ABSTRACT

ÜBER-CLAꝎS: UNSUPERVISED PATTERN CLASSIFICATION FOR MULTI-UNIT EXTRACELLULAR NEURONAL BURST EXTRACTION

by
Rama Natarajan

To further an understanding of how a neuronal population generates patterns of rhythmic activity, the temporal dynamics of the group of neurons must be formalized. Essential to this pursuit, is the ability to reliably detect and separate the classes of single-unit neuronal activity from multi-unit extracellular signals recorded in a single channel. This study proposes a unified approach to automatically detect and classify single-unit bursts, and to observe the precise onset and offset of burst activity. Existing approaches to the problem fundamentally depend on the statistics of spike waveform variability, both extrinsic and intrinsic to the neuron. In contrast, the proposed approach depends on statistics that characterize the burst variability. An unsupervised learning procedure is implemented using hierarchical clustering to derive a complete and natural description of the variability in terms of clusters of bursts that possess strong internal similarities. Redundant solution vectors are used to parameterize each cluster, and a fuzzy classification approach assigns each burst to a class. Accuracy of the technique is demonstrated on in vivo and in vitro recordings of the tri-phasic pyloric rhythm in stomatogastric ganglion of crab *Cancer borealis*. The results, evaluated against a widely used manual classification approach, show that the technique performs detection and classification with comparable accuracy and quantifiable certainty, and is robust to background activity and noise.

ÜBER-CLAωS: UNSUPERVISED PATTERN CLASSIFICATION FOR MULTI-UNIT EXTRACELLULAR NEURONAL BURST EXTRACTION

by
Rama Natarajan

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Science

Department of Computer Science

May 2003

APPROVAL PAGE

ÜBER-CLAϢS: UNSUPERVISED PATTERN CLASSIFICATION FOR MULTI-UNIT EXTRACELLULAR NEURONAL BURST EXTRACTION

Rama Natarajan

Dr. Jason Tsong-Li Wang, Thesis Advisor — Date
Professor of Computer Science, NJIT

Dr. Farzan Nadim, Committee Member — Date
Associate Professor of Mathematics, NJIT
Associate Professor of Biological Sciences, Rutgers University, Newark

Dr. Jorge Golowasch, Committee Member — Date
Associate Professor of Mathematics, NJIT
Associate Professor of Biological Sciences, Rutgers University, Newark

BIOGRAPHICAL SKETCH

Author: Rama Natarajan

Degree: Master of Science in Computer Science

Date: May 2003

Undergraduate and Graduate Education:

Master of Science in Computer Science,
New Jersey Institute of Technology, Newark, NJ, 2003

Bachelor of Science in Computing (Hons),
University of Portsmouth, United Kingdom, 2000

Diploma in Information Technology,
Nanyang Polytechnic, Singapore, 1999

Major: Computer Science

Presentations and Publications:

Rama Natarajan and Farzan Nadim,
"Uber-Claws: Unsupervised Burst ExtRaction, Classification and Analysis Software,"
23rd Annual East Coast Nerve Net, Woods Hole, MA, USA, Mar 2003.

There is, it seems to us,

At best, only a limited value

In the knowledge derived from experience.

The knowledge imposes a pattern, and falsifies,

For the pattern is new in every moment

And every moment is a new and shocking

Valuation of all we have been.

- T.S. Elliot. *Four Quartets*, II East Coker (1940)

To my beloved parents.

ACKNOWLEDGMENT

Many thanks are due to my advisors Dr. Jason Wang, Farzan Nadim and Jorge Golowasch. They have been tremendously supportive all through the program, providing valuable advice and encouragement. I am thankful to Yair Manor for meticulously proofreading this document. His feedback improved the quality of exposition considerably.

This work would not have been possible without the friends at the Stomatogastric Ganglion Lab, with whom I have enjoyed the opportunity to interact. Talking to them helped to articulate many of the basic concepts I tried to elucidate.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

A study of the electrical stimulus-response properties of a group of neurons provides an insight into the physiological mechanisms that control and coordinate the generation of neuronal activity, and their contribution to the production of behavior. Single-electrode extracellular recordings capture action potentials from several neurons in one stream. This multi-unit signal must be sorted into single-unit activity and corresponded with the neuron or group of neurons that generated it.

Several lines of research have lead to novel spike sorting algorithms using Bayesian modeling, neural networks, hierarchical and fuzzy clustering procedures, software-based filtering and wavelet analysis methods. The algorithms separate single-unit activity based on the characteristic differences in waveform shape generated by individual neurons. However, when neurons generate bursts of spikes as opposed to single spikes, several aspects of the spike waveform variability are observed even within the single-unit bursts. The distribution of the inter-spike interval statistics of burst activity is also highly variable. This necessitates a different approach to spike sorting.

We propose a comprehensive framework for burst detection and classification by making intuitive a priori assumptions of the form of variability among bursts. Distinct burst patterns are detected by representing the geometry of bursts in terms of the burst amplitude, frequency, duration and inter-burst interval. Our approach is especially suited for correlated and simultaneous burst firing activity.

Two studies that specifically address the sorting of burst activity are those of Fee et al., (Fee, Mitra et al. 1996), and Snider and Bonds (Snider and Bonds 1998). The basic principle is to first group the detected spikes into several preliminary clusters, and then systematically aggregate them to form putative clusters that represent single-unit activity.

Fee et al. (Fee, Mitra et al. 1996), hypothesized that two preliminary clusters result from the activity of the same neuron, if the combined ISI statistic does not differ significantly from the ISI statistic of the parent clusters. This is made under the assumption that each neuron is characterized by a certain refractory period. This method is applicable only to independently firing neurons. When several neurons in a local group of neurons fire simultaneously to form complex burst patterns, the hypothesis will not be valid and the clusters will not be combined. Also, when the activity of two different neurons is correlated, the algorithm will falsely combine clusters from the individual neurons.

Snider and Bonds (Snider and Bonds 1998) proposed another hierarchical clustering method to cater to correlated firing patterns. Their approach relies on the non-stationarity of neuronal activity, when waveform shapes change gradually over time due to various factors such as electrode drift, or other intrinsic biophysical properties. Under such assumptions, clusters belonging to the same neuron will be connected by a trail of points. Discontinuities between clusters indicate activity from separate neurons. This approach limits the applicability of this algorithm to situations when the signal has been cleaned of stochastic interferences for efficient processing. Also, this algorithm does not handle correlated firing from a single group of neurons.

CHAPTER 2

A PRIMER TO NEURONAL SIGNAL TRANSMISSION

This chapter introduces selective, basic concepts of neuronal signal transmission and processing. One of the fundamental issues in neuroscience is understanding how neuronal activity is generated, and how it contributes to the production of behavior. Essential to this pursuit, is the ability to study the electrical stimulus-response properties of an ensemble of neurons. Extracellular recordings of these properties capture simultaneous activity from several neurons in one single stream. The resulting multi-unit signal must be demultiplexed into component signals, each corresponding to the neuron that generated it.

The computational problem of looking for *patterns* in the neuronal signals will be defined within the context of these notions. This study is motivated by the need to formalize the temporal dynamics of a group of neurons, to further an understanding of how the neuronal population generates patterns of rhythmic activity and how the patterns are modified and regulated. By observing the activity of the neurons in concert, it becomes possible to quantify the collective computation of the sub-system. The discussion in this chapter will help to see why a formalism of the fine details of coincidental neuronal firing patterns, is crucial to the understanding of the signaling mechanism.

2.1 Elements of Neuronal Signal Transmission and Processing

Figure 2.1(A) shows a drawing by Ramón y Cajal (Cajal 1909), of a portion of the mammalian cortex. The elementary processing units of a nervous system are the *neurons*, which are connected to each other in an intricate pattern.

The neuron, as a dynamic electrical unit, is typically in one of two functional states – quiescent or transmitting an electrical signal. At any given time, a neuron is quiescent at a negative voltage, or producing stereotyped voltage pulses called *action potentials* that propagate long distances (Gerstner and Kistler 2002).

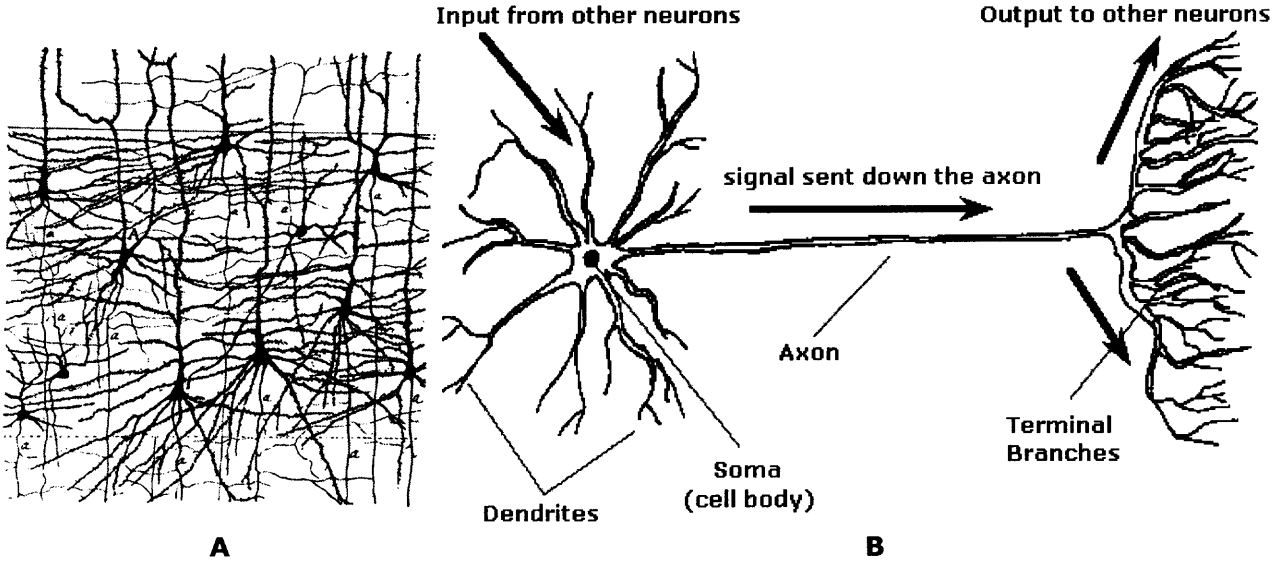


Figure 2.1 Elements of Neuronal Signal Transmission and Processing. (A) Camera Lucida drawing of a portion of mammalian cortex, as observed under a light microscope after golgi staining (Cajal 1909) (B) Functionally distinct parts of a neuron, and their role in neuronal signaling. Figure adapted from (Goren 2001).

Figure 2.1(B) illustrates the three functionally distinct parts of a neuron. Neurons are connected to each other at junctions called *synapses.* Signals are transmitted chemically, across the synapse (The reader is referred to (Cowan, Südhof et al. 2000) for details on synaptic communication). The neuron sending an action potential is commonly referred to as the *presynaptic neuron* and the receiving neuron as the *postsynaptic neuron.* When an action potential reaches a synaptic terminal, a series of complex chemical and electrical processes are triggered, which result in the production of a small voltage signal in the postsynaptic

neuron. The dendrites of a postsynaptic neuron act as an input device, receiving signals from the presynaptic neurons and transmitting them to the soma. The soma is the central processing unit, which performs a nonlinear processing step. When the total input to the postsynaptic neuron exceeds a certain threshold, the soma "fires", generating an action potential. This is the all-or-none signaling mechanism of the nervous system. The axon behaves like a cable and the terminal branches act like an output device that transmits the action potential to other neurons (Gerstner and Kistler 2002) and (Rieke, Warland et al. 1999).

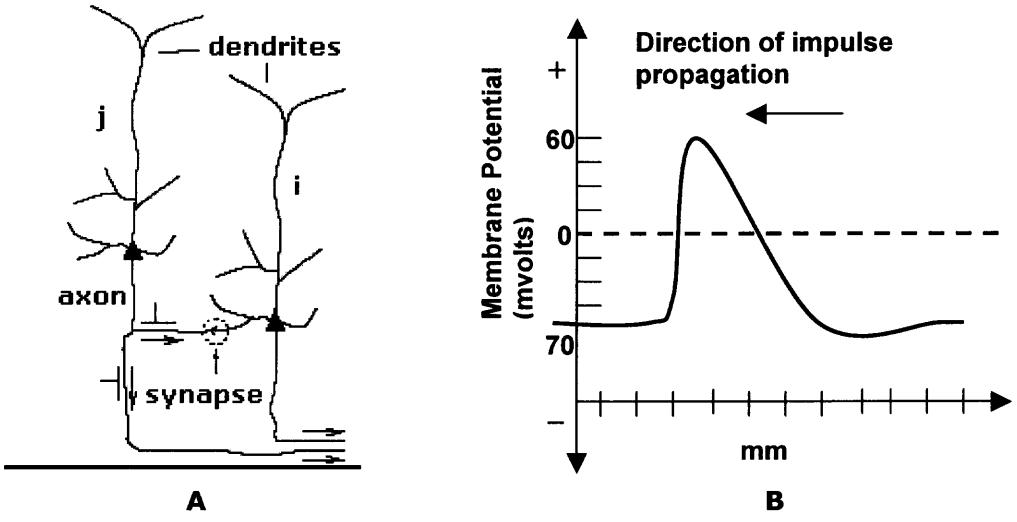


Figure 2.2 (A) Schematic representation of synaptic communication from a presynaptic neuron j to a postsynaptic neuron i. The synapse is marked by the dashed circle. Figure modified from (Gerstner and Kistler 2002) (B) The membrane potential along an axon at a singular time point. The voltage is shown as a function of space.

In certain systems, synapses are unreliable at signaling the arrival of single presynaptic action potentials to the postsynaptic neuron (Lisman 1997). However, bursts of spikes are reliably signaled. Recent theoretical studies backed by experimental evidence

((Cooper 2002), (Goense, Ratnam et al. 2003)), show that a burst of spikes, rather than single isolated spikes, increase the reliability of synaptic communication. Several studies are interested in bursting neurons to understand how the oscillations contribute to synchronization and neuronal processing (Singer 1999).

The effect of an input action potential on a postsynaptic neuron can be recorded *intracellularly* or *extracellularly*, using recording devices such as electrodes. The waveform of the action potential as recorded by the device, is called a *spike*. The sequence of action potentials generated by a neuron constitutes a neuronal *spike train.*

In general, action potentials fired by a single neuron are similar in shape. (However, in real neuronal recordings, extracellular waveforms of action potentials exhibit varying shapes. This is discussed in Section 3.2). Thus, a neuronal spike train can be characterized as a discrete series of temporal events, based on the spike firing times. This is the time series that is transmitted down the axon to the postsynaptic neurons. The spike trains of several neurons recorded simultaneously represent the spatio-temporal pulse pattern illustrated in Figure. 2.3.

