Neural Representation of Auditory Temporal Structure

Thesis by

Michael Samuel Lewicki

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted January 22, 1996)

Abstract

Neurons in the songbird forebrain nucleus HVc are highly sensitive to auditory temporal context and have some of the most complex auditory tuning properties yet discovered. HVc is crucial for learning, perceiving, and producing song, thus it is important to understand the neural circuitry and mechanisms that give rise to these remarkable auditory response properties. This thesis investigates these issues experimentally and computationally.

Extracellular studies reported here compare the auditory context sensitivity of neurons in HVc with neurons in the afferent areas of field L. These demonstrate that there is a substantial increase in the auditory temporal context sensitivity from the areas of field L to HVc. Whole-cell recordings of HVc neurons from acute brain slices are described which show that excitatory synaptic transmission between HVc neurons involve the release of glutamate and the activation of both AMPA/kainate and NMDA-type glutamate receptors. Additionally, widespread inhibitory interactions exist between HVc neurons that are mediated by postsynaptic $GABA_A$ receptors. Intracellular recordings of HVc auditory neurons *in vivo* provides evidence that HVc neurons encode information about temporal structure using a variety of cellular and synaptic mechanisms including syllable-specific inhibition, excitatory post-synaptic potentials with a range of different time courses, and burst-firing, and song-specific hyperpolarization.

The final part of this thesis presents two computational approaches for representing and learning temporal structure. The first method utilizes computational elements that are analogous to temporal combination sensitive neurons in HVc. A network of these elements can learn using local information and lateral inhibition. The second method presents a more general framework which allows a network to discover mixtures of temporal features in a continuous stream of input.

Acknowledgements

I thank my advisor Mark Konishi for his steadfast support and guidance. Without his encouragement, and quiet insistence that computationally-minded students learn how real brains work, this thesis would not have been possible. I am also indebted to the people in Mark's lab for teaching me the ever complicated and often arduous experimental techniques required for neurophysiology. Jamie Mazer was especially helpful for both experimental and scientific advice, not to mention his invaluable programming skills. Allison Doupe showed great patience and support while I learned about zebra finches inside and out. I also thank my collaborators, Ben Arthur, for the experiments in chapter 2, and Eric Vu, for those in chapter 3. Jamie Mazer, Marc Schmidt, Rich Jeo, Ben Arthur, David MacKay, Geneviève Sauvé, Tom Annau, and Bruno Olshausen all provided valuable critiques on various parts of this manuscript.

The CNS program at Caltech has been a thoroughly enriching experience and was crucial for providing an environment that fostered and encouraged the free pursuit of ideas. Conversations with David MacKay and Bruno Olshausen were especially stimulating and inspiring. I am also grateful to John Hopfield and Gilles Laurent for providing valuable advice and guidance. Finally, I would like to thank Geneviève who has been a constant source of love and support. Her cheerful presence has made these final hectic months of graduate student life ever more enjoyable.

Contents

Chapter 1 Introduction

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. One of the reasons these subjects are difficult to study is that, unlike vision, very few animals have auditory perceptual skills that are comparable to our own.

Songbirds are one of the few non-human animals that demonstrate the capacity to produce and recognize complex acoustic sequences. The complexity of birdsong varies tremendously across species. Vocal repertoires range anywhere from a single song like in the zebra finch, the species studied in this thesis, to hundreds in birds like the marsh wren (Canady et al., 1984). Each song is composed of a sequence of acoustic segments called syllables; the type, number, and order of syllables determine the differences among the songs. The song repertoire is acquired early in development when the young bird hasn't begun to sing and when it is exposed to singing adults. After a set of songs have been encoded, the birds can learn to sing these songs entirely from memory (Konishi, 1965; Marler and Peters, 1981; Price, 1979; Bohner, 1983). This remarkable behavior makes the songbird auditory system attractive for investigating the neural representation and learning of complex acoustic sequences.

1.1 Song Learning

There are two stages to the song learning process. The first stage is called the sensory stage. In this stage, the young birds listen the songs they hear in their environment and store them in memory. Usually they hear the songs of their father, but they can also learn songs of other birds in the area. In zebra finches, the sensory period lasts about 15 days, from approximately 20-35 days post-hatch (Bohner, 1990). Birds in the sensory stage make no song-like vocalizations, so they cannot practice the songs they hear. Thus, the memory stored by birds is based entirely on auditory experience. The memory of the songs is called the auditory template (Konishi, 1965), because the songs the young birds eventually learn

to produce are matched to this memory. The song template is critical for providing models of normal adult song, since birds that are raised in isolation produce very abnormal song (Konishi, 1965; Marler and Peters, 1981). The memory of the songs is so good that the birds do not need any further tutoring once past the sensory stage. Song birds can learn to produce the songs entirely from memory. The encoding of the memory can be extremely efficient. In one experiment, a nightingale was tutored with 21 different songs heard 10 times each over a period of 5 days (Hultsch and Todt, 1989). This bird learned to produce 19 of the 21 songs with no further tutoring.

In the second stage of song learning, the sensorimotor stage, the birds begin to sing. At first, they produce very soft vocalizations that sound nothing like fully developed bird song, much like the babbling of babies that are learning to speak. Over the course of weeks of practice, the young birds get better at producing the sounds of adult birds, and the acoustic structure of the songs is gradually refined until it matches the songs memorized in the sensory stage. After this, the songs produced by the bird are crystallized: the songs no longer change and are very stereotyped. In zebra finches, songs are fully crystallized by 80-100 days (Price, 1979; Bohner, 1983) which is when the birds reach adulthood and have fully matured. During the sensorimotor stage, auditory feedback is critical for normal song development. The birds must be able to hear their own vocalizations in order to match the memorized songs. Birds that are deafened during the sensorimotor stage do not develop normal adult song (Konishi, 1965).

An example of zebra finch song learning is shown in figure 1.1. The acoustic structure of zebra finch song can be seen in the sonogram which plots frequency versus time. Typical zebra finch song syllables have a complicated acoustic structure. They can contain tones, harmonics, and noise patterns that can all have amplitude and frequency modulations. A typical zebra finch song contains 5-10 syllables which are delineated by regions of zero amplitude. The song on the left is sung by the father and was the only song heard by the young offspring in the father's nest. The song on the right was learned by one of the sons. Both the acoustic structure of many of the syllables and their order were learned. Typically, songbirds do not make exact song copies and often modify some of the syllables or improvise new ones.

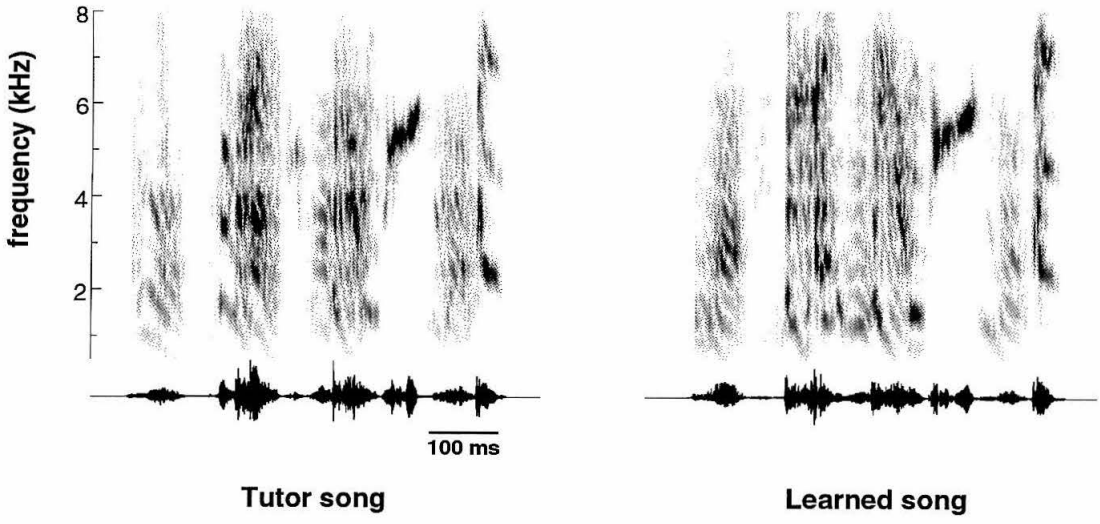


Figure 1.1: Two examples of zebra finch songs. The top graph in each panel shows the sonogram of the song (frequency vs time). The lower graph shows the song oscillograph (amplitude vs time). Birdsong is composed of a sequence syllables which are separated by regions where the song has zero amplitude. Each syllable has a characteristic spectro-temporal pattern which can be seen in the sonogram. The song on the left is the tutor song which in this case is the song of the father. The song on the right was learned by the son.

1.2 Brain Areas of the Song System

A remarkable fact of the songbirds is that the brain areas subserving song learning and song production is composed of a discrete set of nuclei that are easily visible in stained brain sections (Nottebohm et al., 1976). These brain areas are collectively called the song system and are shown in figure 1.2. The song system can be divided into two parts: auditory and motor. Auditory input arrives in the songsystem via the forebrain areas called L1, L2, and L3 (Kelley and Nottebohm, 1979; Fortune and Margoliash, 1995). L2 receives auditory input from the thalamus and is analogous to primary auditory cortex in mammals. The auditory areas of the song system are crucial during song learning, since normal song development depends on auditory feedback. Remarkably, however, the auditory areas are not required for song production after the songs have crystallized, since deafened birds produce normal song (Konishi, 1965). Lesions of either area X or L-MAN does not affect song production in adults (Nottebohm et al., 1976), but lesions of the same areas in young birds during the sensorimotor period does prevent the development of normal song production (Bottjer et al., 1984; Scharff and Nottebohm, 1991).

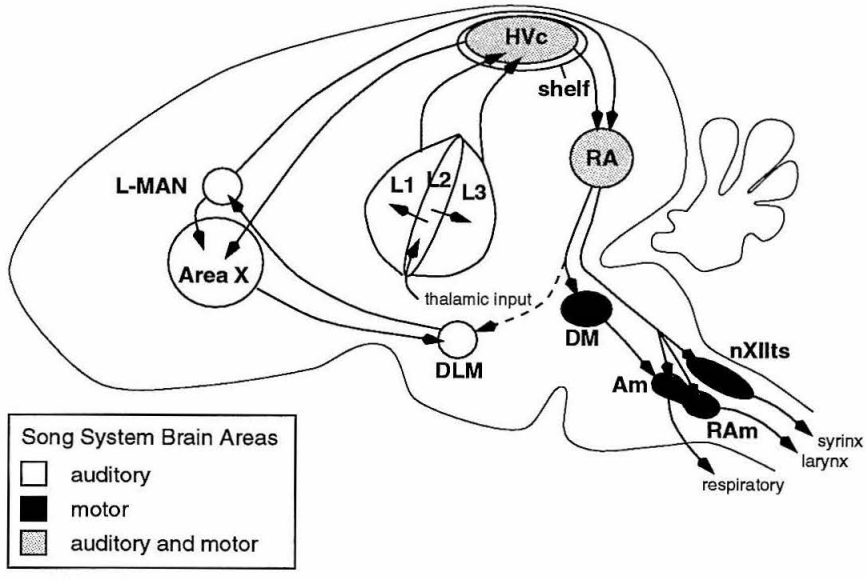


Figure 1.2: A simplified diagram of the song system.

The song system area HVc[1] is at the top of the descending motor pathway controlling song production. (Nottebohm et al., 1976; McCasland, 1987; Vu et al., 1994). HVc projects to RA which in turn innervates the brainstem motor neurons that drive the vocal and respiratory muscles used to produce song.

1.3 Auditory Response Properties of HVc Neurons

Auditory neurons that are sensitive to temporal order have been found in several species, such as the squirrel monkey (Wollberg and Newman, 1972; Newman and Wollberg, 1973; Glass and Wollberg, 1983), guinea fowl (Scheich et al., 1979), and cat (Weinberger and McKenna, 1988; McKenna et al., 1989), but the most complex tuning yet discovered is in the songbird.

Auditory neurons in the songbird forebrain nucleus HVc show a preference for the bird's own song over other songs of its own species of those of other species. These neurons are also

[1] Abbreviations: HVc, *hyperstriatum ventrale pars caudale* also called high vocal center; DLM medial portion of the dorsolateral nucleus of the thalamus; L-MAN, lateral portion of the magnocellular nucleus of the anterior neostriatum; RA, robust nucleus of the archistriatum; nXIIts, syringeal portion of the hypoglossal nucleus; Am, nucleus ambiguous; RAm, nucleus retroambigualis

sensitive to manipulations that affect the song's spectral and temporal structure (McCasland and Konishi, 1981; Margoliash, 1983, 1986) and can integrate auditory information over hundreds of milliseconds (Margoliash, 1983; Margoliash and Fortune, 1992). Studies of these so called "song-specific" neurons have shown that many of them have responses that require the normal sequence of two or three song syllables (Margoliash and Fortune, 1992).

An example of a song-specific HVc is shown in figure 1.3. This cell shows a typical

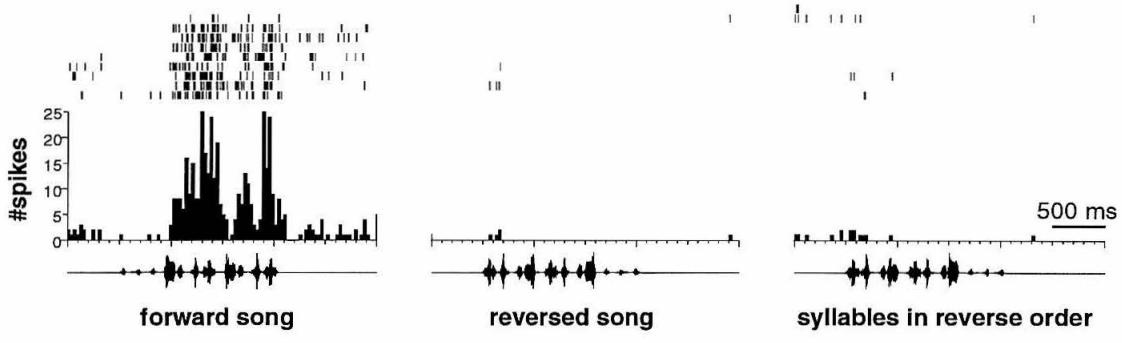


Figure 1.3: The graphs show peristimulus time histograms of the extracellular response recorded from a well-isolated cell in the HVc. The oscillographs of the stimuli are shown below each histogram. The strong response to the bird's own (autogenous) song (left) is completely abolished when the song is played backward (middle). 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 just to the spectral profile of a song syllable but also to the auditory temporal context.

strong response to the bird's own song (figure 1.3, left panel). 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 (figure 1.3, middle panel), 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 (figure 1.3, right panel). 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.

Temporal combination sensitivity

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 (Margoliash, 1983). 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 that test the properties of TCS cells are shown in figure 1.4. 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 (SD) vs 13 ± 11 spikes/sec, $p = 0.0139$, 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 when A and B are separated by intervals ranging from tens to hundreds of milliseconds (Margoliash, 1983).

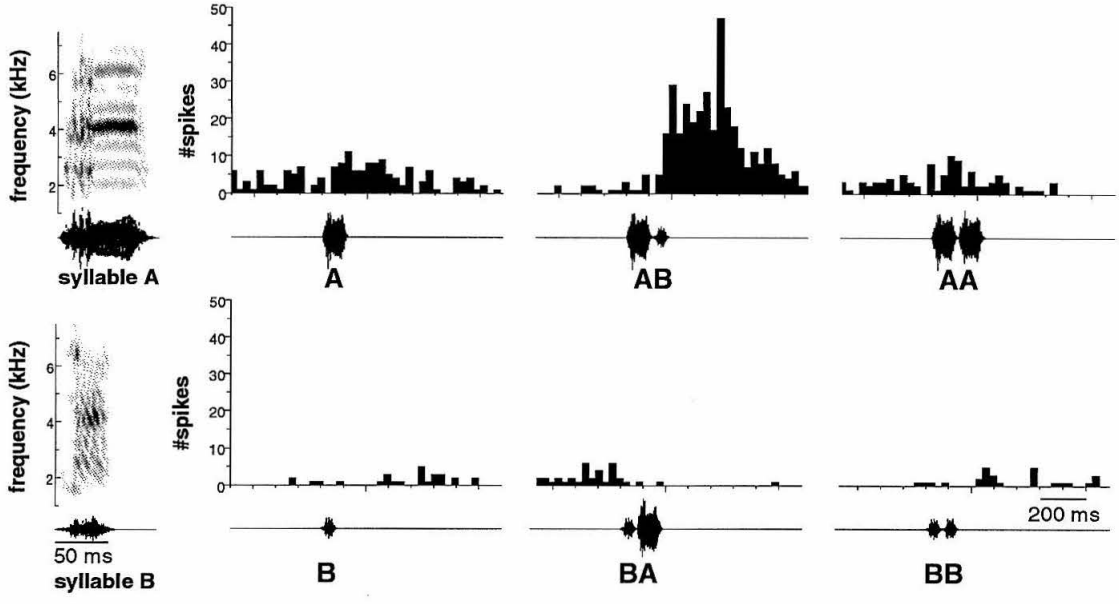


Figure 1.4: Temporal combination sensitivity is illustrated here by the extracellular responses of the HVc cell in figure 1.3 to syllables from the bird's own song. The sonograms and corresponding oscillographs of the two syllables are shown on the left. The oscillograph and syllable labels are plotted below each peristimulus time histogram. The data are taken from ten 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.

Chapter 2 Hierarchical Organization of Auditory Temporal Context Sensivity[1]

2.1 Introduction

The songbird forebrain nucleus HVc (*hyperstriatum ventrale pars caudale*, also called high vocal center) contains auditory neurons that have a variety of complex response properties. Some neurons are highly selective for the bird's own song (hereafter referred to as autogenous song). These neurons respond most strongly to autogenous song and less to songs from the same species and little or not at all to songs of other species (Margoliash, 1983; Margoliash and Konishi, 1985; Margoliash, 1986). The response of these "song-specific" neurons is sensitive to both the song's spectral and temporal structures (Margoliash, 1983; Margoliash and Fortune, 1992).

Other HVc neurons also show a strong response to song but require simpler acoustic features to elicit a response. For example, some neurons require only harmonic combinations of pure tones that are similar to the frequencies contained in autogenous song. Other neurons are sensitive to harmonic combinations of frequency modulated tones. Still others are sensitive to the temporal order of sequences of these acoustic features. Neurons sensitive to the temporal order of acoustic features can integrate auditory temporal context over periods as long as several hundred milliseconds.

A plausible explanation for how the response properties of song-specific neurons arise is that they are the result of the integration of neurons with simpler tuning properties, like those described in the previous paragraph. It has not been established whether the neurons with simpler tuning properties also arise in HVc or are already present in the auditory brain areas afferent to HVc.

HVc receives auditory input primarily from two sources. One area, the HVc *shelf*, is a thin (80 μm) band of neurons on the ventral border of HVc (Kelley and Nottebohm, 1979). The dendritic arbors of HVc neurons extend into the shelf region (Katz and Gurney,

[1] The work in this chapter was done in collaboration with Benjamin J. Arthur.

1981; Fortune and Margoliash, 1995) and could be a source of auditory input to HVc. The second source of auditory input to HVc is a group of brain areas collectively referred to as field L. Field L contains three main regions, L1, L2, and L3. L2, which contains two cytoarchitectonically distinct areas L2a and L2b (Fortune and Margoliash, 1992), receives input from the thalamic auditory area nucleus ovoidalis and projects to L1 and L3 (Karten, 1968; Bonke et al., 1979a; Kelley and Nottebohm, 1979). L1 and L3, in turn, project to the shelf and HVc (Kelley and Nottebohm, 1979; Fortune and Margoliash, 1995). This study focuses primarily on the comparison between neuronal response properties observed in field L and HVc.

Neurons in field L show sensitivity to spectral patterns (Leppelsack and Vogt, 1976; Leppelsack, 1978; Scheich et al., 1979; Langner et al., 1981; Scheich, 1983), amplitude and frequency modulation (Bonke et al., 1979b; Leppelsack, 1983; Muller and Leppelsack, 1985; Hose et al., 1987; Knipschild et al., 1992; Heil et al., 1992), and the spectral and temporal patterns of human speech sounds (Langner et al., 1981; Uno et al., 1991). As a population, field L neurons show no preference for the autogenous song over other songs (Margoliash, 1986). These tuning properties can account for some of the response properties of HVc neurons, but it is not known whether field L contains neurons that show the same capacity to integrate long periods of auditory context that is seen in HVc neurons.

The first set of experiments presented here compares auditory context sensitivity in field L and HVc. Previous studies have shown that the response of song-specific neurons to autogenous song is often abolished if the song is played backward. Song reversal disrupts temporal structure while preserving spectral structure. This would affect the response of any neuron sensitive to the immediate temporal context *e.g.* neurons that are sensitive to frequency modulation, which is a common acoustic feature in birdsong. In the present study, the extent of the temporal context sensitivity was estimated by preserving greater amounts of temporal context and comparing this response to the forward song. For example, presenting the syllables of the song in reverse order (syllables are defined here as points in the song that divide periods of non-zero amplitude from those of zero amplitude) preserves the local spectral and temporal features of each song syllable, which are typically 50-100 msecs in duration, but alters the more global auditory context in which the syllable occurs. If the response at a given point in the song does not depend on the auditory context, then changing the syllable order should not change the response. The amount of context required

for a given response can be assessed by reversing the order of segments of different lengths.

The second part of this study addresses the question of whether the auditory context sensitivity observed in song-specific neurons can be accounted for by the response measured for single syllables and syllable pairs presented in isolation. This provides another method for measuring auditory context sensitivity. A cell is said to be combination sensitive if the response to a syllable pair AB is greater than the sum of the responses to syllables A and B presented in isolation (Margoliash, 1983). Comparing responses to AB and to BA determines whether a cell is sensitive to the order of the syllables. Both of these manipulations measure the dependence of a cell upon auditory context. Previous studies have shown that neurons sensitive to the combination and order of autogenous song syllables are present in HVc (Margoliash, 1983; Margoliash and Fortune, 1992), but it is not known if such neurons are present in field L.

The aim of these experiments is to make a systematic comparison between the response properties of field L and HVc neurons that show a significant response to song in order to determine where the neural response properties underlying the context-sensitive properties of song-specific neurons are first computed.

2.2 Methods

Surgery. Experiments were performed on 25 adult (older than 120 days) male zebra finches (*Taeniopygia guttata*) raised in our own colony. A few days before the experiment, birds were anesthetized with Equithesin (0.03–0.04 ml intramuscular injection; 0.85 g chloral hydrate, 0.21 g pentobarbital, 0.42 g $MgSO_4$, 2.2 ml 100% ethanol, 8.6 ml propylene glycol, filled to a total volume of 20 ml with water, all chemicals were purchased from Sigma, St. Louis, MO), and a small metal post, used to immobilize the head during later physiological recordings, was cemented to the skull with dental cement. One or two days later, the birds were anesthetized with urethane (65-90 μl of 20% solution, Sigma) for physiological recordings.

Electrodes were lowered through a hole in the skull. The hole was made small (400 μm dia.) in order to minimize brain edema and pulsation. If neurons were isolated in field L, the next electrode track was made into HVc and vice versa in order to maximize the number of single neurons from field L and HVc in each bird. Extracellular recordings were obtained

with parylene-coated tungsten electrodes with impedances (at 1.0 kHz) ranging from 1-10 MΩ (AM Systems, Evertt, WA).

The anatomical locations of the recording sites were determined from reference marks consisting of two or more electrolytic lesions (-2 to -3 μA twice for ten seconds each) spaced at least 500 μm apart. At the end of the experiment, birds were perfused transcardially with 0.5% saline followed by 4% paraformaldehyde. Thirty micron frozen sections were cut on a microtome, mounted, and stained with cresyl violet for localization of lesions.

Spike analysis. Extracellular waveforms containing action potentials of different shapes were sorted using a new real time software spike discrimination algorithm (Lewicki, 1994, also appendix A of this thesis) which automatically determines the spike shapes in the extracellular waveform and accurately classifies overlapping action potentials. Otherwise, single units were isolated with conventional methods using a level or window discriminator. Spike classes that were not stable throughout the experiment were omitted from the analyses.

Stimuli. Before each experiment the autogenous song was recorded, digitized, and analyzed on a computer (Sparc station IPX, Sun Microsystems, Mountain View, CA). The bird's own song was used as a search stimulus in both field L and HVc. Well-isolated single neurons were selected for further analysis only if they demonstrated a significant response to song. The electrode was advanced at least 150μm between isolated neurons.

Some of the stimuli used in these experiments were constructed by manipulating the order of syllables and sub-syllables in the autogenous song. Syllable boundaries were defined as points where the song's amplitude falls to zero. Sub-syllable boundaries were defined as places where the sonogram of the song indicated an abrupt change in spectral composition. Typically this was a change in the harmonic pattern or a change in the direction of the frequency modulation. An example of these divisions is shown in figure 2.1. An envelope (3 msecs rise-fall) was placed around each syllable and sub-syllable to remove any transients. All stimuli were presented in free field conditions in a sound attenuating chamber (Acoustic Systems, Austin, TX) with a calibrated speaker (JBL, Northridge, CA). The frequency response of the speaker, as measured from the bird's position in the stereotaxic apparatus inside the chamber, was flat to within 8dB between 500 and 8,000 Hz. Stimuli were presented with a peak amplitude between 60 and 70 dB SPL.

An automated procedure was developed to select syllable pairs for which a neuron would be likely to show order and combination sensitivity. This procedure is illustrated in

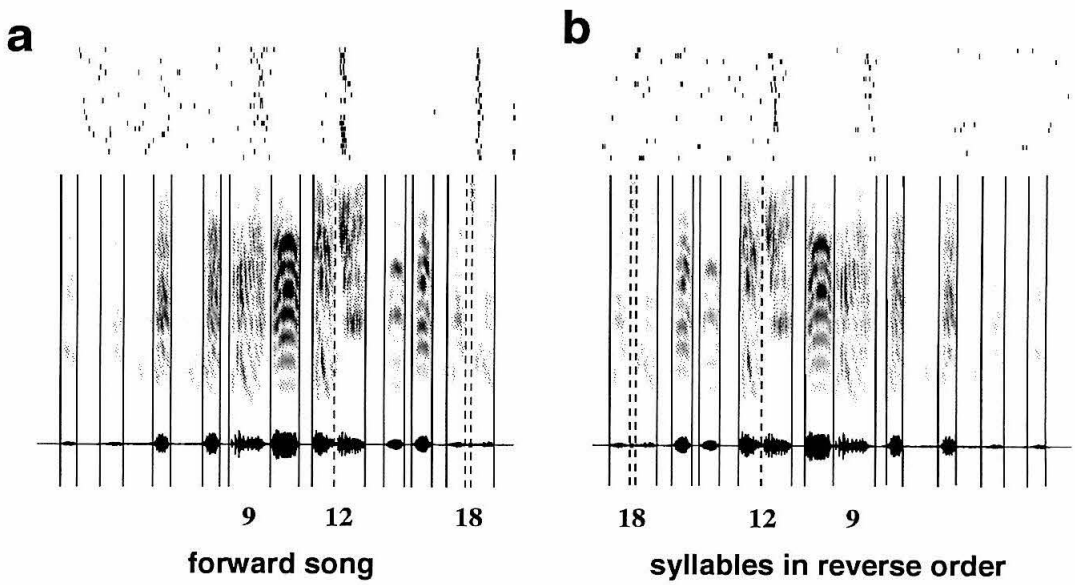
figure 2.1. Syllable pairs were selected by comparing the response of each syllable when presented as part of the forward song to the same syllable as part of a song constructed by playing the syllables in reverse order. The syllables (or sub-syllables) with the greatest statistically different responses (by a paired t-test) were selected to test for temporal combination sensitivity. Syllables at the beginning of the song were not considered, since significant differences can result simply from an onset response. Sometimes more than two syllables were necessary to evoke a response, in which case the set of syllables was divided into two groups and manipulated as above.

Data analysis. The response of a cell to autogenous song and the synthetic songs was measured by the average spike rate during the stimulus presentation minus the spontaneous rate. The variation of the response is reported as plus or minus the standard error of the mean. For shorter stimuli, such as single syllables, the time course of the response of HVc neurons can be highly variable from neuron to neuron, making it difficult to determine exactly when and how much a cell responded. We determined the regions of significant response automatically by calculating where the spike rate differed significantly from background by sliding a 50 msec window from the start of the stimulus (plus latency) to the end of the collection. Because a standard t-test can inaccurately report significant response regions when most or all of the window counts across trials are zero, we determined statistical significance using a Poisson model which takes into account the window size. Excitatory and inhibitory regions were analyzed separately.

The statistical significance of the sensitivity to syllable order was determined by comparing the total spike counts in the significant response regions of syllable pairs AB and BA using a t-test. The significance of the sensitivity to syllable combinations was determined using a t-test to compare the sum of the spike counts in the regions of syllables A and B to the spike counts from the regions of AB.

2.3 Results

We recorded from 52 well-isolated neurons in HVc and 55 neurons in the field L areas that had a significant response to song (shown in figure 2.2). In field L there were 8 well-isolated units in L1, 11 in L2a, 10 in L2b, and 16 in L3, and 11 that bordered two or more field L regions. Neurons that were on the border between field L and non-field L areas, such as



segment#	p val	forward total spikes	reverse order total spikes
12	0.009	58	34
18	0.011	36	19
9	0.072	37	22

Figure 2.1: An illustration of the procedure for selecting syllable pairs for performing the tests for temporal combination sensitivity. The graphs show the response of an HVc cell in response to the forward song (**a**) and to the syllables in reverse order (**b**). The top part of each graph shows the spike rasters. The sonogram and oscillogram of the stimulus are shown below. The solid and dotted vertical lines indicate the syllable and sub-syllable boundaries, respectively. The numbers below the oscillogram indicate the segment numbers referred to in the table. Segments include the silent periods between syllables. The table shows the song segments that had the greatest difference (in terms of the p value of a paired t-test) between the response in the forward song and the response to the song with the syllables in reverse order. Entries with few spikes were omitted. In this example, there were 58 total spikes during segment 18 during the forward song. The same syllable elicited only 34 spikes in the context of the syllable-reversed song. Thus, segments 10-12 would be selected for presentation in isolation to test for temporal combination sensitivity.

caudal neostriatum (NC) and ventral hyperstriatum (HV), were omitted from the analysis. Both phasic and tonic responses were seen in each of the areas. Cells in HVc sometimes responded with bursts of action potentials. Bursting was rarely observed in field L.

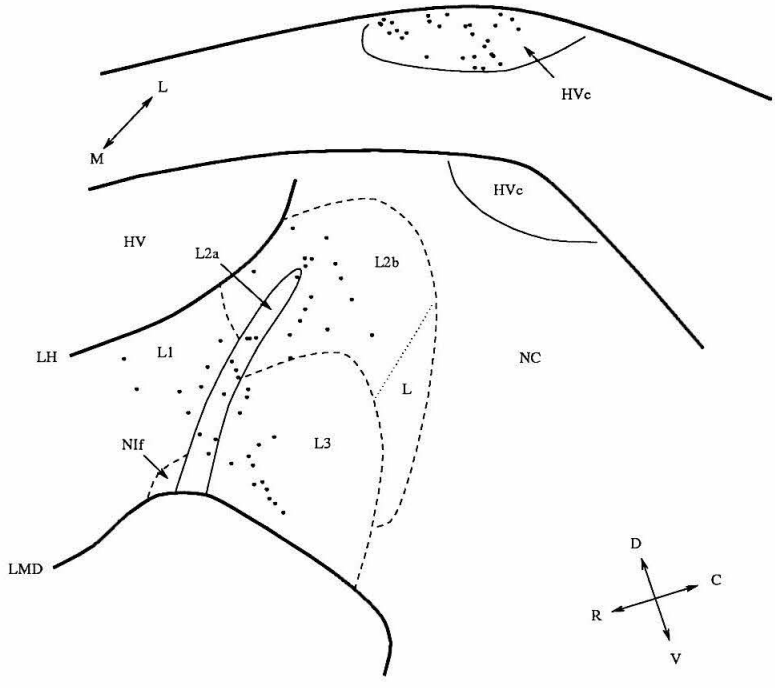


Figure 2.2: The anatomical sites of the neurons analyzed in this study. The top upper diagram shows the sites in HVc, and the lower diagram shows the sites in field L. The average medial-lateral position was 2.3 mm for HVc and 1.5 mm for field L. Area L2a of the field L complex as well as HVc can be clearly identified in cresyl-violet stained sections. The dashed lines indicate the approximate anatomical areas as described by Fortune and Margoliash (1992).

Comparison of Context Sensitivity in HVc and Field L

Sensitivity to auditory temporal context was measured by comparing the response to forward song with the responses to reversed song, sub-syllables in reverse order, and syllables in reverse order. The response of a typical HVc neuron is shown in 2.3a. This particular neuron shows a strong response to forward song (22.78 ± 4.32 spikes/sec), and is slightly inhibited by the reversed song (-2.01 ± 0.09 spikes/sec). The response is also greatly reduced when the order of the sub-syllables or syllables is reversed (-1.30 ± 1.61 and 3.98 ± 2.44 spikes/sec respectively). The differences between the response to forward song and to the

synthetic songs are all statistically significant (p ¡ 0.001, paired t-test). Since the acoustic structure of each syllable or sub-syllable is identical to that in the forward song, this HVc neuron is dependent upon the auditory context which in this case extends beyond a single syllable. Performing this analysis on the population of HVc cells showed that about half of the neurons responded significantly ($p < 0.01$) more to the forward song than to the reverse song (28/52) or to the sub-syllables in reverse or (26/52). About one quarter of HVc neurons responded more strongly to the forward song than to the syllables in reverse order (12/52). About one third (18/52) cells in HVc showed no statistical differences between the response to the forward song and the response to any of the synthetic songs. None of the cells in HVc responded more to any of the three temporally altered songs than to the forward song. This data is summarized in figure 2.4.

