Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 5582–5586, June 1995
Neurobiology

Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order

MICHAEL S. LEWICKI AND MASAKAZU KONISHI

Computation and Neural Systems Program, Division of Biology 216-76, California Institute of Technology, Pasadena, CA 91125

Contributed by Masakazu Konishi, February 22, 1995

ABSTRACT **Neurons in the songbird forebrain area HVc (hyperstriatum ventrale pars caudale or high vocal center) are sensitive to the temporal structure of the bird's own song and are capable of integrating auditory information over a period of several hundred milliseconds. Extracellular studies have shown that the responses of some HVc neurons depend on the combination and temporal order of syllables from the bird's own song, but little is known about the mechanisms underlying these response properties. To investigate these mechanisms, we recorded intracellular responses to a set of auditory stimuli designed to assess the degree of dependence of the responses on temporal context. This report provides evidence that HVc neurons encode information about temporal structure by using a variety of mechanisms including syllable-specific inhibition, excitatory postsynaptic potentials with a range of different time courses, and burst-firing nonlinearity. The data suggest that the sensitivity of HVc neurons to temporal combinations of syllables results from the interactions of several cells and does not arise in a single step from afferent inputs alone.**

Temporal order is an important code in many acoustic signals including speech, music, and animal vocalizations, but little is known about the neural representation of temporal order or its underlying cellular mechanisms. Auditory neurons that are sensitive to temporal order have been found in several species, such as the squirrel monkey (1–3), guinea fowl (4), and cat (5, 6), but the most complex tuning yet discovered is in the songbird.

Neurons in the songbird forebrain nucleus HVc (hyperstriatum ventrale pars caudale or high vocal center) show a preference for the bird's own (autogenous) song over conspecific songs and are sensitive to manipulations affecting the spectral and temporal structure of the song (7–9). These HVc cells can integrate auditory information over hundreds of milliseconds (8, 10). Studies of these "song-specific" neurons have shown that many of them have responses that require the normal sequence of two or three song syllables (10).

Some properties of HVc cells are illustrated in Fig. 1. HVc auditory neurons typically show a strong preference for the bird's own song (Fig. 1*a Left*). The cell's sensitivity to the temporal context can be investigated by manipulating the temporal structure of the song. For example, playing the song backward completely alters the temporal context but preserves the song's spectral structure. This manipulation typically abolishes the response (Fig. 1*a Center*), which indicates that the response cannot be predicted from the spectral characteristics of the song alone but also depends on the temporal pattern. One way to estimate the extent of this dependence is to present the syllables of the song in reverse order (Fig. 1*a Right*). This manipulation preserves the local context but alters the global context. Like backwards song, the reverse-order song also does not evoke a response, indicating that the influence of the temporal context extends across syllable boundaries and that the cell is sensitive to the order of the syllables.

HVc neurons are often driven by a pair of syllables even when they fail to respond to either syllable in isolation. This illustrates another level of auditory context sensitivity called temporal combination sensitivity (8). The response of temporal-combination-sensitive (TCS) cells depends on a combination of syllables from autogenous song presented in a specific order (usually the natural order). Some manipulations to test the properties of TCS cells are shown in Fig. 1*b*. The cell is sensitive to the combination AB, since the response to the pair is much greater than the sum of the responses to A and B in isolation [32 ± 15 vs. 13 ± 11 spikes per sec (mean ± SD); $P = 0.0139$; paired t test]. The cell is also sensitive to the order of the syllables, since it responds to AB but not to BA. The response to AB cannot be explained by a simple facilitation from A, since repeated presentations of the same syllable do not evoke a response. An additional property of TCS neurons not shown here is their ability to respond to the same syllable pair over intersyllable intervals ranging from tens to hundreds of milliseconds (8).

Though the response characteristics of HVc cells have been studied for some time, the mechanisms that give rise to these properties are not known. To test several possible synaptic models that can account for the properties of context-sensitive cells, intracellular recordings were made *in vivo* from HVc neurons. This report provides evidence that the mechanisms used to create temporal combination sensitivity include syllable-specific inhibition and bursting nonlinearity. These results suggest that the response properties of HVc neurons are an emergent property of a network of cells and do not appear to arise in a single step from afferent inputs alone.