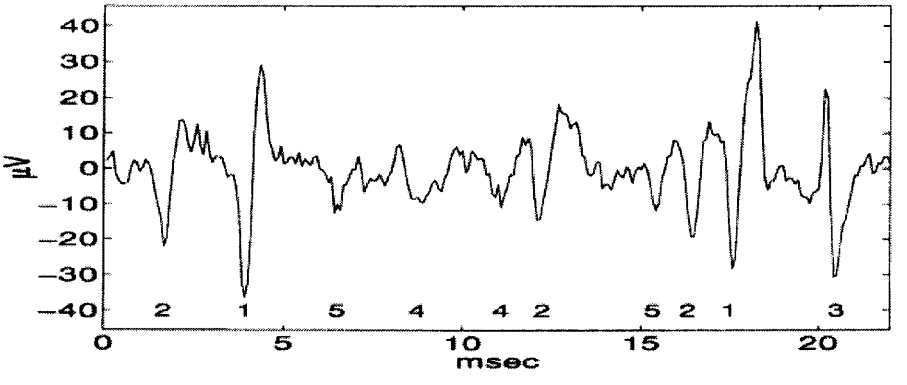


Figure 2.3 Spatio-temporal pulse pattern of 5 different neurons. The spikes of 5 neurons are shown as a function of time. Each spike firing is numbered by the neuron that fired it.

The recorded patterns are typically noisy. The introduction of *noise* into the signal is from many different sources such as the background activity of neighboring neurons, thermal noise, noise from the amplifier etc. (Refer to Section 3.1 for a more detailed explanation of noise sources).

2.2 Characterizing Temporal Dynamics of Neuronal Activity

There exist two major views on how neurons code information to be transmitted. The traditional view, the *rate code*, is that most of the relevant information is coded in the mean firing rate of a neuron. Mean firing rate is defined as a temporal average over a pre-defined time window. This rate can be calculated as the spike count average of a single neuron, the spike density of a neuron over several trials of an experiment, or the firing rate of a population of neurons (Rieke, Warland et al. 1999).

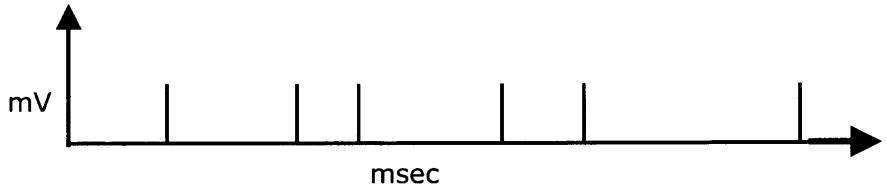


Figure 2.4 Rate Code: Information is coded in the mean firing rate of the neuron, over a pre-defined time window T. The spike firing times are marked by the short vertical bars.

$$\text{Rate code} = \frac{spikecount}{T}$$

However, the mean firing rate has been found to be very simplistic to explain complex neuronal activity. It does not sufficiently explain the fast reaction times of certain nervous systems observed experimentally, for instance, the neurons involved in visual processing. These neurons can complete computation in just 20-30 msecs. Such fast computation may not be possible if a neuron waits for several input action potentials to

calculate the mean firing rate [Rolls and Tovee, 1994]. Several other experimental studies indicate the use of specific firing times by neuronal systems. There is evidence that the spike firing times in the hippocampus of the rat conveys information on the spatial location of the animal, which is not fully accounted for by the firing rate of the neuron [O'Keefe, 1993]. Theoretical considerations indicate that single-neuron firing patterns in the mammalian auditory pathway are too random to effectively represent sound. Correlated activity of multiple neurons may convey information beyond that contained in the firing rate alone. A more recent view is that of the *temporal code* based on the precise firing time of a spike (Gray 1999), (Gerstner and Kistler 2002). Much current work is focusing on population activity, to understand how the coordinated sequences of spikes collectively represent the stimulus-response relationship (Gray 1999).

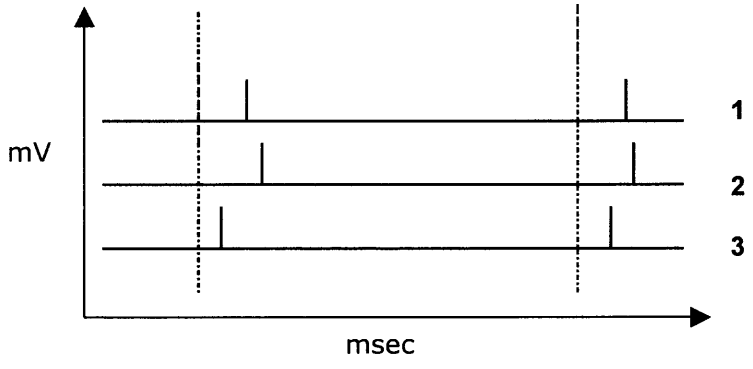


Figure 2.5 Temporal Code: The neurons 1, 2 and 3 fire at different times with respect to the periodic oscillation indicated by the dashed line. The spike firing times are marked by the short vertical bars. Figure adapted from (Gerstner and Kistler 2002)

Information-Theoretic approaches to understanding neuronal signal processing ((Johnson, Gruner et al. 2001), (Borst and Theunissen 1999) and (Barbieri, Quirk et al. 2001) for instance) use statistical models to test specific biological hypotheses. These models quantify the difference between firing patterns of neurons, in terms that advance an

understanding of what information is being transmitted. For this, an accurate statistical representation of the firing patterns is needed. Formal methods are needed to determine potential synchronization of consistent temporal relationship among multiple neurons to evaluate their statistical significance.

2.3 Multi-Unit Extracellular Recording

To observe the specific temporal correlation between spike patterns of multiple neurons, the activity of the neurons must be simultaneously recorded. However, multi-unit recording poses the problem of *isolation*. The neurons of interest (what we can refer to as foreground neurons) need to be properly distinguished from the background neurons, and then using one electrode for each cell, multiple single-unit intracellular recordings can be made. This approach will not scale reasonably with the number of neurons.

A more practical and convenient alternative is extracellular recording. The activity of several neurons is recorded in one single stream. Single-unit neuronal activity must then be reliably separated from the multi-unit signal. This is commonly referred to as spike sorting. An electrode is placed close to a group foreground neurons that are isolated from the rest of the surrounding using a Vaseline well. The activity of the foreground neurons is then significantly dominant over that of neighboring neurons. Figure 2.6 illustrates how neuronal signals can be observed by placing a fine electrode close to the axon.

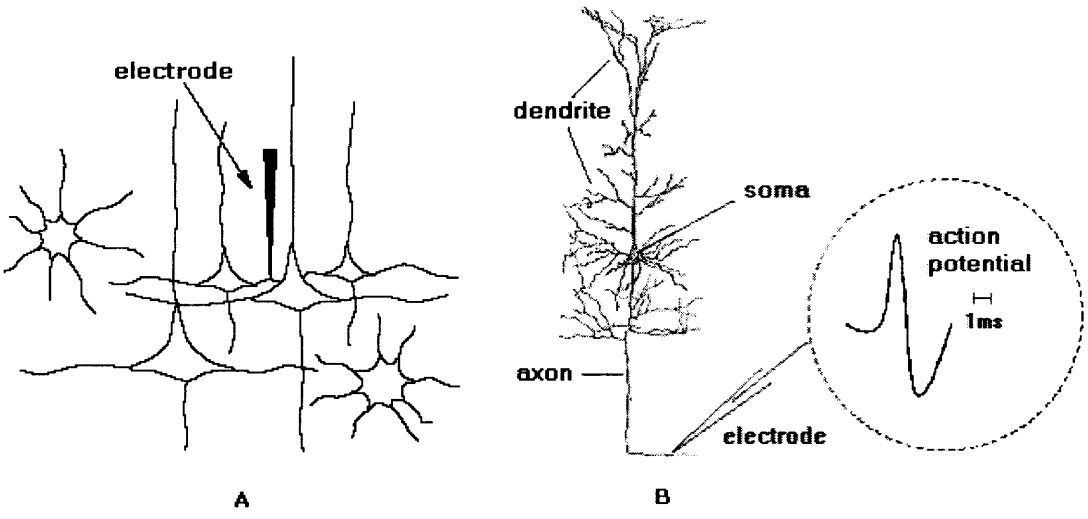


Figure 2.6 Extracellular Recording: (A) Electrical activity of a small population of neurons is recorded using an electrode. (B) Zooming into one neuron in the population of several. The electrode can pick up signals traveling down the axon. The inset shows a voltage trace in time, at a singular location. Figure adapted from (Gerstner and Kistler 2002).

Other popular recording techniques use tetrodes, stereotrodes, or arrays of electrodes. In principal these devices are bundles of electrodes. Each spike from a neuron appears as a stereotyped waveform on the voltage trace recorded by one electrode. A cross correlation of signals across the recorded channels allows disambiguation of spikes from different neurons. This study will deal only with *extracellular recordings* of neuronal signals in a *single channel*.

CHAPTER 3

THE SPIKE SORTING PROBLEM

The separation of single-unit neuronal activity from multi-unit, single-electrode recordings, presents an interesting pattern classification problem. The extracellular spike waveforms of different neurons have characteristically different shapes. This is due to several factors: the position of the neuronal axon on the surface or center of the nerve, the diameter of the axons, and the spatial geometry (for example, fixed versus tapering or irregular diameter). Spikes of different neurons are then sorted based on their characteristic differences in shape. Each unique extracellular waveform can be regarded as a pattern that can be used to visually sort the spikes according to the firing neurons. However, manual sorting methods break down when the patterns are difficult to distinguish visually due to noise, or are highly irregular. Manual sorting is also characterized by bias and inter-observer variability. The task can instead be completed effectively and efficiently by computational and statistical approaches.

The main objective of an automatic spike pattern classification approach is to detect individual neuronal spikes in complex, noisy extracellular recordings. The precise firing times of the spikes must be observed. As compared to manual sorting, automatic approaches are expected to be more reliable and efficient and should ideally perform the sorting in real time. To this end, there are several challenges that need to be overcome.

3.1 Interference due to Noise

Noise is a major deleterious factor in automatic spike-sorting. The recorded signal can be viewed as a linear superposition of signals from foreground neurons with

those from background activity neurons and other noise sources. Noise has two main components –stochastic interferences from background activity and noise from external sources. Activity of the neurons far from the electrode can contribute to high levels of correlated *neuronal noise*. Random significant interferences from background cells can temporarily skew the spike shape. [Figure 3.1(A)].

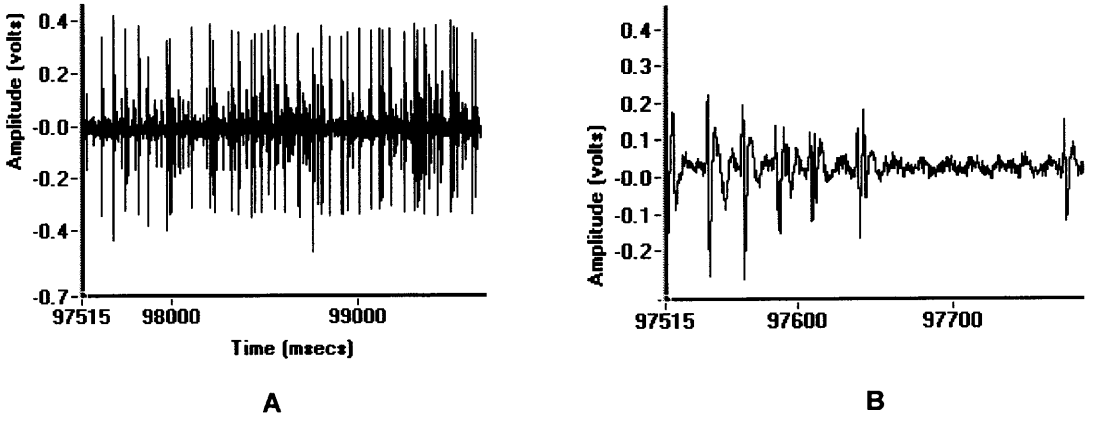


Figure 3.1 Interference due to noise. (A) Stochastic interference from neighboring neurons obscuring the foreground activity making it difficult to detect spikes. (B) Spikes intermingled (star) with electrical noise external sources (circle).

3.1.1 External Noise

Noise is also contributed by external sources, due to capacitive and resistive voltage fluctuations within amplifiers and other recording devices [Figure 3.1(B)].

3.2 Idiosyncratic Waveform Characteristics

When a neuron receives a variety of neuromodulatory input that can affect its input resistance, notable changes in waveform shape are observed. Also, in certain recordings, depending on the electrical setup, the electrode can move systematically over time, away from the original recording site (Lewicki 1998). Such drifts of the electrode position do not

consistently change all waveforms being recorded. This creates complex firing patterns in the trace [Figure 3.2(A)].

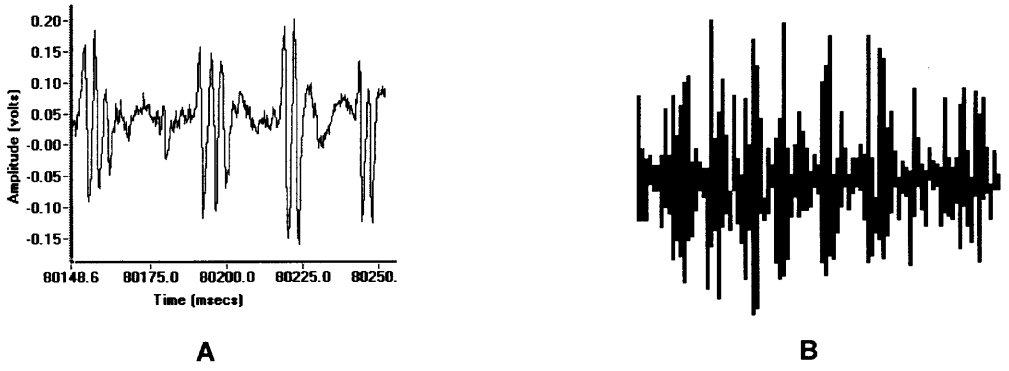


Figure. 3.2 (A) Slightly different waveform shapes of same neuron. (B) Complex burst of spikes.

3.3 Overlapping Spikes

Multiple spikes from different neurons firing at roughly the same time, superimpose, resulting in complex, incomplete waveforms. In multi-unit recordings, spike overlap can occur at high frequency if the firing rates of the neurons are high. Since the superposed waveform shape is not complete, the overlap has to be resolved for accurate spike detection [Figure 3.3].

3.4 Irregularity in *in vivo* Recordings

The relationship between firing patterns of a group of neurons recorded *in vivo* over several trials of an experiment can be seemingly stochastic (Gerstner and Kistler 2002). *In vivo* recordings are characterized by a high degree of irregularity making it difficult to reliably separate spikes from what may seem like noise. When recording *in vivo*, activity of the animal can cause small periodic displacements of the location of electrical activity relative to

the electrode. This can affect the shape and size of the action potential being recorded. Several studies report highly variable recordings in response to repeated stimuli, for example, in cortical neurons recorded *in vivo* (Holt, Softky et al. 1996), (Azouz and Gray 1999).