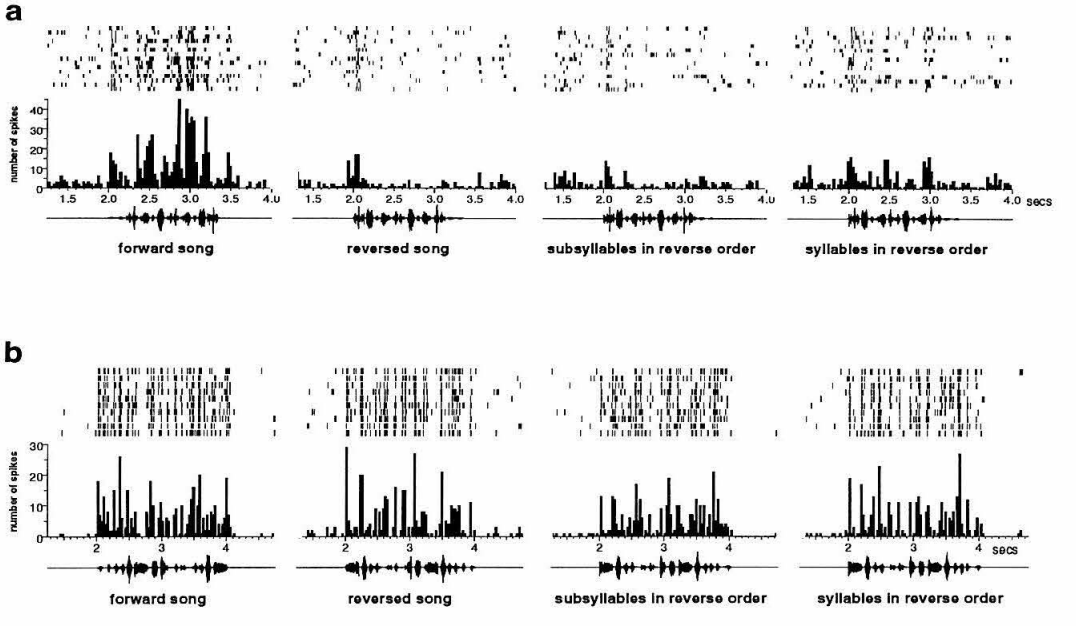


Figure 2.3: The response of representative cells from HVc (**a**) and field L (**b**) to autogenous song and three manipulations of the songs temporal structure. The graphs show the spikes rasters and peristimulus time histograms of the response recorded from two well-isolated units. The oscillograms of the stimuli are shown below each histogram. Cells in HVc showed much greater sensitivity to manipulation of the song's temporal structure than the did cells in field L.

Neurons in field L showed much less sensitivity to manipulations of the auditory temporal context than neurons in HVc. Neurons in all areas of field L responded strongly throughout

the forward song, the reversed song, and to the syllables and sub-syllables in reverse order. The response of a typical field L neuron (in area L3) is shown in figure 2.3b. The differences between the response to forward and the temporally altered songs is much less than in HVc (forward song, 19.07 ± 0.74; reversed song, 16.17 ± 1.18; sub-syllables in reverse order, 15.29 ± 0.76; and syllables in reverse order, 14.10 ± 0.55 spikes/sec). The majority of field L cells (41/56) show no significant difference ($p > 0.01$) between the response to the forward song and the response to any of the three temporally altered songs. These data are summarized in the bar plots in figure 2.4.

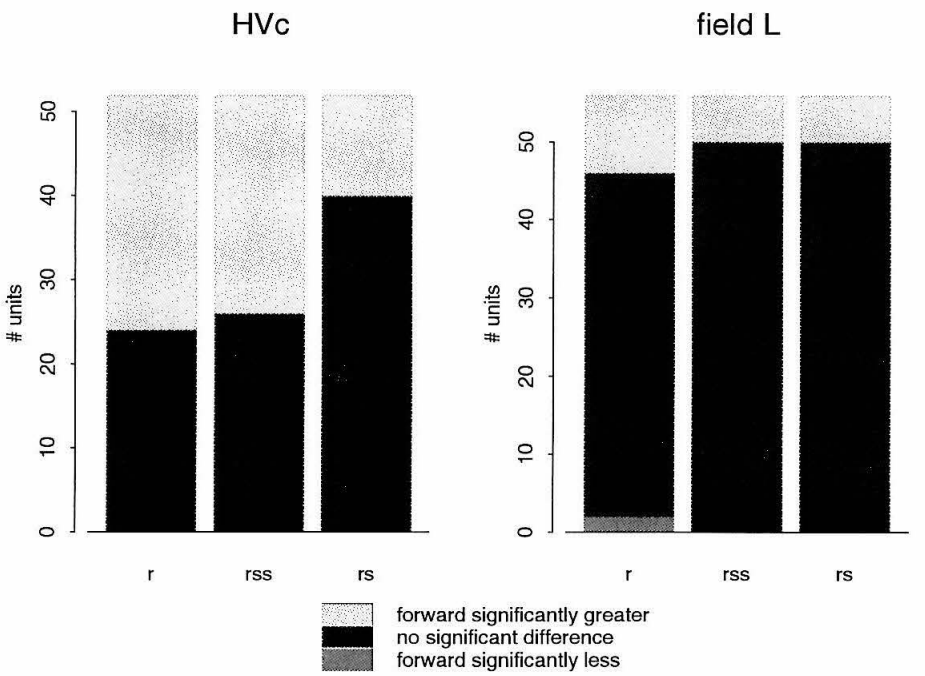


Figure 2.4: The bar graphs show the number of times the response to forward song differed significantly from the response to reversed song (r), to sub-syllables in reverse order (rss), or to syllables in reverse order (rs).

Of the field L subdivisions, only L3 contained neurons that showed a significant difference in response between the forward song and the syllable reversed song (6/16). Both L1 and L3 had neurons that showed a significant difference between the forward song and sub-syllable reversed song (1/8 and 4/16, respectively). All subdivisions of field L had neurons that showed a significant difference between the forward and reversed song (L2a, 3/11; L2b,

2/10; L1, 1/8; L3, 5/16).

Although both field L and HVc contain cells that were significantly dependent on the temporal order, the difference in response rates between the forward song and the altered songs for these cells is much greater in HVc. This difference in the response properties of the two populations can be summarized by plotting the response of the forward song against the response of altered songs. Figure 2.5 shows that responses of the population of field L neurons is largely the same for all four types of stimuli, but the response of many neurons in HVc compared to forward song is reduced or inhibited when the temporal structure of the forward song is altered. Each graph plots the response to forward song against the response to the three temporally altered songs. In HVc, many of the neurons respond much more to the forward song than to the temporally altered songs. This is indicated on the graph in figure 2.5a by the large number of points above the line $y = x$ (solid line). The further a point is above the line, the greater difference in response. If a neuron shows no sensitivity to auditory context, the responses to forward and the temporally altered stimuli should be the same, and the points should fall near the line $y = x$ (solid line). This is indeed the case for the field L data shown in figure 2.5b.

HVc neurons clearly show greater sensitivity to the auditory temporal context than field L neurons. The difference between the HVc and field L responses is statistically significant for forward vs reversed song ($p < 0.001$, unpaired t-test), forward vs sub-syllables in reverse order ($p < 0.001$), and forward vs reverse syllables ($p < 0.05$). A one way analysis of variance indicated no statistical differences ($p > 0.2$, F-test) among any of the field L areas between the forward song and the temporally altered stimuli. Since there are relatively few neurons in each of the field L subdivisions, these tests do not rule out the possibility that more subtle differences among these areas do exist.

Another way to see the difference between HVc and field L response properties, is to look at the time course of the responses. Figure 2.6 shows the average response to the forward song and the three synthetic stimuli for HVc and field L neurons. The plots were generated by computing for each cell the response in 50 msec time windows over the duration of the collection. The average response was computed by normalizing the time axis from 0.0 and 1.0 and then averaging the response rates across cells. The average response plots show that both HVc and field L neurons have an onset response. For HVc cells, the average response to forward song builds up during the course of the song, while the response attenuates

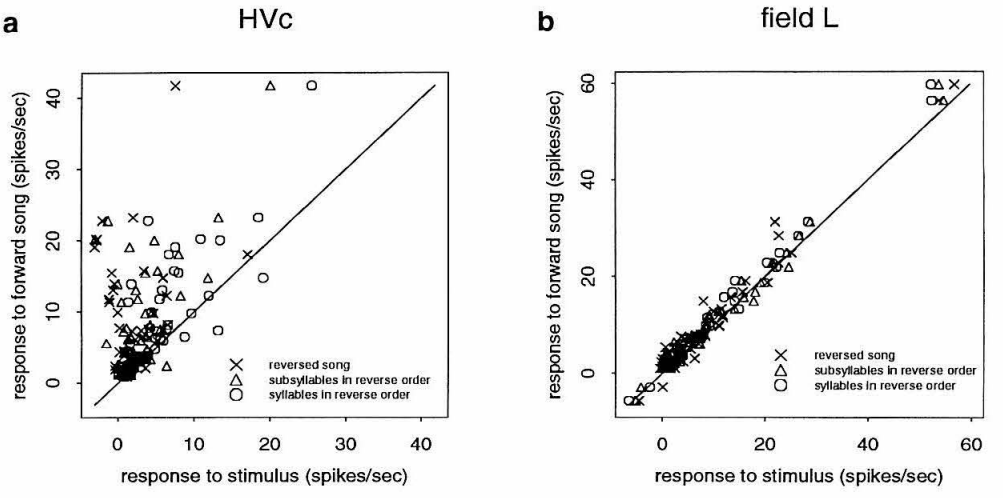


Figure 2.5: The graphs show summary data for HVc (**a**) and field L (**b**). Response is defined as the average spikes per second during stimulus minus the spontaneous rate. The response to the forward song is plotted against the response to the altered song. The solid diagonal line is $y = x$. (**a**) Many neurons in HVc show a large dependence on the song's temporal structure. (**a**) In Field L, however the response to the altered song is typically the same as the response to forward song.

during the course of reverse song and the sub-syllables and syllables in reverse order. For field L neurons, the average response to the song and its temporal manipulations is roughly the same. Neurons typically have a strong onset response and accommodate at the same rate over the course of all four types of stimuli.

Comparison of Order and Combination Sensitivity in HVc and Field L

Previous studies have shown that HVc neurons are sensitive to the order and combination of syllables from the autogenous song, a property called temporal combination sensitivity (TCS) (Margoliash, 1983; Margoliash and Fortune, 1992). Figure 2.7a shows an example of such a neuron. The syllables A and B were selected using the automated procedure described in the methods. Since the response to AB is significantly greater than the response to BA ($p < 0.001$, paired t-test), this cell shows order sensitivity. The cell also shows combination sensitivity, since the response to AB is significantly greater than the sum of the responses to A and B presented in isolation ($p < 0.001$). It is possible that the response to the pair AB could be explained by simple facilitation. For example, syllable A may facilitate

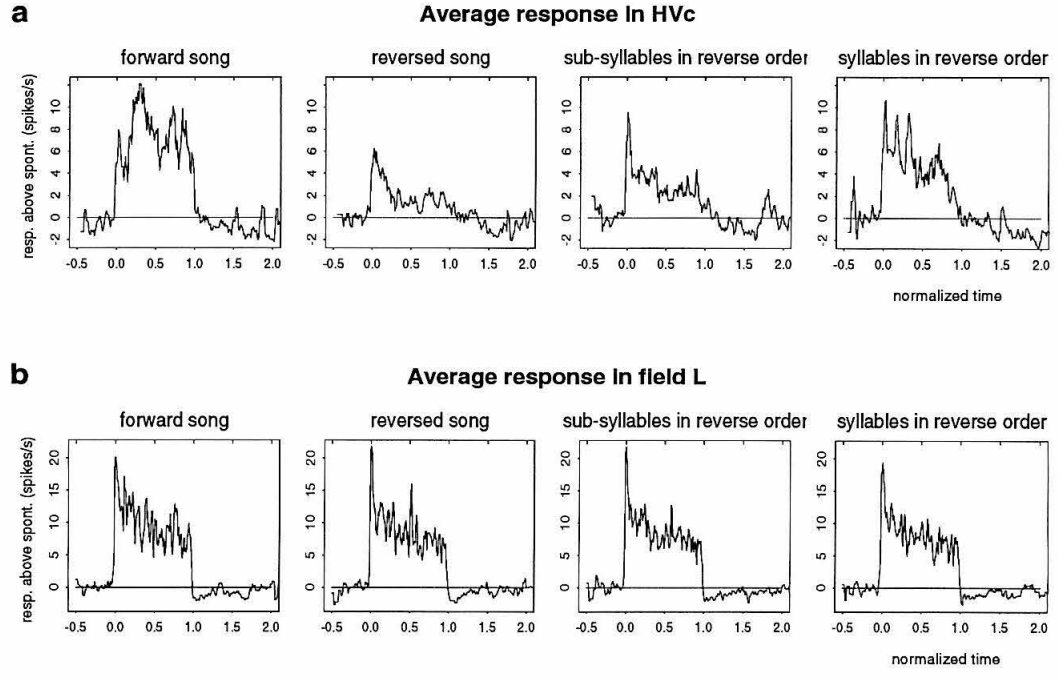


Figure 2.6: The graphs show the average normalized response (see text) for HVc and field L areas. (**a**) The average response of HVc cells builds up during the course of the forward song, but only has an onset response to the other stimuli. (**b**) The average responses in field L were the same for all manipulations of the song with a strong onset response in all cases.

the response to any subsequent stimulus. Conversely, it is also possible that any auditory stimulus facilitates the response to B. To test for these possibilities, the present study also measured the responses to repetitions of each syllable, AA and BB. This cell shows no response to either which provides further evidence that the cell is indeed selective for the syllable combination AB.

In field L, temporal combination sensitivity was also observed despite the lack of strong sensitivity to syllable order of the whole autogenous song as reported in the previous section. Figure 2.7b shows an example of a temporal combination sensitive field L neuron. This cell shows order sensitivity, since the response to the syllable pair AB was (18.03 ± 2.52 spikes/sec) significantly greater ($p < 0.001$) than the response to BA (6.23 ± 2.21 spikes/sec). This cell was also combination sensitive, since the response to AB was significantly greater than the response to the sum of the responses to A and B in isolation ($p < 0.001$). The response to AB cannot be accounted for by facilitation by syllable A since the syllable pair AA produces no response. This cell did, however, show significant facilitation ($p = 0.018$), since the response to BB was greater than the twice the response to syllable B when presented alone. The other two examples of temporal combination sensitivity seen in field L (not shown) showed strong responses to individual syllables and syllable pairs (but still satisfied the criteria in table 2.8). This was not the case for any of the temporal combination sensitive units in HVc.

Tests for order and combination sensitivity were performed on 31 and 42 well-isolated neurons in HVc and field L respectively. All of these cells showed a significant response to autogenous song. Cells which did show a significant response to the selected syllable pair AB when presented in isolation were not analyzed. There were 26 neurons in HVc and 27 in field L that showed a significant response to a syllable pair AB in isolation. Nearly all the syllable pairs selected for HVc neurons produced a significant response (26/29) compared to 27/42 in field L. This difference arises because several field L neurons had a relatively weak, but statistically significant, response to autogenous song. These neurons did not produce a significant response to isolated syllables. Most of the HVc neurons responded more strongly to autogenous song and to the syllables presented in isolation. The results of the syllable tests on the population of HVc and field L neurons are summarized in figure 2.8. A greater percentage of HVc units showed some sensitivity, but in both areas there were instances of order and combination sensitivity.

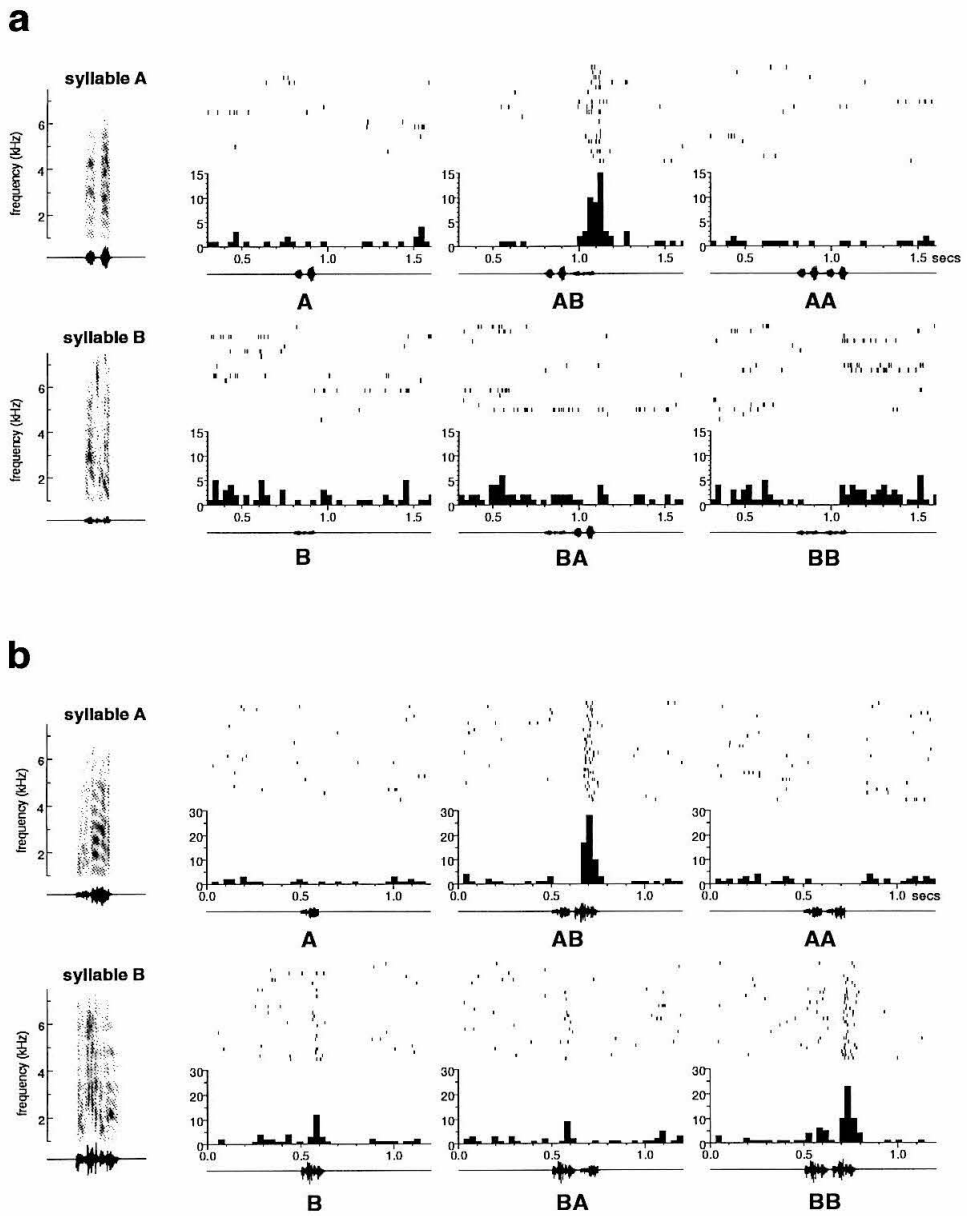


Figure 2.7: To the left of each set of histograms is the sonogram and oscillogram of the stimuli which are syllables selected from the autogenous song. (**a**) A temporal combination sensitive (TCS) neuron from HVc. (**b**) A field L TCS neuron.

The data were also analyzed for significant responses to the syllable pairs in reverse order. 20 cells in HVc and 29 in field L showed a significant response to the syllable pair BA in isolation. Of these, one HVc cell showed significant reverse order sensitivity ($BA > AB$) compared to 4 in field L. In HVc, two showed significant reverse combination sensitivity ($BA > B + A$) compared to 3 in field L. No cells in HVc showed both of these properties whereas one did in field L. The number of HVc cells showing significant order or combination sensitivity for the reverse order, BA, was lower than for the normal order, AB (20 vs 28). In field L, there were roughly equal numbers for both the reverse and normal order (27 vs 29).

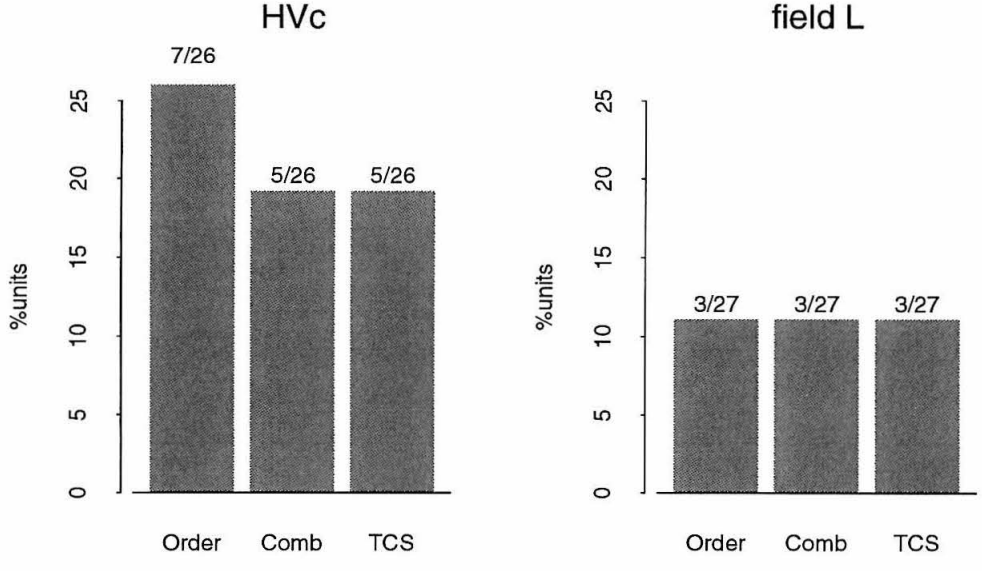


Figure 2.8: The tests for order and combination sensitivity were performed on cells that showed a significant response to forward song. For each set of tests, a pair of syllables (AB) was selected that was likely to show order sensitivity (see methods). The bars indicate the percentage of cells that satisfied the listed conditions given a significant response to the syllable pair AB. Order: the response to AB is significantly greater than the response to BA; Comb: the response to the syllable combination AB is significantly greater than the response to A or B when presented in isolation; TCS (temporal combination sensitivity): the cell showed both order and combination sensitivity. The incidence of TCS is low, but not inconsistent with previous studies.

Not all of the cells that showed auditory context sensitivity also showed order or combination sensitivity. In HVc, tests for order and combination sensitivity were performed on 13 cells that showed a significant difference between the response to the forward song and the response to the syllable- or subsyllable-reversed songs. Of these, only 7 were sensitive

to the temporal order and/or combination of the syllables, in spite of large differences in response between the forward and syllable-reversed songs. In field L, order and combination test were performed on four cells that showed significant sensitivity to the syllable or sub-syllable order. In contrast, to HVc three of these cells showed either order or combination sensitivity and the other showed significant facilitation. This suggests that additional mechanisms are required to explain the strong auditory context sensitivity seen in HVc.

2.4 Discussion

These results show that there is a substantial increase in the auditory temporal context sensitivity associated with the progression from the areas of field L to HVc. Neurons in field L typically respond equally well to autogenous song and to the temporally manipulated versions of the song. In contrast, neurons in HVc are highly dependent on the song's temporal structure. They respond strongly to the forward song, but respond weakly to the reversed song and to the song with the syllables or sub-syllables in reverse order. This extends previous findings that neuronal preference for autogenous song is observed in HVc but not in field L (Margoliash, 1986).

The response of song-specific units in HVc depends on auditory temporal context which extends beyond a single syllable. FM-sensitivity alone is insufficient to account for the context sensitive properties of these neurons. Previous studies comparing the context sensitivity of HVc and field L neurons used only forward and reverse song (Margoliash, 1986; Margoliash et al., 1994). Reversing the order of the syllables or sub-syllables does not change the direction of the frequency sweeps in the song. Thus, any change in response between the forward and syllable or subsyllable-reversed song cannot be attributed to FM-sensitivity and must arise from the integration of the auditory context of the previous syllables.

Neurons in field L are much less sensitive to the auditory temporal context than are HVc neurons. Earlier studies suggested that the responses of field L neurons could be accounted for by the sensitivity to short-term spectro-temporal structure, such as amplitude and frequency modulation (Schafer et al., 1992). The presence of TCS neurons in field L provides new evidence that field L neurons can also encode information about syllable combination and order. This observation agrees with previous findings that field L neurons in Mynah birds can be sensitive to the temporal structure of human vowel sounds (Uno

et al., 1991). Sensitivity to syllable order and combination requires integration of auditory context over longer periods of time, as much a hundred milliseconds. Thus it is not clear how these responses could be accounted for by FM or AM sensitivity which are sensitive to temporal structures on a time scale of a few milliseconds.

One explanation for the temporal context sensitivity of HVc song-specific units is in terms of the neural sensitivity to syllable order and combinations. In this model, sensitivity to the order of the syllables in the autogenous song results from either sensitivity to the order of particular syllable combinations or from integrating the output of such neurons. Several HVc neurons, however, showed strong sensitivity to the order of the syllables in the whole song, but were not sensitive to either the order or combination of syllable pairs when presented in isolation. These data suggest that HVc neurons integrate auditory context over periods greater than the duration of syllable pairs which supports the conclusions reached by earlier studies (Margoliash and Fortune, 1992; Margoliash and Bankes, 1993; Lewicki and Konishi, 1995).

The results presented here fit well with known facts about the anatomy of the song system auditory pathway. A hierarchical arrangement of response properties would be expected, since the L2 areas project to L1 and L3 (Kelley and Nottebohm, 1979) which then project to HVc (Fortune and Margoliash, 1995). All of these areas were sensitive to the local temporal structure, as evidenced from the differences in response between the forward and reversed song. Only L1, L2, and HVc showed sensitivity to the order of the syllables or sub-syllables. Dramatic dependencies on the auditory temporal structure of autogenous song was only observed in HVc. Given the highly recurrent nature of the projection patterns of HVc neurons (Katz and Gurney, 1981; Fortune and Margoliash, 1995), it is possible that song-specific neurons could in fact integrate auditory information over the entire duration of the song.

Chapter 3 Synaptic Mechanisms Subserving the Intrinsic Interactions of HVc Neurons[1]

3.1 Introduction

The telencephalic nucleus HVc represents a critical part of a network of localized brain nuclei in songbirds and is involved in song learning and production (Nottebohm et al., 1976; Konishi, 1989; Vicario, 1991; Doupe, 1993). Lesioning HVc permanently impairs singing behavior (Nottebohm et al., 1976; Simpson and Vicario, 1990) as well as the ability of animals of both sexes to discriminate songs of conspecifics from those of other species (Brenowitz and Arnold, 1990; Cynx, 1993). Neurons in HVc are responsive to auditory stimuli, and many respond preferentially to playback of the bird's own song (Margoliash, 1983, 1986; Margoliash and Fortune, 1992). Multiple-unit recordings in HVc reveal bursts of activity before and during each song vocalization (McCasland and Konishi, 1981; McCasland, 1987). Electrical perturbation of the firing pattern of HVc neurons during singing alters the temporal pattern of the song (Vu et al., 1994). Further evidence that HVc is involved in song production is that young birds begin to vocalize at the time that HVc axons first innervate nucleus RA, which projects directly to the motorneurons that innervate the muscles of the syrinx, the vocal organ (Konishi and Akutagawa, 1987). Thus, HVc is thought to play important roles in song control, perception, and acquisition.

In order to understand better the cellular mechanisms that subserve the various functions of HVc, we have begun to characterize the intrinsic physiological properties of neurons in HVc and their synaptic interactions. This chapter describes the types of neurotransmitter receptors that are activated during synaptic transmission between neurons within HVc. Immunohistochemical studies have demonstrated the presence of N-methyl-D-aspartate (NMDA) receptors on HVc neurons (Aamodt et al., 1992). Autoradiographic studies have shown that HVc contains cell bodies as well as axon terminals containing γ-aminobutyric acid (GABA) (Zuschratter et al., 1987; Grisham and Arnold, 1994). We report

[1]The work in this chapter was done in collaboration with Dr. Eric T. Vu

here results from intracellular recordings of HVc neurons in vitro which demonstrate that antidromic stimulation of RA-projecting HVc neurons activates NMDA and non-NMDA glutamate receptors as well as $GABA_A$ receptors.

3.2 Methods

Acute brain slices were prepared from 37- to 55-day-posthatch male zebra finches (Taeniopygia guttata) obtained from our breeding colony. Animals were deeply anesthetized with ketamine hydrochloride (40 mg/kg, intra-muscular) and Metofane (methoxyfurane; Pitman-Moore, Inc., Mundelein, IL), then decapitated. The brain was removed and placed in ice-cold, oxygenated, artificial cerebrospinal fluid (ACSF; concentrations in mM: NaCl, 134.0; NaHCO3, 25.7; NaH2PO4, 1.3; KCl, 3.0; MgSO4(7H2O), 1.3; CaCl2(2H2O), 2.4; glucose, 12.0). The brain was blocked along the sagittal midline and both hemispheres were sectioned simultaneously with a vibratome (Ted Pella, Inc., St. Louis, MO). Parasagittal slices were cut at 400 μm thickness and transferred to an interface-type holding chamber with a humidified atmosphere of 95% O2 / 5% CO2.

Following a 2- to 3-hr recovery period, individual brain slices were transferred to a submersion-type recording chamber perfused at a rate of 3-5 ml/min with ACSF saturated with 95% O2 / 5% CO2. Slices were maintained and recorded at room temperature ($25 \pm 1°C$). A pair of formvar-insulated tungsten electrodes (1.2 MW; AM-Systems, Seattle, WA) bonded at the shafts served as bipolar stimulating electrodes. The stimulating electrode tips were separated by 300-400 μm. The stimulating electrodes were placed in an area ventral and slightly posterior to HVc, half the distance between HVc and RA which were clearly visible in unstained brain slices. The HVc-to-RA fiber tract was stimulated with 100 msec pulses of isolated constant current (1-40 μA) at 0.033-0.067 Hz using a standard stimulator (Pulsemaster, WPI, Inc., New Haven, CT).

Individual HVc neurons were recorded with standard techniques for whole-cell recording in current-clamp mode. Glass pipette electrodes with 3-7 MΩ resistances were fabricated from borosilicate glass (WPI 1B1001F) with a microelectrode puller (Flaming-Brown, Sutter Instruments, Novato, CA). The solution in the pipette consisted of (in mM): MgCl2, 4; CaCl2, 0.1; NaOH, 11.25; Na2ATP, 3; NaGTP, 2; glucose, 1; Kgluconate, 140; HEPES, 10; EGTA, 1.1. The input resistance of each cell was monitored throughout the recording time.

Pharmacological agents were purchased from Sigma (St. Louis, MO) and dissolved in the perfusing ACSF solution to the stated concentrations.

3.3 Results

We report observations from neurons in HVc that exhibited stable resting potentials more negative than -50 mV with action potential heights greater than 50 mV. Postsynaptic potentials (PSP) were observed in all such neurons (N=46) following electrical stimulation of the brain area known to contain the axon fibers of HVc neurons projecting to RA (3.1). Stimulation in this region of the slice presumably evoked antidromic action potentials in a subset of HVc neurons that project to RA. Consistent with this, antidromic action potentials were observed in three other HVc neurons, the postsynaptic responses of which were not included in this analysis.

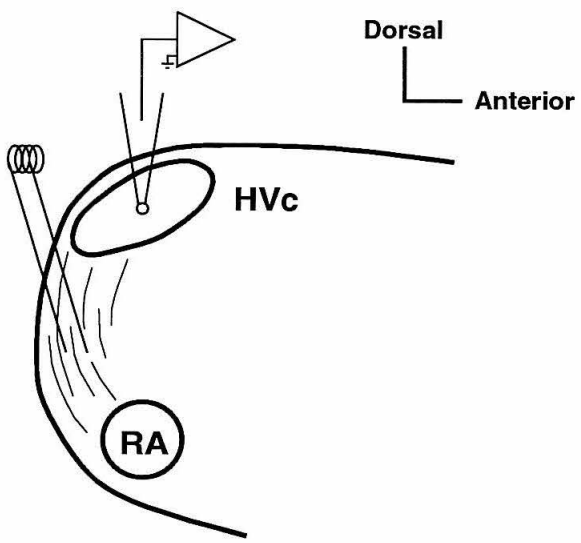


Figure 3.1: Schematic of the slice recording and stimulation setup. Intracellular potentials of neurons in HVc were measured using whole cell patch pipettes. Neurons were stimulated antidromically by stimulating the fiber tract projecting from HVc to RA with bipolar tungsten electrodes.

Lower stimulus intensities evoked a depolarizing PSP in the recorded neuron (Fig. 1A), which increased in size with increasing stimulus amplitudes. However, a second hyperpolarizing phase was observed at higher stimulus intensities. The latter phase tended to begin

prior to the peak of the initial phase, thus limiting the maximum size of the depolarizing phase with further increases in stimulus strength.