MATERIALS AND METHODS

Experiments were performed on adult (older than 120 days) male zebra finches (*Taeniopygia guttata*) raised in our own colony. Before each experiment the bird's own song was recorded, digitized, and analyzed on computer. A few days before the experiment, birds were anesthetized with Equithesin (0.03–0.04 ml i.m.; 0.85 g of chloral hydrate/0.21 g of pentobarbital/0.42 g of $MgSO_4$/2.2 ml of 100% ethanol/8.6 ml of propylene glycol, brought to a total volume of 20 ml with water), and a small metal post that immobilized the head during physiological recordings was cemented to the skull with dental cement. One or 2 days later, the birds were anesthetized with urethane (65–90 μl of a 20% solution) for physiological recordings.

The HVc was first located physiologically with extracellular glass electrodes. Electrodes were lowered through a small (<0.3 mm) hole in the skull to minimize brain edema and pulsation. Intracellular recordings were obtained with sharp electrodes [60–100 MΩ, filled with 4 M potassium acetate (pH

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HVc, hyperstriatum ventrale pars caudale or high vocal center; TCS, temporal combination sensitive; IPSP, inhibitory postsynaptic potential; EPSP, excitatory postsynaptic potential.

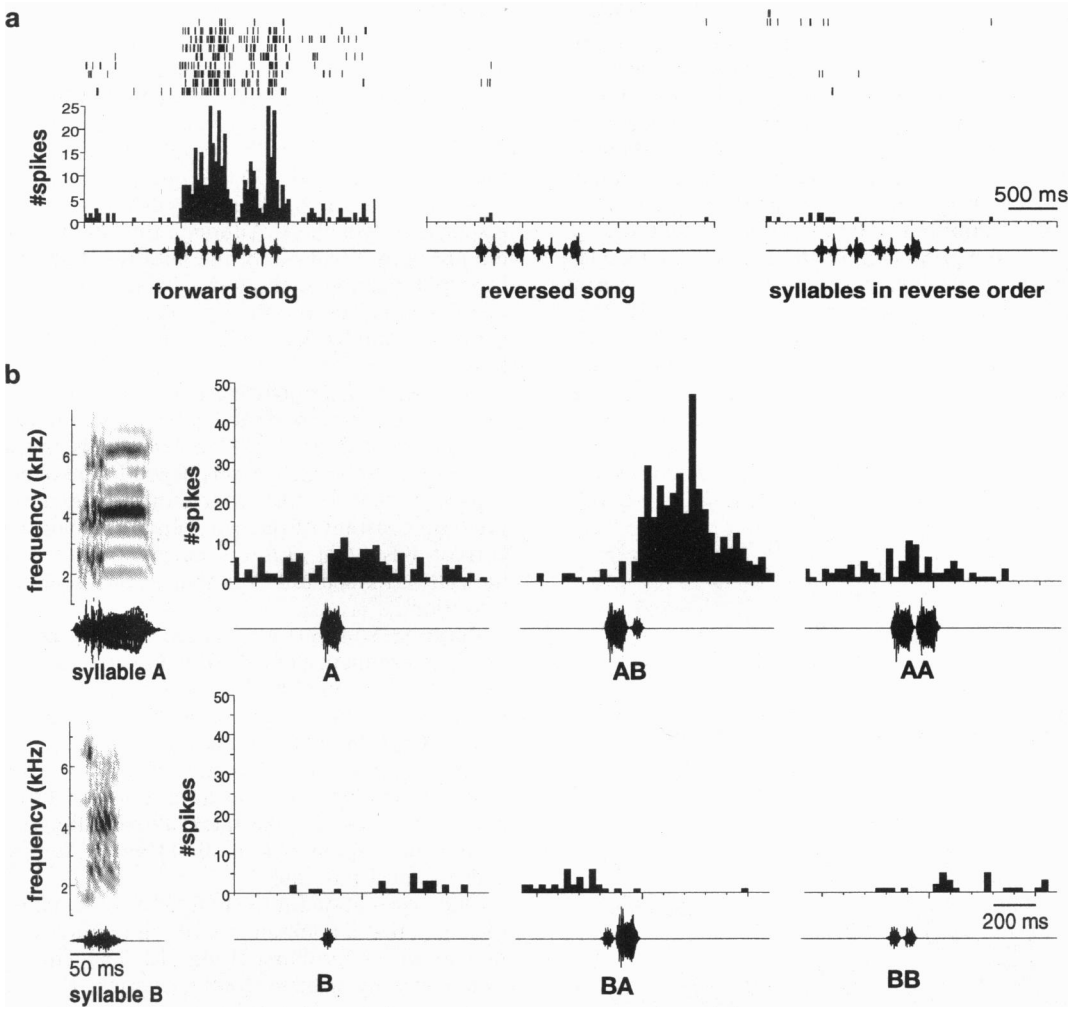

FIG. 1. (*a*) Peristimulus time histograms of the extracellular response recorded from a well-isolated cell in the HVc. The oscillograms of the stimuli are shown below each histogram. The strong response to the bird's own (autogenous) song (*Left*) is completely abolished when the song is played backward (*Center*). This manipulation preserves the spectral structure of the song, but completely alters its temporal structure. The cell also fails to respond when the order of the song syllables is reversed (*Right*), but each syllable still appears as it does in the forward song. This manipulation preserves the local temporal structure within each syllable but alters the global temporal structure of the whole song. These data indicate that the cell is sensitive not only to the spectral profile of a song