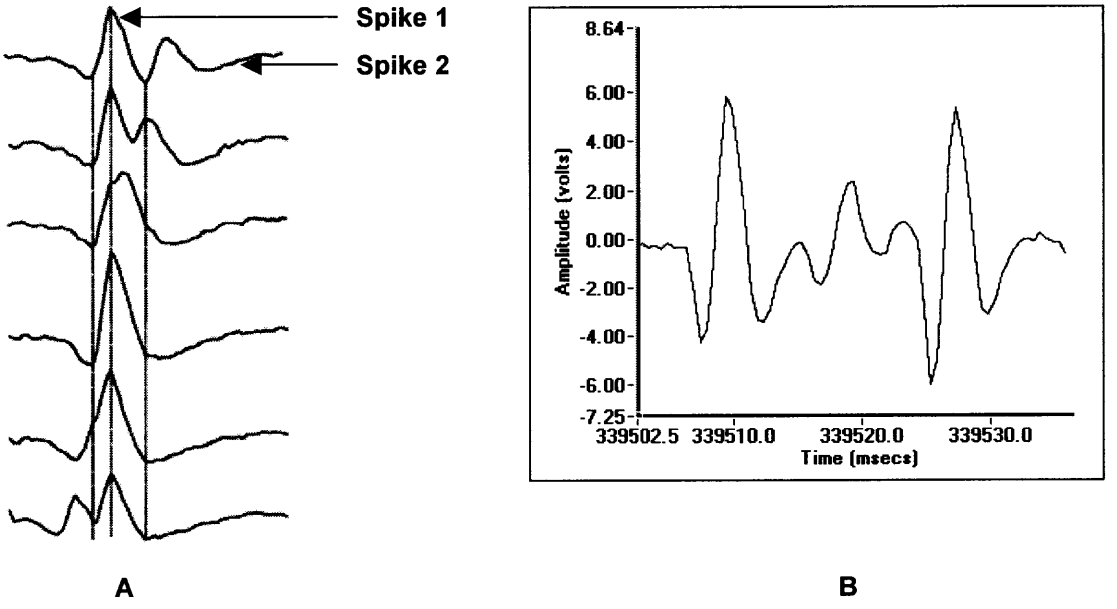


Figure 3.3 (A) Summation effects of two closely firing spikes. Figure adapted from (Zouridakis and Tam 2000) (B) Overlapped incomplete waveforms.

3.5 Complex Bursts of Spikes

Some neurons produce bursts of spikes as opposed to isolated spikes. When two or more neurons in a population burst simultaneously, it gives rise to complex spikes. In some systems, there is a systematic change in spike amplitude, during the course of a burst. For example, recordings from the cricket abdominal nerve cord show that action potentials generated at the beginning of a long burst can sometimes have slightly different amplitude than spikes occurring later in the burst (Gozani and Miller 1994).

3.6 Discussion

Significant deterministic influences can affect the shape of the spike waveforms in multi-unit recordings. The previous section discussed sources of variability that can make a spike train *time variant*. The activity in a recorded signal is non-stationary and has varying spectral characteristics. The probability distributions of the voltage observations can vary with time. The noise (neuronal and electrical components) is not necessarily isotropic with a gaussian distribution. In the example shown in Figure 3.4, classification would not be difficult. The action potential waveforms are quite different and separated in time. Figure 3.5 shows a more complicated section from the same recording that illustrates why spike sorting is a non-trivial problem. It is much more difficult to distinguish the action potentials from two different neurons here and especially hard to discern low-amplitude spikes from noise.

Automatic classification can effectively classify this activity and at the same time improve accuracy of temporal observations. This capability becomes especially important for experimental studies of neuronal signaling mechanism that use firing times of spikes. The following chapter reviews some of the techniques proposed for automatic spike pattern classification.

CHAPTER 4

A REVIEW OF SPIKE SORTING METHODS

Three elaborate reviews on spike sorting techniques are frequently cited in the literature – those of (Wheeler and Heetderks 1982), (Schmidt 1984) and (Lewicki 1998). This chapter will not rehash the approaches that have been discussed in these reviews. Recently, several exciting algorithms that address the inadequacies of earlier approaches have been proposed. Another comprehensive review is due. The principle, applicability, uncertainties and limitations of each of the newer techniques will be evaluated here in detail. The discussion will point out certain situations that can impair the applicability of the existing techniques. This will help to establish the need for a different approach to spike sorting.

The contribution of this review is twofold. First, the general approach to automatic spike sorting is formalized, based on working methodologies applied by several studies. Second, the techniques are broadly categorized according to their general sorting scheme, and an evaluation chart is presented as a summary. The discussion will emphasize novel approaches and only briefly mention the application of well-established theories.

4.1 A General Framework for Automatic Spike Sorting

Spike sorting can be split into four stages: spike detection, feature extraction, learning and classification. This section will present a unified framework grounded in pattern classification theory. Subsequent discussions of spike sorting algorithms will be made in the context of this framework.

4.1.1 Spike Detection

It is seen in a sample recording below (Figure 4.2), that the spikes are manifestly intermingled with noise. Spike detection involves isolating the neuronal activity from noise. This can be achieved using *amplitude-based threshold* discriminators where the signal-to-noise ratio (SNR) is acceptable (Bankman and Janselewitz 1995) or by performing more sophisticated time-frequency analyses that detect temporal increase in amplitude and frequency.

4.1.2 Feature Extraction

The goal of feature extraction is to characterize each spike by measurements whose values will be similar for spikes in one class and different for spikes in other classes. Therefore, it is important to find *distinguishing features*. Spikes can be represented by their complete voltage waveform where each of the time samples is used as a feature [Figure 4.3(A)], or by a wavelet transformation of the waveform [Figure 4.3(B)]. Another common approach is to characterize the shape of a spike using a reduced set of features. By sorting the different shapes, the spike trains of individual neurons are separated. The choice of representation determines both the efficiency and effectiveness of the sorting procedure:

Wavelet Analysis provides a view of the voltage waveform of a spike, in terms of a linear sum of features of different time scales. Wavelet transform coefficients derived from the temporal profiles of spikes have been used as feature extraction parameters (Yang and Shamma 1988). The coefficients that best preserve the characteristic shape of spikes are used for classification [Figure 4.3(B)].

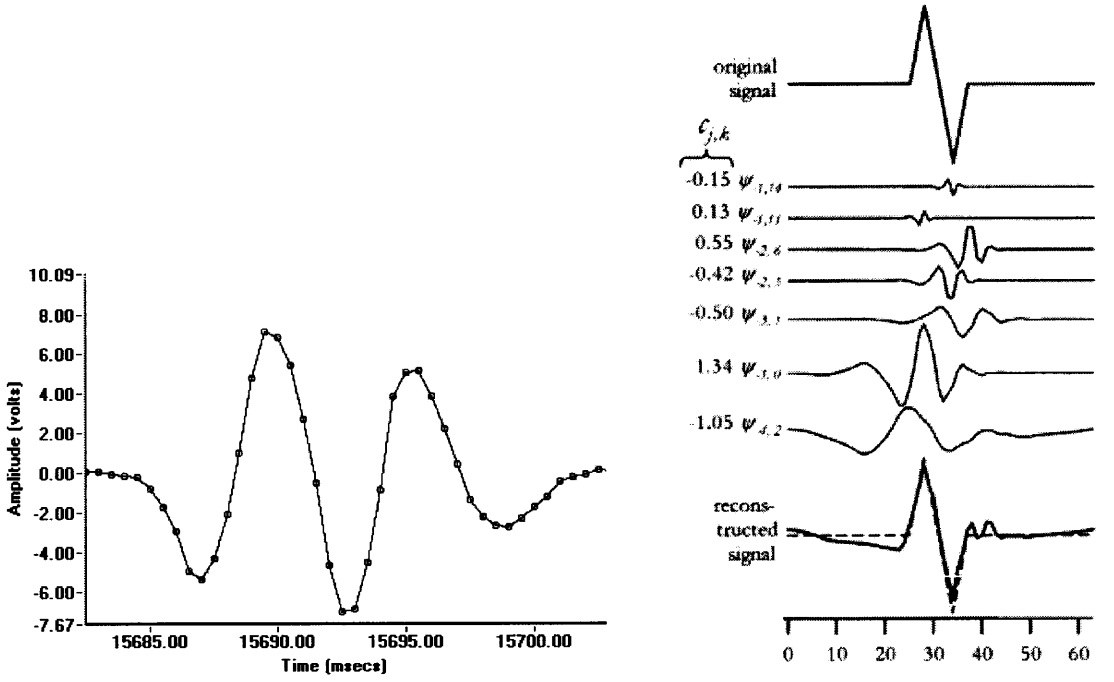


Figure 4.3 Feature Extraction from Spikes. (A) The time samples of a waveform (represented by the circles) can be used as features. (B) The original spike is decomposed by discrete wavelet transform. Decomposed wavelets are plotted below the original signal. The $\psi_{J,k}$ coefficients capture the spiky nature of the signal (Letelier and Weber 2000).

Reduced Feature Vectors are constructed based on a manually selected reduced set of parameters that represent essential variability among spikes belonging to different classes. These features typically include peak-to-peak amplitude, spike width and zero-crossings (Glaser and Marks 1968). A classification function is defined on the vectors built on these and other empirically determined features.

Principal Component Analysis (PCA) is one method commonly applied for choosing the features automatically (Glaser and Marks 1968). PCA is a linear dimensionality reduction technique; it determines an ordered set of orthogonal vectors that capture the directions of largest variation in the spike waveforms. The directions are called principal components and carry significant information about the spatial structure of the spikes in the

order of their contribution to the overall variance of the structure. The top principal components have been found to enhance the differences among the shapes of spikes.

4.1.3 Learning: Training the Classifier

After feature extraction, the spikes have to be classified based on their characteristic differences. The classification can be only as accurate as the initial representation of the variability among the spikes. *Learning* is the process of analyzing the feature values and finding compact representations of this variability. The entire dataset or a sample of it can be used for learning. The process of using the data to determine the classifier function is referred to as *training the classifier*. The dataset used for training, is called the *training sample*.

Supervised Learning: The learning can be supervised, where a labeled training sample (with category label for each pattern in the training set) is used to derive the classifier. Learning involves the estimation of a representative pattern (called *template*) for each class. The template may be selected manually based on visual inspection of the training sample (Bergman and DeLong 1992). However such manual methods are time consuming and suffer from bias, inter-observer variability and ill-defined selection criteria. Automatic techniques such as that of (Jansen and Maat 1992) have been employed, where a model of each class of spikes is automatically estimated as an average of the spike waveforms in the class.

Unsupervised Learning: The training can also be *unsupervised* where there is no explicitly labeled training sample and the classifier is derived from underlying probability distribution of the data, or by determining natural groupings called *clusters*, in the dataset.

4.1.4 Pattern Classification

The role of a classifier is that of a discriminant function, that uses the features derived from the spikes to assign each spike to a putative class. The objective is to partition the feature space into regions, separated by decision boundaries, where all points in one region, belong to one class. Classification can be performed using distance metrics that measure similarity between spikes, or by comparing the spikes with the set of pre-determined templates that are representative of the different classes. In the template-based approach, spikes are classified based on a *goodness-of-fit* measure with each of the templates.

4.2 Recent Developments in Spike Sorting

The studies evaluated in the reviews mentioned earlier, can be regarded as the *first-generation* of spike sorting techniques. They were primarily concerned with automating the spike detection, representation and classification procedures. It was assumed that the spike waveform is stationary and that noise is isotropic with Gaussian distribution. As was discussed in Chapter 3, these assumptions may prove restrictive for real neuronal recordings. Despite the limiting assumptions, the findings of those studies guided the application of more rigorous procedures. A *second generation* of algorithms have been developed and tested, that are robust to a breach of most of the assumptions.

The review chart below summarizes some of the recent approaches to spike sorting. The problems they solve (noise attenuation, spike overlap etc.) and the parameters they meet (real time sorting, efficiency) are noted.

Algorithm	Main Focus	Principle	Applicability / Limitations
Software-Based Filtering			
(Gozani and Miller 1994)	Resolve: 1. Temporal profiles similarity 2. Overlaps	1. Complete voltage waveform representation. 2. Supervised Learning of templates	Multi-channel recordings for accurate sorting
Artificial Neural Networks			
(Chandra and Optican 1997)	Resolve: 1. Complex decision boundaries. 2. Overlaps	All-inclusive training sample – with overlapped spikes and neuronal noise	1. Input includes all data points, not just detected spikes 2. Accuracy - Reliant on labeled training set
(Bohté, Poutré et al. 2002)	Automatic detection of complex clusters	Multilayer Radial Basis Functions. Unsupervised Clustering	Computational model of the physiological spiking neuron model
(K.H.Kim and S.J.Kim 2000)	Spike detection under low signal-to-noise ratios	Learning the underlying probability distribution	Accuracy - Reliance on labeled training set

Table 3.1 Spike Sorting Techniques – Evaluation Chart

4.2.1 Software-Based Linear Filtering

This approach is based on deriving optimal filters that respond maximally to spikes from one neuron and attenuate all other spikes and background noise.

Gozani and Miller (Gozani and Miller 1994) applied software-based linear filters to the problem. A spike is first detected within an empirically determined time window and characterized by its complete voltage waveform. Template waveforms are estimated automatically by modeling the training sample as an optimum number of average waveforms. A linear filter is derived for each template and is optimized to filter out spikes from other neurons as well as background noise. The spikes are then convolved with the set of filters and classified according to the filter that generates the largest response.

Spike overlaps are resolved by comparing the Root-Mean-Square (RMS) distance to each of the templates, with an empirically defined threshold. If a fit is not found, a new template is created.

Algorithm	Main Focus	Principle	Applicability / Limitations
Fuzzy Clustering			
(Zouridakis and Tam 2000)	1. Creation of spike templates 2. Overlap resolution	1. Complete voltage waveform representation. 2. Fuzzy K-means algorithm	Tested on Synthetic spike trains.
Hierarchical Clustering			
(Fee, Mitra et al. 1996)	Sorting in the presence of anisotropic noise	Variability – characterized by neuron's refractory period	Applies to independently firing neurons
(Snider and Bonds 1998)	Cope with non-gaussian changes in waveform shape	Clustering using time samples of spike as features	Considers correlated firing. Offline only
Wavelet Analysis			
(Zouridakis and Tam 1997)	1. Overlap resolution 2. Sorting spikes and determining exact firing times.	1. Supervised learning to derive templates. 2. Shift-invariant wavelet transform	1. Tested on synthetic spike trains 2. Dependent on initial training set
(Letelier and Weber 2000)	Similar temporal profiles for spikes from different neurons.		
(Hulata, Segev et al. 2002)	Effective Feature Extraction		

Table 3.1 Spike Sorting Techniques – Evaluation Chart (Continued)

This technique can resolve linear spike superposition, however the spikes have to be recorded on more than one channel. (Roberts and Hartline 1975) tested a similar approach and concluded that optimal filtering technique is ideal only for multi-channel recordings

where the neuronal activity is recorded on at least two channels. While the procedure is effective in resolving spike superposition, it assumes that the spike train is time invariant and that noise is stationary and uncorrelated with neuronal activity.