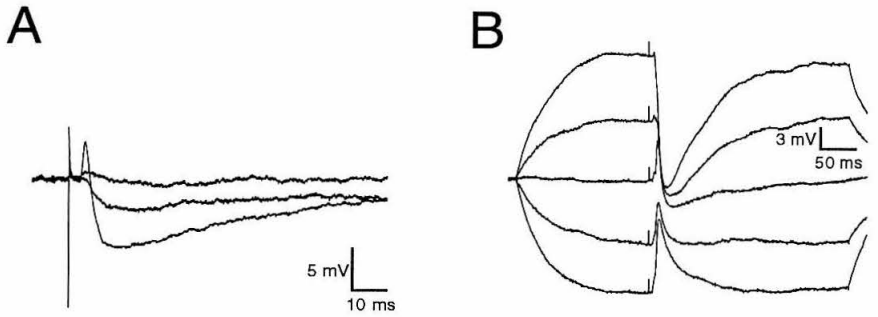


Figure 3.2: (A) Sample PSPs in an HVc neuron following stimulation at different stimulus amplitudes. (B) Effect of membrane polarization on shape of evoked PSP.

The hyperpolarizing phase of the PSP could be reversed by hyperpolarizing the cell membrane potential to between -68 to -72 mV, indicating that this was an IPSP (figure 3.3b). Conversely, the initial depolarizing phase was an EPSP because it did not reverse when summed with membrane depolarization. Furthermore, summing this phase with sufficient membrane depolarization could evoke an action potential at the peak of the EPSP.

To confirm that our usual stimulation site (between HVc and RA) activated primarily axon fibers of HVc neurons projecting to RA, we compared in some preparations the intracellular responses of HVc neurons to stimulation at our usual site with responses to stimulation in the center of nucleus RA, which does not send a direct projection to HVc. The two evoked PSPs had very similar time courses. There was a longer latency to onset for responses to stimulation inside RA, consistent with the greater distance between stimulating and recording electrodes for this case.

To determine the postsynaptic receptors that mediate the IPSP, the specific antagonist of $GABA_A$ receptors, BMI, was delivered via the perfusing solution. In 6 out of 6 cells, the IPSP was reversibly abolished by 5 μM BMI (figure 3.3). In the absence of the IPSP, the EPSP evoked by single-pulse stimulation of the fiber tract could be prolonged and could evoke bursts of action potentials in the postsynaptic neuron (figure 3.3b).

Both the EPSP and the IPSP were abolished by 5 μM CNQX in the perfusing ACSF

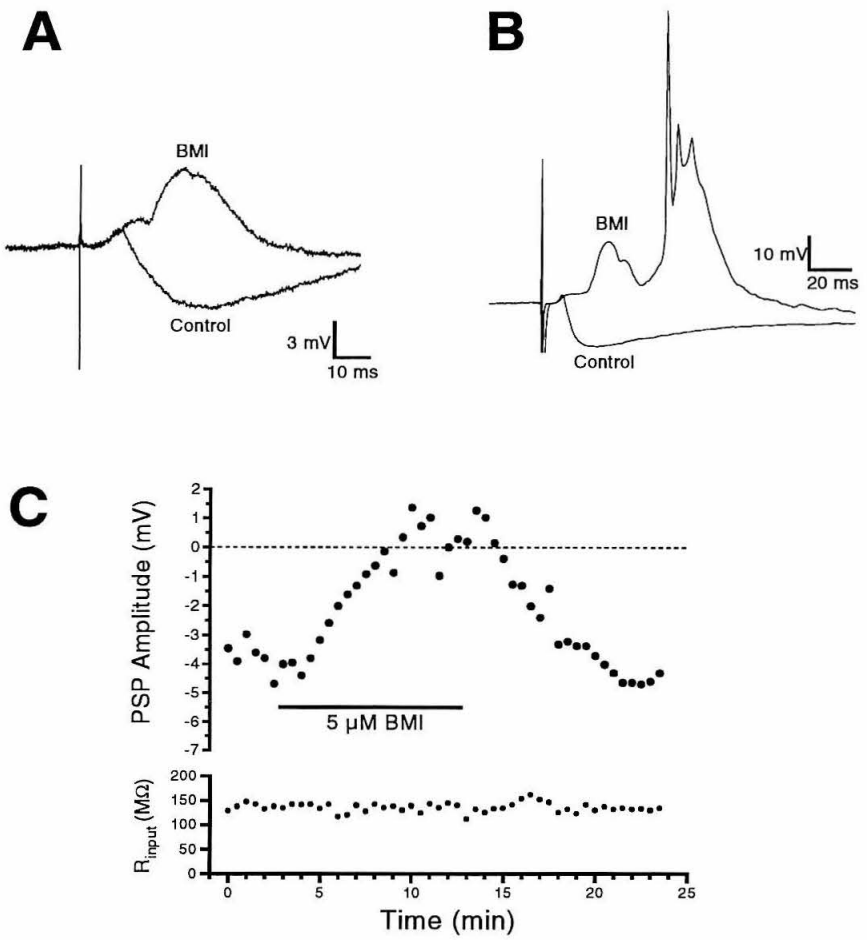


Figure 3.3: (A) Sample traces with and without BMI. (B) Traces from a cell with multiple action potentials following stimulation in presence of BMI. (C) Time course of IPSP in one cell during application of BMI (scatter plot).

(data not shown), indicating that the EPSP is mediated by glutamate receptors and that the IPSP is evoked by activation of glutamatergic synapses on GABAergic interneurons within HVc.

To determine the type of glutamate receptors mediating the EPSP and more specifically to determine whether NMDA receptors are activated during the EPSP, 8 slices were perfused with Mg^{++}-free ACSF containing 5 μM CNQX and 5 μM BMI. This solution was designed to abolish inhibition mediated by $GABA_A$ receptors and excitation mediated by non-NMDA glutamate receptors. The removal of $Mg^{++}$from the perfusing solution ensured that any evoked response that was mediated by NMDA receptors would be observed even at normal resting potentials. Under these conditions, both the IPSP and the fast-rising EPSP were abolished and a slow-rising (time to peak $>$ 10 ms) and long-lasting EPSP was now observed (3.4). To confirm that this EPSP was mediated by NMDA receptors, 10-20 μM DL-APV was then added to the perfusing solution, still in the presence of CNQX and BMI. This reversibly antagonized the slow EPSP in 8 out of 8 cells (figure 3.5).

3.4 Discussion

A schematic of the circuitry suggested by these data is shown in figure 3.6. Antidromic activation of a subset of HVc neurons that project to RA evoked an EPSP followed by an IPSP in neurons recorded in HVc. The EPSP is mediated by activation of both non-NMDA and NMDA glutamate receptors. The IPSP is mediated by activation of $GABA_A$ receptors that presumably gate a chloride current. The IPSP appears to be a polysynaptic response because its onset latency relative to the stimulus was always longer than that of the EPSP.

It is likely that the population sampled in this study contains both X- and RA-projecting neurons as well as intrinsically projecting interneurons. In a few cases, an RA-projecting cell could be confirmed by the presence of an antidromic spike. X-projecting cells could not be confirmed directly, but it is unlikely they were missed, since they constitue 30% of HVc cells (Sohrabji et al., 1989) and tend to be larger than RA-projecting HVc cells (Katz and Gurney, 1981; Paton and Nottebohm, 1984; Paton et al., 1985). A few of the neurons had thin action potential and were non-accommodating, which are common properties of interneurons.

The presence of the EIPSP in all the cells in this study suggests that the population of

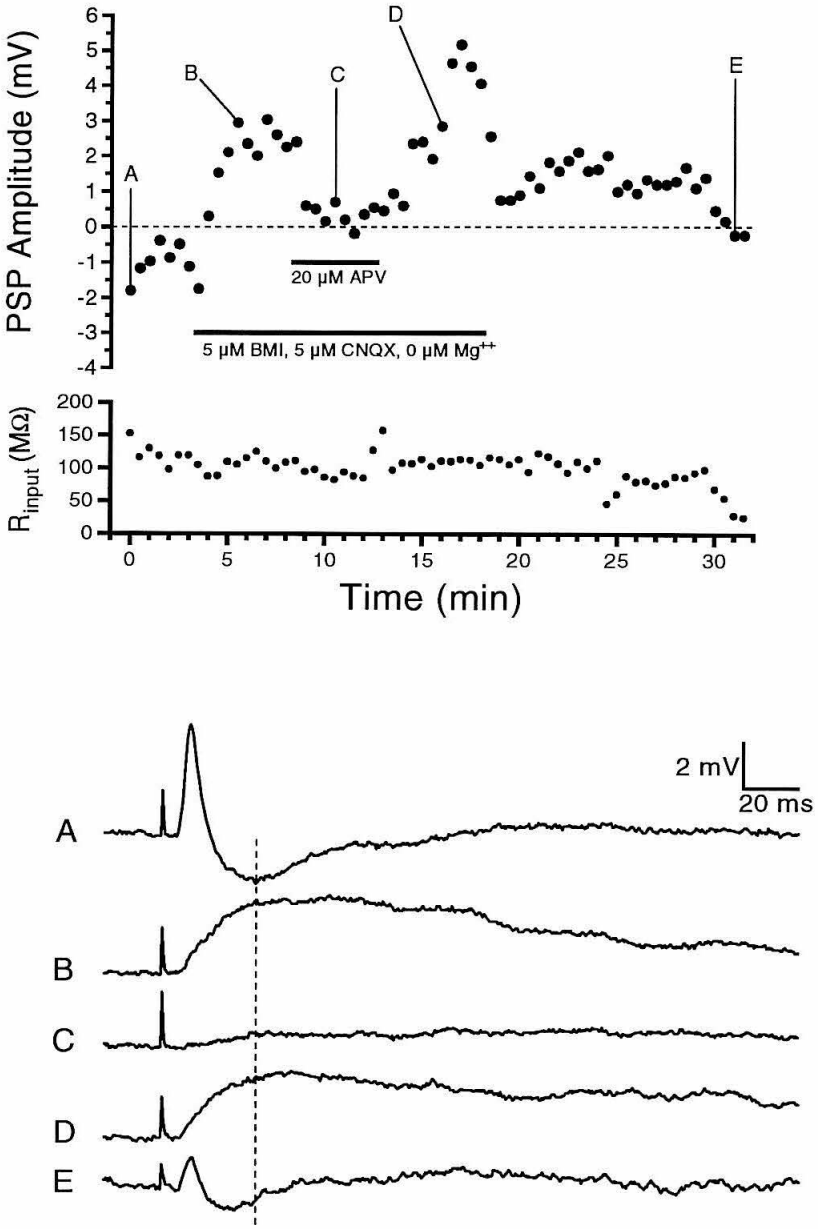


Figure 3.4: Time course of PSP with application of BMI/CNQX, then BMI/CNQX/APV, then BMI/CNQX, then wash (scatter plot). (A-E) Sample traces at representative time points.

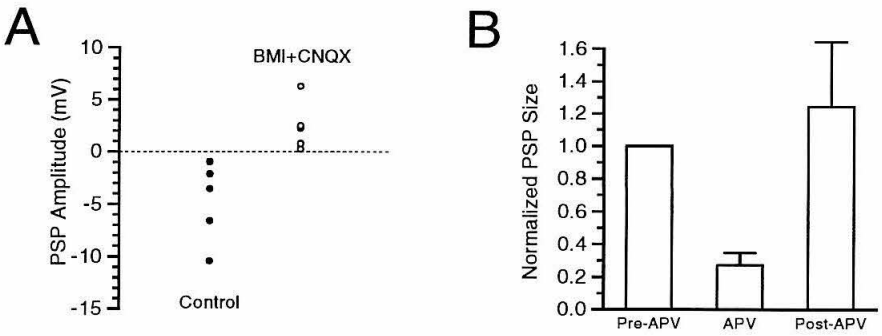


Figure 3.5: Summary data of pharmacological manipulations across 8 cells. (**A**) The amplitude before and after the application of BMI and CNQX in a Mg^{++}-free ACSF. (**B**) Barplots of amplitudes before, during, after the application of DL-APV.

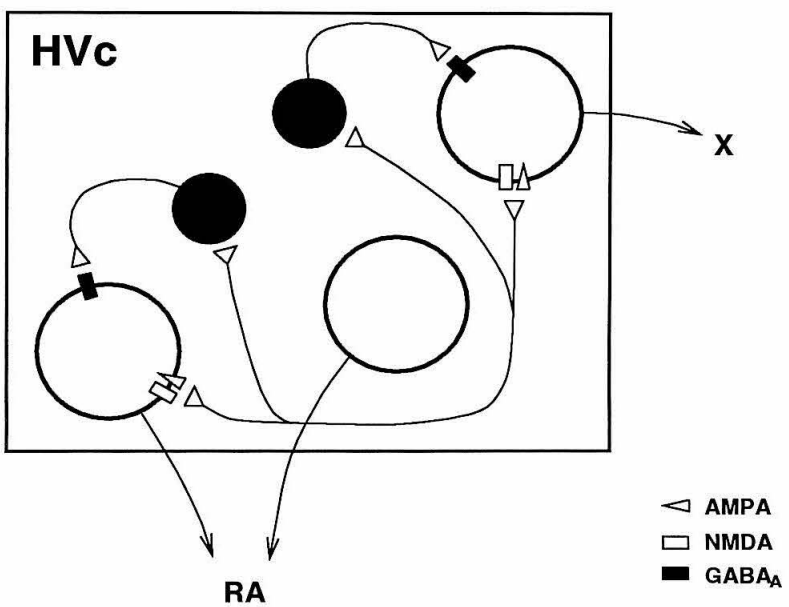


Figure 3.6: A schematic summary of the circuitry and receptor types in HVc suggested by these experiments. The small filled circles represent locally projecting, inhibitory interneurons; The large circles represent projection neurons.

X- and RA-projecting HVc cells are highly interconnected. This observation is consistent with the arborization patterns of the axons of filled HVc cells which show extensive local collaterals (Katz and Gurney, 1981). Since vocal learning is highly dependent on auditory feedback (Konishi, 1965, 1989), the recurrent connections within HVc, specifically between X- and RA-projecting neurons, may subserve the formation of the song production circuitry.

The IPSP could be abolished by bath application of BMI, which suggests it is mediated by the activation of $GABA_A$ receptors. Bath application of CNQX abolished the fast component of the EPSP, suggesting it is mediated by the activation of the AMPA subtype of the glutamate receptor. Application of APV abolished the EPSP's long component, indicating it is mediated by the activation NMDA receptors.

The presence of active NMDA receptors in HVc suggests the possibility that they may be involved in plasticity. Activation of NMDA receptors can lead to long lasting changes in synaptic strength in mamailian vertebrate brain areas such as the hippocampus and visual cortex (Kirkwood et al., 1993; Malenka, 1994) as well as in *Xenopus* and goldfish optic tectum (Cline, 1991). Neural plasticity is a crucial aspect of at least two stages of song development: the sensory stage, where a young bird forms a memory (or template) of the tutor song, and the sensorimotor stage in which the bird learns to produce the memorized song through auditory feedback.

Plasticity in HVc is most likely to take place in the sensory motor stage during the formation of the pattern generation circuitry underlying song production. During the sensorimotor period, the motor program circuitry must adapt in response to feedback resulting from the comparison of immature vocalizations to the learned song template. Over the course of the sensorimotor period the young bird gradually learns to match its song to that stored in the template. The NMDA receptors in HVc could be part of the neural mechanisms underlying the learning of the motor program for song. A similar role has been suggested for NMDA receptors of neurons in RA (Mooney and Konishi, 1991; Kubota and Saito, 1991a; Mooney, 1992), and, like HVc, RA receives auditory input and contains neurons underlying the motor program for song production.

HVc is at the intersection of the auditory and motor pathways and contains neurons that have highly selective auditory responses (Margoliash, 1983; Margoliash and Fortune, 1992) and neurons that show premotor activity specifically to song (McCasland and Konishi, 1981; McCasland, 1987; Yu and Margoliash, 1995). It is thus at location idea for both

the comparison of auditory feedback with the song template and for adaptation of the song motor program. In order for the motor program to adapt, the auditory feedback must be translated into a usable form. Since physiological studies have shown that the song-selective neurons in HVc emerge during the vocal learning process (Volman, 1993; Doupe and Konishi, 1992), the auditory neurons in HVc may reflect the learning of the transformation from the comparison to the motor program. The NMDA receptors may also subserve this learning.

The combination of NMDA- and AMPA-mediated post-synaptic currents may also play an important role in the response properties of HVc cells (Lewicki and Doupe, 1993; Lewicki and Konishi, 1995), and in particular their sensitivity the temporal structure of the bird's own song. IPSPs have also been shown to play an important role in song selectivity. The results of this study provide evidence that these mechanisms are present in HVc and provide building blocks for forming neuron responses that are sensitive to spectrally and temporally complex sounds. These auditory response properties could be important for both the code used by the song template and the translation of auditory feedback into an error signal used by the motor program.

Plasticity in HVc may also be important during the sensory stage when the auditory template is learned. It is possible that HVc may also store some aspects of the template. The presence of song template information in HVc would provide one explanation for how motor circuitry in HVc could adapt in response to comparison of auditory feedback with a learned song template. The presence of active NMDA receptors in HVc may provide a mechanism for this learning. The age of the birds used in this study was 37-55 days which is beyond the sensory stage. It remains to be seen whether there are active HVc NMDA receptors during the sensory stage, and if they show the associative properties that have been observed in other systems.

HVc is believed to be the source of the pattern generation circuitry required for song production (Nottebohm et al., 1976; McCasland, 1987; Vu et al., 1994). In other systems, NMDA receptors have been shown to be involved in the generation and modulation of rhythmic patterns such as swimming in the lamprey (Brodin et al., 1985; Grillner et al., 1991) and in *Xenopus* embryos (Dale and Roberts, 1985). NMDA and AMPA receptors in HVc may subserve the generation of the neural pattern underlying the production of song.

Chapter 4 Intracellular Response Properties of Song-Specific Neurons

4.1 Introduction

Neurons in the songbird forebrain nucleus HVc[1] respond preferentially to the bird's own song (hereafter autogenous song). These "song-specific" cells respond more to autogenous song than to other songs of the same species or to songs of other species (Margoliash, 1983; Margoliash and Konishi, 1985; Margoliash, 1986). Song-specific neurons are also highly sensitive to the temporal structure in the bird's own song and can integrate several hundred milliseconds of auditory context (Margoliash, 1983; Margoliash and Fortune, 1992).

One form of auditory context sensitivity that may subserve the observed response properties of song-specific cells is temporal combination sensitivity (TCS). TCS neurons are sensitive to the temporal order of syllables from the bird's own song (Margoliash, 1983; Margoliash and Fortune, 1992). These cells are also found in HVc and respond only to a pair of syllables in the proper temporal order and do not respond to either syllable when presented in isolation. Some TCS cells require three or more syllables in the proper sequence to elicit a response.

Though the response properties of HVc cells have been studied for some time, the mechanisms that give rise to these properties are not known. Physiological studies in brain slices have suggested several synaptic and intrinsic properties that may contribute to the response properties of song-specific neurons and TCS neurons. Kubota and Saito (1991b) found of Na-dependent outward conductances in HVc cells that persisted for several seconds. In addition, they reported that large current injections into HVc neurons can elicit an initial high-frequency burst of action potentials followed by tonic firing. Vu and Lewicki (1994) found active glutamate receptors of both the NMDA and AMPA/kainate subtype as well as active GABAergic receptors. Schmidt and Perkel (1995) reported evidence of

[1]Abbreviations: HVc, *hyperstriatum ventrale pars caudale*, also called high vocal center; NMDA, N-methyl-D-aspartate; AMPA, alpha-amino-3-hydroxy-5-methyl-isoxazole-4-propionic acid; GABA, γ-aminobutyric acid

synaptically activated $GABA_A$ and $GABA_B$ receptors in HVc. It is not known what role these mechanisms play during an auditory response.

In order to investigate the mechanisms underlying the response properties of song-specific neurons, intracellular recordings were made *in vivo* from HVc neurons. The study had two aims: The first was to determine whether there are intracellular response properties that are unique to song-specific neurons; the second was to distinguish among several possible synaptic models that can account for the properties of temporal combination sensitivity (Margoliash, 1983; Lewicki and Doupe, 1993).

4.2 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 a computer (Sparc station IPX, Sun Microsystems, Mountain View, CA). A few days before the experiment, birds were anesthetized with Equithesin (0.03–0.04 ml i.m.; 0.85 g chloral hydrate, 0.21 g pentobarbital, 0.42 g $MgSO_4$, 2.2 ml 100% ethanol, 8.6 ml propylene glycol, filled to a total volume of 20 ml with water), and a small metal post which immobilized the head during physiological recordings was cemented to the skull with dental cement. For physiological recordings, the birds were anesthetized with urethane (65-90 μl of 20% solution).

Nucleus HVc was first located physiologically with extracellular glass electrodes. Electrodes were lowered through a small hole (0.3 mm dia.) in the skull in order to minimize brain edema and pulsation. Intracellular recordings were obtained with sharp electrodes (60–100 MΩ, filled with 4 M potassium acetate, pH 7.4) or whole cell patch electrodes (6-12 MΩ, filled with solution of 140 mM K-gluconate, 10 mM HEPES, 4 mM $MgCl_2$, 0.1 mM $CaCl_2$, 1.1 mM EGTA, 3 mM Na_2-ATP, 2 mM Na-GTP, pH 7.4, and adjusted to 300-330 mosm). 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 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. Cells that did not have a stable resting potential for more than 2 minutes or had an action potentials height less than

40mV were omitted from the analysis.

Some of the stimuli used in these experiments involved manipulations of the order of syllables taken from the bird's own song. Syllable boundaries were defined as points where the song's amplitude falls to zero. 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 dB SPL.

The anatomical location of the recordings were determined by making reference marks made by electrolytic lesions using extracellular tungsten electrodes (AM systems, Evertt, WA), and by filling the single neurons with biocytin. At the end of the experiment the bird was perfused with saline followed by 4% paraformaldehyde for histological analysis. Electrolytic lesions were located on 30 μm frozen sections stained with cresyl violet.

4.3 Results

Stable intracellular recordings were obtained from 97 cells. The mean duration of intracellular recording time was 16 minutes. The mean initial resting potential was -61 ± 9 mV, and the mean action potential height was 58 ± 13 mV. Out of 97 the cells, 29 showed some auditory response, and 6 of these 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 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 a maximum interspike interval less than 6 ms. Although song-specific cells were relatively rare, this is consistent with the frequency reported in previous extracellular studies in the zebra finch HVc Margoliash and Fortune (1992). Song-specific cells showed no apparent differences from other cells in terms of their resting potential, action potential shape, or holding times.

Properties of intracellular responses to song

Extracellular studies have shown that song-specific neurons can have both phasic and tonic responses to song which often contain bursts of action potentials (Margoliash and Fortune, 1992; Sutter and Margoliash, 1994). These properties were also found to be present in the intracellular records of song-specific cells.

Tonic excitation

Figure 4.1 shows an intracellular recording of a song-specific HVc cell which is tonically excited throughout much of the autogenous song. The response also contains many action potential bursts, which is characteristic of HVc neurons. The median membrane potential (n.b. the median reduces the large influence of the action potentials on the summary waveform) shows that there is some hyperpolarization following the forward song (figure 4.1, left panel) and after the middle of the syllable-reversed song (figure 4.1, middle panel). No such hyperpolarization is present in the response to the reversed song. The response to both the reversed and syllable-reversed song is less than to the forward song. One explanation for the reduced excitation and absence of hyperpolarization to the reversed song is that this neuron is integrating the output of neurons which respond to forward song but not to reversed song. The hyperpolarization to the syllable-reversed song could result from active inhibition from other HVc cells or from mechanisms in the cell itself which are sensitive to the syllable order. The present data do not allow us to distinguish between these two possibilities.

Another tonically excited cell was also depolarized throughout most of the forward song (data not shown). That cell showed no hyperpolarization during the syllable-reversed song, but was simply less depolarized. It also had little depolarization to the reversed song.

Phasic hyperpolarization during tonic excitation

Phasic hyperpolarization was seen in two of the tonically excited song-specific cells. The median response to forward song (figure 4.1, left panel, bottom trace) shows a prominent hyperpolarization near the end of the song. A different cell is shown in figure 4.2 which shows a large hyperpolarization in the middle of the song.

To rule out the possibility that this hyperpolarization was an after-hyperpolarization

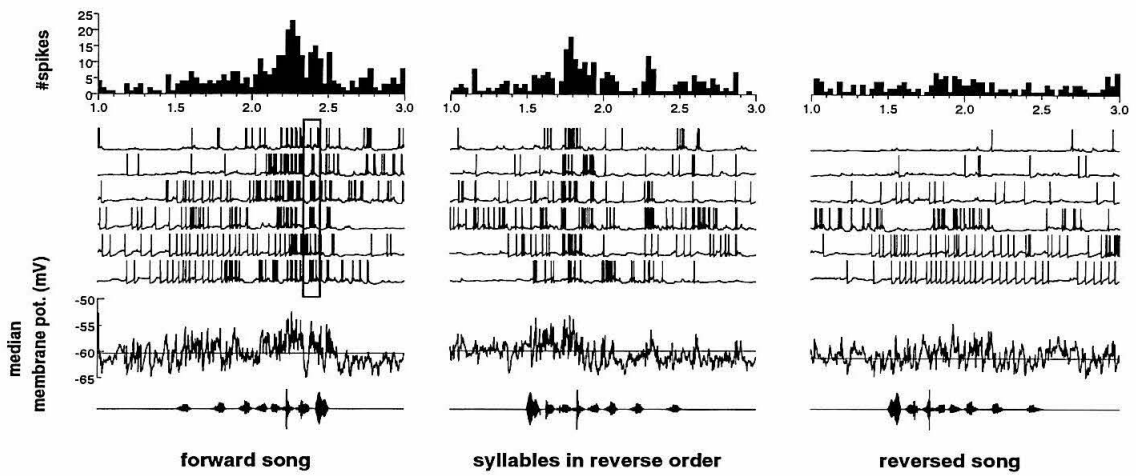


Figure 4.1: An intracellular recording of a tonically excited song-specific HVc cell. Each panel shows the peri-stimulus time histogram (*Top*). The traces below show the intracellular membrane potential for six collections (*Upper traces*). The collections for each stimulus were interleaved. The traces inside the box are plotted in figure 4.6. Below each set of waveform rasters is the median of the individual traces (*Lower trace*); the average resting membrane potential is shown by the horizontal line. The oscillogram of the stimulus is plotted below (*Bottom*). The response to the forward song is greater than to the song with the syllables in reverse order and to the reverse song, indicating that this is a song-specific cell. In this example, the median membrane potential shows that the cell is hyperpolarized to the song with the syllables in reverse order but not to the reversed song.

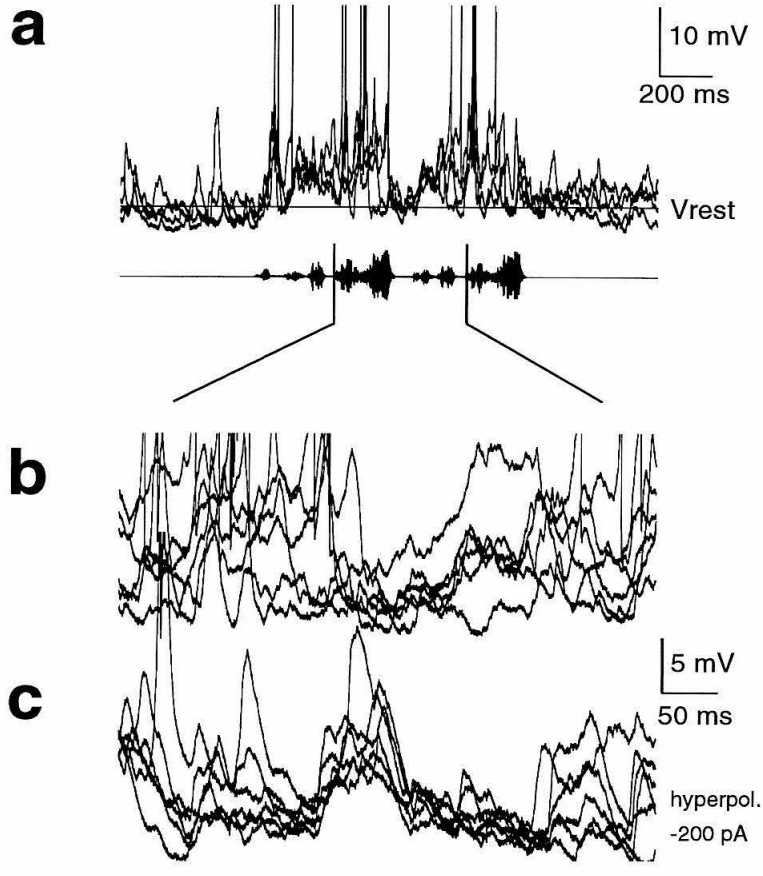


Figure 4.2: (**a**) The graph shows overlaid intracellular recordings of a different HVc neuron in response to song. The straight horizontal line indicates the cell's resting potential, which was −70 mV. Note the general depolarization throughout the song except in the middle where there is hyperpolarization. (**b**) The response at resting potential on an expanded time scale. (**c**) The response, on the same time scale as (**b**) while the cell was hyperpolarized to −85 mV with a −200 pA current injection throughout the duration of the stimulus. During the current injection, the hyperpolarization in the upper traces was reversed, suggesting it is a GABAergic inhibitory post-synaptic potential. The traces in (**c**) are offset by −5 mV to better separate them from the traces in (**b**).

evoked by previous action potentials, 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 (figure 4.2c). The reversal of the hyperpolarization near –70 mV is consistent with GABAergic, Cl^--mediated, IPSPs.

Phasic excitation

An example of a song-specific cell which is phasically excited is shown in figure 4.3. This cell was recorded from the same bird as the neuron in figure 4.1. In this paper, a phasic response refers to a neuron that is excited at very similar temporal positions in repeated presentations of the stimulus. This is in contrast to the tonically excited cells discussed in the previous section which spike throughout the stimulus, but show little or no regularity in the temporal positions of the action potentials. The phasic response to the forward song is not present in the response to either syllable-reversed song or the reversed song. Note also that only the response to forward song contains consistent hyperpolarizations which can be seen in the overlaid waveform rasters (figure 4.1, left panel, bottom traces). Although this cell responsed with roughly equal numbers of action potentials for all three stimuli, there were significantly more action potential bursts in response to forward song than to either the syllables in reverse order ($p < 0.05$, paired t-test) or to the reversed song ($p < 0.001$). The bursts to forward song are outlined by the boxed region in figure 4.3.

Figure 4.4 shows the response of another phasically excited HVc cell. This was classified as a song-specific neuron, since it responded equally well to the forward and syllable-reversed song but showed no response to the reversed song. The response to the syllable-reversed song is shown figure 4.4. The spike bursts of this neuron are aligned to within 5 milliseconds and even shows remarkable consistency in the variation of the sub-threshold membrane potential.

Hyperpolarization during phasic excitation

The induction of long-lasting hyperpolarizing currents after high frequency firing have been reported from *in vitro* studies of HVc (Kubota and Saito, 1991). We have observed similar currents *in vivo* in response to presentation of the bird's own song. Figure 4.5 shows that the hyperpolarization is greatest during forward song when the cell's response is the strongest (figure 4.5a). The membrane potential is less hyperpolarized when the order of the syllables

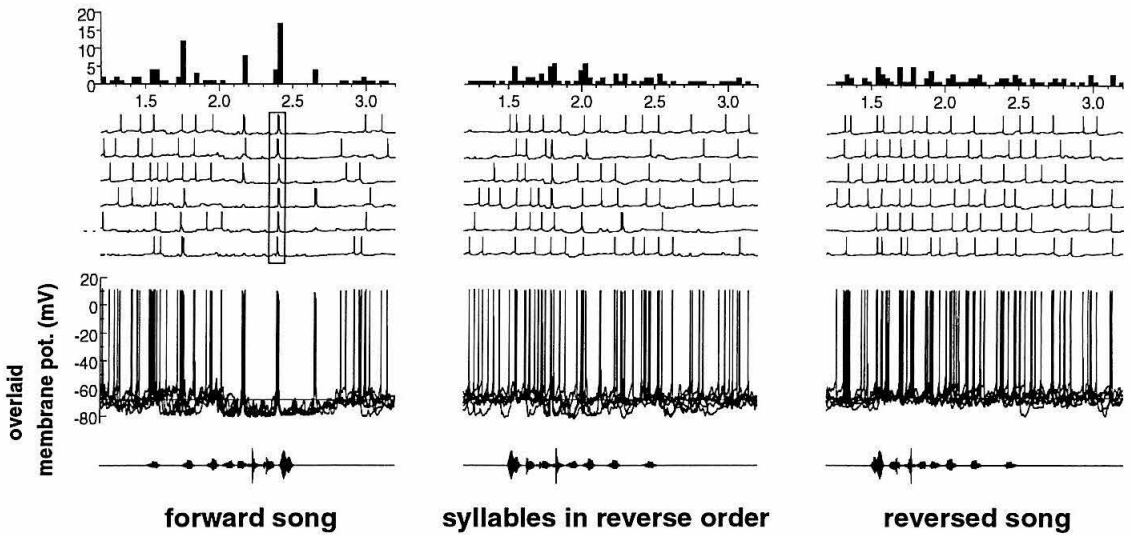


Figure 4.3: An intracellular recording of a phasically excited song-specific HVc cell. The conventions are the same as in figure 4.1 except that in lower traces the waveform rasters are overlaid to show the consistency of both the hyperpolarizations and of the temporal positions of the phasic bursts. The boxed region indicates where this cell responded phasically with bursts of action potentials to the forward song. The phasic response is lost when the syllable order is reversed or when the entire song is reversed.