syllable but also to the auditory temporal context. (*b*) Temporal combination sensitivity is illustrated here by the extracellular responses of the same HVc cell to syllables from the bird's own song. The sonograms and corresponding oscillograms of the two syllables are shown on the left. The oscillogram and syllable labels are plotted below each peristimulus time histogram. The data are taken from 10 interleaved presentations. The cell is combination sensitive because the response to the syllable pair AB is greater than the sum of the response to A and B alone. The cell is also sensitive to the temporal order of the stimuli, since it shows no response to the pair BA. The response is also not simple facilitation, because there is also no response to AA or to BB.

7.4)] or whole-cell patch electrodes (6–12 MΩ, filled with 140 mM potassium gluconate/10 mM Hepes/4 mM $MgCl_2$/0.1 mM $CaCl_2$/1.1 mM EGTA/3 mM Na_2ATP/2 mM NaGTP, pH 7.4, and adjusted to 300–330 milliosmolar). Both intracellular and patch electrodes were pulled on a Flaming–Brown model P-87 micropipette puller (Sutter Instruments, Navato, CA). In some experiments, 1.75% biocytin was added to the pipette solution to stain the cells. Intracellular potentials were amplified with an Axoclamp 2A amplifier (Axon Instruments, Foster City, CA), filtered at 10 kHz, and digitized at a sampling rate of 32 kHz for computer analysis.

The stimuli (autogenous song and its manipulations, white noise, and pure tones) were presented in free field conditions with a calibrated speaker (JBL, Northridge, CA) in a sound attenuation chamber (Industrial Acoustics, Bronx, NY). The peak amplitude of the stimuli was between 60 and 70 decibels sound pressure level.

The anatomical location of the recordings was determined from electrolytic lesions made by extracellular tungsten electrodes (AM Systems, Everett, WA) and by filling the single neurons with biocytin. At the end of the experiment the bird was perfused with saline followed by 4% (wt/vol) paraformaldehyde for histological analysis. Electrolytic lesions were located on 30-μm frozen sections stained with cresyl violet.