4.2.2 Artificial Neural Networks

This class of techniques applies a supervised connectionist neural network classifier to the spike-sorting problem. Chandra and Optican proposed an algorithm to simultaneously detect and classify spikes, by training a neural network on samples that included overlapped spikes and neuronal noise (Chandra and Optican 1997). The training phase enables the neural network to resolve superposition using just single-channel recordings. The hidden units perform nonlinear classification, and since all overlapped spike combinations are learnt in the training phase, this stage also resolves superposition.

When the spikes form complex clusters in multidimensional feature space, a linear discriminant function is inadequate. The non-linear classification capabilities of neural networks are therefore appropriate for the spike-sorting problem.

Bohté et al., (Bohté, Poutré et al. 2002) presented an elegant time-based artificial neural network to perform unsupervised clustering using multilayer Radial Basis Functions (RBF). This approach can be viewed as an "analysis by synthesis", where knowledge of the domain is used to build a model of the spiking neuron. By formalizing the physiological model of the firing neuron, spikes are classified based on how they are generated. Thus, the classifier exploits physiological rather than descriptive time-domain parameters.

The neural network is trained to optimize one radial basis function for each of the classes of spikes. The spikes are characterized using their precise firing times. When an

input spike is presented to the network for classification, only the node that performs the corresponding RBF will fire. The performance of time-based neural networks is dependent on accurate measurement of the spike firing times.

K.H.Kim and S.J.Kim (K.H.Kim and S.J.Kim 2000) used supervised neural networks to learn the underlying *probability distribution of training samples*. They presented a technique for spike detection and classification under very low signal-to-noise ratios when typical amplitude-based threshold methods fail. A spike is detected based on the temporal increase in signal amplitude and frequency when a spike is fired. Time-frequency analysis is performed to determine instances where the signal amplitude and frequency exceed a pre-determined detection threshold. The threshold is manually adjusted based on some sample spikes. The time samples of the detected spike waveform are then input to classifier. Multi-Layer Perceptron and Radial Basis Function Networks are used as classifiers, due to their adaptive capabilities in learning the probability distribution of input data.

The classification performance of neural networks is heavily reliant on the availability of an appropriate labeled training set. Also, when the spectral characteristics of neuronal noise are similar to that of the spikes, the detection and classification performance is affected.

4.2.3 Hierarchical Clustering

The basic approach is to "over-cluster" the data into many small groups and then resolve which of these clusters can be combined to represent the activity of single neuron.

Fee et al., (Fee, Mitra et al. 1996) proposed a technique that could classify spikes in the presence of anisotropic noise and non-gaussian variability. A spike train is characterized

based on its neuron's refractory period. (The refractory period is a physiological property that prevents the neuron from firing twice within a certain time period.) An empirically selected subset of the spikes is first split into several means. The remaining spikes are assigned to the clusters thus formed, based on their distance to the means. The clusters are then aggregated pair-wise, to form single-unit clusters. The aggregation is done based on a "connection strength" between the pairs of clusters and inter-spike interval (ISI) statistics of the spikes in each of the two clusters. The connection strength is a function of the distance between the spikes in each cluster and an *energy operator.* This operator is chosen empirically, based on the average pair-wise distance between the spikes in each cluster. The pair of clusters with the maximum connection strength is combined, provided a hypothesis based on the ISI statistic is valid. The hypothesis is that the distribution of the ISI of the combined cluster should have as good a refractory period as the distribution of ISI of each of the pair of clusters.

Since the aggregation criterion in this approach is based on the average ISI observations over a long time period, the statistics may not reflect local changes in firing patterns. Also, since connection strength is calculated between pairs of clusters, the approach may prove to be computationally intensive.

Snider et al., (Snider and Bonds 1998) presented an off-line "join-the dots" approach to classify waveforms that change their shape linearly, over time. This approach is applicable to neurons whose firing activity is correlated. The strategy is to aggregate pairs of clusters based on continuities in the density of points falling between the clusters. The underlying hypothesis is that when the shape of a spike waveform evolves over time, its projection into a multi-dimensional waveform space leaves a trail of points that can be followed from one

cluster to the next. The drawback of this approach is that the sorting decision requires a priori knowledge of all spike waveforms in a given recording.

4.2.4 Fuzzy Clustering

Zouridakis and Tam (Zouridakis and Tam 2000) applied K-means fuzzy clustering to automatically extract spike templates for use in template-based classification. Each spike is represented by the time samples of its waveform. The fuzzy approach permits fractional membership of a spike in many different classes, instead of a crisp membership (with say, value of 1 for assignment to one particular class, and 0 for all other classes). The grouping is evaluated using a distance-based similarity metric that optimizes the inter-class and intra-class similarities. Overlaps are resolved by assigning fractional membership values to the spikes. A template is estimated for each cluster, as a weighted average of all the points in a cluster. The weights are equivalent to the membership value of each point in the cluster.

The variability among spikes, and the general scheme of overlap resolution is characterized by a fuzzy membership value that is derived as a function of empirically determined fuzziness index and Euclidean distance. The validity of this representation and its applicability to real recordings is questionable. K-means is a *lazy*, computationally intensive algorithm. Different groupings can be derived for the same data set, based on the initialization conditions. The cluster means has to be recomputed at every iteration of the algorithm and this can impede real-time processing.

4.2.5 Wavelet analysis

Spectral analysis is a very popular technique in signal processing, to identify the different frequency components in a signal. But in neuronal recordings where the spectral features of the signal are transient, a time-frequency analysis is more appropriate since it determines spectral components in the signal during a given window of time. Wavelet analysis decomposes a signal into different resolutions, based on its localized frequency components. The energy or power of a signal is given by a two-dimensional expansion function of frequency and time. This expansion function has two parameters called *scale* and *translation*, which are interpreted physically as the inverse of frequency and time respectively (Strang 1993). In the application of wavelet analysis to spike sorting, spikes are represented as localized functions called *wavelets* that span the frequency and time domains. The transformation coefficients of the wavelets characterize the variability among spikes and provide separation ability.

Zouridakis and Tam (Zouridakis and Tam 1997) proposed a shift-invariant discrete wavelet transformation to *resolve spike overlaps*. Template spikes were learnt from a sample of the dataset by applying fuzzy clustering techniques. Two vectors were created, representing the amplitude and phase of a discrete wavelet transform of the template spikes and its time-shifted versions. The result is that the time-shifted versions of a transient signal have identical amplitude vectors, but different phase vectors. The phase of the arbitrarily shifted sequence of the transient spikes is corrected in the wavelet domain by replacing its phase vector with that of the original template. An inverse wavelet transform of the resulting sequence of spikes restores the phase of the time-shifted signal to that of the original. A

fuzzy classification procedure is applied to the phase corrected spikes to detect the constituent spikes and estimate their firing time.

The performance of this technique depends on the initial selection of spike templates. The training sample, from which the templates are learnt, has to account for variability of spikes in the entire spike train.

Letelier and Weber (Letelier and Weber 2000) used a pyramidal decomposition algorithm to decompose signals in the time-frequency space, to address issues of *similarity between temporal profiles* and reliable distinction of signal from noise. The classification is however performed manually, by means of a graphic software interface that allows the user to visually delimit the clusters. The discrete wavelet transform coefficients derived from temporal spike-profiles are used as characterizing features. The objective is to estimate the smallest possible number of coefficients that represent the distinctive features of spikes. The technique is particularly noise resistant, since it eliminates the data points contained in the noise frequency range.

Application of wavelet decomposition techniques to spike sorting can fail due to inadequate resolution: for large frequency values, the time resolution of wavelets can be greater than the width of the spike. With time windows larger than the spike width, overlaps cannot be resolved and even regular spikes cannot be detected properly. Wavelets resolved thus, actually correspond to multiple spikes instead of one.

Hulata et al., (Hulata, Segev et al. 2002) provided a solution to this problem by varying window size as a function of frequency. They automated feature extraction by using wavelet packet decomposition to decompose the time-frequency space into a more adaptive set of localized functions. If the wavelets can be seen as filters spanning time and frequency

domains with one filter for each frequency, wavelet packets are multiple filters of equal bandwidth, spanning the entire frequency domain. Then the time-frequency domain is over-represented by overlapping filters. Each spike is represented using a set of non-overlapping filters spanning the time-frequency domain that optimally localizes the features of the spike. The most discriminating packets are selected for detection and sorting. The union of the packets is viewed as a highly optimized filter for neuronal spikes that preserves its shape and filters out most of the background noise. The different spike shapes in a window are preserved even when they overlap. K-means algorithm is then used to automatically separate the clusters.

4.3 Discussion

The general decomposition approach taken by the techniques discussed in the previous section can result in an *overfitting* situation. As efforts are focused on finding a model that *best fits* the data, the accuracy of classification in novel situations is compromised.

4.3.1 Overfitting Models

As can be observed from the evaluation chart [Table 4.1], many of the techniques employ supervised learning to derive template spikes. A classifier that is trained using labeled samples, runs the risk of being *over-tuned* to a particular training set rather than representing underlying characteristics of the spikes that have to be classified. The templates that are learnt may result in perfect classification of the training sample itself but would yield poor performance on "unseen" patterns. Complex models that are developed based on training data, may provide decision boundaries that are non-linear and complicated. But the model

will not provide good generalization, since the performance will be degraded for patterns that are not in the training sample. When trial-to-trial variability is also factored in, a specialized model will certainly compromise classification accuracy. Yet another drawback of supervised learning is that the decision boundaries that are learnt are only as accurate as the initial labeling.

4.3.2 Dealing with Neuronal Noise

None of the techniques have explicitly handled outliers due to neuronal noise. Variations due to the background activity have been characterized, but no proactive measure is taken to remove the noise before clustering or classification. In situations where background activity is frequent, the noisy inputs to the classifier will degrade computational efficiency and can also affect the learning and classification accuracy.

4.3.3 Bursting Neurons

The latest published review on spike-sorting methods (Lewicki 1998) discussed four situations in which the *first generations* algorithms perform poorly – high-frequency burst firings, transience due to electrode shift, non-stationary background noise and spike overlaps. The *second generation* of spike sorting algorithms has extensively addressed three of the four the issues: transient signal, non-stationarity and spike overlaps. However, situations in which neurons generate high-frequency bursts have not been paid enough attention.

Section 3.1 discussed the variability in a signal due to complex bursts. In the clustering procedures, this variability results in smearing or elongation of the clusters.

"When a neuron generates an irregularly shaped spike, such as a bursting neuron, many clustering algorithms often fit the resulting data with two or more classes" (Lewicki 1998).

4.3.4 Overlap Resolution

The second-generation of algorithms deal with resolving spike overlaps, to accurately measure firing times of all the spikes. The algorithms simultaneously track changes in spike waveforms and cope with factors that contribute to non-stationarity. The overlap decomposition algorithms are sensitive to the detected spike peaks. If several time samples occur during an overlapped spike peak, a peak centered at each of the samples can give slightly different waveform shapes. A caveat to the subtraction-based overlap resolution approaches is that it is possible to introduce false spike shapes if the spike occurrence time is not accurately estimated (Lewicki 1998). Also, when the rate of superposition is high, spike overlap resolution can get computationally demanding. Given these pitfalls, the overlap resolutions techniques will be inappropriate in circumstances where the neuronal recording is characterized by correlated bursting.

4.4 The Case for a Burst Extraction Approach

Several studies show that the statistical structure of the spike train can enhance the synaptic input detection by indicating the presence of a signal with a burst of activity rather than isolated spikes ((Goense, Ratnam et al. 2003), [Izhikevich, 2003 #25]). Their observations suggest that bursting neurons not only increase reliability of synaptic transmission, but also provide effective mechanisms for selective communication between neurons.

Lisman (Lisman 1997) showed that shorts bursts of high-frequency firing may have special importance in brain function. The study demonstrated experimental evidence that bursts of spikes provide more precise information to postsynaptic neuron than isolated spikes, thereby suggesting that the best stimulus for exciting a cell is coincident bursts.

Rhythmic motor systems are often generated by self-sustained patterned activity from networks of neurons called Central Pattern Generators (CPGs) that produce rhythmic burst patterns without any sensory input (Marder and Bucher 2001). To understand the fundamental mechanisms by which the rhythmic circuits generate and regulate patterns, the temporal dynamics of the neurons, as evinced in the burst activity, must be characterized.

The neurons in the brain reward circuit that code for various cognitive aspects of goal-directed behavior have specialized properties that can make them switch between spiking and high frequency burst-firing modes under certain conditions. Researchers are interested in the study of this conditional output, to determine whether bursting mode is used to increase synaptic potentiation (Cooper 2002).

In such situations where neurons generate complex high-frequency bursts as opposed to isolated spikes, a density mixture model cannot adequately characterize the statistical structure of the spike train. If the procedures discussed earlier are applied to these situations, the accuracy of classification may degrade. An overfitting model of the subtle variations of neuronal spike shapes will falsely assign the dissimilar spikes of a burst to different classes.

Wavelet analysis has been applied by (Li, Magnuson et al. 2000) to determine an energy profile for the signal in a way that reflects the burst activity of the neurons. The rhythmic busting activity that they set out to separate is characterized by low-frequency (as opposed to the other) activities in the signal. Their approach will not be applicable to

situations where there is sequential burst activity of roughly the same frequency, from different neurons. Also, their approach extracts bursts but does not classify them.

The hierarchical clustering method of Fee et al., (Fee, Mitra et al. 1996) is applicable to bursting neurons, but if the spike shape variability within a burst is high, the clustering fails. The algorithm also relies on accurate detection of individual spikes within the burst.

Tam (Tam 2002) proposed an algorithm to detect bursts based on the statistics of inter-spike intervals. The algorithm may not be applicable to situations where the neuron switches between spiking and bursting modes, since it cannot detect single-spike bursts.

Given the above limitations, this study presents an approach where the decomposition can be achieved based on statistics that characterize burst pattern variability. Where rhythmic burst activity is concerned, the accurate measurement of the onset and offset times of burst activity rather than firing times of spikes in the burst, is important. The guiding principle is to formalize the same features that make visual distinction of the burst patterns possible. The next chapter presents an overview of the proposed approach. The contributions made in this study are highlighted.

CHAPTER 5

ÜBER-CLAϢS: AN OVERVIEW

This chapter presents an overview of the proposed algorithm. The problem is studied in the context of the *pyloric* neuromuscular system of the crab *Cancer borealis*. The design of the algorithm is motivated by the visual inspection procedure used by human-observers, in classifying the bursts according to the characteristic differences in their *patterns*. The guiding principle of this algorithm is to formalize the same features that make visual distinction of the burst patterns possible.