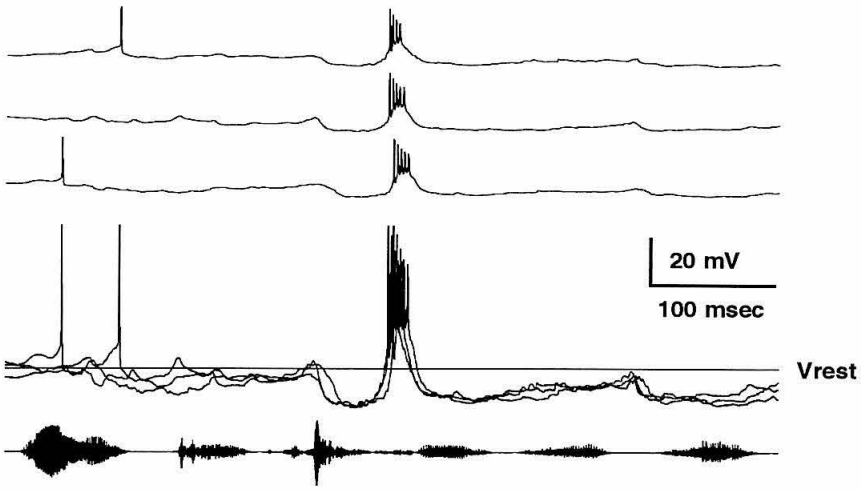


Figure 4.4: Intracellular recording of a phasically excited HVc cell. The horizontal line indicates the average resting potential which was −59mV. Some phasic cells show high regularity in the firing times of the bursts and in the subthreshold membrane potential. The action potentials in these bursts also show attenuation like those in figure 4.6b.

is reversed (figure 4.5b) and is not hyperpolarized at all in response to the reversed song (figure 4.5c). This hyperpolarization is also long-lasting: the recovery to the average resting potential after the end of the forward song (figure 4.5a) takes several hundred milliseconds. Long-lasting after-hyperpolarizations could also be evoked in non-auditory neurons (data not shown) with strong (0.75nA) current injections.

Bursting during tonic and phasic excitation

Some HVc cells were capable of firing 3 or more action potentials in a single high frequency burst. Bursting was present in both tonically and phasically excited song-specific cells. The detailed structure of the responses from the boxed regions of figures 4.1 and 4.3 are shown in figure 4.6. Both responses show bursts of action potentials, but those in figure 4.3b are consistently attenuated over the course of the burst, whereas no attenuation is evident in the bursts in figure 4.6a. This attenuation can be quite dramatic as in figure 4.4 where the last spike in each burst is about half the height of the first. The spike bursts in the phasic responses were also more precisely timed than those in the tonic responses. The bursts in

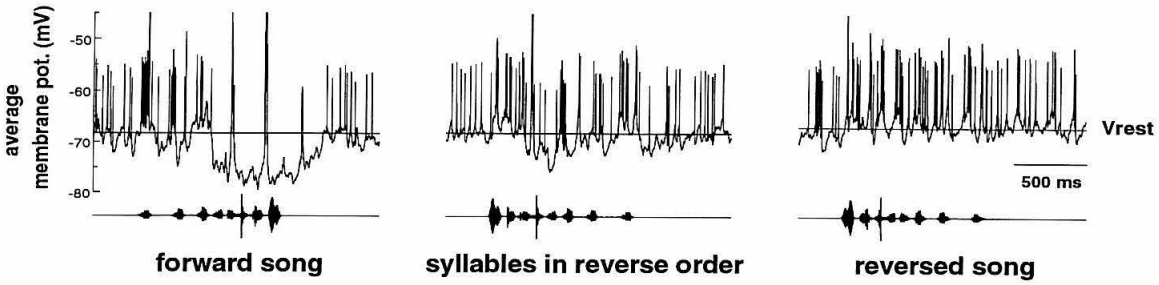


Figure 4.5: Hyperpolarization during a phasically excited song-specific cell. Each trace is the average of the traces shown in figure 4.3. Hyperpolarization is greatest during the forward song, less when the syllables are presented in reverse order, and not present when the song is reversed.

figure 4.6a show no temporal alignment, but those in figure 4.6b temporally aligned to within 8 msecs. These bursts also show consistent prior hyperpolarization. The average membrane potential in the 50 msecs period prior to the first spike in each burst is -6.8 ± 2.1mV below average resting potential (-68 mV). All of the phasically excited cells showed a similar pattern of bursting: attenuation and lack of full repolarization of the action potentials in addition to hyperpolarization prior to each spike burst.

Bursting can also occur at multiple times during the song. One such cell is shown in figure 4.7. Bursting occurs most frequently to forward song where the bursts are phase locked to a particular syllable which occurs three times during the song (figure 4.7a). For this cell, the bursting is less regular than in the previous examples, but like the response shown in figures 4.3, bursting is less frequent when the order of the syllables is reversed (figure 4.7b), and no bursts are present during presentation of the reversed song (figure 4.7c). This cell also bursts most most frequently when the it is hyperpolarized which is shown in figure 4.8.

The correlation between busting and hyperpolarization can be further analyzed by comparing the pre-potential, or the average membrane potential prior to a spike, with the number of subsequent spikes. The action potentials in response to auditory stimuli (during stimulus and background periods) were sorted according to the pre-potential, defined here as the average value of the membrane potential relative to the average resting potential in the 50 ms epoch prior to each spike. Spikes following within 30 ms of an already considered

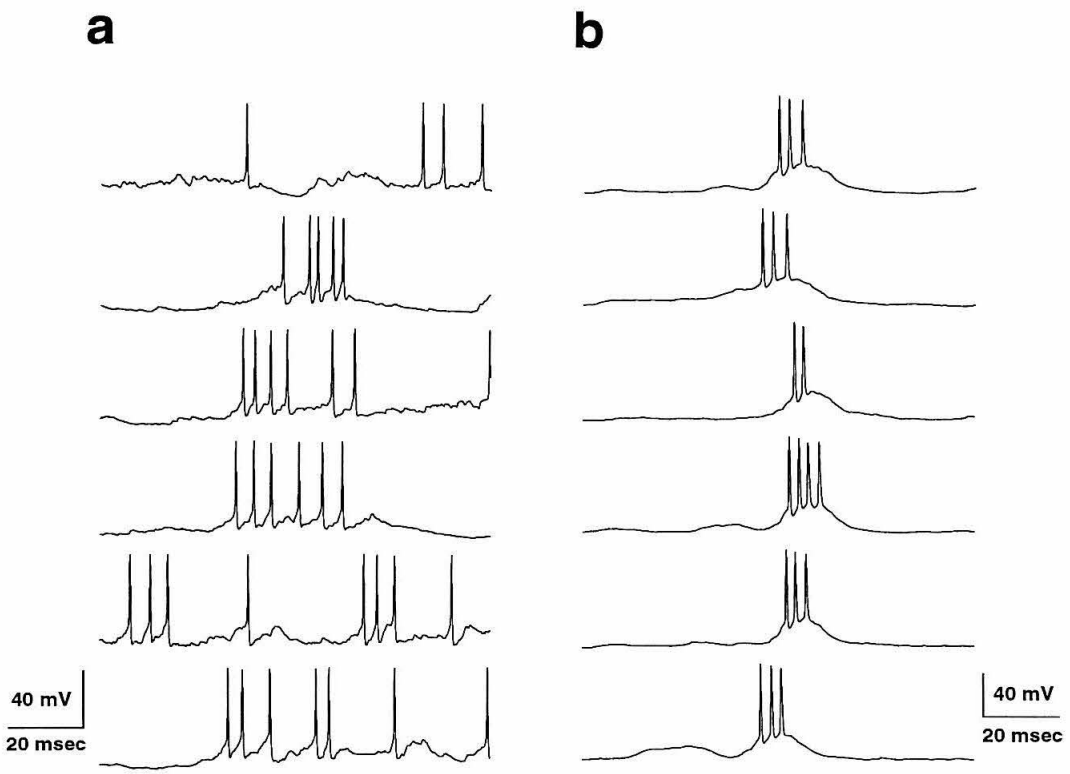


Figure 4.6: Traces from tonically excited song-specific cell shown in figure 4.1 (**a**) and the phasically-excited song-specific cell shown in figure 4.3 (**b**) with expanded time scales. Both responses contain bursts of action potentials, but the bursts from the phasic cell are temporally aligned and show consistent hyperpolarization before each burst. Also, these bursts show a consistent attenuation of action potential height, which is not seen in the bursts of the tonically excited cell.

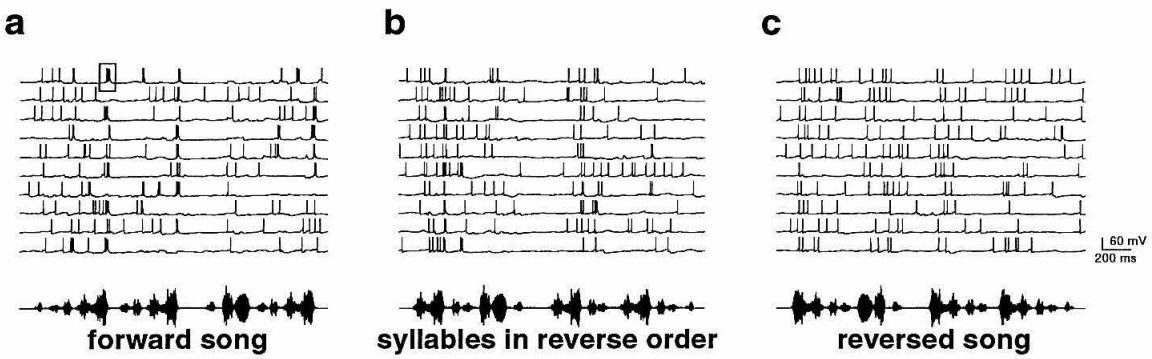


Figure 4.7: Each panel shows the intracellular waveform raster of a song-specific HVc cell. The collections were interleaved. (**a**) The intracellular record shows that this cell bursts (one such burst is outlined by the box) regularly after the fourth syllable in the forward song. (**a**) Less bursting is seen when the syllables are presented in reverse order, and no bursting is seen when the song is reversed (**c**).

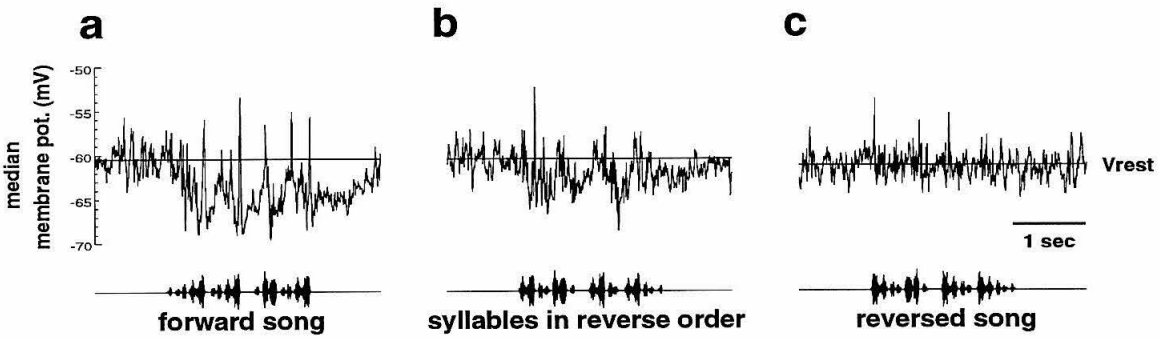


Figure 4.8: The frequency of bursting is correlated with the amount of hyperpolarization. Each trace is the median of the traces shown in figure 4.7 (note the different time scales). The horizontal lines show the resting potential of the cell. As in figure 4.5, hyperpolarization is greatest during the forward song, less when the syllables are presented in reverse order, and not present when the song is reversed.

spike were excluded. The pre-potential and number of following spikes were measured for three groups of neurons: song-specific cells that had a phasic response to the forward song (n=3), song-specific cells that had a tonic response throughout the forward song (n=3), and non-auditory cells that showed spontaneous bursting (n=5). A one-way analysis of variance was performed for each group to determine if there was a statistically significant change in pre-potential.

Figure 4.9 summarizes the distributions of pre-potentials prior to each spike or spike burst for the three different groups. Song-specific neurons that responded phasically to

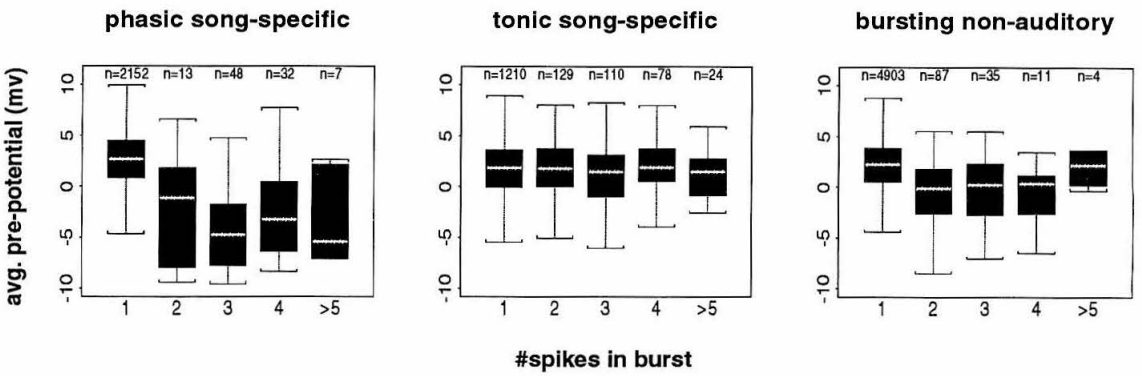


Figure 4.9: Each graph shows a plot of the distribution of pre-potentials for single spikes and spike bursts containing the listed number of action potentials. Each box shows the middle half of each set of pre-potentials. The horizontal line inside each box shows the median. The outer lines indicate range of 99% of the data. Phasic song-specific cells showed significant hyperpolarization prior to each spike burst. Song-specific cells that were tonically excited also showed spike bursting (*e.g.*, figure 4.1), but there was no significant change in the pre-potential as a function of the number of spikes in a burst. Non-auditory cells that generated spontaneous spike bursts did show significant prior hyperpolarization.

song were significantly hyperpolarized prior to each action potential ($p < 0.001$, F-test), but tonically excited song-specific cells showed no significant change in pre-potential versus the number of following spikes ($p > 0.5$, F-test). The spike bursts produced the non-auditory cells were visually similar to those of the phasically-excited song-specific cells, and, like those cells, also showed a significant hyperpolarization ($p < 0.001$, F-test) prior to bursting.

An example of the trend seen in the analysis of the three phasically-excited song-specific cells is shown in figure 4.10. The traces show the waveform patterns following the ten most

negative pre-potentials of all the data collected in response to the forward song (shown in figure 4.7). Nearly every trace is followed by a burst. The most positive pre-potentials, however, are all followed by single action potentials. Conversely, one can examine the

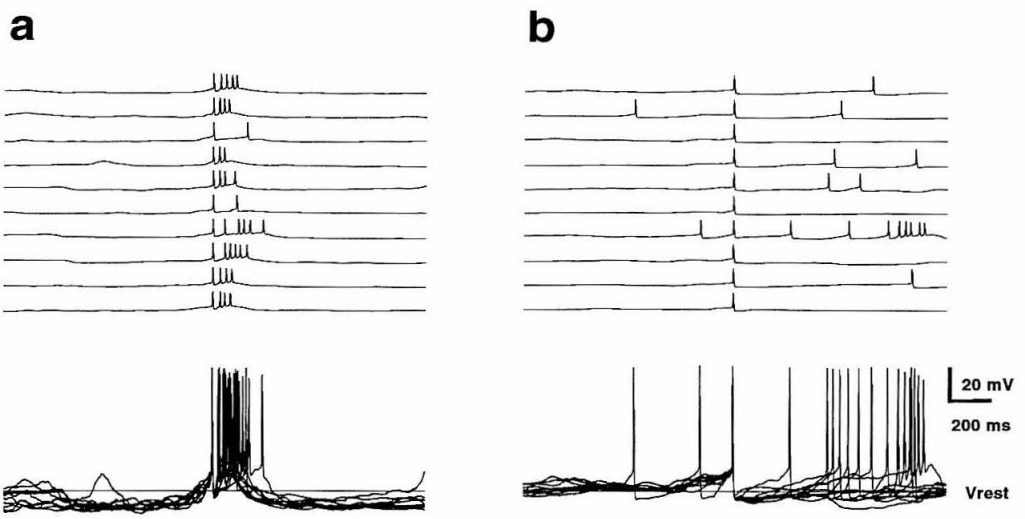


Figure 4.10: The action potentials in response to auditory stimuli (during stimulus and background periods) were sorted according to the average potential 50 ms before each single spike or the first spike in each burst. Spikes with the lowest pre-potential were usually followed by a burst of action potentials (**a**); spikes with the highest pre-potential were not. (**b**).

membrane potential prior to a burst with a certain number of spikes. Figure 4.11a shows four waveforms aligned on the first action potential of bursts containing six spikes. These are preceded by a consistent hyperpolarization, whereas the bursts containing only three spikes are not. These data indicate that a strong action potential burst is often preceded by a hyperpolarization.

Properties of intracellular responses to song syllables

Long-lasting depolarizations

In some cases, the response to stimuli persisted beyond stimulus offset by about 50 to 100 ms, but we did observe depolarizations lasting more than several hundred milliseconds beyond the stimulus (figure 4.12). It is possible that the time course of the depolarization reflects the time constant of the cell membrane, but the responses to current pulses (figure 4.12, inset)

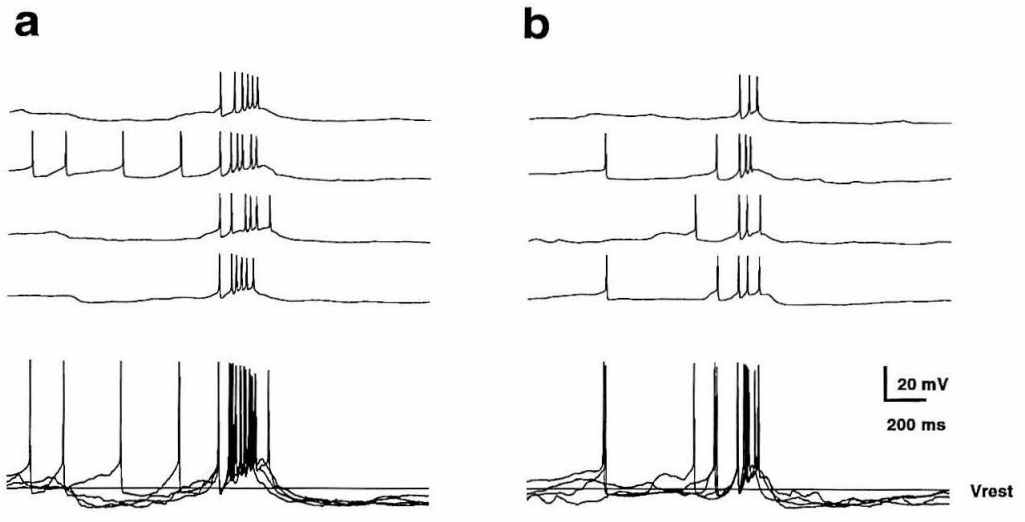


Figure 4.11: The data shown here were extracted from the same stimulus and background periods that were used in the previous figure. In this analysis, a group of spikes was classified as a burst if the first two spikes were less than 8ms apart. Each burst was sorted according to the number of spikes in the 30ms window following the first spike. Burst with a greater number of spikes (**a**) tended to be preceded by a lower pre-potential than burst with fewer spikes (**b**).

indicate that the time constant of the cell was much less than that of the stimulus-driven depolarization.

Syllable-specific inhibition

Inhibition plays a central role for some models of temporal combination sensitivity (Margoliash, 1983; Lewicki and Doupe, 1993), thus it is important to establish whether specific stimuli can differentially evoke inhibitory post-synaptic potentials (IPSPs). An example of syllable-specific inhibition from extracellular records is shown in figure 4.13. 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.

We have also observed rebound from inhibition (figure 4.14). The depolarization following the inhibition is consistent with the inhibitory rebound model of temporal combination sensitivity proposed by Margoliash (1983). Thus far, however, we have not observed a case where rebound has played a direct role in temporal combination sensitivity. Data shown in figure 4.15 shows syllable specific inhibition without rebound. These data suggest that the

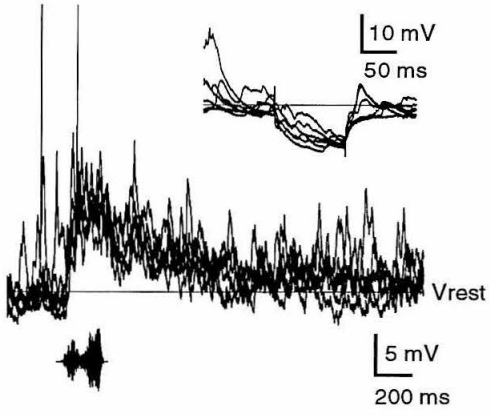


Figure 4.12: Long-lasting depolarization in response to a single syllable in the bird's own song. The inset shows the membrane response to −100 pA current pulses made prior to each stimulus presentation (note different time scales). The action potentials are clipped at −40 mV. The horizontal line indicates the resting potential of the cell, which was −75 mV.

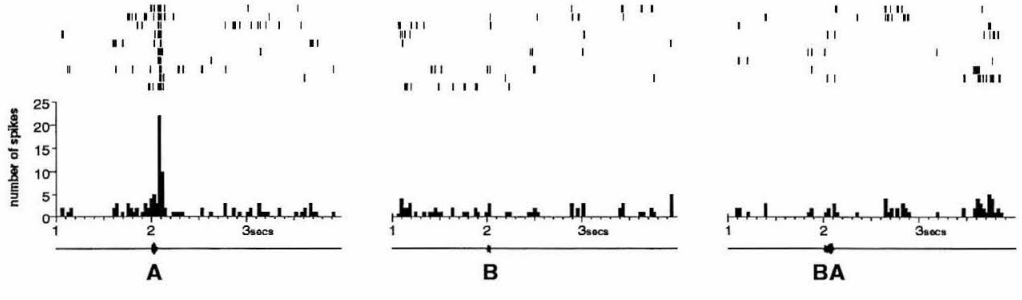


Figure 4.13: Syllable-specific inhibition. The data were recorded extracellularly from a single HVc neuron with song syllables A, B, and BA interleaved. Action potential rasters from individual trials are shown above the PST histograms. Inhibition specific to syllable B completely abolished the response normally evoked by A.

inhibition produced by syllable **A** may be a prerequisite for the strong response produced by the syllable pair **AB**.

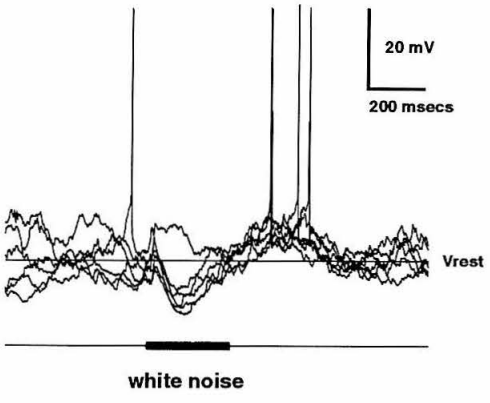


Figure 4.14: *In vivo* intracellular recording from a neuron ventral to HVc. The cell shows rebound from inhibition which is consistent with model of temporal combination sensitivity proposed by Margoliash (1983).

Temporal combination sensitivity

An intracellular recording of a temporal combination sensitive neuron is shown in figure 4.15a. This 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 figure 4.15a is that a combination of inhibition followed by excitation produces the burst firing (Jahnsen and Llinas, 1984; Steriade and Deschenes, 1984). 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 if the firing pattern was altered. The cell in figure 4.15a was given a series of hyperpolarizing current injections ranging from –100 pA to a maximum of –800 pA with a duration ranging from 150 to 200 msecs. Each hyperpolarizing current injection was

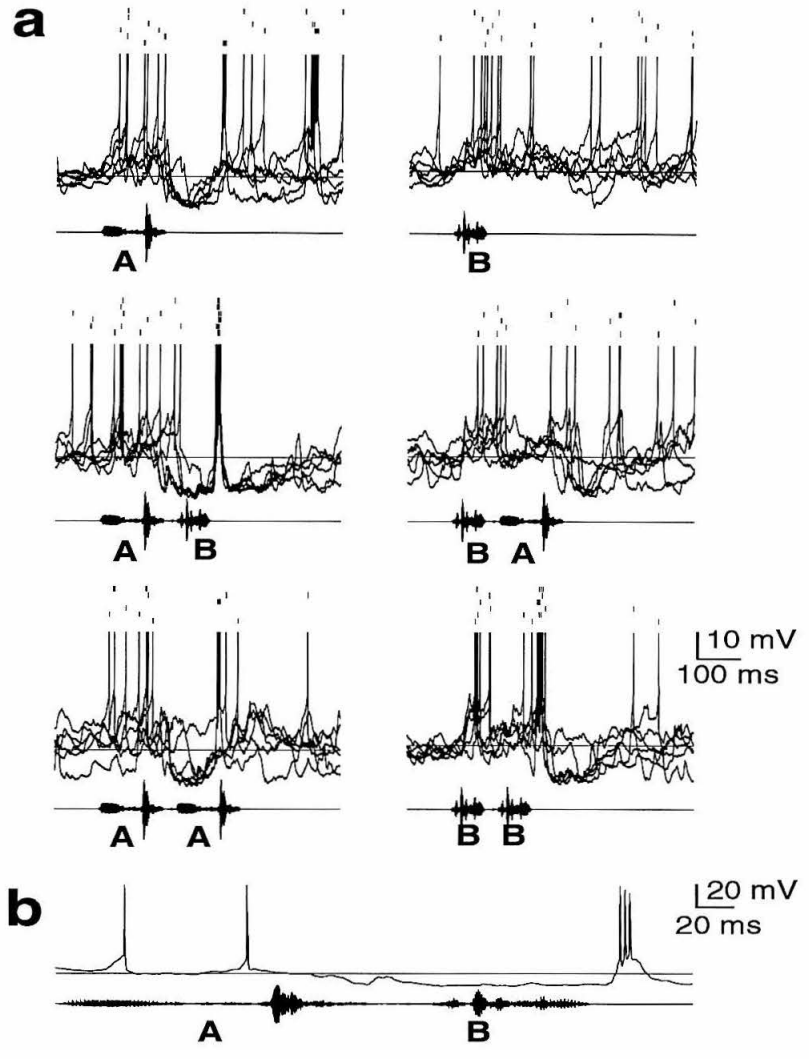


Figure 4.15: (**a**) An *in vivo* intracellular recording of a temporal combination sensitive HVc neuron. Each panel shows the spike raster of six trials (top), the overlaid intracellular traces (middle), and the oscillogram of the stimulus (bottom). Action potentials are clipped to −35 mV. The horizontal 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**) An example of the burst of action potentials which was produced after every presentation of the syllable pair AB.

followed by a depolarizing current injection ranging from 100 to 400 pA with a duration of 100 msecs. In none of these tests was it possible to elicit burst firing. One example is shown in figure 4.16. 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.

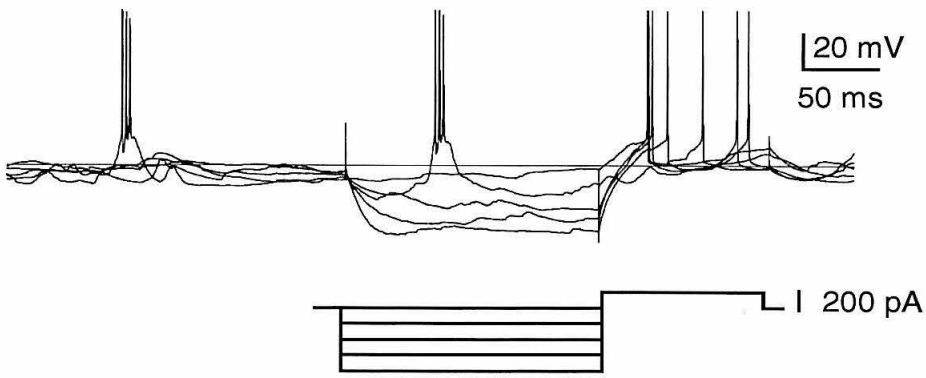


Figure 4.16: Hyperpolarization followed by suprathreshold depolarization does not elicit bursting even though spontaneous bursts continue to occur. The traces show the cell in figure 4.15 which generated spike bursts only in response to a pair of syllables. The lower diagram illustrates the current injection profile. A number of injection profiles were tried (see text) but none were able to elicit spike bursts.

The current injection results are consistent with those of Kubota and Saito (1991b) 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 figure 4.15b, 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. Figure 4.15a, 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.

Morphological description and projections of HVc cells

Three cells were successfully filled with biocytin and stained with avidin-HRP. All had clear axonal projections that could be traced to their targets and are shown in figure 4.17. One of these cells, shown in figure 4.17a, was song-specific (figure 4.3) and sensitive to temporal combinations (figure 4.15). The cell had a soma diameter of approximately 15 μm, and thin, spinous dendrites (shown in figure 4.17b), and a dendritic arborization of about 125 μm. It had a clear axonal projection to area X. This morphology corresponds most closely to a cell of the TD class based on Golgi stains (Nixdorf et al., 1989) which were found to be sexually dimorphic in canaries.

The cells shown in figure 4.17c and d had clear projections to nucleus RA. These cells had no auditory response. Both cells had similar morphology. The somatic diameter was approximately 15 μm. The diameter of their dendritic arborization was about 100 μm.

These observations are consistent with Katz and Gurney (1981) who reported that HVc auditory neurons project to area X and non-auditory neurons project to RA. With such small numbers, they do not rule out the possibility that some auditory neurons project to RA which would be expected from physiological evidence that RA contains auditory neurons via projections from HVc (Doupe and Konishi, 1991; Vicario and Yohay, 1993).

4.4 Discussion

Hyperpolarization during song

Perhaps the most surprising result contained in these data is that some song-specific cells are hyperpolarized during forward song, stimuli which also generate the greatest response. This hyperpolarization was also accompanied by phasic bursts of action potentials in all three cells observed. The reason for the hyperpolarization is unclear, but one possibility is to increase the reliability of the spike timing. Mainen and Sejnowski (1995) reported

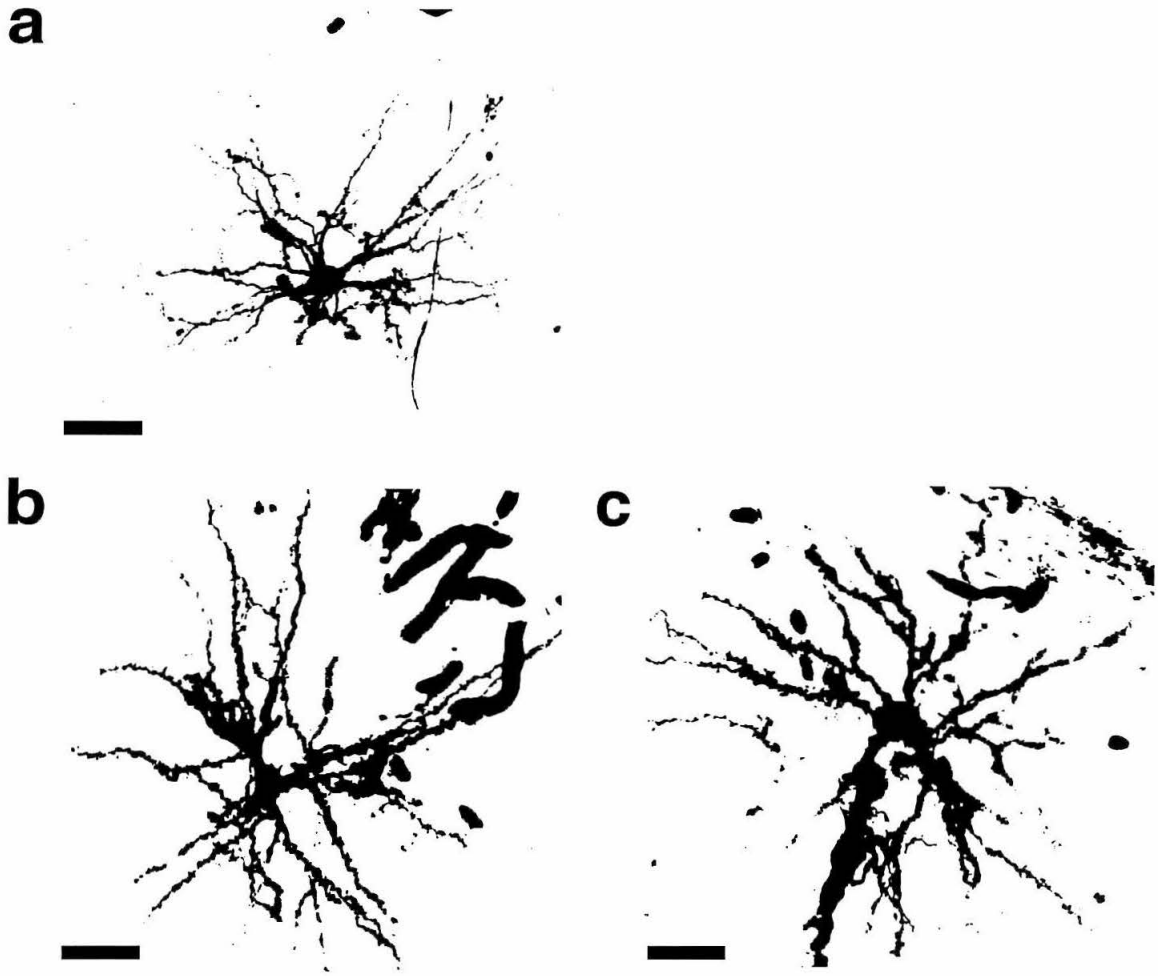


Figure 4.17: Photomicrographs of avidin horseradish peroxidase-stained neurons. (**a**) A song-specific and temporal combination sensitive HVc cell (figures 4.3 and 4.15). The axon of this cell could be traced to area X. (**b** and **c**) Two non-auditory HVc cells. Both of these cells had clear projections to RA. All scale bars are 20μm.

that neurons spike more reliably when starting from a hyperpolarized state than from rest. The temporal alignment of the phasic bursts observed in song-specific cells can be within 5 msecs. High regularity is even seen in the sub-threshold membrane response. Precise timing is likely to be an important feature of both the auditory and motor code of the song system. The hyperpolarization may subserve the creation and/or preservation of this code.