RESULTS

Stable intracellular recordings were obtained from 97 cells. The mean holding time was 16 min. The mean initial resting potential was -61 ± 9 mV, and the mean action potential height was 58 ± 13 mV. Cells not exhibiting a stable resting potential for >2 min were omitted from the analysis. Twenty-nine cells showed some auditory response, and 6 of these cells were classified as song-specific cells. A song-specific cell was defined as a neuron that produced significantly more action potentials during forward song than during either reversed song or the song syllables presented in reverse order. Also counted as song-specific cells were cells that showed no significant difference in terms of the number of action potentials but did show a significantly different number of action

potential bursts. A burst was defined as a sequence of at least two action potentials in a period of 30 ms with no adjacent action potentials separated by >6 ms. Although song-specific cells were relatively rare, their frequency in this intracellular study is consistent with that reported in previous extracellular studies in the zebra finch HVc (10). Song-specific cells showed no apparent differences from other cells in terms of their resting potential, action potential shape, or holding times.

Syllable-Specific Inhibition. Inhibition plays a central role for some models of temporal combination sensitivity (8, 11), thus it is important to establish whether specific stimuli can differentially evoke inhibitory postsynaptic potentials (IPSPs). An example of syllable-specific inhibition from extracellular records is shown in Fig. 2*a*. When syllable B was presented alone, it appeared not to affect the cell. When syllable A was preceded by B, however, the response normally evoked by A was completely abolished.

Intracellular evidence for syllable-specific inhibition is shown in Fig. 2 *b* and *c*. To rule out the possibility that the hyperpolarization during the middle of the song (Fig. 2*b*) resulted from the action potentials during the prior excitation, a current of −200 pA was injected into the cell during the song to hyperpolarize it to −85 mV. This manipulation prevented action potentials and also reversed the hyperpolarization (Fig. 2*c*). The reversal of the hyperpolarization near resting potential is consistent with GABAergic Cl^--mediated IPSPs.

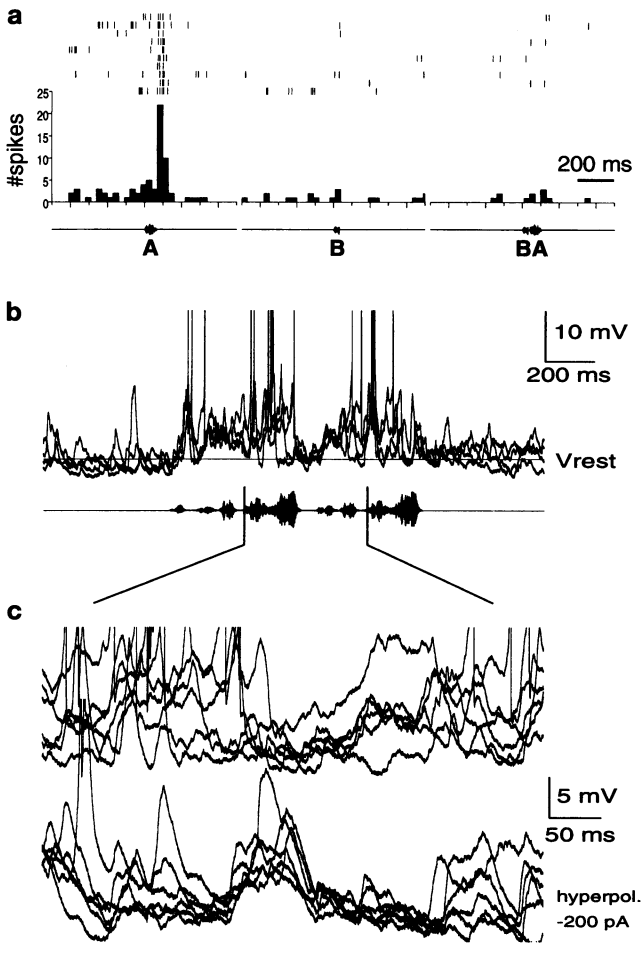

FIG. 2. (*a*) Data were obtained from a single HVc cell with song syllables A, B, and BA interleaved. Action potential rasters from individual trials are shown above the peristimulus time histograms. Inhibition specific to syllable B completely abolished the response normally evoked by A. (*b*) The graph shows overlaid intracellular recordings of a different HVc neuron in response to song. The straight horizontal line indicates the cell's resting potential, which was −70 mV. Note the general depolarization throughout the song except in the middle where there is hyperpolarization. (*c*) Upper traces show the response at resting potential. Lower traces show the response while the cell was hyperpolarized to −85 mV with a −200-pA current injection (during the stimulus). During the current injection, the hyperpolarization in the upper traces was reversed, suggesting it is a GABAergic IPSP. Lower traces are offset by −5 mV to better separate them from the upper traces.