5.1 The Stomatogastric Nervous System

The Stomatogastric Nervous System (STNS) of decapod crustaceans such as lobsters, crabs, crayfish and shrimp, is a well-studied network. It serves as a model system in contributing to the general understanding of neuronal circuit operation (Nusbaum and Beenhakker 2002). The stomatogastric ganglion (STG) controls muscles of the animals' foregut, which is responsible for storage, chewing, and filtering of food for digestion. It contains some 25-30 neurons that generate distinct rhythmic burst patterns, simultaneously innervating muscles in different regions of the animal's foregut. The rhythmic output is controlled by *Central Pattern Generators;* the generation of patterns is dependent on mechanisms that operate on different time scales.

The STG is divided into two sub-networks - the pyloric and the gastric networks, with some interconnections between them. The gastric and pyloric rhythms can be recorded extracellularly using a single electrode, from the Lateral Ventricular Nerve (*lvn*) [Figure 5.1] through which many of the neurons project their axons. The extracellular activity of the

gastric rhythm is the relatively long-duration, rhythmic bursts recorded from the lateral gastric (LG) neuron. The pyloric rhythm is a faster tri-phasic rhythm, and is characterized by the sequentially repeating bursts of spikes in the lateral pyloric (LP), pyloric constrictor (PY) and pyloric dilator (PD) neurons (Nusbaum and Beenhakker 2002). Recordings can be carried out over several hours or even days, and the signals are saved continuously in preset time windows.

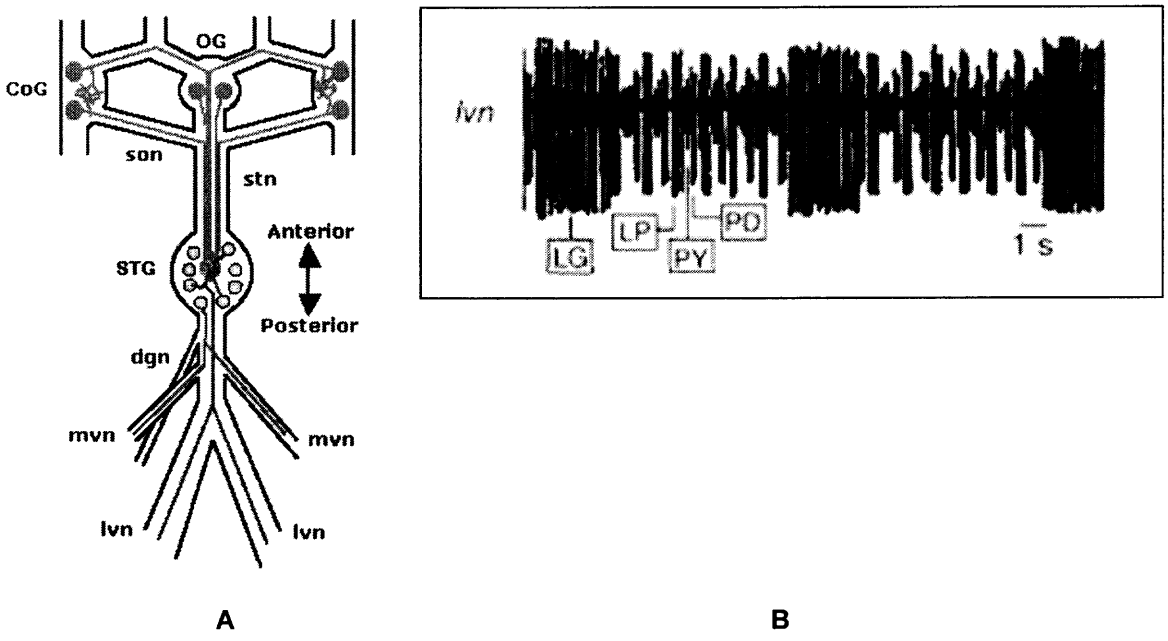
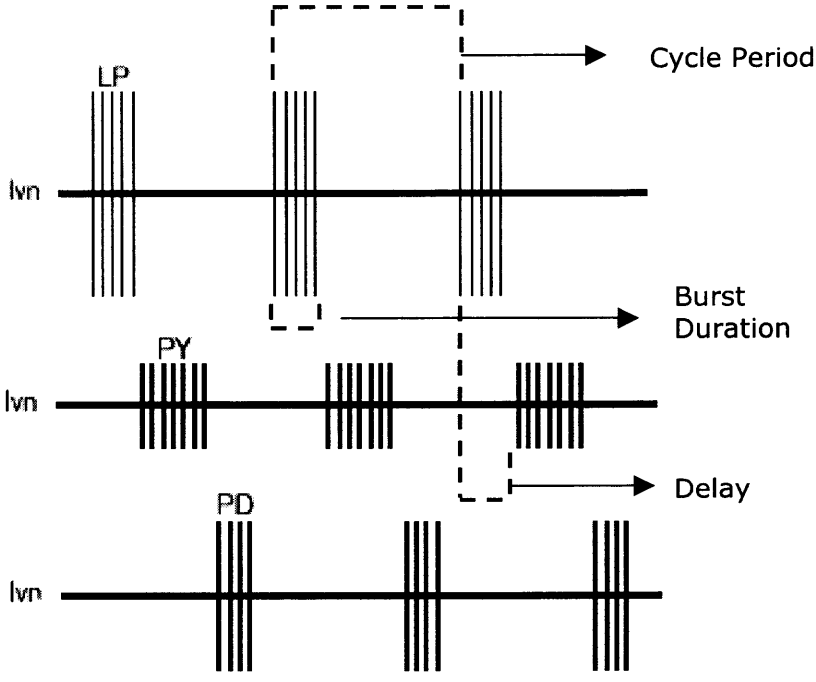


Figure 5.1 (A) Schematic of the STNS in the crab *Cancer borealis*. Abbreviations: commissural (CoG) and oesophageal (OG) ganglia; Lateral (*lvn*) or Medial (*mvn*) Ventricular Nerves; Dorsal Gastric Nerve (*dgn*). (B) The extracellular action potentials of three neurons LP, PY and PD on the *lvn*. Adapted from [Nusbaum and Beenhakker, 2001].

The three neurons fire with relative delays, rendering the spatio-temporal pattern in Figure 5.1(B). The parameters used to quantify the rhythmic patterns include cycle period, burst duration, duty cycle, and phase of firing of an individual neuron. These measurements are defined in Figure 5.2. Experimental considerations have shown that the burst duration and

inter-neuronal firing delay change proportionally with changes in cycle period (Marder and Bucher 2001). This phenomenon is called phase maintenance, where phase $\phi = \frac{Delay}{Cycle\,Period}$.

The objective of the algorithm is to measure the time courses of the three neuronal bursts, to advance an understanding of the intrinsic phase regulation mechanism.



Phase is given by $\phi = \frac{Delay}{Cycle\,Period}$; Duty Cycle = $\frac{Burst\ Duration}{Cycle\ Period}$

Figure 5.2 The coincident activity of the three neurons of the pyloric network is shown, along with the time lag between them. Cycle period, burst duration and inter-neuronal delay are illustrated.

5.1.1 STG Pyloric Rhythm Characteristics

The recordings are typically characterized by high SNR, which allows reliable detection of spikes in the presence of electrical noise. Variability in the signal is mainly due to correlated background activity, high rate of spike superposition within PD and PY bursts and occasional sporadic activity of the bursts.

During each gastric burst (LG neuron burst), the activity of the LP, PY and PD neurons is obscured [Figure 5.4(B)]. Such background activity of gastric network is considered to be neuronal noise. There are other sensory neurons that project their axons along *lvn,* that are normally silent. However, these neurons may fire occasionally, contributing to. There are two PD neurons and often, they fire spikes at roughly the same time. Therefore there is a high rate of superposition of spikes in PD bursts. Spikes differ in shape within the same PD burst and also between different PD bursts in the same recording. Similarly, there are about eight PY neurons and they generate complex bursts as well. The PY neurons can also fire during LP time. It is therefore likely that the first few spikes of the PY burst overlap with the spikes at the end of the LP burst. The recordings are also characterized by sporadic burst activity. Sometimes, the LP neuron has been observed to fire at relatively low frequencies, resulting in single-spike bursts [Figure 5.3(A)] or no bursts at all [Figure 5.3(B)].

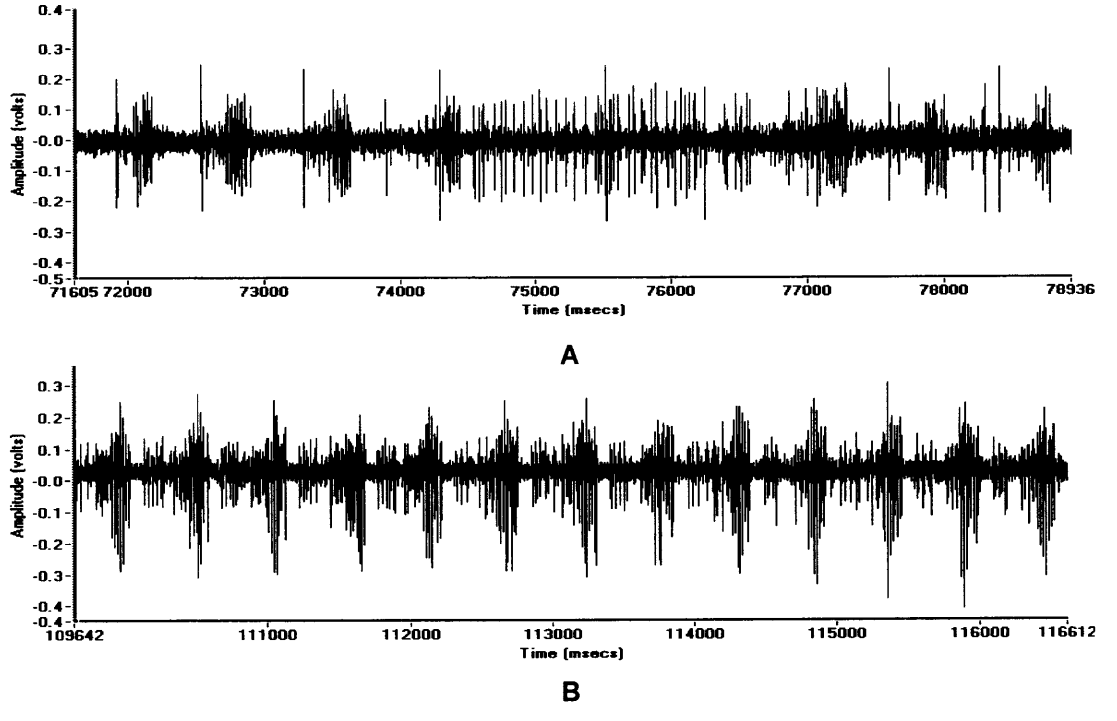


Figure 5.3 (A) A recording where LP is firing only one spike or missing. (B) LP bursts are completely missing in this recording.

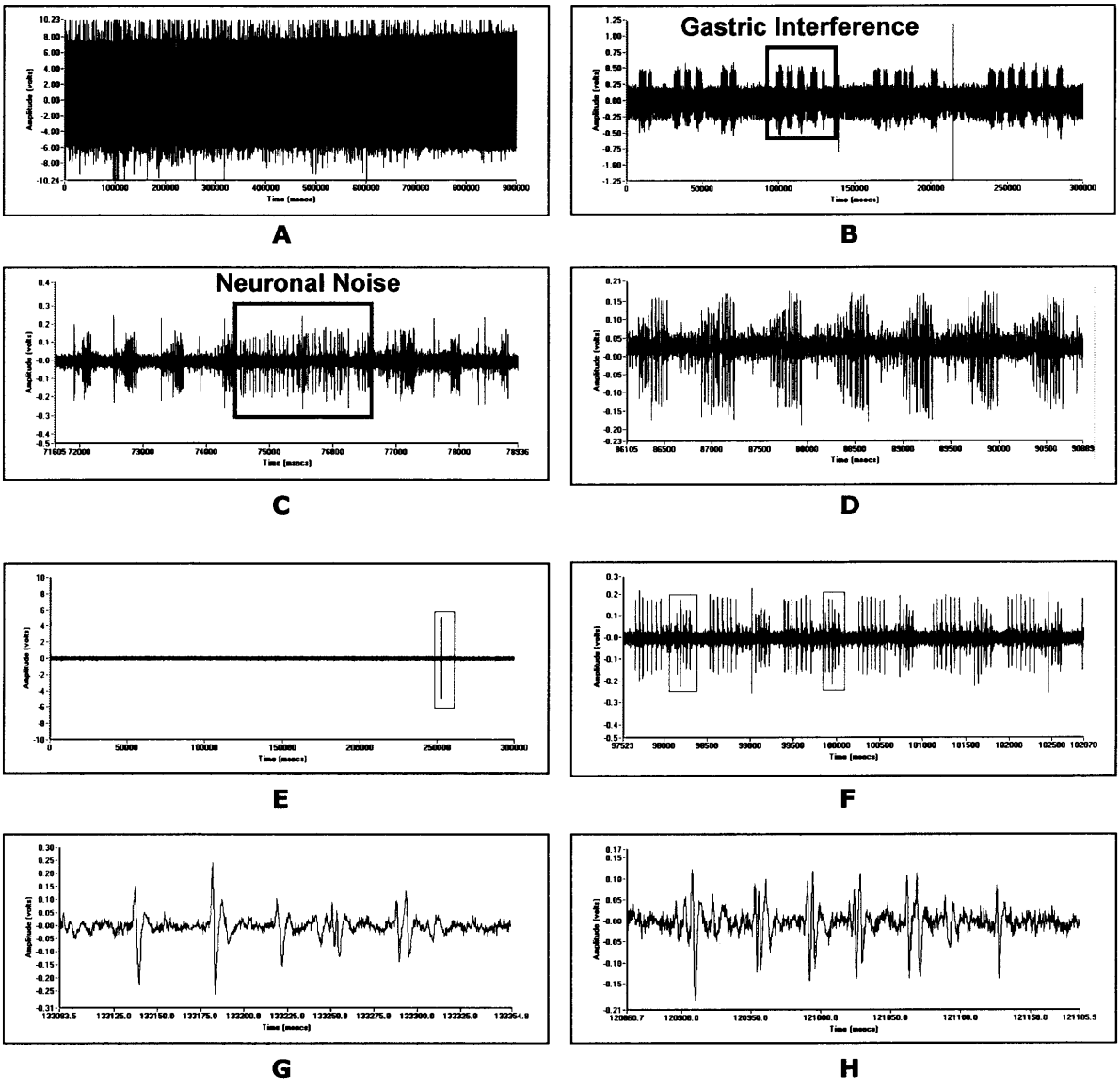


Figure 5.4 (A) A 5-minute extracellular recording of the pyloric rhythm from *lvn*. The rhythm does not typically have any significant shifts in location or scale over time. (B) Neuronal Noise: The gastric interference obscuring pyloric activity. The gastric bursts are characterized by durations higher than that of the pyloric bursts. (C) Neuronal Noise of roughly same amplitude as the pyloric bursts. (D) A recording where even visual distinction of the LP, PY, PD patterns is difficult. The LP bursts are completely missing. (E) Outliers: The pyloric activity is reduced to a baseline in this figure to highlight the high-amplitude outlier in the recording. (F) The rectangle highlights two PD bursts that look very similar. A zoomed-in view of the two bursts is given in figures G and H. (G, H) These figures highlight the variability in shape of the spike waveforms of two PD bursts that look very similar in figure F.