Another function of the hyperpolarization is to elevate the spiking threshold. This could be to ensure that the neuron does not spike at inappropriate times and only spike when there is a large depolarizing input. In this sense, a cell that is hyperpolarized would increase its signal to noise ratio during the song. This predicts that depolarization of the membrane potential during the forward song should decrease the reliability of the spike timing.

The cells hyperpolarized during forward song all showed less hyperpolarization during syllable reversed song and little or none during reversed song. This observation underscores the unique nature of song-specific neurons and also suggests a mechanism by which these cells can integrate auditory context over hundreds of milliseconds. Since time constant of the hyperpolarization is on the order of several hundred milliseconds, acoustic cues which could induce it could have a long-lasting effect on the state of the cell. Since the amount of hyperpolarization is much greater in forward song than to the syllable reversed song, these data suggest that hyperpolarization is one of the mechanism underlying syllable order sensitivity.

These data also provide evidence that the hyperpolarization seen in song-specific cells can have several time courses. Hyperpolarization lasting only 100 msecs was seen in response to a single syllable (figure 4.15). Longer lasting hyperpolarization was seen in response to a syllable pair, but only after a burst of action potentials. This suggests that the spikes themselves may be involved in eliciting the long lasting hyperpolarization.

Experiments performed in HVc brain slices suggests that high frequency firing, such as in spike bursts, activates long-lasting hyperpolarizing currents (Na-activated K^+ currents) (Kubota and Saito, 1991b). A hyperpolarization of similar time course was activated in response to forward song (figures 4.5 and 4.8) and could also be invoked *in vivo* with current injection (figure **??**). This current, however, does not explain the hyperpolarization seen in figure 4.5, which is present even though there are no more spikes during the forward song than during the reversed song where there is no hyperpolarization. One explanation is that the induction of the hyperpolarization is related to the spike rate, which is much higher

during a busrt. Another explanation is that it is synaptically driven. Several receptors have been shown to cause hyperpolarization in HVc. These include $GABA_A$ $GABA_B$ and metabotropic glutamate receptors (Schmidt and Perkel, 1995). One way to test these possibilities would be to manipulate the membrane potential with current injection during presentation of the song.

Bursting

The significance of the bursting accompanied by the hyperpolarization is unclear. Bursting has been suggested to play a role in visual processing in the lateral geniculate nucleus serving as a non-linear amplification of the incoming visual signals (Lu et al., 1993). The role bursting cells observed in HVc, however, would seem to be different, since we have not observed cells that alternate between to different modes for the same stimulus. Phasic burst firing is a plausible code for the motor program underlying song production, since muscle movements required to produce song must be precisely timed. Bursts of action potentials may help to ensure the precise timing of efferent cell firing. The process of song learning requires auditory feedback for normal song production (Konishi, 1965). Since song-specific cells emerge during vocal learning (Volman, 1993; Doupe and Konishi, 1992), it is possible that they play a role in this process. It is reasonable that the auditory feedback, which carries information about the bird's own vocalizations, must also be precisely timed. The phasic song-specific cells may represent a neural code that integrates auditory feedback with the motor program underlying song production.

Extracellular studies have reported that neither precisely-timed firing nor burst firing is seen during forward song in the auditory brain areas afferent to HVc (Lewicki and Arthur, 1995). These afferent areas also do not contain song-specific cells Margoliash (1986); Lewicki and Arthur (1995). This suggests that precisely-timed spike bursting may arise in HVc along with song-specificity. One function of song-specific cells may be to refine the less precisely timed afferent inputs.

Song-specific cells that generate spike bursts are difficult to study extracellularly, because the rapidly attenuating action potentials make it difficult to isolate single neurons. The intracellular recordings in this study underscore how much the action potential shape can change. The suggests that, unless isolation is very good, extracellular recordings may be biased against bursting neurons. Techniques such as "loose patch" extracellular record-

ing may circumvent this problem and allow easier study of these highly selective auditory neurons.

Song-specificity

Tonically excited song-specific cells showed strong depolarization to forward song and weaker depolarization to the reversed or syllable reversed song. Although these data allow no means to distinguish between a lack of excitation and balanced of excitation and inhibition, a simple explanation for this observation is that these cells receive less excitatory input during the reversed song or during the syllable reversed song. This suggests that these song-specific cells are integrating the output of neurons that already show some preference to forward song.

Neurons that show some preference to forward song are known to be present in field L, a group of auditory forebrain areas afferent to HVc (Margoliash, 1986). The selectivity of these neurons can be accounted for by their sensitivity to the direction of frequency modulation (FM) (Bonke et al., 1979b; Leppelsack, 1983; Muller and Leppelsack, 1985; Hose et al., 1987; Knipschild et al., 1992). Downward FM is a prominent feature in the syllables of zebra finch song; reversing the song changes downward FM to upward FM. Thus, an HVc cell could respond more to forward song than to reverse song by integrating the output of field L neurons sensitive to downward FM. This explanation is consistent with absence of depolarization to reversed song observed in the intracellular recordings of song-specific neurons presented here.

Temporal Combination Sensitivity

Sensitivity to FM does not explain the lack of response to the syllable reversed song in which the acoustic structure of each syllable is identical to that in the forward song. The data presented here show that the response of song-specific cells is also dependent on the syllable order which is in agreement with extracellular studies (Margoliash, 1983; Margoliash and Fortune, 1992). Temporal combination sensitivity is one way to account for the sensitivity of song-specific cells to the temporal order of song syllables. These data provide new insights into the mechanisms that may underlie this response property.

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 non-linear 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 non-linear response could be generated by either the spiking threshold or by burst-firing.

Margoliash (1983) proposed a model for temporal combination sensitivity which 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. This does occur, but often, as in figure 1.4, 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 in order 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 (Casseday et al., 1994) 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 (Kubota and Saito, 1991b). 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 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 (Jagadeesh et al., 1993). *In vitro* experiments in the songbird have demonstrated the presence of NMDA, AMPA, and $GABA_A$ potentials in the HVc (Vu and Lewicki, 1994). 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 figure 4.18. Syllable A evokes no response in the TCS 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 around 50-100 ms, so a $GABA_A$ 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.

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 figure 4.13a. 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 figure 4.18 but with the roles of cell A and cell B_1 reversed. Many of the HVc cells reported by Margoliash and Fortune (1992) also had simpler response properties and often exhibited a preference for particular syllables.

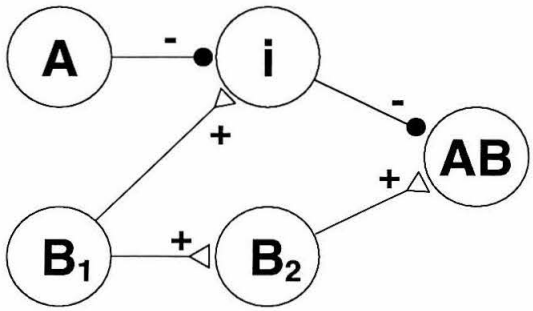


Figure 4.18: The figure illustrates how observed synaptic mechanisms can be used in a neural circuit to obtain temporal combination sensitivity. This network was constructed to model a typical temporal combination sensitive cell. Solid circles represent inhibitory inputs. The triangles represent excitatory inputs with a time course roughly equal to the inhibitory inputs.

Concluding remarks

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 by 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 HVc of the adult bird is not clear, there is evidence that song-specific neurons arise during the song learning process (Doupe and Konishi, 1992; Volman, 1993), suggesting that song-specific neurons may play some role in the song learning process which requires auditory feedback for normal song to develop. HVc projects to area X which is essential during song learning (Bottjer et al., 1984; Scharff and Nottebohm, 1991), but is not required for normal song production in adult birds Nottebohm et al. (1976). The results of biocytin fills presented here indicate that song-specific cells can project to area X, which is consistent the hypothesis that they play a role in song learning. Understanding the mechanisms underlying the encoding of temporal order may elucidate aspects of the vocal learning process as well as provide insight into how the songbirds learn and memorize songs from auditory experience.

Chapter 5 Computational Models for Representing and Learning Temporal Structure

5.1 Introduction

A fundamental problem solved by biological neural systems is discovering good representations for complex sensory patterns. Often, as in the case of young birds memorizing songs from auditory experience, the biological system receives no obvious feedback about which features in the data are important and which are not. The system must discover that for itself. Understanding the computational principles underlying these processes is the main goal of theoretical research in unsupervised learning.

Classical approaches to unsupervised learning algorithms involves finding an efficient representation of the data, for example by vector quantization or principal component analysis (Duda and Hart, 1973; Gray, 1984; Sanger, 1989). More recent work has investigated approaches that try to discover the underlying structure or features that make up the patterns (Barlow, 1989; Foldiak, 1989; Redlich, 1993; Saund, 1995; Hinton et al., 1995). These approaches have been successfully applied to spatial patterns, but less work has been done on temporal patterns.

Much of the previous work on learning temporal structure has focused on algorithms that require supervision. These include speech recognition algorithms based on hidden Markov models (Rabiner, 1989), dynamical networks (Kleinfeld, 1986; Sompolinsky and Kanter, 1986; Tank and Hopfield, 1987), and time-delay backpropagation networks (Waibel, 1989). Unsupervised algorithms for learning temporal structure have also been proposed. For example, Dehaene et al. (1987) described a network composed of elements called "synaptic triads" which detect a transition of activity from one neuron to another. Competitive learning schemes have also been applied to temporal patterns (Chappell and Taylor, 1993; Granger et al., 1994). Other approaches have used oscillations to encode temporal sequences (Wang and Mccormick, 1993) or have encoded time implicitly in the dynamics of an attractor neural network (Amit and Brunel, 1995).

The goals of this chapter are to develop algorithms that discover temporal structure without supervision using well-defined computational goals. Here, we present two computational approaches for representing and learning temporal structure. The first assumes the times of the events of interest are known and that events are real-valued vectors. The second part develops temporal representations based on stochastic binary features. This representation allows for the learning distributed representations of temporal structure in which the event times are inferred. This framework is developed first for spatial patterns and then extended for temporal patterns.

5.2 Event-Based Representations of Temporal Structure

Temporal structures have a rigidity which varies along a continuum from functions of time, $f(t)$, to sequences of discrete events, such as letters in text, where the only structure is in the order of the events. We can obtain intermediate forms of temporal structure by relaxing some of the rigidity. For example, instead of events which are of fixed patterns, we might have events which are described by an underlying distribution. Instead of temporal structure being described by the event order, the time intervals between the events might also be described by a distribution. By varying the spread of the underlying distributions, we can obtain varying degrees of temporal rigidity. In this section, we consider the problem of representing and learning this form of temporal structure.

Continuous hidden Markov models with variable duration

We assume the input data is composed of a sequence of discrete events which are real vectors drawn from unknown underlying distributions. The learning task is to infer from the data both the statistical structure of the distributions and the order relations and relative timing structure in the events. We assume that it is not necessary to detect the occurrence of an event, but that the data is simply a list of events each with an associated time of occurrence.

The spatial distribution of events is modeled as a Gaussian mixture. The model of an event $\mathbf{x}$ from class C is defined by $P(\mathbf{x}|\mu, C)$, which is assumed to be a spherical Gaussian with center μ. When the event class is unknown, the probability of an event is determined

by marginalizing over all possible Gaussian clusters

$$P(\mathbf{x}|\mu_{1:\mathrm{K}}) = \sum_k P(\mathbf{x}|\mu_k, C_k)P(C_k) \tag{5.1}$$

where $P(C_k)$ is the a priori probability of an event originating from cluster C_k, and $\mu_{1:\mathrm{K}}$ specifies the cluster centers for clusters $C_1, \ldots, C_K$.

A simple model of the temporal structure among the events is to describe the probability of transitions among the event clusters. Making the Markov assumption that the current event depends only on the previous event, one can obtain a simple expression for the probability of a sequence of events $\mathbf{x}_{1:\mathrm{N}}$

$$P(\mathbf{x}_{1:\mathrm{N}}) = P(\mathbf{x}_1) \prod_{n=2}^{N} P(\mathbf{x}_n|\mathbf{x}_{n-1}) \tag{5.2}$$

Incorporating the mixture model that describes the events $\mathbf{x}$, we have

$$\begin{aligned} P(\mathbf{x}_n|\mathbf{x}_{n-1}) &= \sum_j P(C_j^{(n-1)}|\mathbf{x}_{n-1})P(\mathbf{x}_n|C_j^{(n-1)}, \mathbf{x}_{n-1}) \\ &= \sum_j P(C_j^{(n-1)}|\mathbf{x}_{n-1}) \sum_i P(C_i^{(n)}|C_j^{(n-1)})P(\mathbf{x}_n|C_i^{(n)}) \end{aligned} \tag{5.3}$$

where $C_i^{(n)}$ indicates the ith cluster for the nth event.

In addition to modeling the transition probabilities between event clusters, one can also model the delays between pairs of events. The probability of the delay separating the pair of events $\mathbf{x}_a$ and $\mathbf{x}_b$ is $P(d_{ab}|C_{ab})$ which is modeled as a gamma distribution. We assume the distributions describing the event classes and event delays are independent. This type of model corresponds to a generalized hidden Markov model with continuous states and variable durations (Levinson, 1986; Ljolje and Levinson, 1991).

A network implementation

A parallel form of the algorithm can be implemented in a network model. Instead of representing all possible event transitions, the model fits the data with a fixed number of event pairs. This type of representation allows for a sparse representation of the transition space. Furthermore, we allow each unit to model both the distribution of the current event and of the previous event. This is a departure from the probabilistic model described above,

since the cluster densities are overrepresented. We use this approximation to simplify the calculations and to keep the representation of event pairs local to a unit. This corresponds to a probabilistic model in which the data are assumed to be independent event pairs.

Each unit in the network represents a pair of events separated by a variable delay. The activity of each unit y_k is proportional to the likelihood of each event pair, $\mathbf{x}_a$, followed by $\mathbf{x}_b$ after a delay of d_{ab}

$$P(\mathbf{x}_a, d_{ab}, \mathbf{x}_b | y_k, \mathbf{w}) = P(\mathbf{x}_a | y_k, \mathbf{w}) P(d_{ab} | y_k, \mathbf{w}) P(\mathbf{x}_b | y_k, \mathbf{w}) \tag{5.4}$$

where $\mathbf{w} = \{\mathbf{a}_k, \hat{d}_k, \mathbf{b}_k\}$. The centers for events $\mathbf{x}_a$ and $\mathbf{x}_b$ are $\mathbf{a}_k$ and $\mathbf{b}_k$ respectively, and $\hat{d}_k$ is the mean of $P(d_{ab}|y_k, \mathbf{w})$. In the network, we assume all transition probabilities to be equal. A graphical depiction of the model for a pair of events is shown in figure 5.1.

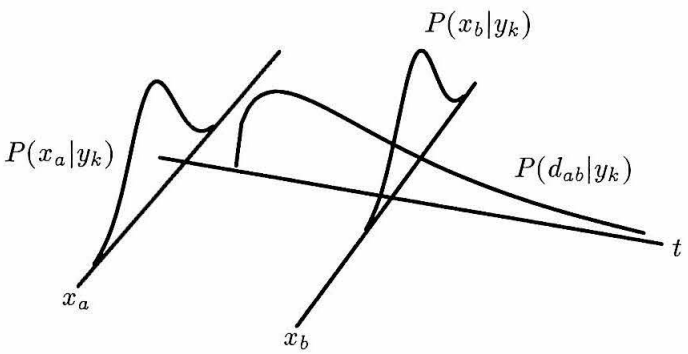


Figure 5.1: Each unit, y_k, in the network represents an event pair by a probabilistic description of the two spatial patterns x_a and x_b, which are, respectively, the previous and current input. $P(x_a|y_k)$ and $P(x_b|y_k)$ represent the distribution of inputs for a pair of events. The model represents the distribution of temporal intervals between the events with $P(d_{ab}|y_k)$. The output a unit is proportional to the product of these likelihoods.

The model is adapted by maximizing a function analogous to (5.2).

$$\mathcal{L} = \prod_{n=2}^{N} P(\mathbf{x}_{n-1}, d_{n-1,n}, \mathbf{x}_n | \mathbf{w}) \tag{5.5}$$

Since we are assuming a mixture model, we have

$$P(\mathbf{x}_{n-1}, d_{n-1,n}, \mathbf{x}_n|\mathbf{w}) = \sum_k P(\mathbf{x}_{n-1}, d_{n-1,n}, \mathbf{x}_n|y_k, \mathbf{w})P(y_k|\mathbf{w}) \tag{5.6}$$

where $P(y_k|\mathbf{w})$ is the a priori probability of unit y_k which we assume to be equal for all units.

The gradient for the cluster center representing the previous event of unit k is given by

$$\frac{\partial \log \mathcal{L}}{\partial \mathbf{a}_k} \propto \sum_n P(y_k|\mathbf{x}_a, d_{ab}, \mathbf{x}_b, \mathbf{w})(\mu_a - \mathbf{x}_a) \tag{5.7}$$

The equation for $\partial \log \mathcal{L}/\partial \mathbf{b}_k$ has identical form, and adaptation of the parameters defining the delay distributions is also straightforward. These equations allow the parameters of the model to be optimized with a standard gradient descent procedure. It is also simple to derive re-estimation formulas for the cluster centers for batch training. In this case, the cluster means are estimated directly after each pass through the dataset and have much faster convergence.

It is important to note that the network can adapt using only local information and lateral inhibition. This can be seen by inspecting the components of the gradient. The direction of change for each weight is determined by $\mu_a - \mathbf{x}_a$. Thus, each weight adapts according to the difference between its value and the value of the input. Since we have assumed the a prior probabilities of the units are equal, the weight change is scaled by

$$P(y_k|\mathbf{x}_a, d_{ab}, \mathbf{x}_b, \mathbf{w}) = \frac{P(\mathbf{x}_a, d_{ab}, \mathbf{x}_b|y_k, \mathbf{w})}{\sum_k P(\mathbf{x}_a, d_{ab}, \mathbf{x}_b|y_k, \mathbf{w})} \tag{5.8}$$

The term in numerator, $P(\mathbf{x}_a, d_{ab}, \mathbf{x}_b|y_k, \mathbf{w})$, is the activity of the unit. The term in the denominator is a normalizing coefficient, which scales the weight change by the activity of the whole network. If the units in the network respond equally to an event pair, the weight change will be determined by $\mu_a - \mathbf{x}_a$. If one unit responds more than others, then the weight change will be large for that unit, and small for the others. This results in a soft form of "winner-take-all" learning and can be viewed as a temporal extension of maximum likelihood competitive learning (Nowlan, 1990). We illustrate the learning of the network with a simple example shown in figure 5.2.

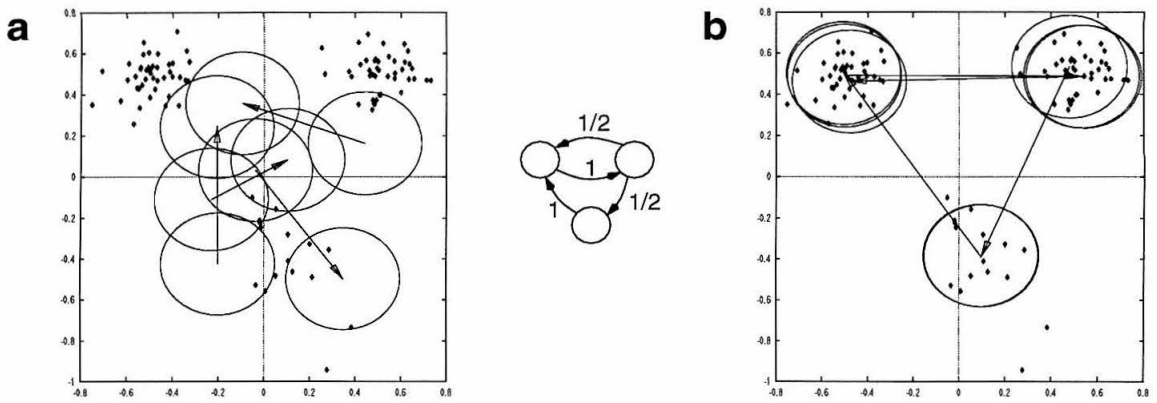


Figure 5.2: Each pair of circles represents a single unit in the network. The arrow between a pair of circles indicates the order of events represented by the unit. Events are indicated by the diamonds. The dataset consists of a series of events sampled from three Gaussian clusters. The order in which the clusters are samples is specified by the small transition diagram with transitions according to the specified probabilities. (**a**) Initial state of the network. The network consists of four units whose initial state is random. (**b**) Finial state of the network after learning. The network has learned the clusters in the dataset and the temporal structure of transitions between the clusters.

5.3 Learning Patterns of Stochastic Binary Features

The models describing sequences of real vectors are limited, because it is assumed that the event times are known. In many situations, however, the temporal events in a data stream are not known a priori and their location and structure must be inferred from the data. In this section, we introduce models that solve these problems. We first introduce a class of models for fitting local representations of patterns of binary features. This is extended to allow distributed representations. Finally, both these cases are extended to allow representations of temporal patterns.

Learning local representations

The assumption of the model is that the input data is composed of patterns that can be well modeled by combinations of binary features. What a feature represents is arbitrary. For example, a feature could be a pixel in an image, or a feature could represent a high level concept like "has two legs". Each pattern is characterized by a set of features, but an instance of a pattern need not contain all of the features in this set. For example a

horizontal line segment may not always contain the pixels at the edges, or not all birds may fly. Thus, each feature has a probability of being present given a certain pattern. Furthermore, the features themselves can be noisy. Whether a feature is present in a given pattern is represented by a probability. This representational scheme and the notation used in this section is illustrated in figure 5.3.

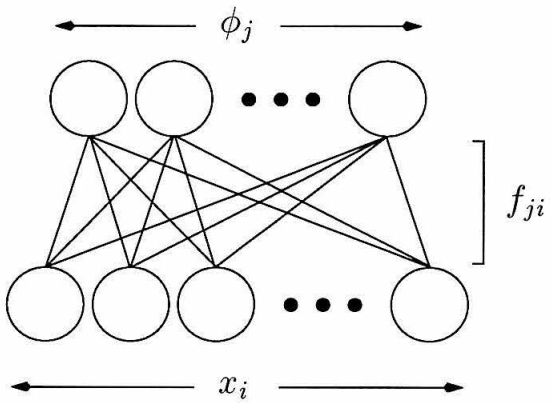


Figure 5.3: The representation used by the network. Each input x_i encodes the likelihood that the ith binary input feature is present. Each unit ϕ_j outputs the probability that a particular combination of features it represents is present. The set of features represented by ϕ_j is defined by the weights f_{ji} which encode the probability of feature i being in an instance of ϕ_j.

The input to the network is a collection of normalized likelihoods, $\mathbf{x}$, which provide information about the underlying binary categories. Each input is real-valued and varies between 0 and 1. The probability of a pattern $\mathbf{x}$, or a particular combination of features given a particular pattern ϕ_j is given by

$$P(\mathbf{x}|\phi_j, \mathbf{f}_j) = \prod_i [x_i f_{ji} + (1 - x_i)(1 - f_{ji})] \tag{5.9}$$

where f_{ji} is the probability that input feature i is present given ϕ_j.

The learning objective is to maximize the probability of a set of patterns $\mathbf{x}_{1:N}$:

$$\mathcal{L} = P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J}) = \prod_n \sum_j P(\phi_j) P(\mathbf{x}_n|\phi_j, \mathbf{f}_j) \tag{5.10}$$

$P(\phi_j)$, the a priori probability of feature combination ϕ_j, is assumed to be unknown and is initialized to be equal across units. This formulation is a mixture model and only one

feature is assumed to be present at a time. This gives a local representation of the input.

Unconstrained optimization is performed using $f = 1/(1+\exp(-w))$ and optimizing w. The gradient of the (log) objective function is

$$\frac{\partial \log \mathcal{L}}{\partial w_{ji}} = \sum_n \frac{(1-2x_{ni})f_{ji}(1-f_{ji})}{x_i f_{ji} + (1-x_i)f_{ji}} \cdot P(\phi_j|\mathbf{x}_n, \mathbf{f}_{1:J}) \tag{5.11}$$

The a priori probabilities of the features are also learned and are re-estimated after each iteration using $P(\phi_j) = \sum_n P(\phi_j|\mathbf{x}_n, \mathbf{f}_j)/N$. A more detailed Bayesian approach to fitting models of this form can be found in (Neal, 1992).

We illustrate some of the properties of the model's representation with a simple example illustrated in figure 5.4. The network is trained on data consisting of a set of patterns each

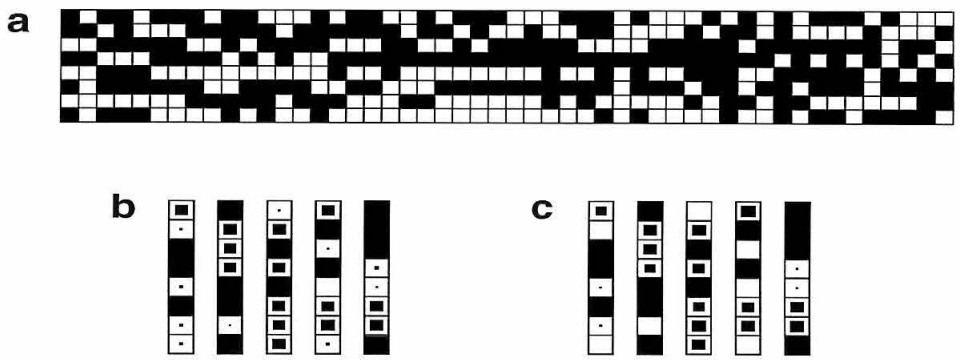


Figure 5.4: An example of training a network using a local representation of stochastic binary features. (**a**) A subset of the patterns used to train the network. Each column represents one pattern. The squares indicate the presence or absence of features in each pattern. (**b**) The patterns used to generate the input data. Each column represents one pattern. The size of a filled square represents the probability of a feature being in a given pattern. A small dot indicates that there is a small probability of that particular feature being present in a pattern; a black square indicates that a feature is always present. (**c**) The patterns learned by the network. These correspond to $\phi_1, \ldots, \phi_5$. Each square represents the value of f_{ji}.

of which is characterized by a set of features which occur with a certain probability. Each pattern j contains feature i with probability $P(i|j)$. A value of 1.0 means a feature is always present in a pattern, 0.0 means it is never present, and 0.5 means it is present half of the time. The value of x_i is the likelihood that the underlying binary feature is present. In this example, there is no uncertainty so $x_i \in \{0, 1\}$. A subset of the 1000 patterns used in the

training data is shown in figure 5.4a. The features used to generate the training data are shown in figure 5.4b.

The network is trained by maximizing (5.10) using (5.11) with conjugate gradient optimization. The resulting solution is shown in figure 5.4c. The value of $\log \mathcal{L}$ for the training dataset when the network parameters are set to the true solution is -4566.1. The value of $\log \mathcal{L}$ for the solution shown in figure 5.4c is -4538.0. The learned model has a higher log likelihood, because that model adapts to the actual frequencies of the patterns and features in the dataset.

Learning distributed representations

Often patterns contain several independent components. Images, for example, often contain lines and sounds are often composed of different tones. If these patterns had to represent by a single ϕ_j as in the previous section, we would need a unit for each pattern. One way to alleviate this problem is to allow patterns to be represented by multiple feature combinations. In this case, the representation is then distributed, because the model can allow for more than one feature combination ϕ_j to be present in a pattern. A set of features combinations $\{\phi_{j_1}, \phi_{j_2}, \ldots\}$ is represented by Φ_k. The probability of the data is again computed by marginalizing, but now over $\mathbf{\Phi}_{1:\mathrm{K}}$

$$\mathcal{L} = P(\mathbf{x}_{1:\mathrm{N}}|\mathbf{f}_{1:\mathrm{J}}) = \prod_n \sum_k P(\Phi_k) P(\mathbf{x}_n|\Phi_k, \mathbf{f}_{1:\mathrm{J}}) \tag{5.12}$$

Following Saund (1995), we assume an OR superposition rule for combinations of different ϕ_j's. The probability of a feature x_i being present given multiple ϕ_j's is given by the probability that the feature is not in any of the individual ϕ_j's.

Since there are 2^{K-1} possible combinations of K ϕ_j's it is infeasible to use arbitrary mixtures. One useful approximation is to limit the maximum number of ϕ's for the set of Φ's: $\Phi_k = \{\phi_{j_1}, \ldots, \phi_{j_M}\}$. For the examples given here, we restrict M to the maximum number independent components of the patterns in the dataset, in which case the approximations are exact.

The probability of a pattern $\mathbf{x}$ is given by

$$P(\mathbf{x}|\Phi_k, \mathbf{f}_{1:\mathrm{J}}) = \prod_i [x_i g_{ki} + (1 - x_i)(1 - g_{ki})] \tag{5.13}$$

where $g_{ki} = 1 - \prod_j (1 - f_{ji})$ and $j = j : \phi_j \in \Phi_k$. The gradient for the distributed case is given by

$$\frac{\partial \log \mathcal{L}}{\partial w_{ji}} = \sum_n \sum_{l:\phi_j \in \Phi_l} \frac{(1 - 2x_{ni}) f_{ji} (1 - g_{li})}{x_i g_{li} + (1 - x_i) g_{li}} \cdot P(\Phi_l | \mathbf{x}_n, \mathbf{f}_{1:J}) \tag{5.14}$$

The a priori probabilities are re-estimated in the same manner as in the local case: $P(\Phi_k) = \sum_n P(\Phi_k | \mathbf{x}_n, \mathbf{f}_{1:J})/N$. The a priori probabilities for the individual ϕ_j's are obtained by marginalization of $P(\Phi_k)$.

We illustrate the algorithm with the lines problem (Foldiak, 1989). The patterns in the dataset are composed of horizontal and vertical lines as illustrated in figure 5.5a. For these examples, the feature probabilities are all set to either one or zero. Is important to note that, although the datasets are displayed as lines on a 2-D grid, the network makes no assumptions about topography. Since each unit receives that same input, all spatial arrangements of the inputs are identical. The features learned by the network are shown in figure 5.5b.

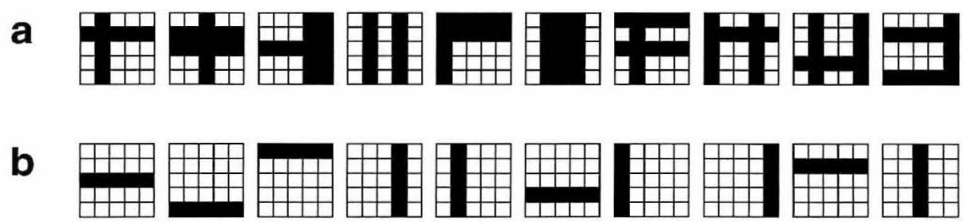


Figure 5.5: An example of training a network using a distributed representation of stochastic binary features. (**a**) A subset of the patterns used to train the network. (**b**) The features learned by the network. A square represents the value of f_{ji}.

Representing temporal structure

We now consider the problem of extending the model to represent temporal features, first addressing the case of local representations. To represent the temporal patterns, we extend the notion of a feature combination ϕ_j to include combinations of features x_i that occur at particular (relative) times. The input to the network is a sequence of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$. The task of the model is to infer what temporal patterns are present in the data. It must also infer the temporal position of the pattern relative to the current time step. Incomplete

data is allowed, so that if only part of the pattern is present, the units ϕ_j still compute the correct probability of that feature being present given the available data.

