7, 8, 16, 19, 22, 25, 27 and 31 ms. These values were chosen to replicate as closely as possible the methods used in a previous behavioral assessment of categorical perception in swamp sparrows[7] and to resolve the location of the categorical boundary. The resulting synthetic syllable contained a replacement note in place of the target note, and all other notes and intervals were identical to the natural song. This synthetic syllable was assembled into a trill with a trill rate that was identical to the natural song and a total duration that closely approximated the duration of the natural song type (increasing or decreasing individual note durations changed the overall duration slightly). In a subset of cells, additional synthetic stimuli were presented in which either the rate of frequency modulation or the frequency bandwidth was controlled while the duration of the note was varied (**Fig. 5a–c**).

Experimental protocol. All experiments were performed using awake and freely behaving birds. Individual neurons were isolated using a microdrive device[24] implanted under isoflurane anesthesia. All birds were allowed to recover for 3 d following implantation before recording began. Neurons in HVC were identified as HVC_X, HVC_{RA} or HVC_{INT} cells either using antidromic stimulation methods or according to their electrophysiological and auditory response properties (**Supplementary Methods**). All neural data were amplified, filtered (band pass 500 Hz to 10 kHz) and digitized (25 kHz) to computer file (LabView) and all action potentials of individual units were discriminated using amplitude discrimination of the largest unit in a record (custom software) or discrimination on the basis of waveform characteristics (Wave-Clus). In both cases, single-unit isolation was verified using an interspike interval histogram to test for the presence of a refractory period.

When a single unit had been isolated and identified, stimulus presentation was immediately initiated (10-s quiet interval between each song presentation, stimuli presented in randomized order). Songs were played to the sparrow at 70 dB (peak r.m.s., A-weighted) through a speaker placed 20–35 cm away in the chamber (distance varied according to the bird's location in the cage). Playback of the bird's entire song repertoire, as well as their synthetic variants (see **Supplementary Methods**), was used to assess the auditory response of each neuron described above. Natural song types (unaltered from the original recordings) were played through a speaker placed inside the recording chamber to assess the auditory selectivity of each neuron (that is, to identify the primary song type). Extracellular recordings of action potentials in response to these stimuli were collected from 29 individual HVC_X (5 birds) and 18 HVC_{INT} units (same 5 birds).

Quantification of auditory response. Auditory activity in HVC_X neurons was taken as significant if the response to any natural song stimulus exceeded the mean + 5 s.d. of the baseline firing rate. Auditory activity in HVC interneurons was tested for significance using response strength ($P < 0.05$), a metric that also compares the response and baseline conditions[37]. The responses of HVC_X neurons and HVC_{INT} cells were normalized using the strongest response of each cell to any member of the set of synthetic stimuli, enabling comparison of auditory responses across cells (for example, **Fig. 3b**) and birds (for example, **Fig. 3c**).

Neural assessment of categorical responsiveness. Noting that the dataset of auditory responses in HVC_X cells tended to include transitions from strong responses (normalized responses closer to 1) to little or no responses (normalized responses closer to 0), we used interpolation to compute the note duration at which the auditory response crossed 0.5. In cases in which interpolated data crossed 0.5 multiple times (for example, **Fig. 3d**), the transition point in the cell was represented by the mean of the note durations corresponding to those multiple crossing points. This method revealed a putative boundary in the auditory responsiveness of these cells (~ 21 ms) and this value was used to test

whether cells expressed categorical responses to stimuli on either side of this boundary. In off-line analysis, neural responses were compared to well-established criteria[7,25,27] to determine whether responses were categorical in nature (**Supplementary Methods**). In short, responses were said to be categorical if activity was similar among members of a group of stimuli, but different between groups of stimuli, indicating a greater sensitivity to stimulus category than to the physical properties of the stimulus.

Behavioral assessment of categorical perception. Behavioral methods (Crawford County, Pennsylvania) closely paralleled a previous study of categorical perception in a New York population of swamp sparrows[7]. Song stimuli were created in the same manner described for neurophysiological testing (see above) and were presented in pairs (4:8 ms, 8:16 ms, 16:32 ms; one pair of songs to each bird; presentation sequence randomized) through a speaker placed inside of the bird's breeding territory. Perception of song features was assessed by quantifying territorial wing-wave responses to song stimuli in a habituation/dishabituation procedure that was previously employed to test auditory perception in swamp sparrows[7].

Note: Supplementary information is available on the Nature Neuroscience website.