5.2 Overview of the Algorithm

The proposed algorithm characterizes a burst pattern using features such as burst amplitude, frequency, duration and inter-burst interval. Bursts having similar feature values cluster together in the high-dimensional feature space. An unsupervised learning technique is suggested, that does not use labeled samples to train the classifier. No prior forms of underlying density functions of the data are assumed and no *a priori* assumptions are made about distribution of bursts or noise. The unlabeled set of bursts is first grouped into clusters that and then labeled based on domain-specific knowledge.

A typical extracellular recording of the pyloric rhythm on the *lvn* is given in Figure 5.4(A). The signal can be observed as a univariate time series $v(t)$ where voltage observations are recorded sequentially over time increments Δt [Figure 5.5(B)].

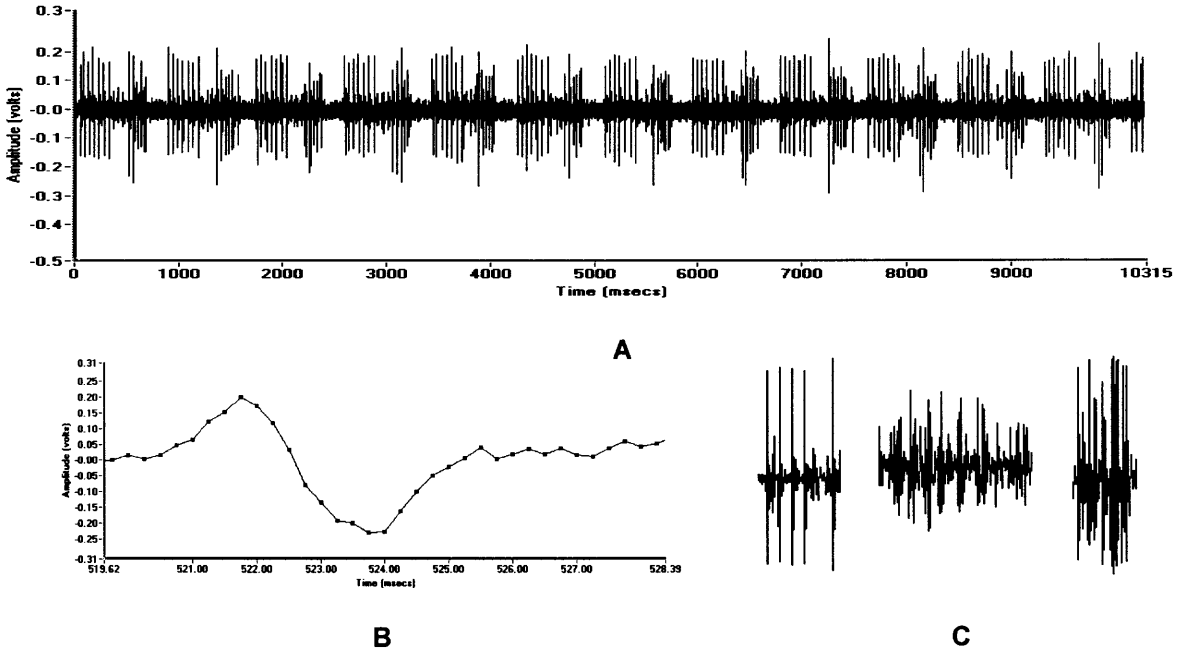


Figure 5.5 The Pyloric Rhythm (A) Extracellular recordings of the nerve *lvn*. (B) Voltage observations (represented by the dots) recorded sequentially over equal time increments. (C) The burst patterns of neurons LP, PY and PD

This technique processes the data "as-is" in the time-domain, without performing any transformations. The entire signal is used as a training sample of the different classes of bursts. The first step is to preprocess the data using the secondary statistics (mean and variance) of the entire signal, to control the effect of extreme values on subsequent calculations and to attenuate the effects of low-amplitude noise. Two amplitude thresholds (one positive and another negative) are estimated to detect spike peaks within an empirically determined time window. A spike is detected only when the action potential rises above positive threshold value and falls below negative threshold value. Each spike is represented by the amplitude and time of occurrence of its positive and negative peak. This procedure can identify only those overlaps that have identifiable peaks. Once the spikes are detected, the high-amplitude neuronal noise is detected based on the variance in amplitude in the detected spikes. If the background activity has significantly higher amplitude than the foreground activity, it is removed. If the rate of neuronal noise activity is higher than a pre-determined threshold, the pre-processing and detection steps are repeated again, over the clean signal. This is done to avoid any undue influence of high-amplitude noise on the detection parameters.

5.2.1 Burst-Edge Detection

Accurate detection of burst edges is central to effective performance of the algorithm. Bursts are detected using differences in ISIs between the spikes. The ISI between the ending of a burst and the beginning of the next burst is greater than the average ISI between the spikes within either burst. This procedure detects single-spike "bursts" as well.

5.2.2 Feature Extraction

Each burst is characterized using four features that are selected intuitively based on general characteristics of neuronal data - the amplitude, frequency, duration and inter-burst interval. In this step, the bursts are reduced to a feature vector in the 4-dimensional feature space.

5.2.3 Unsupervised Clustering

In comparing similarity between two feature vectors, the Manhattan Distance metric is used. A single-pass variable-length bin estimation algorithm partitions the set of bursts into a preliminary set of clusters. Each bin represents a cluster with bursts in the same cluster being more similar to each other than they are to bursts in other clusters. The bins are systematically aggregated to form putative clusters of bursts from individual neurons. This procedure is recursively applied to different combinations of the four features, forming different sets of clusters. The grouping that minimizes the sum-of-squared-error criterion is used as the final set of clusters. This way, the maximal subset of the features that best characterizes the variability among the bursts is automatically selected. Redundant solution vectors are found for each cluster such that they are complete representations of the entire set of bursts in the cluster. The solution vectors are labeled according to the neuron that fired the bursts represented by the vector. Heuristics based on prior domain knowledge, are formalized to generate the conditions for labeling.

5.2.4 Fuzzy Classification

The feature space is now partitioned into regions such that all vectors in one region belong to one class. The bursts are classified based on the minimum distance of their vector representations, to each of the solution vectors. The activity of one neuron is chosen as a time reference. Ties are resolved probabilistically. A phase plot is rendered, to show the relative timing of the burst activity of each neuron, with respect to a reference neuron.

The following chapter presents a detailed description of each of these steps. This section introduces the clustering procedure to automatically learn the structure and natural groupings in the data. This technique processes the data as-is in the time-domain, without performing any transformations. The goal is to classify the patterns according to the neurons that fired them, and derive the precise bursting times of the neurons, to calculate and plot the differences in phase. The problem of burst classification is thus reformulated to one of partitioning the data set into clusters. The principle of the approach is that the same few features that make two burst patterns look different to the human eye will also differentiate them in the time domain.

The problem is defined as learning the structure of multidimensional patterns from a set of unlabeled samples. Viewed geometrically, these samples form clouds of points in a d-dimensional space.

CHAPTER 6

AUTOMATIC BURST DETECTION

This chapter presents the procedures for automatic noise removal, spike detection and burst extraction. Given a group of temporal observations $V = \{v_i \mid i = 1...N\}$ where v_i is univariate, the neuronal bursts are automatically detected and characterized.

6.1 Preprocessing the Data

The data in set $V = \{v_i \mid i = 1...N\}$ is preprocessed to control the effect of extreme values on subsequent calculations and to attenuate the effects of electrical noise. The data is first normalized by way of Mean Centering. Each point is centered around the grand mean as follows: $v_i = v_i - \mu_g$ where the grand mean $\mu_g = \frac{1}{N}\sum_{i=1}^{N} v_i$ (1)

Algorithm 1. (***Pre-processing the Data***)

begin initialize

    calculate grand mean μ_g and deviation σ_g for all data points

    for each data point

        adjust the value using Eq. 1

    end for

    return *noise threshold* σ_g

end

In the case of the pyloric rhythm, the recordings are characterized by a reasonable signal-noise ratio that permits acceptable isolation using amplitude thresholds. To attenuate the effects of low-amplitude noise components on the recorded signal, an amplitude threshold σ_g is estimated based on the variance of the entire recording [Algorithm 1].

The shape of the action potential is such that it has a positive and a negative peak. Therefore, two thresholds ($+\sigma_g$ and $-\sigma_g$) are applied to the signal. All the points falling between $+\sigma_g$ and $-\sigma_g$ are discarded as noise.

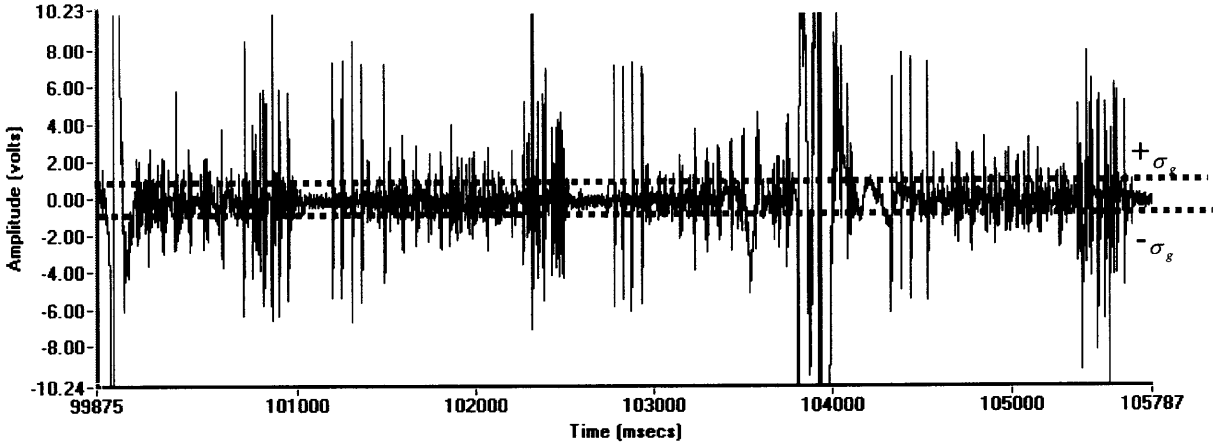


Figure 6.1 Automatic removal of low-amplitude noise based on variance of the entire signal

6.2 Spike Detection

After low-amplitude noise has been removed, the next step is to detect the spikes. A second amplitude threshold - the *detection* threshold - is estimated for spike peak detection. In this approach, each spike is characterized by the amplitude and time of occurrence of its positive and negative peak. These two parameters give the peak-to-peak amplitude of a spike and a measure of the inter-spike interval (ISI). All spikes can then be compared based on these common reference parameters.

Previously proposed amplitude threshold procedures failed due to two main reasons – the "true" peak was obscured by several local maxima and several false peaks were detected. Figure 6.2 from (K.H.Kim and S.J.Kim 2000) illustrates the argument against an amplitude-based threshold for spike peak detection.

The erroneous detection is due to the fact that the threshold is applied only in one direction. The proposed algorithm addresses these problems by setting two detection thresholds (one positive and another negative), and then looking for a global maximum within a predefined time window [Algorithms 2a, 2b].

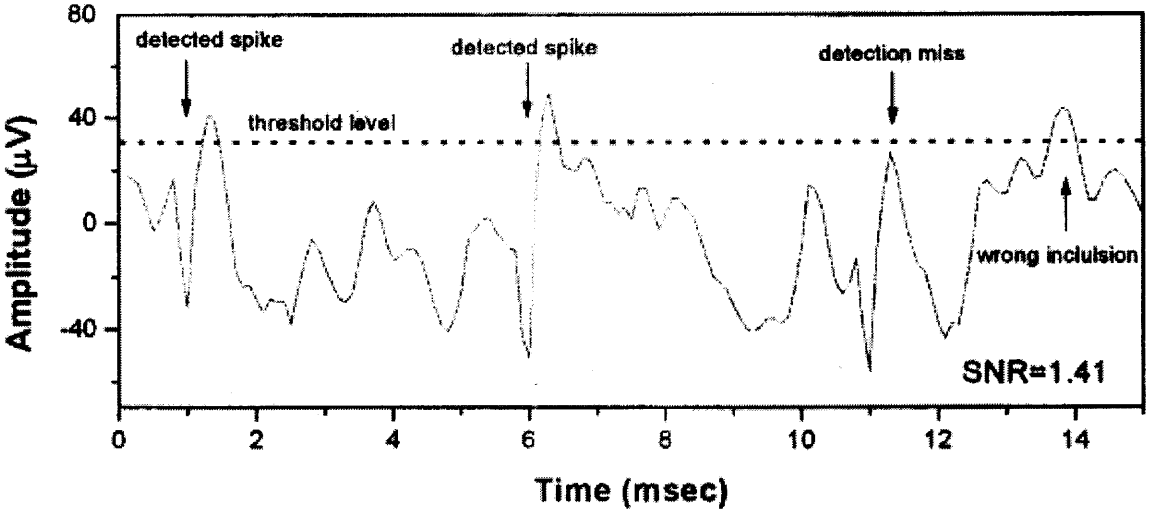


Figure 6.2 The problem of spike detection using amplitude threshold is illustrated. A noise peak is erroneously included and a valid spike is missed (K.H.Kim and S.J.Kim 2000).

The problems illustrated in Figure 6.2 will be resolved by using two thresholds, one for positive peak and another for negative peak detection. This algorithm assumes that the foreground activity is rare enough such that the threshold measurement is dominated by the background activity. Only the crossings where an action potential rises above positive threshold value and falls below negative threshold value are accepted as valid spikes.

6.2.1 Estimating the Spike Detection Threshold

A positive detection threshold μ_p that is a mean of the points above the noise threshold $+\sigma_g$ and a negative threshold μ_n that is a mean of the points below $-\sigma_g$ are estimated.

Reference points just above and below μ_p and μ_n are detected [Algorithm 2a], to begin looking for spike peaks.