Syllable-specific inhibition is also evident (see Fig. 4*a*) in response to syllable A. Although the holding time for this cell did not permit reversal of the putative IPSP, it is unlikely the hyperpolarization seen after syllable A was due to the weak depolarization during the syllable, since a similar amount of depolarization by syllable B did not evoke any hyperpolarization.

Long-Lasting Depolarizations. Typical depolarizations in response to stimuli persisted beyond stimulus offset by 50–100 ms, but we did observe depolarizations lasting more than several hundred milliseconds beyond the stimulus (Fig. 3). It is possible that the time course of the depolarization reflects the time constant of the cell membrane, but the responses to current pulses (Fig. 3 *Inset*) indicate that the time constant of the cell was much less than that of the stimulus-driven depolarization.

Bursting. Some HVc cells fired three to six action potentials in a high-frequency burst. Bursting occurs most frequently to forward song, and the bursts are often synchronized with particular syllables in the song. Burst firing was also seen in some TCS cells. One example is shown in Fig. 4*a*. The cell produced a burst of action potentials after every presentation of the syllable pair AB, but never burst in response to A or B alone. In terms of spike rates, all stimuli except AA show a significant response ($P < 0.01$) when compared with the background firing rate.

One explanation for the temporal combination sensitivity in Fig. 4*a* is that a combination of inhibition followed by excitation produces the burst firing (12, 13). This hypothesis was tested directly with current injections. First a depolarizing current level was found that produces regular spiking. Then, prior to the depolarizing pulse, a series of hyperpolarizing current pulses was injected into the cell to see whether the firing pattern was altered. The cell in Fig. 4*a* was given a series of hyperpolarizing current injections ranging from −100 pA to

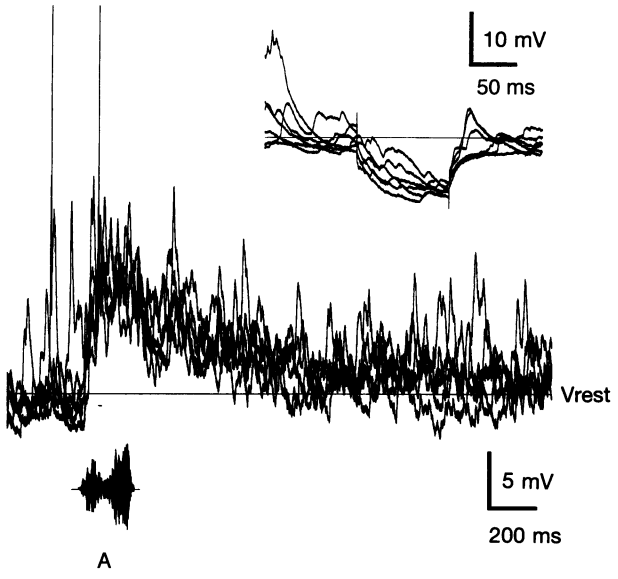

FIG. 3. Intracellular response of a neuron to a single syllable in the bird's own song. (*Inset*) Membrane response to −100-pA current pulses made prior to each stimulus presentation (note different time scales). The action potentials are clipped at −40 mV. The horizontal line indicates the resting potential of the cell, which was −75 mV.