Since the dataset in the temporal case is a sequence of patterns, the probability of the dataset, $P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J})$, cannot be conveniently written as the product of the probabilities of the patterns at each time step, because the patterns at successive time steps are not independent. If the occurrence time of each temporal pattern were known, $P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J})$ could be computed, since the successive temporal patterns are assumed to be independent. Exact calculation of $P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J})$ requires marginalizing over all possible sequences, S_r, that describe the temporal positions of the features:

$$P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J}) = \sum_r P(S_r|\mathbf{f}_{1:J})P(\mathbf{x}_{1:N}|S_r, \mathbf{f}_{1:J}) \tag{5.15}$$

There are obviously a large number of possible sequences for any reasonable length of data, but (5.15) can be approximated by calculating the sequences S_r over a limited history at each time step. Then we have

$$\mathcal{L} = P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J}) \approx \prod_n \sum_r P(S_r^{(n)}|\mathbf{f}_{1:J})P(\mathbf{x}_n|S_r^{(n)}, \mathbf{x}, \mathbf{f}_{1:J}) \tag{5.16}$$

where $S_r^{(n)}$ specifies a sequence of ϕ's at particular times over the next T time steps. We calculate (5.16) by extending a set of probable sequences from the beginning of the dataset and treating $\mathbf{x}_n$ as the beginning of the time window. As the features in the dataset are learned, the approximation becomes exact. Note that in noisy data, or when the temporal patterns are unknown, there are a large number of equally probable explanations of the data sequence. We make this computation efficient by keeping track of only the most probable sequences, S_r, given the data. Keeping track of multiple sequences simultaneously allows for accurate computation of (5.16), but no longer affords an obvious network implementation. A further approximation could be made by keeping track of only the most probable sequence, which could be represented by the states of the ϕ_j units.

The gradient equation is similar to the local case (equation 5.11), except that the equations are now scaled by the probability of each sequence, $S_r^{(n)}$:

$$\frac{\partial \log \mathcal{L}}{\partial w_{jti}} = \sum_n \sum_r \frac{(1-2x_{ni})f_{jti}(1-f_{jti})}{x_i f_{jti} + (1-x_i)f_{jti}} \cdot P(S_r^{(n)}|\mathbf{x}_{n:n+T-1}, \mathbf{f}_{1:J}) \tag{5.17}$$

For the examples described here, we set T to be one time step longer than the duration of features represented by $\phi_{1:J}$.

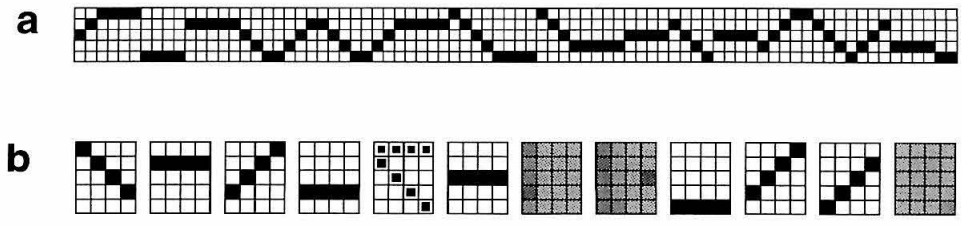


Figure 5.6: An example of training a network using a local representation of stochastic binary temporal features. (**a**) A subset of the sequence of patterns used to train the network. The temporal positions of the features are unknown and are inferred by the network as it learns. (**b**) The patterns learned by the network. A square represents the value of f_{jti}. The grayed features were assigned a priori probability zero by the network and are effectively removed from the network. A sub-optimal solution is shown to illustrate the behavior of the model when one unit is forced to model two features. The 4th feature indicates that 43% of the time the upper inputs are present, and the lower are present 57% of the time. This reflects exactly the relative frequency of these features in the dataset.

An example of training a network on a dataset composed of temporal patterns is shown in figure 5.6. The temporal position of the patterns is arbitrary with the constraint that no two features overlap. In this example, we have not included any gaps between successive features, but the model can also represent gaps as well as temporal features with different durations. The network can learn the optimal solution, but we show a solution at a local maximum figure 5.6b to illustrate the behavior of the model when the number of features in the model is less than the number of features in the data. In this case, the 4th feature represents two separate features in the data. The values of $\mathbf{f_{jti}}$ for that feature indicate that the upper features are present 43% of the time and the lower features are present 57% of the time, which is exactly the ratio of these two features in the dataset. The network also learns the a priori probabilities of each feature, thus if the features do not represent structure in the data, they are assigned probability zero. As the rest of the network learns the structure in the data, the units that are not representing any structure are effectively switched off.

Learning distributed representations of temporal structure

The model for temporal features can also be extend to use distributed representations, by assuming that data in the sequence are generated by multiple ϕ_j's. Like before, we assume an OR superposition rule. We let Φ_k represent combinations of ϕ_j's over different relative positions. The representation of time is the same as in the local condition, except that multiple features may be present at once, at arbitrary relative temporal positions. The probability of the data sequence in the distributed case is similar to the local case (5.15), but each sequence S_r must now describe the positions of all the features for each Φ_k. The gradient of $\mathcal{L} = P(\mathbf{x}_{1:N}|\mathbf{f}_{1:J})$ becomes:

$$\frac{\partial \log \mathcal{L}}{\partial w_{ji}} = \sum_n \sum_{l:\phi_j \in \Phi_l} \frac{(1-2x_{ni}) f_{ji}(1-g_{li})}{x_i g_{li} + (1-x_i) g_{li}} \cdot P(S_r^{(n)}|\mathbf{x}_{n:n+T-1}, \mathbf{f}_{1:J}) \tag{5.18}$$

As in the local case, T is set to be one time step longer than the duration of the features represented by $\mathbf{\Phi}_{1:K}$.

An example of training a network on a dataset composed of combinations of temporal patterns is shown in figure 5.7. The dataset was constructed such that no more than two patterns overlapped. Both the features and their temporal positions must be inferred by the network. The network was initialized with three more units than there were features in the dataset. The network correctly inferred all of the features that make up the dataset and assigned probability zero to the remaining units (figure 5.7a).

Each unit in the network outputs the probability of its feature being present in the data sequence. The probabilities of the units are calculated using all data available, and the units do not require the whole pattern to be present in order to output the correct probability. For example, at the start of the data sequence, it is not known whether particular features are starting, ending, or in the middle of a pattern. The outputs of the units precisely indicates this uncertainty. As more patterns come along, the network uses this additional information to restrict the number of possibilities, which limits the activity to a few units at a time.

The output shown in figure 5.7c was calculated using a flat prior across $\mathbf{\Phi}_{1:K}$, but priors that are better matched to the data will make more accurate predictions. Figure 5.8 shows the output of the network with $P(\mathbf{\Phi}_{1:K})$ inferred from the training data set. The additional knowledge provided by the prior knowledge greatly restricts the number of probable possi-

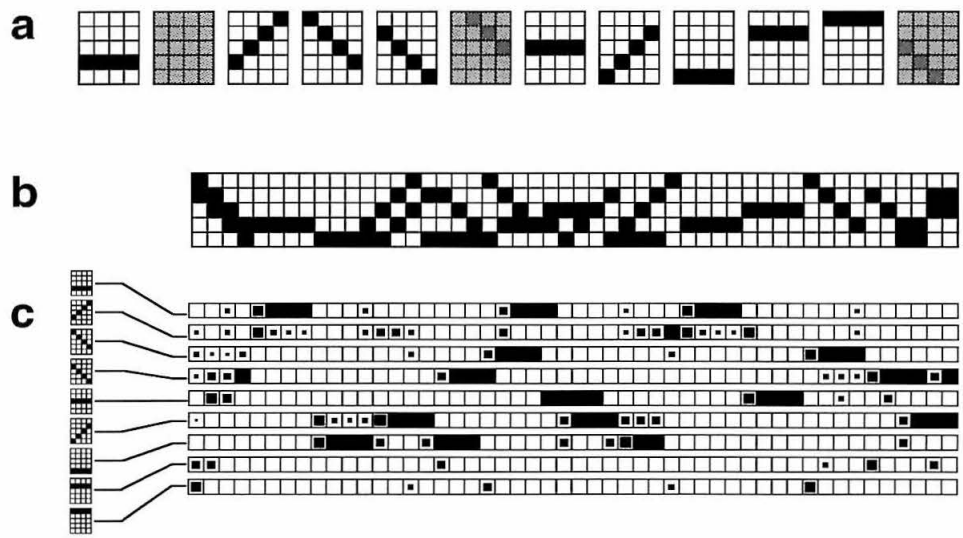


Figure 5.7: (**a**) The temporal features learned by the network. Temporal position is along the horizontal axis. The network makes no assumptions about spatial topography. The grey features were assigned probability zero by the network and never generate output. (**b**) An example of the patterns in the input data used by the network to learn the features. The network was presented with only one pattern at a time without any information about feature occurrence times. (**c**) The output of each unit in the network to the input data in (**b**). The feature represented by the units is shown on the left. Units with zero probability were omitted from the graph. Each unit outputs the probability that its feature is present in the current time step. The area of the filled square in each box is proportional to the unit's output.

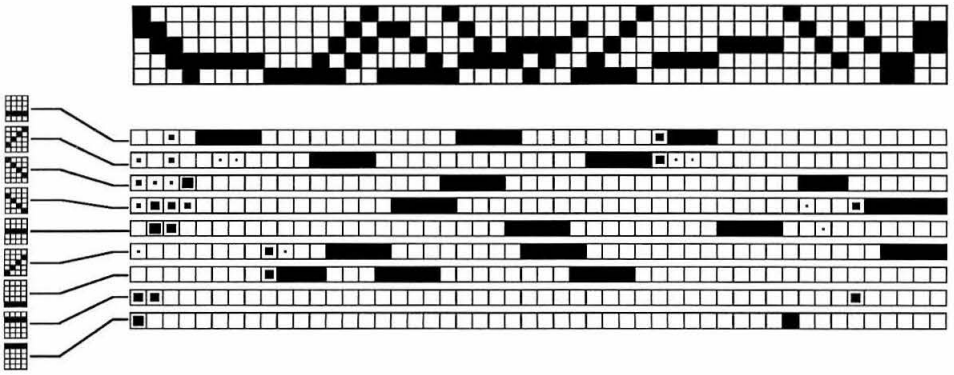


Figure 5.8: The output of the network using learned feature and feature combination occurrence probabilities. In addition to the joint probabilities of features, the network also learns the probabilities of feature combinations at different relative times. Each unit outputs the probability that its feature is present in the current time step. With prior information, the network can be much more certain about which features are present than with the flat priors used in figure 5.7, but it can make bad predictions where the priors are too strong.

bilities. Since $P(\mathbf{\Phi}_{1:K})$ encodes the probability of both the combination of features and their relative occurrence times, the network can in many cases predict which feature is present given only the first pattern in that feature. Since there is relatively few patterns compared to the number of different relative temporal positions of the features, the model's estimation of the a prior values of $\mathbf{\Phi}_{1:K}$ can easily overfit the data. If the priors are too strong the network will make bad predictions, but the incorporation of some prior information allows for much more accurate predictions. One approach for obtaining more accurate estimates would be to introduce a Dirichlet prior on $P(\mathbf{\Phi}_{1:K})$ (Neal, 1992).

Computational issues

The binary feature model can discover a large number of underlying features in patterns of data. This is useful in the case of abstract features that have no implicit spatial topography, but can be limiting if the input patterns are 2D images or a 1D spectrum. For example, the last three examples all have an obvious topographic structure which is completely ignored by the model. Utilizing such information would significantly simplify the computations and make the model better matched to real data.

The temporal features of the binary feature model were developed in the most gen-

eral setting by allowing a different value of f_{ji} for each temporal position in the feature. This allows the network to discover a large number of possible temporal patterns, but also introduces a large amount of complexity in the computation of the probabilities and optimization of the network. Another way that the temporal features could be implemented is to extend beyond simple delay lines by allowing the units to represent feature combinations over a range of temporal intervals rather than precise delays. This would better capture the structure of many realistic temporal patterns, such as natural sounds.

The binary temporal feature model represents temporal context with a distributed representation. This type of representation encodes temporal context much more efficiently that a conventional hidden Markov model in which the history must be represented by a single state (Williams and Hinton, 1991; Ghahramani and Jordan, 1995).

The representation of the input is the likelihood that a feature is present. In the case of the models using distributed representations, multiple patterns of feature combinations are allowed to be present simultaneously. The computation performed by each unit in the network is to compute the probability that a combination of features is present. Thus, the input representation is the same as the output representation. This presents the possibility of arranging the networks hierarchically. Hierarchical Gaussian mixture models have been successfully applied to images (Luttrell, 1994). These methods could be applied to allow a hierarchical network of the models described here to learn higher order feature combinations.

5.4 Discussion

The work is this chapter is motivated by two goals. The first is to understand the computational problems underlying the representation and learning of temporal structure. The second is to understand the representations of temporal structure used by the biological system and how they might be learned. The models presented in this chapter provide two different ways to represent and learn temporal structure.

Models of neural selectivity are by themselves insufficient to explain how a network of neurons learns to represent information. We are interested in not only the mechanisms underlying temporal combination sensitivity, but also in how representations of temporal patterns, based on these mechanisms, are learned.

The event based model presented section 5.2 is interesting from a biological perspective because each unit has a representation of temporal structure, similar to that of neurons in the songbird forebrain which respond to temporal combinations of sounds. The representation of temporal structure is flexible, because a single unit network can respond to a pair of events over a range of temporal delays. Also of interest, is that the network can learn using Hebbian-like adaptation in combination with lateral inhibition. The relevance to biological learning has been noted before with other competitive learning schemes (Ambrosingerson et al., 1990; Coultrip et al., 1992; Granger et al., 1994).

The model based on stochastic binary features presents a new algorithm that can discover temporal features even when there are multiple features present simultaneously. The model is admittedly far removed from biological networks, but both networks solve a similar problem. It will be interesting to use this framework to develop models that provide greater insight into the function of real neural networks.

Chapter 6 Conclusions

In this thesis, we have investigated the organization and mechanisms underlying song-specific cells. The organization of the neural circuitry underlying these cells is hierarchical. We showed that simpler forms of auditory context sensitivity, like sensitivity to temporal combinations, are present in field L. The ability of song-specific cells to integrate large amounts of auditory temporal context, however, is only present in HVc.

Hierarchical organization is nearly universal in the perceptual areas of the brain. The selectivity of song-specific cells, however, does not result simply from the integration of neurons with simpler tuning properties. These data presented here indicate that song-specific neurons achieve their selectivity through a variety of special mechanisms which are not present in afferent areas. Bursting is common in HVc, but was not observed in field L. HVc cells also show a long lasting hyperpolarization both during and following the song. The extracellular data show no indication that such mechanisms are present in field L.

The intracellular recordings of song-specific cells presented in this thesis have given us new insights into mechanisms underlying neuronal coding and selectivity. The temporal pattern of the neuron response of song-specific cells can be highly regular, even at the level of the subthreshold membrane potential. The presence of such precise timing is especially remarkable when we consider that HVc is several synapses efferent to the auditory thalamus. Furthermore, these studies suggest that the temporal precision of spike timing is refined within HVc, since the spike timing in field L is not as precise as that in HVc. In the case of HVc, precise timing is correlated with phasic bursting, and it is thus possible that phasic bursting and its associated hyperpolarization subserve this temporal code. A code that utilizes precise spike timing has obvious utility in the context of song learning and production.

Many lines of evidence point toward network-level computations in HVc (Margoliash et al., 1994; Sutter and Margoliash, 1994). These include the dramatic increase in temporal context sensitivity from field L to HVc, the extensive intrinsic connections of HVc axonal arborizations, and the presence of "higher order" neurons which require several syllables in the correct sequence in order to elicit a response. This thesis has provided much new

information on the properties of single neurons. This type of experimental investigation, however, offers only a limited view of how these neurons function in the context of the overall network. Understanding the network function will require an understanding of the underlying computations.

The computational models present in this thesis represent preliminary steps toward understanding some aspects of how networks of neurons can learn representations of temporal structure. This computation is a crucial aspect of song learning during the sensory period. We have shown the useful representations of temporal structure can be learned using Hebbian adaptation in combination with lateral inhibition. Although these models offer an explanation for the formation of the circuitry subserving song-specific units, a full understanding will require an understanding of the principles underlying the auditory feedback process during song learning and how it relates to song production.

A principle of neuroethology states that neural function is best understood in the context of animal behavior. Intermediate between neuron and behavior is computation. My belief is that combined investigations of neural function and computation will lead ultimately to an understanding of the principles underlying neural representation and learning.

Appendix A Bayesian Modeling and Classification of Neural Signals

A.1 Introduction

Waveforms of extracellular neural recordings often contain action potentials (APs) from several different neurons. Each voltage spike in the waveform shown in figure A.1 is the result of APs from one or more neurons. An individual AP typically has a fast positive component and a fast negative component and may have additional slower components depending on the type of neuron and where the electrode is positioned with respect to the cell. Determining what cell fired when is a difficult, ill-posed problem and is compounded by the fact that cells frequently spike simultaneously which results in large variations in the observed shapes.

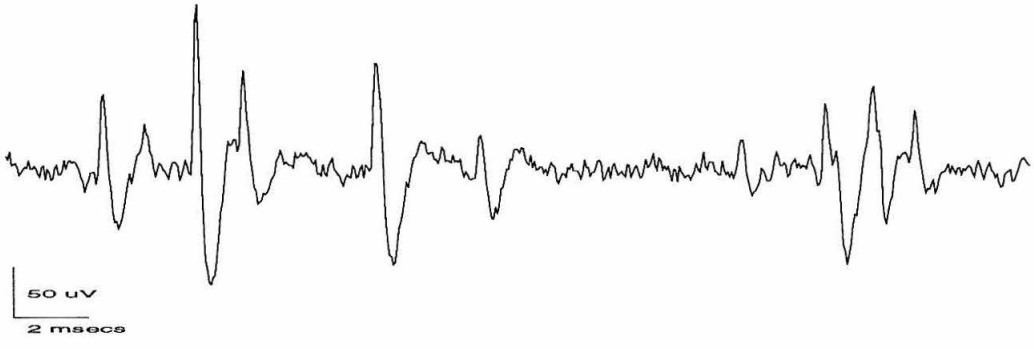


Figure A.1: The extracellular waveform shows several different action potentials generated by an unknown number of neurons. Note the frequent presence of overlapping APs which can, in the case of the right most group, completely obscure individual spikes. The waveform was recorded with a glass-coated platinum iridium electrode in zebra finch nucleus lMAN (courtesy of Allison Doupe, Caltech).

Identifying and classifying the APs in a waveform, which is commonly referred to as "spike sorting," has three major difficulties. The first is determining the AP shapes, the

second is deciding the number of distinct shapes, and the third is decomposing overlapping spikes into their component parts. In general, these cannot be solved independently since the solution of one will affect the solution of the others. Algorithms for identifying and classifying APs (see Schmidt (1984) for a review) fall into two main categories: feature clustering and template matching.

Feature clustering involves describing features of APs, such as the peak value, spike width, slope, *etc.*, and using a clustering algorithm to determine distinct classes in the set of features. Using a small set of features, although computationally efficient, is often sufficient only to discriminate the cells with the largest APs. Increasing the number of features in the clustering often yields better discrimination, but there still remains the problem of how to choose the features, and it is difficult with such techniques to handle overlapping spikes.

In template matching algorithms, typical action potential shapes are determined, either by an automatic process or by the user. The waveform is then scanned and each event classified according to how well it fits each template. Template matching algorithms are better suited for classifying overlaps since some underlying APs can be correctly classified if the template is subtracted from the waveform each time a fit is found. The main difficulty in template matching algorithms is in choosing the templates and in decomposing complex overlap sequences.

The approach demonstrated in this paper is to model the waveform directly, obtaining a probabilistic description of each action potential and, in turn, of the whole waveform. This method allows us to compute the class conditional probabilities of each AP which quantifies the certainty with which an AP is assigned to a given class. In addition, it will be possible to quantify the certainty of both the form and number of spike shapes. Finally, we can use this description to decompose overlapping APs efficiently and to assign probabilities to alternative spike model sequences.

A.2 Modeling Action Potentials

First we consider the problem of fitting a model to events from a single cell. Let us assume that the data from the event we observe (at time zero) is a result of a fixed underlying spike

function, $s(t)$, plus noise:

$$d_i = s(t_i) + \eta_i. \tag{A.1}$$

A computationally convenient form for $s(t)$ is a continuous piece-wise linear function:

$$s(t) = y_j + \frac{v_j}{h}(t - x_j), \qquad x_j \leq t < x_{j+1}, \tag{A.2}$$

where $h = x_{j+1} - x_j, j = 1 \ldots R$, and $v_j = y_{j+1} - y_j$. We will treat R and the x_j's as known. The noise, η, is modeled as Gaussian with zero mean and standard deviation σ_η.

The posterior for the model parameters

From the Bayesian perspective, the task is to infer the posterior distribution of the parameters, $\mathbf{v} = \{v_1, \ldots, v_R\}$, given the data from the observed events, D, and our prior assumptions of the spike model, M. Applying Bayes' rule we have

$$P(\mathbf{v}|D, \sigma_\eta, \sigma_w, M) = \frac{P(D|\mathbf{v}, \sigma_\eta, M)\, P(\mathbf{v}|\sigma_w, M)}{P(D|\sigma_\eta, \sigma_w, M)}. \tag{A.3}$$

$P(D|\mathbf{v}, \sigma_\eta, M)$ is the probability of the data for the model given in (A.2) and is assumed to be Gaussian:

$$P(D|\mathbf{v}, \sigma_\eta, M) = \frac{1}{Z_D(\sigma_\eta)} \exp\left[-\frac{1}{2\sigma_\eta^2}\sum_{i=1}^{I}(d_i - s(t_i))^2\right], \tag{A.4}$$

where $Z_D(\sigma_\eta) = 1/(2\pi\sigma_\eta^2)^{I/2}$. The time of the ith data point, d_i, is taken to be relative to the corresponding event, *i.e.*, $t_i = t_i^{(n)} - \tau^{(n)}$. By convention, $\tau^{(n)}$ is the time of the inferred AP peak. The data range over the predetermined extent of the action potential.[1]

$P(\mathbf{v}|\sigma_w, M)$ specifies prior assumptions of the structure of $s(t)$. Ideally, we want a distribution over $\mathbf{v}$ from which typical samples result only in shapes that are plausible APs. Conversely, this space should not be so restrictive that legitimate AP shapes are excluded. We adopt a simple approach and use a prior of the form

$$P(s(t)|\sigma_w, M) \propto \exp\left[-\int du\, s^{(m)}(u)^2/\sigma_w^2\right], \tag{A.5}$$

where the superscript (m) denotes differentiation. $m = 1$ corresponds to linear splines, $m = 2$ corresponds to cubic splines, *etc.* The smoothness of $s(t)$ is controlled through the

[1] For the examples shown here, this range is from 1 msec before the spike peak to 4 msec after the peak.

parameter σ_w with small values of σ_w penalizing large fluctuations. A prior simply favoring smoothness ensures minimal restrictions on the kinds of functions we can interpolate, but it doesn't buy us anything either. If we had a more informative prior, we would require less data to reach the same conclusions about the form of $s(t)$. Any reasonable prior should have little effect on the shape of the final spike function if there are abundant data. Even though the prior may have little effect on the shape, it still plays an important role in model comparison which will be discussed in section A.4.

The components of the posterior distribution for $\mathbf{v}$ are now defined. There still remains, however, the problem of determining σ_η and σ_w. An exact Bayesian analysis requires that we eliminate the dependence of the posterior on σ_η and σ_w by integrating them out:

$$P(\mathbf{v}|D,M) = \int d\sigma_\eta \, d\sigma_w \, P(\mathbf{v}|D,\sigma_\eta,\sigma_w,M) \, P(\sigma_w,\sigma_\eta|M). \tag{A.6}$$

In this paper, we use the approximation $P(\mathbf{v}|D,M) \approx P(\mathbf{v}|D,\sigma_w^{\mathrm{MP}},\sigma_\eta^{\mathrm{MP}},M)$. The most probable values of $\mathbf{v}$, σ_w, and σ_η were obtained using the methods of MacKay (1992) which we briefly summarize here. First, we transform $\mathbf{v}$ to a basis in which the Hessian of $\log P(\mathbf{v}|\sigma_w,M)$ is the identity. For splines, this is the Fourier representation:

$$s(t) = a_0 + \sum_{j=1}^{\frac{R}{2}-1} \left(a_j \frac{\sqrt{2}\cos 2\pi jt}{(2\pi j)^m} + b_j \frac{\sqrt{2}\sin 2\pi jt}{(2\pi j)^m} \right) + a_{R/2} \frac{\cos 2\pi Rt}{(\pi R)^m} \tag{A.7}$$

using the prior

$$P(\mathbf{w}|\sigma_w,M) = \frac{1}{Z_W(\sigma_w)} \exp\left[-\frac{1}{2\sigma_w^2} \sum_r w_r^2 \right], \tag{A.8}$$

where $\mathbf{w} = \{\mathbf{a},\mathbf{b}\} - a_0$. The term a_0 is set to the known DC level (the offset of the A/D converters). In the limit $R \to \infty$, $\frac{1}{2}\sum_r w_r^2 = \int_0^1 (s^{(m)}(u))^2 du$ (Wahba, 1990) which is the splines regularizer. We take $m = 1$ for linear splines.

The most probable parameter values, $\mathbf{w}^{\mathrm{MP}}$, were determined as follows. Let $E_D = \frac{1}{2}\sum_i (d_i - s(t_i))^2$ and $E_W = \frac{1}{2}\sum_r w_r^2$. Letting $\mathbf{B} = \nabla\nabla E_D$ and $\mathbf{C} = \nabla\nabla E_W$ (around $\mathbf{v}^{\mathrm{ML}}$), we obtain $\mathbf{w}^{\mathrm{MP}} = \frac{1}{\sigma_\eta^2}\mathbf{A}^{-1}\mathbf{B}\mathbf{w}^{\mathrm{ML}}$, where $\mathbf{A} = \frac{1}{\sigma_w^2}\mathbf{C} + \frac{1}{\sigma_\eta^2}\mathbf{B}$. The maximum likelihood values, $\mathbf{v}^{\mathrm{ML}}$, can be determined efficiently by inverting a tridiagonal matrix. The Fourier coefficients can be computed efficiently with the fast Fourier transform.

The most probable values of σ_η and σ_w were obtained using the re-estimation formulas

$\sigma_\eta^2 = 2E_D/(I-\gamma)$ and $\sigma_w^2 = E_W/\gamma$, where $\gamma = \sum \lambda_r/(\lambda_r + \sigma_w^{-2})$ and λ_r is the rth eigenvalue of $\frac{1}{\sigma_\eta^2}\mathbf{B}$. In terms of $\boldsymbol{\lambda}$, $w_r^{\mathrm{MP}} = \lambda_r w_r^{\mathrm{ML}}/(\lambda_r + \sigma_w^{-2})$.

Note that we could at this point apply the methods described by (MacKay, 1992) and discussed later on in section A.4 to compare alternative spike models, in essence to determine the *most probable* spike model given the data. For example, we might choose cubic splines instead of piece-wise linear functions or choose priors that better represented our knowledge about spike shapes. The piece-wise linear spike models discussed here can be made to fit any fixed shape, since they can contain arbitrarily many segments. With 75 segments, the spike models have been descriptively sufficient for all the data we have observed. Situations for which this is not the case will be discussed in section A.9. Figure A.2a shows the result of fitting one spike model to data consisting of 40 APs.

Checking the assumptions

Before proceeding to the more complicated cases of multiple spike models and overlapping spikes, we must check our assumptions on real data. Equation (A.1) assumes that the noise process is invariant throughout the duration of the AP, but in principle this need not be the case. For example, the noise might show larger variation at the extremes. The spike model residuals, $\eta_i = d_i - s(t_i)$, shown in figure A.2a, give no indication of an amplitude-dependent noise process.

A second assumption we have made is that the noise is Gaussian. Figure A.2b shows a Gaussian distribution with the inferred width σ_η overlaid on a normalized histogram of the residuals from figure A.2a. The most significant deviation is in the tails of the distribution which reflects the presence of overlapping spikes. In this case, the overlaps are evenly distributed over the range of the fitted event so they have little effect on the model's form in the limit of large amounts of data. The model would be poorly inferred, however, if the overlaps were not uniformly distributed over the interval, for example if one cell tended to fire within a few milliseconds of another. This is a common problem in practice and will be addressed in section A.5.

An assumption which has not been tested is whether the residuals are independent. Figures A.2c and A.2d show that the noise in these data is slightly correlated. This has little effect on the fit of the models but does affect the accuracy of the probabilities discussed in the later sections. A convenient way of reducing the correlation is to sample close to the

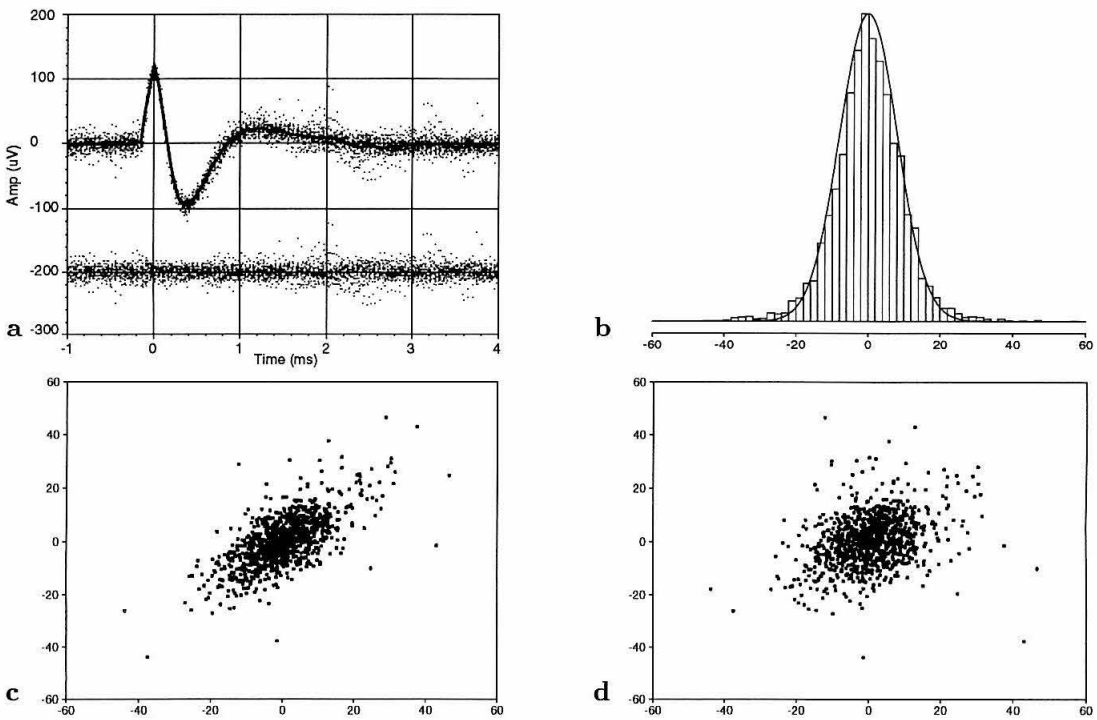


Figure A.2: (**a**) Spike model fit to data consisting of 40 APs. The solid line is a 75 segment piece-wise linear model. Each AP is aligned with respect to the inferred spike peak. Each dot is one sample point. The residual error for each sample, $\eta_i = d_i - s(t_i)$, is offset by -200μV and plotted below. The flat residuals indicate that the data is well-fit by the model. (**b**) Normalized histogram of the residuals from **a**. The curve is the Gaussian inferred with the methods discussed in the text. The outliers result from overlapping APs which can be seen in the data in **a**. (**c** and **d**) Lagged scatter plot of a sample of the residuals in **a**. (**c**) η_i vs η_{i+1}. (**d**) η_i vs η_{i+2}. These graphs indicate that there is some correlation between η_i and η_{i+1} (**c**), but little between η_i and η_{i+2} (**d**). This is expected for these data because the sampling rate (20kHz) was higher than the Nyquist rate (14kHz).

Nyquist rate to avoid correlation introduced by the amplifier filters.

A.3 Multiple Spike Shapes

When a waveform contains multiple types of APs, determining the spike shapes is more difficult because the classes are not known *a priori*. We cannot infer the parameters for one spike model if we don't know what data is representative of its class. Furthermore, if two spike models are similar, it is possible that an observed event could have come from either class with equal probability. The uncertainty of which class an event belongs to can be incorporated with a mixture distribution (Duda and Hart, 1973).