Algorithm 2a. (***Find Reference Points***)

begin initialize

calculate mean μ_p and deviation σ_p of all points above $+\sigma_g$
calculate mean μ_n and deviation σ_n of all points below $-\sigma_g$

for each point p above μ_p

$$v_{Pref} = v_p \mid v_p > \mu_p, v_{p-1} \leq \mu_p$$

end for

for each point n below μ_n

$$v_{Nref} = v_n \mid v_n > \mu_n, v_{n-1} \leq \mu_n$$

end for

return $\{v_{Pref}\}$ and $\{v_{Nref}\}$

end

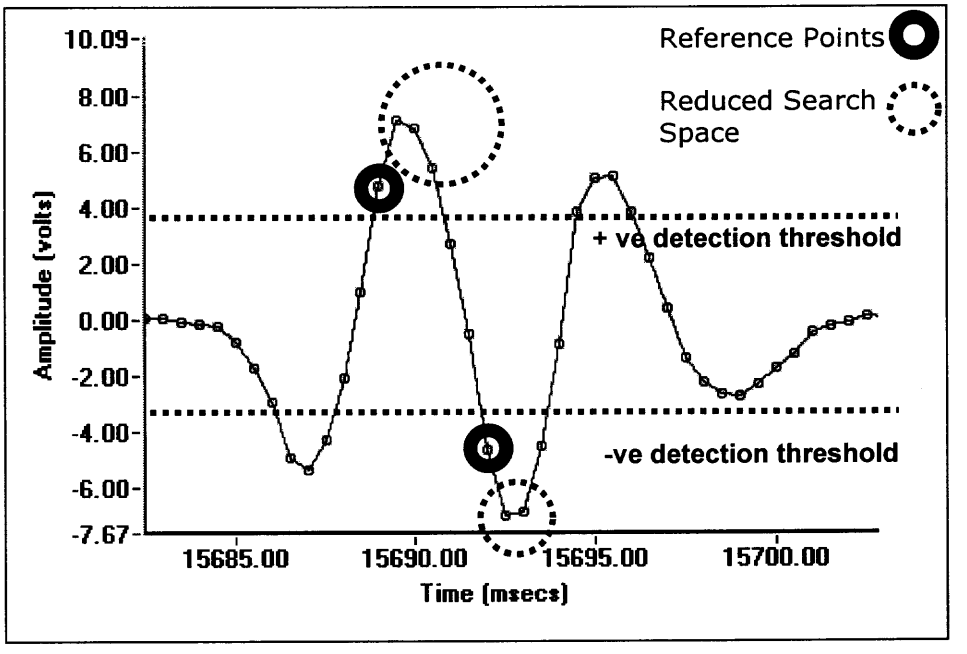


Figure 6.3 Estimating spike detection thresholds and reference points for peak detection.

6.2.2 Peak Detection

Spikes are detected by looking for peak data points in a window (θ_w) of 2 milliseconds from the time of occurrence of the reference points v_{Pref} and v_{Nref}. Like the subtraction methods mentioned in Section 4.1, this algorithm can only identify those overlaps that have identifiable peaks. With this approach it is important to make sure that the positive and negative peaks detected, indeed belong to the same spike. The negative peak is therefore chosen conditionally, such that it lies within a window of $2 * \theta_w$ from the detected positive peak. The input signal, viewed as a continuous function of time $v(t)$ is now converted to a sequence of discrete spikes at times $\{t_1,\ldots,t_n\}$ [Algorithm 2b].

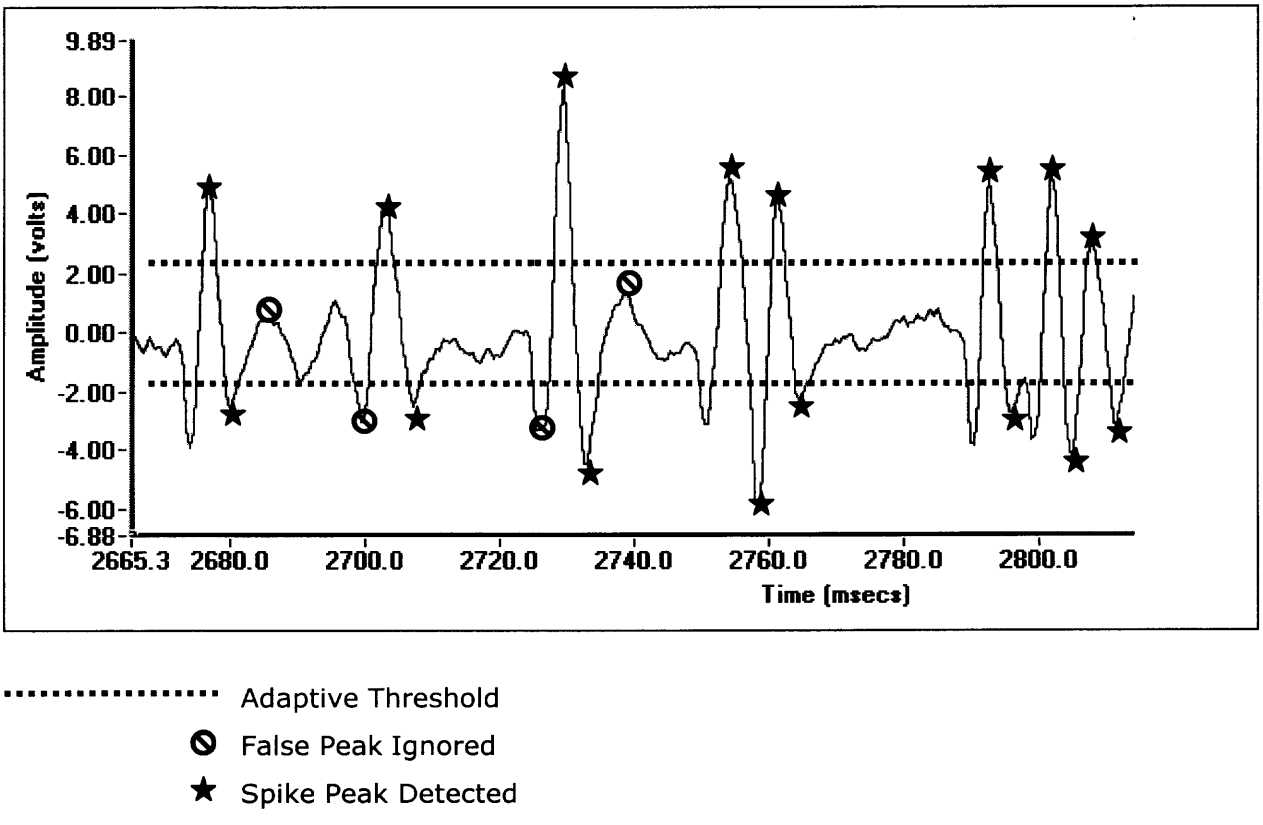


Figure 6.4 Spike peak detection using two thresholds. Only those action potentials that rise above the positive threshold and fall below the negative threshold are accepted as valid spikes.

A function of time describing the spikes is given by:

$$\rho(t)=\sum_{i=1}^{n}\delta\left(t-t_i^{\,p}\right) \tag{2}$$

Spikes are thus reduced to points in time.

Algorithm 2b. ***(Peak Detection)***

begin initialize number of spikes i

for each reference point $Pref$ in $\{v_{Pref}\}$

begin search in window of θ_w points starting from $Pref$

find maximum point $v_{\max}$

end search

temporary +ve peak $\dot{v}_{Ppeak} = v_{\max}$

for each reference point $Nref$ in $\{v_{Nref}\}$

if 1. $t_{Nref} > \dot{t}_{Ppeak}$ *and* $t_{Nref} < t_{Pref+1}$

2. $t_{Nref} - \dot{t}_{Ppeak} < 2 * \theta_w$

begin search in window of θ_w points from $Nref$

find minimum point $v_{\min}$

end search

positive peak $v_{Ppeak} = \dot{v}_{Ppeak}$

negative peak $v_{Npeak} = v_{\min}$

$s_i^{Pv} = v_{Ppeak}$; $s_i^{Nv} = v_{Npeak}$

$s_i^{Pt} = t_{Ppeak}$; $s_i^{Nt} = t_{Npeak}$

incr i

end if

end for

end for

return $\{s_i\}, i$

end

With this technique, spike detection is not sensitive to sudden amplitude shifts in the recording. Significant changes in the form of the spike do not degrade the detection accuracy.

6.3 Cleaning the Spikes

This step cleans the detected spikes, based on uncharacteristic variations in spike amplitude. High-amplitude stochastic interference from background neurons can be detected and removed, if the amplitude of a spike is greater than a certain threshold $\theta_{Sa} * \mu_{Sa}$. The threshold is a function of the mean amplitude μ_{Sa} (calculated as demonstrated in Algorithm 3) and an empirically determined cutoff $\theta_{Sa} = 1.6$.

Algorithm 3. **(Spike Cleaning)**

begin initialize $\mu_{Sa} = 0; n_{Sirreg} = 0; flag = 0$

    for each spike i in $\{s_i\}$
        $s_i^a = s_i^{Pv} - s_i^{Nv}$
        compute mean amplitude μ_{Sa} of previous spikes

        if $s_i^a > \theta_{Sa} * \mu_{Sa}$
            mark for deletion
            incr n_{Sirreg}
            do not include s_i^a in μ_{Sa} calculation
        end if
    end for

    if $n_{Sirreg} > \theta_{Sirreg}$ and $flag = 0$
        delete irregular spikes
        set *flag* to repeat preprocessing
    end if

    return *flag*
end

The mean amplitude μ_{Sa} is calculated adaptively; if an uncharacteristic spike is detected, it is not included in the calculation of the mean. The counter n_{Sirreg} keeps track of the number of uncharacteristic spikes. If the number is greater than a predetermined threshold $\theta_{Sirreg} = 0.6 * (\#\{s_i\})$, it is reasonable to deduce that this irregularity unduly influenced the calculation of amplitude thresholds for noise level and spike detection. Therefore under such circumstances, the preprocessing and detection steps are repeated to ensure correct detection on valid spikes that had lower amplitudes than the detection thresholds. Thus, additional computational costs are incurred only if the percentage of irregular spikes in the entire recording will affect the initial values.

With this algorithm, correlated high-amplitude neuronal interference that obscures pyloric activity can be removed, leaving the pyloric bursts intact. However, there is a caveat: if a valid spike from a pyloric burst overlaps with the noise to form a spike of uncharacteristically high amplitude, it will also be removed.

6.4 Burst Edge Detection

Central to effective classification is the correct detection of burst edges. The onset and offset times of the burst are detected using ISI statistics. After the spikes are detected, an adaptive mean ISI μ_{Sisi} is calculated for the entire data set [Algorithm 4]. The beginning of a burst is marked by the spike that is at a threshold distance of $\theta_{Sisi} * \mu_{Sisi}$ milliseconds from its previous spike. The ending of a burst is marked by the spike that is of the same threshold distance to the next spike. The inter-spike distance threshold $\theta_{Sisi} = 3$ is an empirically determined value.

Algorithm 4. (***Burst Edge Detection***)

begin initialize number of bursts $n = 0$

for each spike i in $\{s_i\}$

$$s_i^{isi} = \frac{s_{i+1}^{Pt} + s_{i+1}^{Nt}}{2} - \frac{s_i^{Pt} + s_i^{Nt}}{2}$$

update mean spike interval μ_{Sisi}

if $s_i^{isi} > \theta_{Sisi} * \mu_{Sisi}$

do not include s_i^{isi} in μ_{Sisi} calculation

end if

end for

for each spike i in $\{s_i\}$

if $\left(\frac{s_i^{Pt} + s_i^{Nt}}{2} - \frac{s_{i-1}^{Pt} + s_{i-1}^{Nt}}{2} \right) > \theta_{Sisi} * \mu_{Sisi}$

Burst Beginning $b_n^{begin} = i$

for each spike j in $\{s_j\}$, starting from $j = i$

if $s_i^{isi} > \theta_{Sisi} * \mu_{Sisi}$

Burst Ending $b_n^{end} = j$

incr n

$i = j$

break from loop

end if

end for

end if

end for

return n

end

Burst onset and offset times are thus determined by the average times of occurrence of the positive and negative peaks of the beginning and ending spikes respectively.

6.5 Feature Extraction

A set of four manually selected parameters is measured for each burst. Thus, each burst can be regarded as a point in a four-dimensional space called the feature space. The four features are selected intuitively based on general characteristics of neuronal data:

- Amplitude: average amplitude of spikes in a burst - F_b^1
- Duration: time between onset and offset of a burst - F_b^2
- Frequency: number of spikes in the burst, in the given duration - F_b^3
- Inter Burst Interval (IBI): the duration between the ending of a burst and the beginning of the next burst - F_b^4

In this case, the feature set can be regarded as a feature vector $\vec{f}$, where $\vec{f}$ is the 4-dimensional column vector $f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix}$.

The feature vector of each burst can then be defined as: $f_b^k = \frac{1}{k} \sum_{i=1}^{k} \left\| F_b^i \right\|$ (3)

The feature extraction process caters to single-spike bursts as well. During the unsupervised clustering phase (Section 7.3), this feature set is evaluated to find the right subset that can be employed for learning. The clustering algorithm is therefore said to be *wrapped* into the attribute selection process.

Algorithm 5. **(Feature Extraction)**

begin initialize $i=0, j=0, m=0$

for each burst n in $\{b_n\}$

$$j = b_n^{end}; \quad i = b_n^{begin}; \quad m = b_{n+1}^{begin}$$

if $j - i = 0$

$$b_n^{dur} = 2 * \theta_W$$

else

$$b_n^{dur} = \frac{\left(s_j^{Pt} + s_j^{Nt}\right)}{2} - \frac{\left(s_i^{Pt} + s_i^{Nt}\right)}{2}$$

end if

$$b_n^{amp} = \frac{1}{(j-i)+1} \sum_{k=i}^{j} \frac{\left(s_k^{Pv} - s_k^{Nv}\right)}{2}$$

$$b_n^{freq} = \frac{j-i}{b_n^{dur}}$$

$$b_n^{ibi} = \frac{\left(s_m^{Pt} + s_m^{Nt}\right)}{2} - \frac{\left(s_j^{Pt} + s_j^{Nt}\right)}{2}$$

end for

end

6.6 Cleaning the Bursts

The guiding principle for this preprocessing step is that the classifier should not have to take unnecessary efforts in repairing inconsistencies that could have been fixed by proper detection and cleaning. Uncharacteristic bursts, characterized by very large duration and amplitude values, can be reasonably assumed as interference from neighboring neurons.