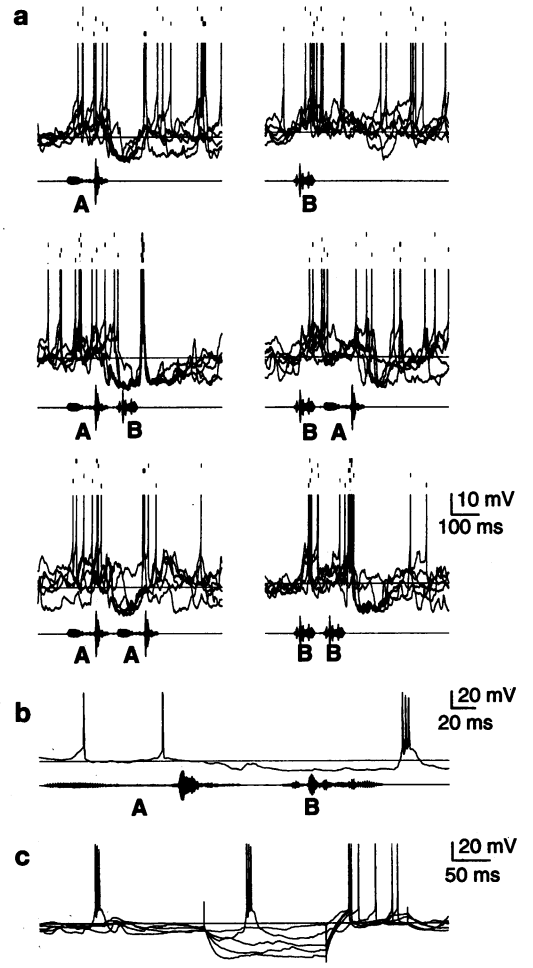

FIG. 4. (*a*) *In vivo* intracellular recording of a TCS HVc neuron. Each panel shows the spike raster of six trials (*Top*), the overlaid intracellular traces (*Middle*), and the oscillogram of the stimulus (*Bottom*). Action potentials are clipped to −35 mV. The horizontal line indicates the resting potential of the cell, which was −69 mV. All trials were interleaved. Song syllable A in isolation evokes a weak excitation followed by inhibition. Syllable B evokes only a weak excitation. The syllable pair AB, however, produces a much stronger response than either A or B. In addition to weak excitation followed by inhibition, AB produces a burst of action potentials during the hyperpolarization in six out of six trials. Reversing the order of the pair (BA) results in only weak excitation followed by inhibition. (*b*) Example of the burst of action potentials produced after every presentation of the syllable pair AB. (*c*) Hyperpolarization followed by suprathreshold depolarization does not elicit bursting even though spontaneous bursts continue to occur.

a maximum of −800 pA with a duration ranging from 150 ms to 200 ms. Each hyperpolarizing current injection was followed by a depolarizing current injection ranging from 100 to 400 pA with a duration of 100 ms. In none of these tests was it possible to elicit burst firing. One example is shown in (Fig. 4*b*). It was also evident that the mechanisms underlying the bursting were still intact, since the cell continued to burst spontaneously. Similar tests were performed on other HVc cells (n = 11) that showed burst firing, but in no case was it possible to elicit bursting with hyperpolarization followed by suprathreshold depolarization. One possible explanation for this is that the mechanisms underlying burst firing are located in the distal parts of the dendritic tree, and space clamp limitations preclude control of those mechanisms from the recording site in the soma.

The current injection results are consistent with those of Kubota and Saito (14) who reported that the burst firing seen in the HVc is of the high-threshold type by showing that bursts could be evoked by strong current injection but did not require prior hyperpolarization. The auditory responses of the cell in Fig. 4*b*, however, are inconsistent with high-threshold bursting, because when the cell bursts in response to the syllable pair AB, the amount of depolarization resulting from syllable B when it is preceded by A (i.e., hyperpolarized) should be less than when syllable B is presented alone. Thus, if the cell bursts in response to AB, it should also burst in response to B. Fig. 4*a*, however, shows no sign of bursting in response to syllable B. One possible explanation is that this results from network interactions, for example, if the additional excitation required to elicit a burst is suppressed when syllable B is presented alone but is present when B is preceded by A. One way the excitation could be suppressed is if the recorded cell is inhibited by a cell that responds to syllable B but is also inhibited by syllable A.

DISCUSSION

Temporal combination sensitivity can be described by two basic components: order sensitivity and combination sensitivity. Order sensitivity requires a mechanism for context preservation by which the first syllable can affect the response to a subsequent syllable. Combination sensitivity requires a nonlinear response mechanism for the neuron to be activated by the syllable pair but not by either syllable in isolation or in repetition. The results presented here are consistent with the idea that both excitatory and inhibitory currents subserve the preservation of context. A nonlinear response could be generated by the spiking threshold or by burst firing.