The probability of a particular event, $\mathbf{D}_n$, given all spike models, $M_{1:\mathrm{K}}$, is

$$P(\mathbf{D}_n|\mathbf{v}_{1:\mathrm{K}}, \boldsymbol{\pi}, \sigma_\eta, M_{1:\mathrm{K}}) = \sum_{k=1}^{K} \pi_k P(\mathbf{D}_n|\mathbf{v}_k, \sigma_\eta, M_k), \tag{A.9}$$

where π_k is the *a priori* probability that a spike will be an instance of M_k ($\sum \pi_k = 1$). The joint probability for $\mathbf{D}_{1:\mathrm{N}} = \{\mathbf{D}_1 \ldots \mathbf{D}_N\}$ is simply the product

$$\mathcal{L} = P(\mathbf{D}_{1:\mathrm{N}}|\mathbf{v}_{1:\mathrm{K}}, \boldsymbol{\pi}, \sigma_\eta, M_{1:\mathrm{K}}) = \prod_{n=1}^{N} P(\mathbf{D}_n|\mathbf{v}_{1:\mathrm{K}}, \boldsymbol{\pi}, \sigma_\eta, M_{1:\mathrm{K}}). \tag{A.10}$$

The posterior for multiple spike models is then

$$P(\mathbf{v}_{1:\mathrm{K}}, \boldsymbol{\pi}|\mathbf{D}_{1:\mathrm{N}}, \sigma_\eta, \boldsymbol{\sigma}_w, M_{1:\mathrm{K}}) = \frac{P(\mathbf{D}_{1:\mathrm{N}}|\mathbf{v}_{1:\mathrm{K}}, \boldsymbol{\pi}, \sigma_\eta, M_{1:\mathrm{K}})\, P(\mathbf{v}_{1:\mathrm{K}}|\boldsymbol{\sigma}_w, M_{1:\mathrm{K}})\, P(\boldsymbol{\pi}|M_{1:\mathrm{K}})}{P(\mathbf{D}_n|\sigma_\eta, \boldsymbol{\sigma}_w, M_{1:\mathrm{K}})}. \tag{A.11}$$

We use $P(\mathbf{v}_{1:\mathrm{K}}|\boldsymbol{\sigma}_w, M_{1:\mathrm{K}}) = \prod_k P(\mathbf{v}_k|\sigma_{wk}, M_k)$ and take $P(\boldsymbol{\pi}|M_{1:\mathrm{K}})$ to be flat over $[0,1]^K$ subject to the constraint $\sum_k \pi_k = 1$.

Note that we have implicitly assumed that the spike occurrence times are Poisson in nature with mean firing rates proportional to π_k. This assumes as little as possible about the temporal structure of the spikes. A more powerful description, *e.g.*, modeling the distribution of the inter-spike interval, would be obtained by incorporating this information into (A.10).

Maximizing the posterior

We proceed as before to find the maxima of the posterior which will give us the most probable values for the whole set of spike models. The conditions satisfied at the maxima of $\mathcal{L}$ given in (A.10) are obtained by differentiating $\log \mathcal{L}$ with respect to $\mathbf{v}_k$ and equating the result to zero,

$$\frac{\partial \log \mathcal{L}}{\partial \mathbf{v}_k} = \sum_{n=1}^{N} P(M_k|\mathbf{D}_n, \mathbf{v}_k, \boldsymbol{\pi}, \sigma_\eta) \frac{1}{\sigma_\eta^2} \sum_i [d_{n,i} - s_k(t_i - \tau_n\,; \mathbf{v}_k)] \frac{\partial s_k(t_i\,; \mathbf{v}_k)}{\partial \mathbf{v}_k} = 0, \quad \text{(A.12)}$$

where τ_n is the occurrence time of $\mathbf{D}_n$. Thus we obtain a soft clustering procedure in which the error for each event, $\mathbf{D}_n$, is weighted by the probability that it is an instance of M_k:

$$P(M_k|\mathbf{D}_n, \mathbf{v}_k, \boldsymbol{\pi}, \sigma_\eta) = \frac{\pi_k P(\mathbf{D}_n|\mathbf{v}_k, \sigma_\eta, M_k)}{\sum_k \pi_k P(\mathbf{D}_n|\mathbf{v}_k, \sigma_\eta, M_k)}. \quad \text{(A.13)}$$

Although (A.12) can be solved exactly, it is still expensive to compute, because it uses all of the data. We adopt the approach of estimating each $\mathbf{v}_k$ by fitting each model to a reduced event list allowing the possibility of an event being in the lists of multiple models. These lists are obtained by sampling events from the whole data set and including an event in a model's reduced event list with probability proportional to $P(M_k|\mathbf{D}_n, \mathbf{v}_k, \boldsymbol{\pi}, \sigma_\eta)$. We apply the techniques used in the previous section to determine the values for $\boldsymbol{\sigma}_w$, and in turn the most probable values of $\mathbf{v}_{1:\mathrm{K}}$.

Differentiating (A.10) and finding the condition satisfied at the maximum, we obtain the re-estimation formula

$$\pi_k = \frac{1}{N} \sum_n P(M_k|\mathbf{D}_n, \mathbf{v}_k, \boldsymbol{\pi}, \sigma_\eta). \quad \text{(A.14)}$$

For each model, σ_η can be estimated using the methods of the previous section. The mixture model estimate for σ_η is obtained by a weighted average of the individual estimates using weight π_k.

Selecting events from the data

For these demonstrations, any peak in the waveform that deviated from DC level by more than four times the estimated RMS noise level was labeled as an event, $\mathbf{D}_n$. Once an event

is located, it is important to obtain accurate estimates of the occurrence time (with each spike model) by maximizing (A.4) over τ_n. For the largest models, deviations from the optimal value as little as one-tenth the sampling period will introduce misfit errors greater than σ_η. The τ_n's must be re-estimated as the spike models change for optimal results. An efficient way to perform this optimization is to use the k-d trees discussed in section A.5.

Initial conditions

Since the re-estimation formulas derived here will find *local* maxima, it is critical to use good initial conditions for the spike models. Poor fits will result if there are too few spike models representing what are in fact several distinct APs. Conversely, if there are more spike models than distinct APs, not only will there be excess computational overhead, but there is no guarantee that each AP will be represented, since some spike functions may converge to represent the same AP class. Ideally, we want all potential spike shapes to be represented in the initial spike function set, $s_{1:\mathrm{K}}(t)$. One approach toward obtaining an even representation of the AP shapes is to initialize each spike function to single events so that $\max_t s(t) - \min_t s(t)$ is evenly distributed with a separation proportional to the estimated waveform RMS noise. This approach works well for present purposes, because the height of an AP captures much of the variability among classes. By erring on the side of starting with too many spike models, we can obtain a good initial representation of the AP shapes. There is still a need to decide if two different models should be combined and if one class should be split into two. How to choose the number of spike models objectively will be demonstrated in the next section.

A.4 Determining the Number of Spike Models

If we were to choose a set of spike models which best fit the data, we would wind up with a model for each event in the waveform. We might think of heuristics which would tell us when two spike models are distinct and when they are not, but *ad hoc* criteria are notoriously dependent on particular circumstances, and it is difficult to state precisely what information the rules take into account. A solution to this dilemma is provided by probability theory (Jeffreys, 1939; Jaynes, 1979; Gull, 1988).

To determine the *most probable* number of spike models, we need to derive the proba-

bility of a set of spike models, denoted by $S_j = \{M_{1:K}^{(j)}\}$, conditioned only on the data and information known *a priori*, which we denote by H. From Bayes' rule, we obtain

$$P(S_j|\mathbf{D}_{1:N}, H) = \frac{P(S_j|H)P(\mathbf{D}_{1:N}|S_j, H)}{P(\mathbf{D}_{1:N}|H)}. \tag{A.15}$$

The only data dependent term is $P(\mathbf{D}_{1:N}|S_j, H)$ which is called the *evidence* for S_j. If we assume all the hypotheses $S_{1:J}$ under consideration are equally probable, $P(\mathbf{D}_{1:N}|S_j, H)$ ranks alternative spike sets, since it is proportional to $P(S_j|\mathbf{D}_{1:N}, H)$. With equal priors, the ratio $P(\mathbf{D}|S_i, H)/P(\mathbf{D}|S_j, H)$ is equal to the Bayes factor in favor of hypothesis S_i over hypothesis S_j which is the standard way to compare hypotheses in the Bayesian literature.

The evidence for S_j is obtained by integrating out the nuisance parameters in (A.11):

$$P(\mathbf{D}_{1:N}|S_j) = \int d\mathbf{v}_{1:K} d\boldsymbol{\pi} d\sigma_\eta d\boldsymbol{\sigma}_w \, P(\mathbf{D}_{1:N}|\mathbf{v}_{1:K}, \boldsymbol{\pi}, \sigma_\eta, S_j) \, P(\mathbf{v}_{1:K}|\boldsymbol{\sigma}_w, S_j) \, P(\boldsymbol{\pi}|S_j) \, P(\sigma_\eta, \boldsymbol{\sigma}_w|S_j). \tag{A.16}$$

This integral is analytically intractable, but it is often well-approximated with a Gaussian integral which for a function $f(\mathbf{w})$ is given by

$$\int d\mathbf{w} f(\mathbf{w}) \approx f(\widehat{\mathbf{w}}) \, (2\pi)^{d/2} \, |-\nabla\nabla \log f(\mathbf{w})|^{-1/2}, \tag{A.17}$$

where d is dimension of $\mathbf{w}$, $\widehat{\mathbf{w}}$ is a (local) maximum of $f(\mathbf{w})$, $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$, and the derivatives are evaluated at $\widehat{\mathbf{w}}$. With this we obtain the evidence for spike set S_j,

$$\begin{aligned} P(\mathbf{D}_{1:N}|S_j, H) \;=\; & P(\mathbf{D}_{1:N}|\widehat{\mathbf{v}}_{1:K}, \widehat{\boldsymbol{\pi}}, \widehat{\sigma}_\eta, S_j) \, P(\widehat{\mathbf{v}}_{1:K}|\boldsymbol{\sigma}_w, S_j) \, P(\widehat{\boldsymbol{\pi}}|S_j) \, P(\widehat{\boldsymbol{\sigma}}_w, \widehat{\sigma}_\eta|S_j) \\ & \cdot (2\pi)^{d/2} \, |-\nabla\nabla \log P(\mathbf{D}_{1:N}|\mathbf{v}_{1:K}, \boldsymbol{\pi}, \sigma_\eta, S_j)\,|^{-1/2} \, \Delta \log \widehat{\boldsymbol{\sigma}}_w \, \Delta \log \widehat{\sigma}_\eta \end{aligned} \tag{A.18}$$

where $\Delta \log \widehat{\boldsymbol{\sigma}}_w = \prod_k \sqrt{2/\gamma_k}$, $\Delta \log \widehat{\sigma}_\eta = \sqrt{2/(NI - \gamma)}$, and $d = KR + K + 1$. γ_k is the number of good degrees of freedom for M_k (MacKay, 1992) which can be thought of as the number of parameters that are well-determined by the data. $\gamma = \sum_k \gamma_k$. $P(\boldsymbol{\sigma}_w, \sigma_\eta|S_j)$ is assumed to be separable and flat over $\log \sigma_w$ and $\log \sigma_\eta$. Since the labeling of the models is arbitrary, an additional factor of $1/K!$ must be included to estimate the posterior volume accurately. The Hessian $-\nabla\nabla \log P(\mathbf{D}_{1:N}|\mathbf{v}_{1:K}, \boldsymbol{\pi}, \sigma_\eta, S_j)$ (with respect to $\mathbf{v}_{1:K}$ and $\boldsymbol{\pi}$) was evaluated both analytically and using a diagonal approximation. Both methods

produced similar results, and the latter, being much faster to compute, was used for these demonstrations. Notice that the approximation for the evidence decomposes into the best-fit likelihood for the best fit parameters times the other terms which collectively constitute a complexity penalty called the Ockham factor ((MacKay, 1992)). Since this factor is the ratio of the posterior accessible volume in parameter space to the prior accessible volume, it is smaller for more complicated models. Overly broad priors will introduce a bias toward simpler models. Unless the best-fit likelihood for complex models is sufficiently larger than the likelihood for simple ones, the simple models will be more probable.

A convenient way of collapsing the spike set is to compare spike models pairwise. Two models in the spike set are selected along with a sampled set of events fit by each model. We then evaluate $P(\mathbf{D}|S_1)$ and $P(\mathbf{D}|S_2)$. S_1 is the hypothesis that the data is modeled by a single spike shape; S_2 says there are two spike shapes. Included in the list of spike models should be a "null" model which is simply a flat line at DC. This hypothesis says that there are no events and that the data is a result of only the noise. Examples of this comparison are illustrated in figure A.3. If $P(\mathbf{D}|S_1) > P(\mathbf{D}|S_2)$, we replace both models in S_2 by the one in S_1. The procedure terminates when no more pairs can be combined to increase the evidence.

A.5 Decomposing Overlapping Events

The method of inferring the spike models we have discussed thus far is valid if the event occurrence times can be accurately determined and if the noise is Gaussian and stationary. Often these conditions cannot be met without identifying and decomposing overlapping events. Even if the spike models are good, overlap decomposition is necessary to detect and classify individual events with accuracy.

For a given sequence of overlapping APs, there are potentially many spike model sequences that could account for the same data. An example is shown in figure A.4. We can calculate the probability of each alternative, but there are an enormous number of sequences to consider, not only all possible models for each event but also all possible event times. A brute-force approach to this problem is to perform an exhaustive search of the space of overlapping spike functions and event times to find the sequence with maximum probability. This approach was used by Atiya (1992) in the case of two overlapping spikes with the times

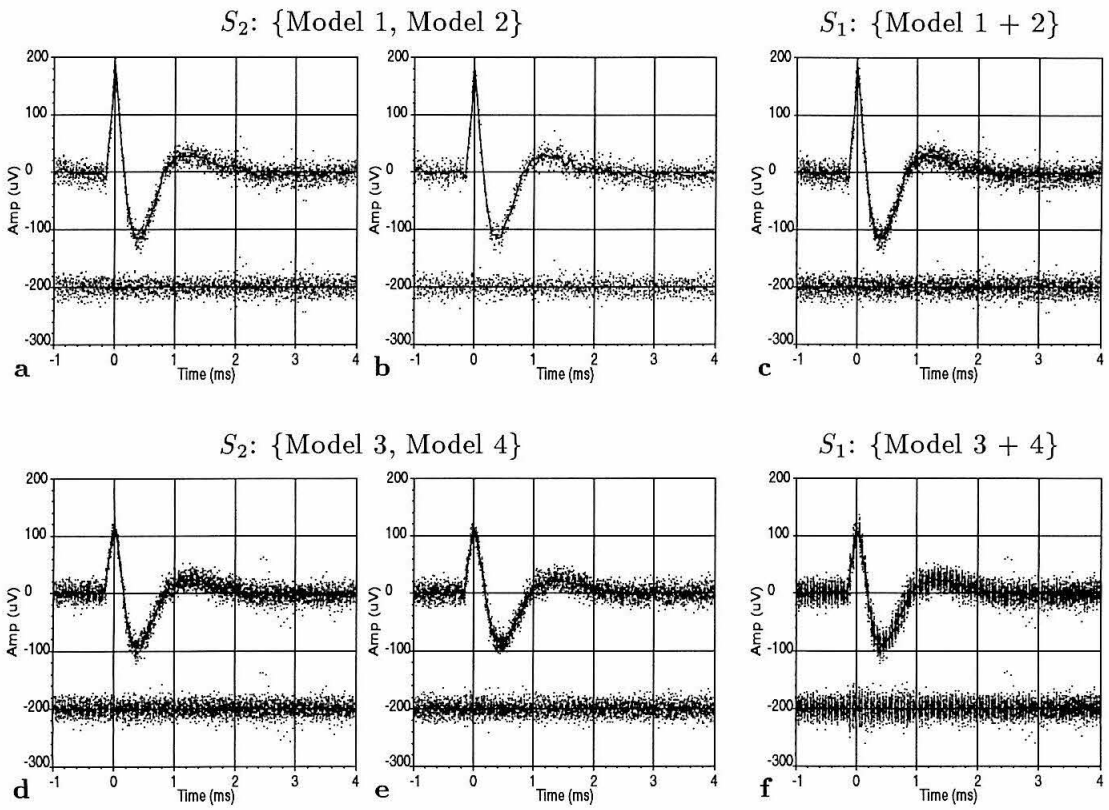


Figure A.3: The most probable number of distinct spike models is determined by evaluating the evidence for alternative hypotheses for a given set of data. Simple hypotheses are generated by selecting similar shapes in a spike set. S_2 is the hypothesis that there are two distinct spike models; the fits of two such models to a sampled set of data are shown in **a** and **b**. S_1 is the hypothesis that there is only one spike model; the fit of this model is shown in **c**. In this case, even though the total misfit is less for S_2, the simpler hypothesis, S_1, is more probable by exp(111) to 1. In the second row, S_2 (**d** and **e**) is more probable than S_1 (**f**) by exp(343) to 1. Note the increase in residual error with the model shown in **f**. The difference between models 3 and 4 is better illustrated in figure A.8 (where they are labeled M_2 and M_3 respectively). The large log probability ratios reported here result mainly from the abundance of data and the non-Gaussian outliers in the noise. A more realistic noise model, such as heavy-tailed Gaussian, would result in more accurate estimates of the true probability ratios.

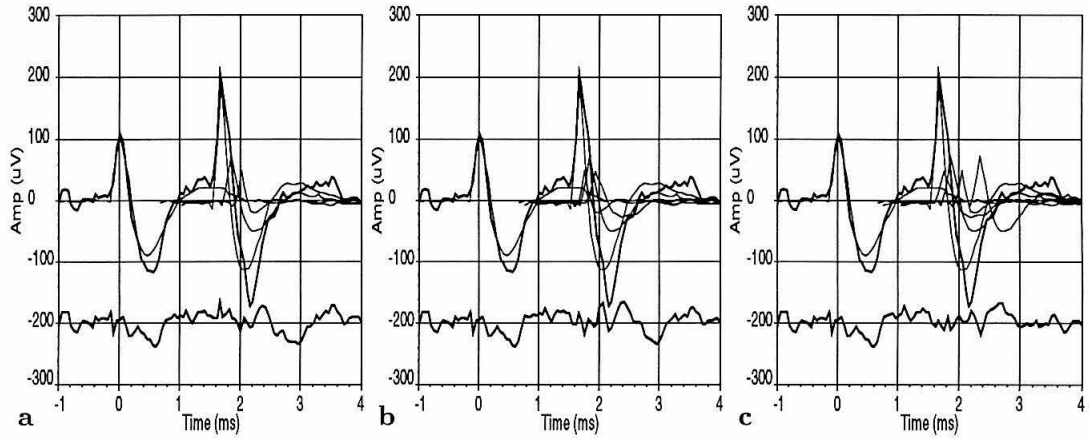


Figure A.4: Overfitting also occurs in the case of decomposing overlapping events. Shown are three of many well-fitting solutions for a single region of data. Thick lines are drawn between the data samples. The thin lines are the spike functions. Note that these examples were taken from the first iteration of the algorithm, so the spike functions are noisy estimates of the underlying AP shapes. The best-fitting overlap solution in this case is not the most probable: the solution with four spike functions shown in **a** is more than 8 times more probable than either **b** (five spike functions) or **c** (six spike functions) even though these fit the data better. The simple approach of using the *best-fitting* overlap solution actually *increases* the classification error especially in the number of false positives for the smaller models. To minimize classification error, it is necessary to find the *most probable* overlap solution.

optimized to one sample period. Unfortunately, for many realistic situations this method is computationally too demanding even for off-line analysis. For overlap decomposition to be practical, we need an efficient way to fit and rank a large number of model potential spike sequences. In addition, we would like to state precisely what hypothesis subspace is searched, so we can say what model combinations *cannot* account for a given region of overlapping events.

We can obtain a more efficient decomposition algorithm by employing two techniques. The first is to consider only AP sequences that occur with non-negligible probability. This allows us to obtain a large, but manageable hypothesis space in which to search. The second is to make the search itself efficient using appropriate data structures and dynamic programming.

Restricting the overlap hypothesis space

The main difficulty with overlapping APs is that there is no simple way to determine the event times. For many overlaps, such as the one in figure A.5a, the event times can be determined directly, because the APs are separated enough so that the models can be fit independently. As the degree of overlap increases, as in figures A.5b and c, accurate classification of one event depends on accurate classification of the surrounding events. In this case, the overlapping models must be fit simultaneously. Moreover, since small misalignments of the model with respect to the event can introduce significant residual error, each model in the overlap sequence must be precisely aligned.

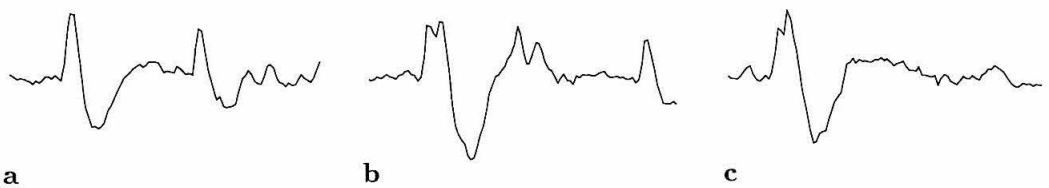


Figure A.5: As the peaks of two action potentials get closer together, it becomes more difficult to classify either one with accuracy. It is necessary in this case (**b** and **c**) to fit multiple models simultaneously.

The continuum of possible event times is the major factor contributing to the multitude

of potential overlap models. We can reduce this space significantly if we consider to what precision the τ_n's must be optimized. For a given spike model, $s_k(t)$, the maximum error resulting from a misalignment of δ_k is given by[2]

$$\epsilon = \delta_k \max_t \left| \frac{ds(t)}{dt} \right| . \tag{A.19}$$

From this we obtain the precision necessary to ensure that the error introduced by the model alone is less than ϵ and only need to choose among a discrete set of points.[3]

Even with this reduction, the number of possible sequences is still exponential in the number of overlapping models. This space can be reduced by considering only sequences that are likely to occur. For example, if there are five units with a Poisson firing rate of 20Hz, the probability of observing three events within half a millisecond is about 0.1%. Eliminating sequence models with more than two peaks within 0.5ms of each other will introduce about 0.1% error. In this manner, the desired trade-off between classification accuracy and computational cost can be determined. In practice, however, spikes often do not fire in a Poisson manner but fire in bursts. The firing rate model in this case should be adapted accordingly so that the expected number of missed events is estimated accurately.

Searching the overlap hypothesis space

Let us first outline the decomposition algorithm. To fit general model sequences, we use the methodology of dynamic programming. The event data is fit in sections from left to right. At every stage, a list is maintained of all plausible sequences[4] from the restricted hypothesis space determined by the methods described above. The length of data fit is extended by computing for each sequence on the list all plausible models that result by fitting the residual structure in the next region. The probabilities for all sequences are then recomputed, discarding any sequences below the probability threshold. The search terminates when no further overlaps are encountered in the most probable sequence model.

We now discuss each step in more detail. The primary operation in the algorithm is that of determining the most probable sequence models for a region of data. For efficiency, we

[2] We ignore the discontinuities in the derivative of the piece-wise linear model.

[3] For these demonstrations we use $\epsilon = 0.5\sigma_\eta$ which results in δ_k's ranging from 0.05 to 0.3 sampling periods.

[4] By plausible sequences we mean sequences with probability greater than a specified threshold.

precompute all possible waveform segments and store the set in a k-d tree (Bently, 1975) with which a fixed-radius nearest neighbor search can be performed in time logarithmic in the number of models (Friedman et al., 1977; Ramasubramanian and Paliwal, 1992). $O(N \log N)$ time is required to construct the tree, but once it is set up, each nearest-neighbor search is very fast. The set of overlap functions for a region from a to b around the spike peak is defined by

$$\Lambda_{k_{1:L},n}(t) = \sum_{j=1}^{L} s_{k_j}(t - n\delta_{k_j}), \quad k_j = 1, \ldots, K, \quad k_1 < \ldots < k_L, \quad a < t - n\delta_{k_j} < b, \quad n \text{ integer}, \tag{A.20}$$

where L is the maximum number of overlapping spike function segments in the peak region $[a, b]$, and δ_{k_j} is the τ-resolution for $s_k(t)$ defined in (A.19). The size of the peak region is somewhat arbitrary; the larger the region, the larger the number of waveform segments that must be considered, but the smaller the number of plausible overlap sequences found. In practice, the size of the peak region is largely limited by the memory required for the k-d tree. For these demonstrations, we take $L = 2$ (up to two overlapping spike functions segments) with a peak region of 0.25ms and include a "noise" model Λ_0 which has constant value equal to the DC voltage level. The number of waveform segments in the set can be reduced by eliminating overlapping spike functions for which the peak would have been (with high probability) detected at a sample position other than that of the data. Even with this reduction, an 11-model spike set results in about 50,000 waveform segments.

Once the best-fitting waveform segments for the first peak region are obtained, each segment is extended up to the next peak in the residuals for that segment. This peak is then fit using the k-d tree which in turn generates additional overlap sequences. As long as the introduction of new waveform segments does not alter our conclusions about the ordering sequence list, for example by fitting structure in a preceding region, we ensure either that one of the overlap sequences is true or that the sequences we are considering cannot account for the data.

After each sequence from the original list has been extended, the probability of each sequence model, $\mathbf{c}_i$, is recomputed. The exact relation is given by

$$P(\mathbf{c}_i|\mathbf{D}, \mathbf{S}) = \int d\boldsymbol{\tau}_i \, \frac{P(\mathbf{D}|\mathbf{c}_i, \boldsymbol{\tau}_i, \mathbf{S}) P(\mathbf{c}_i, \boldsymbol{\tau}_i|\mathbf{S})}{P(\mathbf{D}|\mathbf{S})}, \tag{A.21}$$

where $\mathbf{D}$ is the subset of data common to all sequences, and $\mathbf{S} = \{\mathbf{v}_{1:K}, \boldsymbol{\pi}, \sigma_\eta, M_{1:K}\}$. The form of the probability density function, $P(\mathbf{D}|\mathbf{c}_i, \boldsymbol{\tau}_i)$, is the same as (A.4). Equation (A.21) can be approximated with a Gaussian integral by treating each peak region as a separable component,

$$P(\mathbf{c}_i|\mathbf{D}, \mathbf{S}) \approx \frac{P(\mathbf{D}|\mathbf{c}_i, \widehat{\boldsymbol{\tau}}_i, \mathbf{S})\,(2\pi)^{C/2} \prod_j d_j^{-1/2}\, P(\mathbf{c}_i|\mathbf{S})\, P(\boldsymbol{\tau}_i|\mathbf{S})}{P(\mathbf{D}|\mathbf{S})}, \tag{A.22}$$

where C is the number of total number of spike functions in the sequence, and d_j is the determinant of Hessian of the τ's for the jth peak region. The values needed to compute the Hessians can be obtained directly from the k-d tree. Note that integrating over $\boldsymbol{\tau}_i$ performs the function of Ockham's Razor by penalizing sequences with many spike models. Omitting this would reduce the solution to one of maximum likelihood which chooses the sequence that best fits the data. For example, the solutions shown in figure A.4b and A.4c both fit the data better than in A.4a, but by (A.22), A.4a is more than eight times more probable than the others. Use of the best-fitting solutions would result in an *increase* in the classification error due to the introduction of too many models. Classification error is minimized by using the most probable overlap sequences.

$P(\mathbf{c}_i, \boldsymbol{\tau}_i|\mathbf{S})$ describes the *a priori* probability of the sequence of models in $\mathbf{c}_i$ with associated occurrence times $\boldsymbol{\tau}_i$. For this discussion, we assume $P(\mathbf{c}_i|\mathbf{S})$ to be Poisson with rate proportional to $\langle \pi_k \rangle$ and $P(\boldsymbol{\tau}_i|\mathbf{S})$ to be proportional to $1/\langle \pi_k \rangle$. Useful alternatives for $P(\mathbf{c}_i, \boldsymbol{\tau}_i|\mathbf{S})$ include models which take into account a refractory period or describe different types of spiking patterns.

Once the probabilities for the sequence models have been computed, the improbable models are discarded. The decomposition algorithm iterates until no overlapping structure is found in the most probable model. The search can fail if an outlier is encountered or if the true sequence is outside the hypothesis space. Otherwise, upon termination the search results in a list of all plausible sequence models of the given data along with their associated probabilities. Example decompositions are shown in figure A.6.

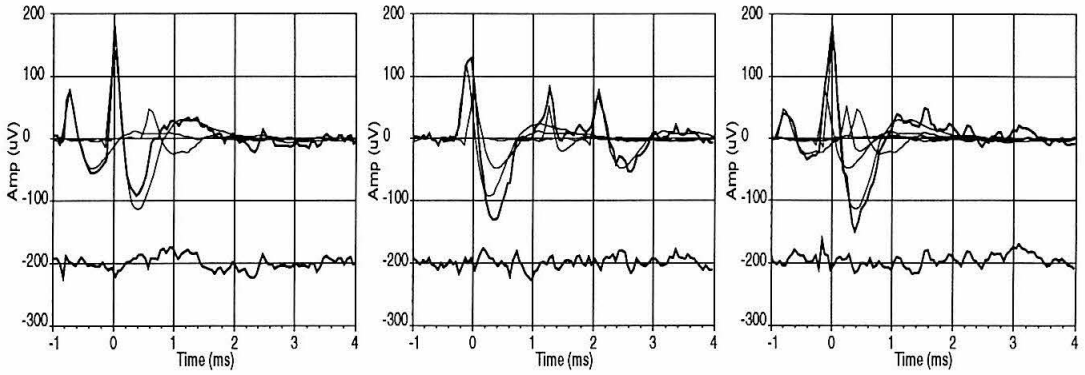


Figure A.6: Example overlap solutions. Thick lines are drawn between the data samples. The thin lines are the spike models. The overlap sequence in **a** has 3 spike functions, **b** contains 4 spike functions, and **c** contains 5 spike functions.

A.6 Performance on Real Data

The algorithm was first tested on real data, a section of which was shown in figure A.1. The whole waveform consisted of 40secs of data, filtered from 300 to 7000Hz and sampled at 20kHz. Three iterations of the algorithm were performed with overlap decomposition after the second (with $L = 1$) and third (with $L = 2$) iterations. Spike models which occurred less than ten times were discarded for efficiency, and the remaining events were reclassified. The inferred spike models are shown in figure A.7. The residuals indicate that these spike models account for almost all events in the 40sec waveform. Out of about 1500 total events, only 6 were not fit to within $5\sigma_\eta$. By eye, these events looked very noisy and had no obvious composition in terms of the spike models. One possibility is that they resulted from animal movement. Such events were not present in the synthesized data set described in section A.7 where all the events were fit with the inferred spike models.

By eye, all the models look distinct except perhaps for M_2 and M_3. One way to see the difference between these two models is to fit the data from model 3 with model 2 as shown in figure A.8. With a single electrode it is difficult to determine whether or not these two shapes result from different neurons, but they are clearly two types of events. One possibility is that these are different states of the same neuron; another is that the shape in model 3 results from a tight coupling between two neurons. Recording with multiple electrodes from a local region of tissue would help resolve issues like this.

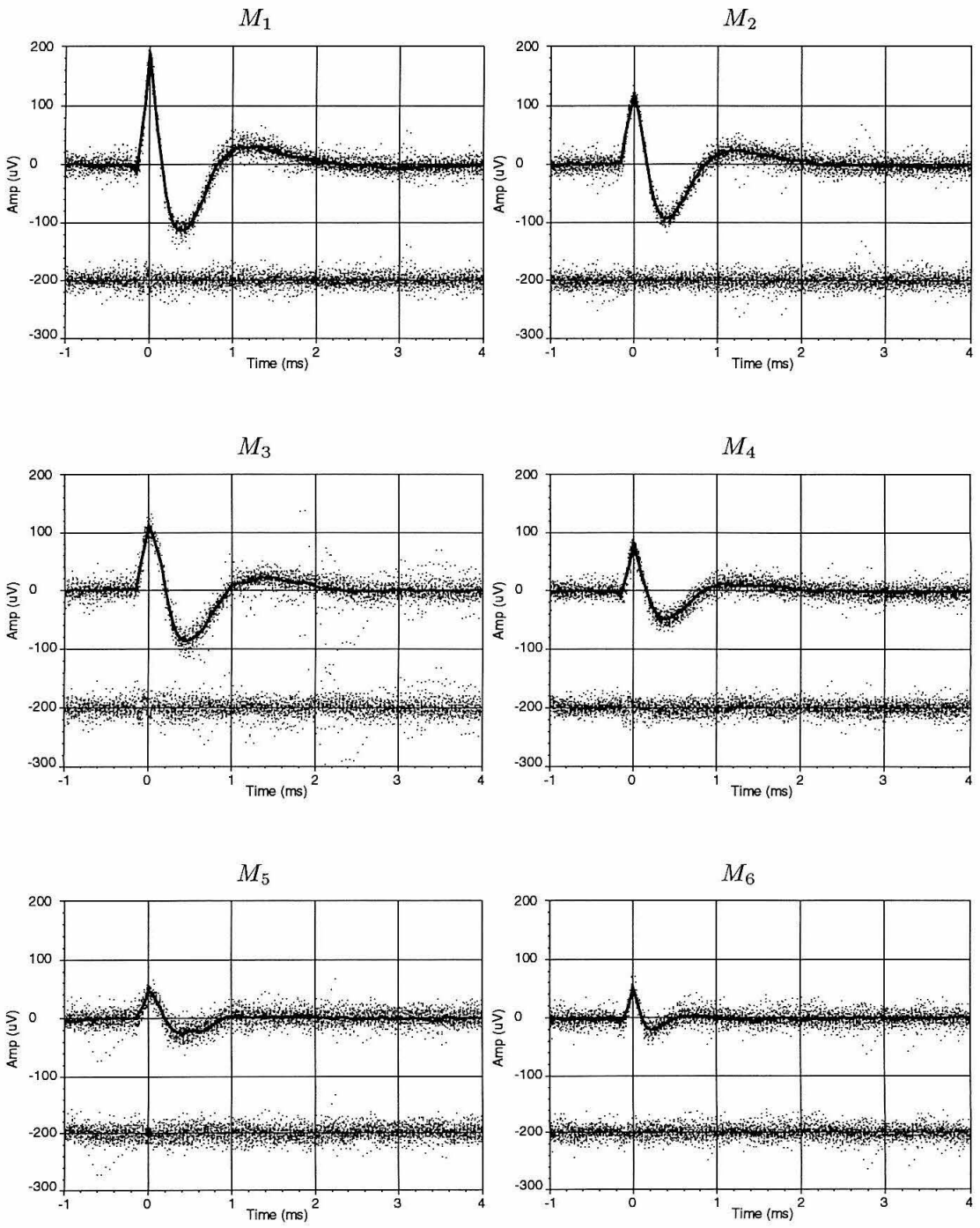


Figure A.7: The solid lines are the inferred spike models. The data overlying each model is a sample of at most 40 events. The residual errors are plotted below each model.