1. Miller, E.K., Nieder, A., Freedman, D.J. & Wallis, J.D. Neural correlates of categories and concepts. *Curr. Opin. Neurobiol.* **13**, 198–203 (2003).
2. Bregman, A. *Auditory Scene Analysis* (MIT Press, Cambridge, Massachusetts, 1990).
3. Liberman, A.M., Harris, K.S., Hoffman, H.S. & Griffith, B.C. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* **54**, 358–368 (1957).
4. Liberman, A.M., Harris, K.S., Kinney, J.A. & Lane, H. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. Exp. Psychol.* **61**, 379–388 (1961).
5. May, B., Moody, D.B. & Stebbins, W.C. Categorical perception of conspecific communication sounds by Japanese macaques, *Macaca fuscata. J. Acoust. Soc. Am.* **85**, 837–847 (1989).
6. Kuhl, P.K. & Miller, J.D. Speech-perception by chinchilla: phonetic boundaries for synthetic VOT stimuli. *J. Acoust. Soc. Am.* **57**, S49–S50 (1975).
7. Nelson, D.A. & Marler, P. Categorical perception of a natural stimulus continuum: birdsong. *Science* **244**, 976–978 (1989).
8. Baugh, A.T., Akre, K.L. & Ryan, M.J. Categorical perception of a natural, multivariate signal: mating call recognition in tungara frogs. *Proc. Natl. Acad. Sci. USA* **105**, 8985–8988 (2008).
9. Nowak, M.A. & Krakauer, D.C. The evolution of language. *Proc. Natl. Acad. Sci. USA* **96**, 8028–8033 (1999).
10. Marler, P. & Peters, S. Sparrows learn adult song and more from memory. *Science* **213**, 780–782 (1981).
11. Catchpole, C.K. & Slater, P.J.B. *Birdsong: Biological Themes and Variations* (Cambridge University Press, New York, 1995).
12. Marler, P. & Pickert, R. Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Anim. Behav.* **32**, 673–689 (1984).
13. Clark, C.W., Marler, P. & Beeman, K. Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology* **76**, 101–115 (1987).
14. Nottebohm, F., Stokes, T.M. & Leonard, C.M. Central control of song in the canary, *Serinus canarius. J. Comp. Neurol.* **165**, 457–486 (1976).
15. Gentner, T.Q., Hulse, S.H., Bentley, G.E. & Ball, G.F. 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 (2000).
16. Brenowitz, E.A. Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science* **251**, 303–305 (1991).
17. Yu, A.C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871–1875 (1996).
18. Bottjer, S.W., Miesner, E.A. & Arnold, A.P. Forebrain lesions disrupt development, but not maintenance of song in passerine birds. *Science* **224**, 901–903 (1984).
19. Hahnloser, R.H., Kozhevnikov, A.A. & Fee, M.S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
20. Prather, J.F., Peters, S., Nowicki, S. & Mooney, R. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* **451**, 305–310 (2008).
21. Mooney, R., Hoese, W. & Nowicki, S. Auditory representation of the vocal repertoire in a songbird with multiple song types. *Proc. Natl. Acad. Sci. USA* **98**, 12778–12783 (2001).

22. Scharff, C., Nottebohm, F. & Cynx, J. Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. *J. Neurobiol.* **36**, 81–90 (1998).
23. Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192 (2004).
24. Fee, M.S. & Leonardo, A. Miniature motorized microdrive and commutator system for chronic neural recording in small animals. *J. Neurosci. Methods* **112**, 83–94 (2001).
25. Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. & Cooper, F.S. Theoretical notes. Motor theory of speech perception: a reply to Lane's critical review. *Psychol. Rev.* **77**, 234–249 (1970).
26. Wyttenbach, R.A., May, M.L. & Hoy, R.R. Categorical perception of sound frequency by crickets. *Science* **273**, 1542–1544 (1996).
27. Freedman, D.J., Riesenhuber, M., Poggio, T. & Miller, E.K. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316 (2001).
28. Diehl, R.L., Lotto, A.J. & Holt, L.L. Speech perception. *Annu. Rev. Psychol.* **55**, 149–179 (2004).
29. Balaban, E. Cultural and genetic variation in swamp sparrows (*Melospiza georgiana*). *Behaviour* **105**, 250–290 (1988).
30. Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. Speech perception in infants. *Science* **171**, 303–306 (1971).
31. Ballentine, B., Searcy, W.A. & Nowicki, S. Reliable aggressive signaling in swamp sparrows. *Anim. Behav.* **75**, 693–703 (2008).
32. Eifuku, S., De Souza, W.C., Tamura, R., Nishijo, H. & Ono, T. Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *J. Neurophysiol.* **91**, 358–371 (2004).
33. Etcoff, N.L. & Magee, J.J. Categorical perception of facial expressions. *Cognition* **44**, 227–240 (1992).
34. Marler, P. & Tamura, M. Culturally transmitted patterns of vocal behavior in sparrows. *Science* **146**, 1483–1486 (1964).
35. Kuhl, P.K., Tsao, F.M. & Liu, H.M. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA* **100**, 9096–9101 (2003).
36. Freedman, D.J. & Assad, J.A. Experience-dependent representation of visual categories in parietal cortex. *Nature* **443**, 85–88 (2006).
37. Coleman, M.J. & Mooney, R. Synaptic transformations underlying highly selective auditory representations of learned birdsong. *J. Neurosci.* **24**, 7251–7265 (2004).
38. Rosen, M.J. & Mooney, R. Synaptic interactions underlying song-selectivity in the avian nucleus HVC revealed by dual intracellular recordings. *J. Neurophysiol.* **95**, 1158–1175 (2006).
39. Rosen, M.J. & Mooney, R. Inhibitory and excitatory mechanisms underlying auditory responses to learned vocalizations in the songbird nucleus HVC. *Neuron* **39**, 177–194 (2003).
40. Mooney, R. & Prather, J.F. The HVC microcircuit: the synaptic basis for interactions between song motor and vocal plasticity pathways. *J. Neurosci.* **25**, 1952–1964 (2005).
41. Theunissen, F.E. & Doupe, A.J. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *J. Neurosci.* **18**, 3786–3802 (1998).
42. Margoliash, D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* **3**, 1039–1057 (1983).
43. Margoliash, D. & Fortune, E.S. Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *J. Neurosci.* **12**, 4309–4326 (1992).
44. Lewicki, M.S. & Konishi, M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl. Acad. Sci. USA* **92**, 5582–5586 (1995).
45. Mooney, R. Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. *J. Neurosci.* **20**, 5420–5436 (2000).