Margoliash (8) proposed a model for temporal combination sensitivity that used the superposition of rebound caused by inhibition from the first syllable and excitation from the second syllable. This model predicts that, under some conditions, there should also be a response to the syllable pair BB, but often, as in Fig. 1, successive presentations of the second (depolarizing) syllable in a combination produce no response. The conditions under which a response to BB would be expected indicate some of the complexity that is possible with these simple models. One condition is that the time course of the response to syllable B must be long relative to the syllable duration for the two currents to add. A second condition is that the depolarization in response to B be large enough for the syllable pair BB to produce a response. Some of the response properties to individual syllables can be deduced from extracellular recordings. Unambiguous information, however, is difficult to obtain without recording intracellularly.

Although in our intracellular recordings we have observed rebound from inhibition in a cell ventral to the HVc (data not shown), thus far, we have not observed a case in which rebound has played a direct role in temporal combination sensitivity. Also, there were no obvious examples of rebound in the cells that showed a response to song. This type of mechanism does exist in the bat inferior colliculus (15) where the coincidence of rebound from inhibition and delayed excitation has been shown to underly a neural sensitivity for sound duration.

Intracellular recordings of HVc cells obtained *in vitro* have thus far shown no evidence of rebound from inhibition (14). This does not rule out the possibility, however, that inhibitory rebound occurs prior to the HVc and, therefore, underlies some form of context sensitivity.

The results presented here indicate that temporal combination sensitivity arises from the interaction of syllable-specific excitatory postsynaptic potentials (EPSPs) and IPSPs, possibly of different time courses. Studies in the cat visual system have demonstrated that the linear summation of excitation and inhibition with a threshold nonlinearity can account for the direction selectivity of simple cells (16). *In vitro* experiments in the songbird have demonstrated the presence of NMDA (*N*-methyl-D-aspartate), AMPA (α-amino-3-hydroxy-5-methyl-

4-isoxazole propionic acid), and $GABA_A$ (type A γ-aminobutyric acid) receptor potentials in the HVc (17). One form of temporal combination sensitivity can be obtained if the first syllable activates a long-duration EPSP and the second syllable activates a short-duration EPSP, but we have thus far observed no clear example of differential activation of short- vs. long-duration synaptic currents. Furthermore, such a model predicts a response to the syllable pair AA, which is inconsistent with some of the data.

One hypothesis that is consistent with these observations is that temporal combination sensitivity arises from the interaction of several cells, in contrast to arising from the convergence of monosynaptic inputs from afferent cells that are selective for particular syllables. While the latter possibility cannot be ruled out, and the existing mechanisms could accommodate such a circuit, the present data are perhaps best explained by an interactive network model. An example of such a circuit is shown in Fig. 5. Syllable A evokes no response in the TCS cell, cell AB. If syllable B is presented alone, an IPSP and an EPSP of a similar time course are evoked in cell AB, which cancel and produce no response. If syllable A is presented before syllable B, the subsequent IPSP in cell AB is removed, thus generating a response in cell AB. In contrast, presenting the syllable pair BA fails to generate a response, because the IPSP evoked in cell i by syllable A does not cancel the IPSP already evoked in cell AB by syllable B. A typical song syllable has a duration of 50–100 ms, so a type A γ-aminobutyric acid receptor IPSP elicited by syllable A could easily last long enough to suppress subsequent excitation of cell i by cell B_1. Another way for cell A to suppress a response from cell B_1 is if cell A responds to the offset of syllable A.