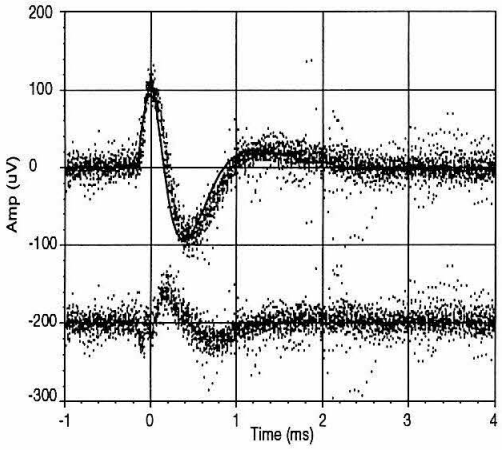


Figure A.8: One way to see the difference between the spike models M_2 and M_3 is to fit the data from M_3 (points) with M_2 (solid line). The residual errors are plotted below. All the data from both spike models is plotted. If the noise level is constant throughout the duration of the AP, the large deviation in the residuals indicates that there are two distinct classes.

In spite of all the math, the algorithm is fast. Inferring the spike set with overlap decomposition takes a few minutes on a Sun Microsystems Sparc IPX. Classification of the 40 second test waveform with overlap decomposition (using $L = 1$) takes about 10 seconds.

A.7 Performance on Synthesized Data

The accuracy of the algorithm was tested by generating an artificial data set composed of the six inferred shapes shown in figure A.7. The event times were Poisson distributed with frequency equal the inferred firing rate of the real data set. Gaussian noise was then added with standard deviation equal to σ_η. The algorithm was run under the same conditions as above.

The algorithm chose 14 initial spike models which were subsequently collapsed to 6 using the methods discussed in the previous section. Note that in this case, the number of inferred models matches the number of true models, but this need not be the case if some true models are too similar to be resolved, or if there are insufficient data to identify two distinct classes. The six-model spike set was preferred over the most probable five-model spike set by exp(34) : 1 and over the most probable seven-model spike set by exp(19) : 1.

A summary of the accuracy of the spike shapes is shown in table A.1.

Table A.1: Results of the spike model inference algorithm on the synthesized data set.

Model	1	2	3	4	5	6
$\Delta_{\max}/\sigma_\eta$	0.44	0.36	1.07	0.78	0.84	0.40
$\max_t s_k(t)/\sigma_\eta$	17.9	11.1	10.6	7.4	4.4	5.0
No. occurrences	39	63	45	238	155	1055

Table A.1: Both the form and number of spike models were determined by the algorithm. The inferred number of spike models matched the true number (6 models). The second row is the maximum absolute difference between the true spike model and the inferred model normalized by σ_η. The third row is the normalized peak of the inferred spike models which is an indication of how far each type of AP is above the noise level. The last row shows the number of times each model occurred in the synthesized data.

The results of inferring and classifying the synthesized data set are shown for the non-overlapping spikes in table A.2 and for the overlapping spikes in table A.3. An event was considered an overlap if the extent[5] overlapped the extent of another event. Perfect performance would have all zeros in the off-diagonal entries and no undetected events. An event can be missed if it is not detected in an overlap sequence or if all its sample values fall below the threshold for event detection ($4\sigma_\eta$). The tables indicate that for the largest four spikes, the performance is nearly perfect, even including the overlapping cases.

Performance is worst in the two smallest spike models where there are a large number of missed events. For these models, there are typically only two or three samples that would be expected to exceed the noise level. As the threshold for event detection is lowered, there is a trade off between the number of real spikes missed and the number of false positives resulting from common instance of when the noise contains a spike-like shape. The number of below threshold missed events can be minimized (with additional computational expense) by computing the probabilities at every sample point instead of only those that cross threshold. It is worth noting that this situation often does not pose a problem in practice, since observed spikes just above the noise level frequently correspond to many different neurons.

[5]The extent of an event is defined as the minimum and maximum values in time at which the best-fitting spike function differs from DC by more than $0.5\sigma_\eta$.

Table A.2: Classification results for the non-overlapping events of the synthesized data set.

True Models	Inferred Models						Missed Events	Total Events
	1	2	3	4	5	6		
1	17	0	0	0	0	0	0	17
2	0	25	1	0	0	0	0	26
3	0	0	15	0	0	0	0	15
4	0	0	0	116	0	0	1	117
5	0	0	0	0	56	0	17	73
6	0	0	0	0	0	393	254	647

Table A.3: Classification results for the overlapping events of the synthesized data set.

True Models	Inferred Models						Missed Events	Total Events
	1	2	3	4	5	6		
1	22	0	0	0	0	0	0	22
2	0	36	1	0	0	0	0	37
3	0	0	20	0	0	0	0	20
4	0	1	0	116	0	1	3	121
5	0	0	0	1	61	1	19	82
6	0	0	0	3	2	243	160	408

Tables A.2 and A.3: Each matrix component indicates the number of times true model i was classified as inferred model j. Events were missed if the true spikes were not detected in an overlap sequence or if all sample values for the spike fell below the event detection threshold ($4\sigma_\eta$). There was 1 false positive for M_5 and 7 for M_6. See text for additional comments.

A.8 Comparisons with Other Approaches

It is instructive to contrast the spike sorting algorithm presented here with other methods by comparing their performances on the synthesized data set used in the previous section. The most common method of classifying APs is through use of a hardware level detector which detects an AP if the voltage exceeds a user-determined level. For the synthesized data set, a level detector is sufficient only to classify the largest AP (M_1) with accuracy. Another common hardware approach is a window discriminator with which APs are detected only if the peak value is within a voltage window. A window discriminator can classify M_1 accurately and classify M_4 with some error since the distribution of the M_4 peak voltages overlaps somewhat with other models, but it is not sufficient to discriminate between M_2 and M_3 or between M_5 and M_6. These discriminations demand more sophisticated methods.

A common software-based method for spike sorting is a feature clustering algorithm such as the one used in the commercial physiological data collection system *Brainwave*. The synthesized waveform was classified independently by an experienced Brainwave user (Matt Wilson). The features used to perform the classification were maximum spike amplitude, minimum spike amplitude, and time from the spike maximum to the spike minimum. Brainwave generates a list of occurrence times for each cluster but not explicit spike functions, so it was not possible to see how close the "inferred spike functions" were to the true spike functions. The occurrence times were compared to the known AP positions. Two separate classifications were performed (one using four clusters and another using six clusters), and the results of the most accurate classification (six clusters) are reported here.

Tables A.4 and A.5 show the classification results for the synthesized data set for the non-overlapping and overlapping action potentials respectively. A total of six clusters were found, but not all of these correspond to the true underlying clusters.

The tables show that true models M_1 and M_4 were accurately identified and classified. True models M_2 and M_3, however, were collapsed into a single cluster. This discrimination is difficult to make without accurately estimating the occurrence time of the APs. Brainwave uses the spike peak for the occurrence time which is accurate to within one sample period and introduces a significant amount of noise into the features. In contrast, the Bayesian approach estimates the spike occurrence times with sub-sampling period accuracy. Note also that with no overlap decomposition, there are significantly more missed events for the

larger APs.

True models M_5 and M_6 were described with three clusters, with clusters four and five roughly corresponding to M_5 and cluster six corresponding to M_6. For these models, the features used make it difficult to choose the correct clusters, since the smaller models are not well separated in the 3-dimensional feature space. There is less separation, because the occurrence times are not estimated accurately and no overlap decomposition is done. Even if the cluster centers were accurate, we would expect the Brainwave classification to be less accurate than the Bayesian approach. Using spike functions to perform the classification utilizes all significant sample points in the waveform which for the smallest two models is between four and eight. In contrast, only three features are used by Brainwave.

Table A.4: Brainwave classification results for the non-overlapping events of the synthesized data set.

True	Cluster Number						Missed	Total
Models	1	2	3	4	5	6	Events	Events
1	17	0	0	0	0	0	0	17
2	0	26	0	0	0	0	0	26
3	0	15	0	0	0	0	0	15
4	0	0	116	0	0	0	1	117
5	0	0	1	24	0	6	42	73
6	0	0	0	22	13	188	424	647

Table A.5: Brainwave classification results for the overlapping events of the synthesized data set.

True	Cluster Number						Missed	Total
Models	1	2	3	4	5	6	Events	Events
1	22	0	0	0	0	0	0	22
2	0	34	1	0	0	0	2	37
3	0	18	1	0	0	0	1	20
4	0	3	106	3	0	2	7	121
5	0	0	1	26	8	10	45	82
6	0	3	9	15	2	108	264	408

Tables A.4 and A.5: Each matrix component indicates the number of times true model i was classified as belonging to Brainwave cluster j. An event was missed if a true AP did not correspond to any of the APs identified by Brainwave. The false positive counts were 2, 3, 4, and 2 for Brainwave clusters 3, 4, 5, and 6 respectively.

A.9 Extensions

There are a number of possible directions for improvements to the general waveform model we have described. At the lowest level there are possibilities for alternative noise models. For example, real extracellular noise tends to be correlated and slightly non-Gaussian. Incorporating this information would make the probabilities more accurately reflect the real world.

The piece-wise linear model we have described is general enough to fit almost arbitrary shapes, but that generality is also part of its shortcoming. Since in the algorithm we have placed minimal restrictions on the form of the spike model, more data is required to infer the shape. Incorporating knowledge about the spike shapes would result in more accurate conclusions with the same amount of data. Overly weak spike shape priors will also result in overly strong Occam factors which will bias the results of model comparisons toward simpler models.

For some types of neurons the shape of an action potential is not constant. Bursting neurons, for example, have spikes that decay dramatically during a burst. Modeling the resulting shape is complicated because the inter-spike intervals during a burst are not constant over different bursts, and the degree of attenuation depends on the intervals. Another way in which APs can change their shape is due to electrode drift which results in a slow change of the spike shapes over time. This can be handled readily by the algorithms since re-estimating previously inferred shapes is very fast.

Another limitation stems not from the algorithm but from the method of recording. Since a single electrode gives little information about a neuron's position, decisions about whether two shapes constitute two neurons must be made based on shape and firing frequency alone. The use of multiple electrodes in a local area resolves this issue by recording the same group of neurons from different sites. Thus even if two neurons have identical shapes when recorded from electrode, it is unlikely that those two neurons will generate the same AP shape when observed simultaneously from a different electrode. A trivial extension of the algorithm would be to run it on each electrode and then look for cross-correlations in the event times, but better results could be obtained by incorporating the information about multiple electrodes into a single model.

A.10 Discussion

Formulating the task as the inference of a probabilistic model made clear what was necessary to obtain accurate spike models. Optimizing the τ_n's is crucial for both inference and classification, but this step is commonly ignored by algorithms which cluster the sample points or derive spike shapes from principal components. The soft clustering procedure makes it possible to determine the spike shapes with accuracy even when they are highly overlapping. Unless the spike shapes are well-separated, hard clustering procedures such as k-means will lead to inaccurate estimates of the spike shapes.

Probability theory also provided an objective means of determining the number of spike models which is an essential reason for the success of this algorithm. With the incorrect number of spike models, overlap decomposition becomes especially difficult. If there are too few spike models, the overlap data cannot be fit. If there are too many, decomposition becomes a very expensive computation. Evaluating the probability of alternative spike sets has proved to be a sensitive method for determining when two classes are distinct. Previous approaches have relied on *ad hoc* criteria or the user to make this decision, but such approaches cannot be relied upon to work under varying circumstances since their inherent assumptions are not explicit. An advantage of probability theory is that the assumptions are explicit, and given those assumptions, the answer provided by the evidence is optimal.

One might wonder if the user, having much more information than has been incorporated into the model, can make better decisions than the evidence about what constitutes distinct spike models. Probability theory provides a calculus for stating precisely what can be inferred from the data given the model. When the conclusions reached through probability theory do not fit our expectations, it is due to a failure of the model or a failure of the approximations (if approximations are made). From the performance on the synthesized data, however, the approximations appear to be reasonable. Thus when the conclusions reached through the evidence are at variance with the user's, information is at hand about possible shortcomings of the current model. In this manner, new models can be constructed, and moreover, they can be compared objectively using the evidence.

Probability theory is also essential for accurate overlap decomposition. It is not sufficient just to fit data with compositions of spike models. That leads to the same overfitting problem encountered in determining the number of spike models and in determining the

spike shapes. The Ockham penalty introduced by integrating out the τ's was key to finding the most probable fits and consequently for achieving accurate classification. Previous approaches have been able to handle only a limited class of overlaps, mainly due to the difficulty in making the fit efficient. The algorithm we have described can fit an overlap sequence of virtually arbitrary complexity in milliseconds.

In practice, the algorithm we have described allows us to extract much more information from an experiment than with previous methods. Moreover, this information is qualitatively different from a simple list of spike times. Having reliable estimates of the action potential shapes makes it possible to study the properties of these classes, since distinct neuronal types can have distinct neuronal spikes (Connors and Gutnick, 1990). With stereotrodes this advantage would be amplified, since it is then possible to estimate somatic size which is another distinguishing characteristic of cell type. Finally, accurate overlap decomposition makes it possible to investigate interactions among local neurons which were previously very difficult to observe.

Bibliography

Aamodt, S. M., Kozlowski, M. R., Nordeen, E. J., and Nordeen, K. W. (1992). Distribution and developmental-change in [h-3] mk-801 binding within zebra finch song nuclei. *J. Neurobiol.*, 23(8):997–1005.

Ambrosingerson, J., Granger, R., and Lynch, G. (1990). Simulation of paleocortex performs hierarchical-clustering. *Science*, 247(4948):1344–1348.

Amit, D. J. and Brunel, N. (1995). Learning internal representations in an attractor neural-network with analog neurons. *Network Computation in Neural Systems*, 6(3):359–388.

Atiya, A. (1992). Recognition of multiunit neural signals. *IEEE Trans. Biomed. Eng.*, 39(7):723–729.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

Bently, J. (1975). Multidimensional binary search trees used for associative searching. *Comm. ACM*, 18(9):509–517.

Bohner, J. (1983). Song learning in the zebra finch (taeniopygia-guttata) - selectivity in the choice of a tutor and accuracy of song copies. *Animal Behav.*, 31(FEB):231–237.

Bohner, J. (1990). Early acquisition of song in the zebra finch, taeniopygia-guttata. *Animal Behav.*, 39(FEB):369–374.

Bonke, B., Bonke, D., and Scheich, H. (1979a). Connectivity of the auditory forebrain nuclei in the guinea fowl (numida meleeaagris). *Cell Tiss Res*, 200:101–121.

Bonke, D., Scheich, H., and Langner, G. (1979b). Responsiveness of units in the auditory neostriatum of the guinea fowl (numida meleagris) to species-specific calls and synthetic stimuli. I. tonotopy and functional zones of field L. *J. Comp. Physiol.*, 132:243–255.

Bottjer, S. W., Miesner, E. A., and Arnold, A. P. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, 224(4651):901–903.

Brenowitz, E. A. and Arnold, A. P. (1990). The effects of systemic androgen treatment on androgen accumulation in song control regions of the adult female canary brain. *J. Neurobiol.*, 21(6):837–843.

Brodin, L., Grillner, S., and Rovainen, C. M. (1985). N-methyl-D-aspartate (nmda), kainate and quisqualate receptors and the generation of fictive locomotion in the lamprey spinal-cord. *Brain Res.*, 325(1-2):302–306.

Canady, R. A., Kroodsma, D. E., and Nottebohm, F. (1984). Population differences in complexity of a learned skill are correlated with the brain space involved. *Proc. Natl. Acad. Sci. USA Biological Sciences*, 81(19):6232–6234.

Casseday, J. H., Ehrlich, D., and Covey, E. (1994). Neural tuning for sound duration - role of inhibitory mechanisms in the inferior colliculus. *Science*, 264(5160):847–850.

Chappell, G. J. and Taylor, J. G. (1993). The temporal kohonen map. *Neural Networks*, 6(3):441–445.

Cline, H. T. (1991). Activity-dependent plasticity in the visual systems of frogs and fish. *Trends in Neurosci.*, 14(3):104–111.

Connors, B. and Gutnick, M. (1990). Intrinsic firing patterns of diverse neocortical neurons. *TINS*, 13(3):99–104.

Coultrip, R., Granger, R., and Lynch, G. (1992). A cortical model of winner-take-all competition via lateral inhibition. *Neural Networks*, 5(1):47–54.

Cynx, J. (1993). Conspecific song perception in zebra finches (taeniopygia-guttata). *J. Comp. Psychol.*, 107(4):395–402.

Dale, N. and Roberts, A. (1985). Dual-component amino-acid-mediated synaptic potentials - excitatory drive for swimming in xenopus embryos. *J. Physiol. London*, 363(JUN):35–59.

Dehaene, S., Changeux, J.-P., and Nadal, J.-P. (1987). Neural networks that learn temporal sequences by selection. *Proceedings of the National Academy of Science U.S.A.*, 84:2727–2731.

Doupe, A. (1993). A neural circuit specialized for vocal learning. *Curr. Opin. Neurobiol.*, 3:104–111.

Doupe, A. and Konishi, M. (1991). Song-selective auditory circuits in the vocal control system of the zebra finch. *Proc. Natl. Acad. Sci. USA*, 88:11339–11343.

Doupe, A. and Konishi, M. (1992). Song-selective auditory neurons emerge during vocal learning in the zebra finch. *Soc. Neurosci. Abstr.*, 18:527.

Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. Wiley, New York.

Földiák, P. (1989). Adaptive network for optimal linear feature extraction. In *Proceedings of the International Joint Conference on Neural Networks*, volume I, pages 401–405, Washington, D.C.

Földiák, P. (1989). Forming sparse representations by local anti-hebbian learning. *Biological Cybernetics*, 64:165–170.

Fortune, E. S. and Margoliash, D. (1992). Cytoarchitectonic organization and morphology of cells of the field-L complex in male zebra finches (*taenopygia guttata*). *J. Comp. Neurol.*, 325(3):388–404.

Fortune, E. S. and Margoliash, D. (1995). Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (*taeniopygia guttata*). *J. Comp. Neurol.*, 360(3):413–441.

Friedman, J., Bently, J., and Finkel, R. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Software*, 3(3):209–226.

Ghahramani, Z. and Jordan, M. (1995). Factorial hidden markov models. Technical Report 9502, MIT Computational Cognitive Science, Cambridge, MA.

Glass, I. and Wollberg, Z. (1983). Auditory-cortex responses to sequences of normal and reversed squirrel-monkey vocalizations. *Brain Behav. and Evol.*, 22(1):13–21.

Granger, R., Whitson, J., Larson, J., and Lynch, G. (1994). Non-hebbian properties of long-term potentiation enable high-capacity encoding of temporal sequences. *Proc. Natl. Acad. Sci. USA*, 91(21):10104–10108.

Gray, R. M. (1984). Vector quantization. *IEEE ASSP Magazine*, pages 4–29.

Grillner, S., Wallen, P., Brodin, L., and Lansner, A. (1991). Neuronal network generating locomotor behavior in lamprey - circuitry, transmitters, membrane-properties, and simulation. *Ann. Rev. Neuro.*, 14:169–199.

Grisham, W. and Arnold, A. P. (1994). Distribution of GABA-like immunoreactivity in the song system of the zebra finch. *Brain Res.*, 651(1-2):115–122.

Gull, S. (1988). Bayesian inductive inference and maximum entropy. In Erickson, G. and Smith, C., editors, *Maximum Entropy and Bayesian Methods in Science and Engineering, vol. 1: Foundations.* Kluwer.

Heil, P., Langner, G., and Scheich, H. (1992). Processing of frequency-modulated stimuli in the chick auditory-cortex analog - evidence for topographic representations and possible mechanisms of rate and directional sensitivity. *J. Comp. Physiol. A*, 171(5):583–600.

Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. M. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1161.

Hose, B., Langner, G., and Scheich, H. (1987). Topographic representation of periodicities in the forebrain of the myna bird - one map for pitch and rhythm. *Brain Res.*, 422(2):367–373.

Hultsch, H. and Todt, D. (1989). Memorization and reproduction of songs in nightingales (*luscinia megarhynchos*) - evidence for package formation. *J. Comp. Physiol. A*, 165(2):197–203.

Jagadeesh, B., Wheat, H. S., and Ferster, D. (1993). Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual-cortex. *Science*, 262(5141):1901–1904.

Jahnsen, H. and Llinas, R. (1984). Electrophysiological properties of guinea-pig thalamic neurons - an invitro study. *J. Physiol. (London)*, 349(APR):105–226.

Jaynes, E. (1979). Review of *inference, method, and decision* (r. d. rosenkrantz). *J. Am. Stat. Assoc.*, 74:140.

Jeffreys, H. (1939). *Theory of Probability.* Oxford University Press (3rd revised ed 1961).

Karten, H. (1968). The ascending auditory pathway in the pigeon (*columbia liva*) ii. telencephalic projections of the nucleus ovoidalis thalami. *Brain. Res.*, 11:134–153.

Katz, L. and Gurney, M. (1981). Auditory responses in the zebra finch's motor system for song. *Brain. Res.*, 211:192–197.

Kelley, D. and Nottebohm, F. (1979). Projections of a telencephalic auditory nucleus - field L - in the canary. *J. Comp. Neur.*, 183:455–470.

Kirkwood, A., Dudek, S. M., Gold, J. T., Aizenman, C. D., and Bear, M. F. (1993). Common forms of synaptic plasticity in the hippocampus and neocortex invitro. *Science*, 260(5113):1518–1521.

Kleinfeld, D. (1986). Sequential state generation by model neural networks. *Proceedings of the National Academy of Sciences, USA*, 83:9469–9473.

Knipschild, M., Dorrscheidt, G. J., and Rubsamen, R. (1992). Setting complex tasks to single units in the avian auditory forebrain .1. processing of complex artificial stimuli. *Hearing Res.*, 57(2):216–230.

Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift fur Tierpsychologie*, 22:771–783.

Konishi, M. (1989). Birdsong for neurobiologists. *Neuron*, 3:541–549.

Konishi, M. and Akutagawa, E. (1987). Hormonal-control of cell-death in a sexually dimorphic song nucleus in the zebra finch. *Ciba Foundation Symposia*, 126:173–185.

Kubota, M. and Saito, N. (1991a). NMDA receptors participate differentially in two different synaptic inputs in neurons of the zebra finch robust nucleus of the archistriatum invitro. *Neurosci. Letters*, 125(2):107–109.

Kubota, M. and Saito, N. (1991b). Sodium-dependent and calcium-dependent conductances of neurons in the zebra finch hyperstriatum-ventrale pars caudale invitro. *J. Physiol. (London)*, 440(AUG):131–142.

Langner, G., Bonke, D., and Sheich, H. (1981). Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. *Exp. Brain Res.*, 43:429–436.

Leppelsack, H. (1978). Unit responses to species-specific sounds in the auditory forebrain center of birds. *Fed Proc*, 37:2336–2341.

Leppelsack, H. (1983). Analysis of song in the audiotry pathway of songbirds. In Ewert, J., editor, *Advances in vertabrate neuroethology*, pages 783–800. Plenum Press, New York.

Leppelsack, H. and Vogt, M. (1976). Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *J. Comp. Physiol.*, 107:263–274.

Levinson, S. E. (1986). Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language*, 1(1):29–45.

Lewicki, M. (1994). Bayesian modeling and classification of neural signals. *Neural Computation*, 6:1005–1030.

Lewicki, M. and Arthur, B. (1995). Sensitivity to auditory temporal contex increases significantly from field l to hvc. *Soc. Neurosci. Abstr.*, 21:958.

Lewicki, M. and Doupe, A. (1993). Synaptic activity of neurons in zebra finch song nucleus HVc in response to auditory stimuli. *Soc. Neurosci. Abstr.*, 19:1016.

Lewicki, M. and Konishi, M. (1995). Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl. Acad. Sci. USA*, 92:5582–5586.

Ljolje, A. and Levinson, S. (1991). Development of an acoustic-phonetic hidden markov model for continuous speech recognition. *IEEE Transactions on Signal Processing*, 39(1):29–39.

Lu, S. M., Guido, W., and Sherman, S. M. (1993). The brain-stem parabrachial region controls mode of response to visual-stimulation of neurons in the cats lateral geniculate-nucleus. *Visual Neurosci.*, 10(4):631–642.

Luttrell, S. P. (1994). Partitioned mixture distribution - an adaptive bayesian network for low-level image-processing. *IEE Proceedings Vision Image and Signal Processing*, 141(4):251–260.

MacKay, D. J. C. (1992). The evidence framework applied to classification networks. *Neural Computation*, 4(5):720–736.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506.

Malenka, R. C. (1994). Synaptic plasticity in the hippocampus - ltp and ltd. *Cell*, 78(4):535–538.

Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.*, 3(5):1039–1057.

Margoliash, D. (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.*, 6(6):1643–1661.

Margoliash, D. and Bankes, S. C. (1993). Computations in the ascending auditory pathway in songbirds related to song learning. *Am. Zoologist*, 33(1):94–103.

Margoliash, D., Fortune, E., Sutter, M., Yu, A., Wren-Hardin, B., and Dave, A. (1994). Distributed representation in the song system of oscines: Evolutionary implications and functional consequences. *Brain Bhav. and Evol.*, 44:247–264.

Margoliash, D. and Fortune, E. S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch HVc. *J. Neurosci.*, 12(11):4309–4326.

Margoliash, D. and Konishi, M. (1985). Auditory representation of autogenous song in the song system of white-crowned sparrows. *Proc. Natl. Acad. Sci. USA*, 82:5997–6000.

Marler, P. and Peters, S. (1981). Sparrows learn adult song and more from memory. *Science*, 213:780–782.

McCasland, J. (1987). Neuronal control of bird song production. *J. Neurosci.*, 7(1):23–39.

McCasland, J. and Konishi, M. (1981). Interaction between auditory and motor activities in an avian song control nucleus. *Proc. Natl. Acad. Sci. USA*, 78(12):7815–7819.

McKenna, T. M., Weinberger, N. M., and Diamond, D. M. (1989). Responses of single auditory cortical-neurons to tone sequences. *Brain Res.*, 481(1):142–153.

Mooney, R. (1992). Synaptic basis for developmental plasticity in a birdsong nucleus. *J. Neurosci.*, 12(7):2464–2477.

Mooney, R. and Konishi, M. (1991). Two distinct inputs to an avian song nucleus activate different glutamate receptor subtypes on individual neurons. *Proc. Natl. Acad. Sci. USA*, 88(10):4075–4079.

Muller, C. and Leppelsack, H. (1985). Feature extraction and tonotopic organization in the avian forebrain. *Exp. Brain Res.*, 59:587–599.

Neal, R. M. (1992). Connectionist learning of belief networks. *Artificial Intelligence*, 56(1):71–113.

Newman, J. and Wollberg, Z. (1973). Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res.*, 54:287–304.

Nixdorf, B., Davis, S., and DeVoogd, T. (1989). Morphology of golgi-impregnated neurons in hyperstriatum ventralis, pars caudalis in adult male and female canaries. *J Comp Neur*, 284:337–349.

Nottebohm, F., Stokes, T., and Leonard, C. (1976). Central control of song in the canary, *serinus canarius*. *J. Comp. Neur.*, 165:457–486.

Nowlan, S. J. (1990). Maximum likelihood competitive learning. In Touretzky, D. S., editor, *Advances in Neural Information Processing Systems*, volume 2, pages 574–582, San Mateo. (Denver 1989), Morgan Kaufmann.

Paton, J. A. and Nottebohm, F. N. (1984). Neurons generated in the adult brain are recruited into functional circuits. *Science*, 225(4666):1046–1048.

Paton, J. A., Oloughlin, B. E., and Nottebohm, F. (1985). Cells born in adult canary forebrain are local interneurons. *J. Neurosci.*, 5(11):3088–3093.

Price, P. (1979). Developmental determinants of structure in zebra finch song. *J. Comp. Physiol. Psychol*, 93:260–277.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77:257–286.

Ramasubramanian, V. and Paliwal, K. (1992). Fast k-dimensional tree algorithms for nearest neighbor search with application to vector quantization encoding. *IEEE Trans. Signal Proc.*, 40(3):518–531.

Redlich, A. N. (1993). Redundancy reduction as a strategy for unsupervised learning. *Neural Computation*, 5(2):289–304.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

Saund, E. (1995). A multiple cause mixture model for unsupervised learning. *Neural Computation*, 7(1):51–71.

Schafer, M., Rubsamen, R., Dorrscheidt, G. J., and Knipschild, M. (1992). Setting complex tasks to single units in the avian auditory forebrain .2. do we really need natural stimuli to describe neuronal response characteristics. *Hearing Res.*, 57(2):231–244.

Scharff, C. and Nottebohm, F. (1991). A comparative study of the behavior deficits following lesions of various parts of the zebra finch song system - implications for vocal learning. *J. Neurosci.*, 11(9):2896–2913.

Scheich, H. (1983). Two columnar systems in the auditory neostriatum of the chick: evidence from 2-deoxyglucose. *Exp. Brain Res.*, 51:199–205.

Scheich, H., Langner, G., and Bonke, D. (1979). Responsiveness of units in the auditory neostriatum of the guinea fowl (numida meleagris) to species-specific calls and synthetic stimuli. II. discrimination of iambus-like calls. *J. Comp. Physiol.*, 132:257–276.

Schmidt, E. (1984). Computer separation of multi-unit neuroelectric data: a review. *J. Neurosci. Meth.*, 12:95–111.

Schmidt, M. and Perkel, D. (1995). Characterization of slow inhibitory potentials in nucleus HVc of adult male zebra finches. *Soc. Neurosci. Abstr.*, 21:958.

Simpson, H. B. and Vicario, D. S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *J. Neurosci.*, 10(5):1541–1556.

Sohrabji, F., Nordeen, E., and Nordeen, K. (1989). Selective impairment of song development in zebra finches following lesions of the song control nucleus, area X. *Soc. Neurosci. Abstr.*, 15:618.

Sompolinsky, H. and Kanter, I. (1986). Temporal association in asymmetric neural networks. *Physical Review Letters*, 57:2861–2864.

Steriade, M. and Deschenes, M. (1984). The thalamus as a neuronal oscillator. *Brain Res. Reviews*, 8(1):1–63.

Sutter, M. L. and Margoliash, D. (1994). Global synchronous response to autogenous song in zebra finch HVc. *J. Neurophysiol.*, 72(5):2105–2123.

Tank, D. W. and Hopfield, J. J. (1987). Neural computation by time compression. *Proceedings of the National Academy of Sciences, USA*, 84:1896–1900.

Uno, H., Ohno, Y., Yamada, T., and Miyamoto, K. (1991). Neural coding of speech sound in the telencephalic auditory area of the myna bird. *J. Comp. Physiol. A*, 169(2):231–239.

Vicario, D. S. (1991). Neural mechanisms of vocal production in songbirds. *Curr. Opin. Neurobiol.*, 1:595–600.

Vicario, D. S. and Yohay, K. H. (1993). Song-selective auditory input to a forebrain vocal control nucleus in the zebra finch. *J. Neurobiol.*, 24(4):488–505.

Volman, S. (1993). Development of neural selectivity for birdsong during vocal learning. *J. Neurosci.*, 13(11):4737–4747.

Vu, E. and Lewicki, M. (1994). Intrinsic interactions between zerbra finch HVc neurons involve NMDA-receptor mediated activation. *Soc. Neurosci. Abstr.*, 20:166.

Vu, E. T., Mazurek, M. E., and Kuo, Y. C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. *J. Neurosci.*, 14(11):6924–6934.

Waibel, A. (1989). Modular construction of time-delay neural networks for speech recognition. *Neural Computation*, 1:39–46.

Wang, Z. and Mccormick, D. A. (1993). Control of firing mode of corticotectal and corticopontine layer-V burst-generating neurons by norepinephrine, acetylcholine, and 1s,3r-acpd. *J. Neurosci.*, 13(5):2199–2216.

Weinberger, N. M. and McKenna, T. M. (1988). Sensitivity of single neurons in aduditory cotex to contour: Toward a neurophysiology of music perception. *Music Perception*, 5:355–390.

Williams, C. and Hinton, G. (1991). Mean field networks that learn to discriminate temporally distorted strings. In Touretzky, D., Elman, J., Sejnowski, T., and Hinton, G., editors, *Connectionist Models: Proceedings of the 1990 Summer School*, pages 18–22, San Mateo. Morgan Kaufmann.

Wollberg, Z. and Newman, J. (1972). Auditory cortex of squirrel monkey: Response patterns of single cells to species-specific vocalizations. *Science*, 175:212–214.

Yu, A. and Margoliash, D. (1995). Function hierarchy defined by single units in singing birds: HVc represents syllables and RA represents notes. *Soc. Neurosci. Abstr.*, 21:958.

Zuschratter, W., Braun, S., and Scheich, H. (1987). Co-localization of parvalbumin, calbindin and GABA in avian vocal motor system. *Neurosci. Suppl.*, 22:S114.