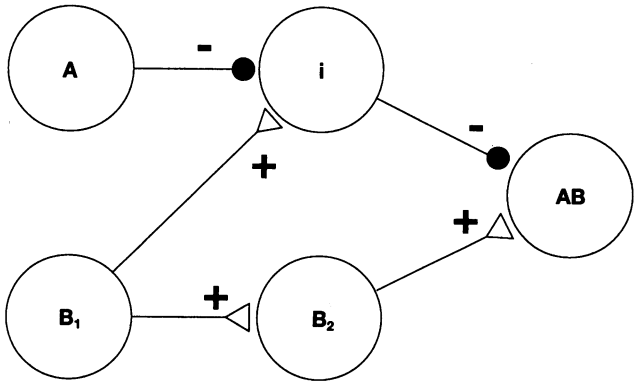

FIG. 5. Observed synaptic mechanisms can be used in a neural circuit to obtain temporal combination sensitivity. This network was constructed to model the data in Fig. 1*b*. Solid circles represent inhibitory inputs. The triangles represent excitatory inputs with a time course roughly equal to the inhibitory inputs.

If the response properties of song-selective cells and TCS cells are indeed constructed from simpler building blocks, we would expect to see cells in the HVc that have simpler types of response. In fact, such a cell was shown in Fig. 2*a*. This cell had a strong response to syllable A but was inhibited by syllable B, which is identical to the response of the model cell i in Fig. 5 but with the roles of cell A and cell B_1 reversed. Many of the HVc cells reported in ref. 10 also had simpler response properties and often exhibited a preference for particular syllables.

All of the mechanisms described here could be combined in various ways to yield a wide variety of sensitivities to temporal context. It is likely that many of the responses observed cannot be explained simply through temporal combination sensitivity, and the complexity of the tuning reflects a larger-scale network-level code. Although the functional significance of the response properties observed in the HVc of the adult bird is not clear, there is evidence that song-specific neurons arise during the song learning process (18, 19), suggesting that song-specific neurons may play some role in the song learning process which requires auditory feedback for normal song to develop. Understanding the mechanisms underlying the encoding of temporal order may elucidate aspects of the vocal learning process and provide insight into how the songbirds learn and memorize songs from auditory experience.

We thank James Mazer, David Perkel, and Allison Doupe for helpful comments on the manuscript. This work was supported by a National Institutes of Health research training grant and an Engineering Research Center fellowship.

1. Wollberg, Z. & Newman, J. (1972) *Science* **175,** 212–214.
2. Newman, J. & Wollberg, Z. (1973) *Brain Res.* **54,** 287–304.
3. Glass, I. & Wollberg, Z. (1983) *Brain Behav. Evol.* **22,** 13–21.
4. Scheich, H., Langner, G. & Bonke, D. (1979) *J. Comp. Physiol.* **132,** 257–276.
5. Weinberger, N. M. & McKenna, T. M. (1988) *Music Percept.* **5,** 355–390.
6. McKenna, T. M., Weinberger, N. M. & Diamond, D. M. (1989) *Brain Res.* **481,** 142–153.
7. McCasland, J. & Konishi, M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7815–7819.
8. Margoliash, D. (1983) *J. Neurosci.* **3,** 1039–1057.
9. Margoliash, D. (1986) *J. Neurosci.* **6,** 1643–1661.
10. Margoliash, D. & Fortune, E. S. (1992) *J. Neurosci.* **12,** 4309–4326.
11. Lewicki, M. & Doupe, A. (1993) *Soc. Neurosci. Abstr.* **19,** 1016.
12. Jahnsen, H. & Llinas, R. (1984) *J. Physiol. (London)* **349,** 105–226.
13. Steriade, M. & Deschenes, M. (1984) *Brain Res. Rev.* **8,** 1–63.
14. Kubota, M. & Saito, N. (1991) *J. Physiol. (London)* **440,** 131–142.
15. Casseday, J. H., Ehrlich, D. & Covey, E. (1994) *Science* **264,** 847–850.
16. Jagadeesh, B., Wheat, H. S. & Ferster, D. (1993) *Science* **262,** 1901–1904.
17. Vu, E. & Lewicki, M. (1994) *Soc. Neurosci. Abstr.* **20,** 166.
18. Doupe, A. & Konishi, M. (1992) *Soc. Neurosci. Abstr.* **18,** 527.
19. Volman, S. (1993) *J. Neurosci.* **13,** 4737–4747.