Algorithm 6. **(*Burst Cleaning*)**

begin initialize $\Delta shift = 0$, $m = 0$

calculate μ_{Bamp}, σ_{Bamp}, μ_{Bfreq}, σ_{Bfreq}, μ_{Bdur}, σ_{Bdur}

for each burst n in $\{b_n\}$

if $b_n^{dur} > \mu_{Bdur} + (2 * \sigma_{Bdur})$ and $b_n^{amp} > \mu_{Bamp} + \sigma_{Bamp}$

mark b_n for deletion

$\Delta shift \mathrel{+}= b_n^{dur} + b_n^{ibi}$

end if

if $b_n^{amp} \leq (\mu_{Bamp} + \sigma_{Bamp})$ and $b_n^{amp} \geq (\mu_{Bamp} - \sigma_{Bamp})$ and $b_n^{dur} > \mu_{Bdur} + (2 * \sigma_{Bdur})$

get *freak* spike s_f according to Algorithm 6a

split b_n into B_m and B_{m+1} at s_f

get B_m and B_{m+1} features per Algorithm 5

incr m

mark b_n for deletion

end if

if b_n is not marked for deletion

$B_m = b_n$

if $\Delta shift \neq 0$

$B_m^{begin} = b_n^{begin} - \Delta shift$

$B_m^{end} = b_n^{end} - \Delta shift$

end if

incr m

end if

end for

end

Two valid bursts may have been detected as one due to an interfering spike [Figure 6.5]. This interfering spike (*freak* spike) could be two or more overlapping spikes of a PY burst, or some stochastic interference from another neuron.

The resulting combined burst will then have an uncharacteristically long duration, but a valid amplitude. Instead of marking off such "double" bursts for deletion, the interfering spike is detected and the burst is split into its two valid components.

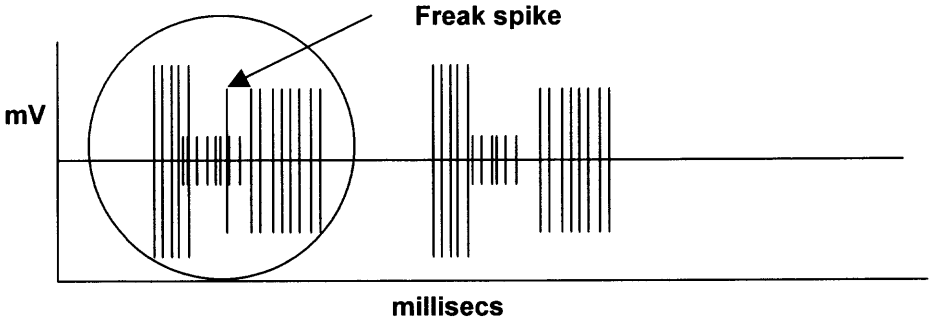


Figure 6.5 "Double" Bursts, indicated by the circle, are detected as one due to an interfering freak spike

---

Algorithm 6a. **(*Freak Spike Detection*)**

begin initialize $b_n \leftarrow$ *input from Algorithm 6*

find maximum s_i^{isi} of in b_n

$\dot{s}_f = s_i$

if $\left(\frac{\dot{s}_f^{Pt} + \dot{s}_f^{Nt}}{2} - \frac{\dot{s}_{f-1}^{Pt} + \dot{s}_{f-1}^{Nt}}{2} \right) > s_i^{isi}$

$s_f = \dot{s}_{f+1}$

end if

return s_f

end

When uncharacteristic bursts are deleted, all the other valid bursts that follow are shifted backward in time, so the IBI is maintained. At this stage valid bursts have been extracted and

characterized. Outliers have been detected and removed. The next chapter presents the unsupervised learning algorithm.

CHAPTER 7

UNSUPERVISED LEARNING BY CLUSTERING

Clustering may be defined colloquially as finding natural groupings in data but the similarity metrics and grouping criterion have to be quantified. The objective of the clustering algorithm is to group the feature vectors (characterizing the bursts) into putative clusters, each of which will represent single-unit activity. Particularly, in the case of the pyloric rhythm, the goal is to find three clusters of bursts, generated by the LP, PY and PD neurons. This chapter will discuss the choice of the similarity measure and cluster evaluation procedures.

Each of the feature vectors defines a point in feature space $|R^D$ and the similarity between vectors is measured by the Manhattan Distance metric. The set of feature vectors (henceforth referred to as data points) constitutes an "unlabeled" data set since the category of each vector is not known at this stage. The data points are first clustered into groups and then labeled based on domain-specific knowledge. The criterion for the assignment of d-dimensional observations $\{x_i^d\}$to k clusters is defined.

7.1 Distance Metric for Similarity

In comparing similarity between two samples, Euclidean distance is a commonly used metric. Similarity calculated using the Euclidean distance is justified if the feature space is isotropic and the data are spread roughly evenly along all directions. In this case, the four original features (Section 6.5) are meant to measure different burst characteristics but they are influenced by some common mechanism and tend to vary together. For instance, the LP bursts typically have a have a higher amplitude and lower frequency than the PD bursts.

Also, frequency is defined as a function of the number of firings in a given burst duration. Such correlation among the features may degrade the performance of a classifier based on Euclidean distance. This algorithm uses the L_1 norm of the Minowski distance, also called the Manhattan similarity measure. This distance is the absolute value sum of the orthogonal distances:

$$d(D_a, D_b) = \|D_a - D_b\|_1 \tag{4}$$

7.2 Standardizing the Feature Scales

Since the different features have different scales, it is important to prevent certain features from dominating distance calculations merely because they have large numerical values. A typical procedure to standardize each feature value is to subtract the mean and divide by standard deviation so that each feature has zero mean and unit variance. But dividing by the standard deviation is not appropriate if the spread in the data is mainly due to normal random variation (Duda, Hart et al. 2000). In this approach feature vectors are standardized as follows: $x_i = \frac{\max - x_i}{\max - \min}$

(5)

where max and min are the maximum and minimum values of the feature.

The standardized values are used in measuring the similarity between the data points. This distance has the important property that it is scale invariant. That is, the units for the various features will have no effect on the resulting distances, and thus no effect on the final classification.

7.3 Wrapper Method for Automatic Feature Selection

An ideal feature set is one that will produce similar feature vectors for all patterns in the same class, and different feature vectors for patterns in different classes.

A feature vector constructed using all the four features, is not always the best representation of variability. Principled methods are applied to determine which features to jettison and which ones to keep, such that the within-class variability is small relative to the between-class variability. To this end, feature vectors of different dimensions are constructed using different combinations of the four features. The clustering procedure is applied to the different sets of feature vectors thus obtained.

Algorithm 7. **(Feature Vector Construction)**

begin initialize

for each feature set f in $\{rfs_f\}$

for each burst n in $\{B_n\}$

$$x_n^f = \bar{x}_n^f = \frac{1}{f^d}\sum_{i=1}^{f^d}\left\| B_n^{f^i} \right\|$$

end for

end for

end

7.4 Unsupervised Learning

The learning problem is defined as follows: the set $D = \{x_1, x_2, x_3 \ldots x_n\}$ of n vectors is the entire set of bursts that need to be partitioned into disjoint bins $D_i, \cdots, D_k$ where k is not explicitly specified. Each bin is to represent a cluster with data points in the same cluster being more similar to each other than they are to points in other clusters.

7.4.1 Initial Clustering

A minimum-variance criterion function that measures the clustering quality of any partition of the data is defined. The partition that extremizes the criterion function is the final grouping. [Algorithm 8a] demonstrates the initial clustering, based on a simple change point analysis procedure. The feature vectors are sorted in descending order and the mean $\overline{\mu}_x$ and variance $\overline{\sigma}_x$ of the entire set of bursts is calculated. Exploratory analysis of the burst patterns was performed using scatter plots and histograms, to understand the distribution of the data and the outliers. The graphical analysis suggested that the data is from a multi-modal distribution. Typically, each of the modes has a rough bell-shaped component. Thus there are three or more similar but separate sub-processes.

Five empty bins $\{\dot{D}_c\}$ are first initialized, one each for LP, PY, PD bursts, outliers with high features values and outliers with low feature values. The first empty bin is initialized with the first burst from the sorted feature vector list $\lfloor x_i^d \downarrow \rfloor$. Subsequent vectors are assigned to a bin such that the compactness of the bin (as defined by the variance of the points already in the bin) is not affected.

If the insertion of a data point into a bin will change its variance beyond the variance of all the points in the data set, the point is assigned to the next bin. If a data point does not belong to any of the five anticipated bins, a new bin is created. Thus, an ordered set of bins of varying lengths is formed, each of which has the points that are closer to each other than to points in other bins.

Algorithm 8a. **(Unsupervised Learning: *Initial Clustering*)**

begin initialize $c \leftarrow$ # initial clusters; *flag* $\leftarrow$ false

for all combinations of features
$\left\lfloor x_i^d \downarrow \right\rfloor$ = sort *desc* $\left\{x_i^d\right\}$
calculate $\overline{\mu}_x$ and $\overline{\sigma}_x$
for each vector i in $\left\lfloor x_i^d \downarrow \right\rfloor$
for each cluster c in $\left\{\dot{D}_c\right\}$
if $\dot{D}_c^{size} = 0$
$\dot{D}_c^{pt} = x_i^d$; *incr* $\dot{D}_c^{size}$
break from loop
else if $\left(\left\|\mu_{\dot{c}j} - x_i^d\right\|_1 + \sigma_{\dot{c}j}\right) \leq \overline{\sigma}_x$
$\dot{D}_c^{pt} = x_i^d$; *incr* $\dot{D}_c^{size}$
calculate $\mu_{\dot{D}c}$ and $\sigma_{\dot{D}c}$
break from loop
else
if $c = \#\left\{\dot{D}_c\right\}$
set *flag* to *true*, create new cluster
end if
end if
end for

if *flag* set to *true*
incr c
create *new* empty cluster $\dot{D}_c$
$\dot{D}_c^{pt} = x_i^d$; *incr* $\dot{D}_c^{size}$
flag = false
end if
end for
end for
end

The bin thus derived, are then aggregated to form putative clusters that represent single-unit activity.

7.4.2 Bin Aggregation

The working hypothesis is that if the bins are separated by a distance lesser than an implicitly derived threshold θ_d, and the variance of burst distribution in the resulting combined bin is not significantly greater than that of each of the original bins, the bins are combined. The hypothesis implies that the two combined bins contain bursts have similar characteristics and are therefore generated by the same neuron.

Algorithm 8b. **(Unsupervised Learning: *Aggregation*)**

begin initialize $c \leftarrow$ # initial clusters

for all combinations of features

$$\theta_d = \frac{1}{c-1}\sum_{m=1}^{c-1} d(\dot{D}_m, \dot{D}_{m+1})$$

$merged = 0$

for i=1 to c-1

if $merged = 1$

$$\theta_d = \frac{1}{c-i}\sum_{m=i}^{c-1} d(\dot{D}_m, \dot{D}_{m+1})$$

calculate $\sigma_{\dot{D}i}$

end if

if $\left(d(\dot{D}_i, \dot{D}_{i+1}) < \theta_d\right)$ and $\left(\mathrm{E}_{i\oplus i+1} \le \frac{\mathrm{E}_{\dot{D}i} + \mathrm{E}_{\dot{D}(i+1)}}{2}\right)$

$$\dot{D}_{i+1} = \left(\dot{D}_i \oplus \dot{D}_{i+1}\right)$$

$$\dot{m}_{i+1} = \{\dot{m}_i, \dot{m}_{i+1}\}$$

$merged = 1$

else

$$D_k = \dot{D}_i$$

$$m_k = \{\dot{m}_i\}$$

$merged = 0$

end if

end for

calculate $J_{e_{rfs}}$ as per Eq (8)

end for

end

The average distance between one bin and the next in the hierarchy of bins derived by [Algorithm 8a], defines the distance threshold θ_d. The merge operator is indicated by the symbol $D_a \oplus D_b$ and simply means combining the points of bins a and b together in one common bin. The mean-variance criterion is satisfied by minimizing the error function

$$E_{D_i} = \sum_{x \in D_i^{PI}} \|x - m_i\|_1 \tag{6}$$

where the mean vector m_i of the bin is $m_i = \frac{1}{D_i^{size}} \sum_{x \in D_i^{PI}} x$ (7)

At this stage, putative clusters are derived for each of the feature sets.

7.4.3 Finding Redundant Solution Vectors

The unsupervised learning procedure has tessellated the feature space into decision regions. All feature vectors in a decision region belong to the same category. The decision regions are not exactly spherical can be connected at one or more points. Therefore, redundant solution vectors are derived to represent bursts in different sections of a cluster. The solution vectors of a cluster are simply the mean vectors m_i (given by Eq. 7) of the bins that were combined to form the cluster.

7.4.4 Determining the Final Clusters

The Sum-of-Squared-Error Criterion finds the partition that renders clusters of minimum variance. This criterion in especially suited in this case (Duda, Hart et al. 2000), since the groupings of the data (LP, PY and PD) have very similar number of points. The sum-of-squared errors J_e is calculated in [Algorithm 8b] and is defined by:

$$J_e = \sum_{i=1}^{c} \sum_{x \in D_i} \| x - m_i \|_1 \tag{8}$$

where m_i is the mean solution vector.

The partition that gives the $min(J_e)$ is the final set of clusters. Thus, the features that provide little improvement or possibly even degrade the clustering performance are not included. The clusters derived thus, reflect the dynamics of the system that causes bursts of one neuron to bear a stronger similarity with one another than to other bursts.

CHAPTER 8

FUZZY CLASSIFICATION OF BURST PATTERNS

The classifier can now be designed based on the structure of the data as revealed by the clusters. To classify all the points, solution vectors $\{m_i\}$ for each cluster were derived in [Algorithm 8b]. The solution vectors represent critical boundary instances for each class. A Minimum-Distance classifier with a fallback heuristic is defined. Feature vectors that fall in the overlap of boundaries are assigned a category based on minimum-error classification. By definition, the learning algorithm groups together those bursts that are highly similar to each other. The following sections describe how the mean vectors are labeled, and the bursts are classified.

8.1 Labeling the Solution Vectors Using Fuzzy Logic

Since the number of clusters is not specified during clustering, it is reasonable to assume that the clusters with the largest number of points are those that represent single-unit activity. Other smaller clusters, if any, could be outliers that were left undetected during the preprocessing stages, or valid bursts with extreme values.

The solution vectors are labeled using fuzzy logic. Heuristics based on prior domain knowledge are formalized to generate conditions that label the mean vectors. The feature values of the mean vectors are compared as follows:

If the frequency of the bursts (corresponding to the set of solution vectors for one cluster) is greater than that of the others and the amplitude is lower, these vectors are labeled as PY. If the amplitude of the bursts is between that of the bursts labeled PY and the other bursts, and the IBI is the largest, the corresponding vectors are labeled PD. The vectors that

have the maximum amplitude are labeled LP. These formalized heuristics will break down when the bursting patterns deviate from normal and are irregular. Under such circumstances, user intervention is required to label the vectors correctly.

Algorithm 9. ***(Labeling the Solution Vectors)***

begin initialize

$$\left[D_1, \cdots, D_k \downarrow\right] = \text{sort } \{D_i\} \text{ by } max\left(D_i^{size}\right)$$

from $\left[D_1, \cdots, D_k \downarrow\right]$ select top 3 D_i

$$\{m_i\}^{\omega} = PY \mid i = min\left(M_1, M_2, M_3\right)^{amo} \text{ and } max\left(M_1, M_2, M_3\right)^{freq}$$

$$\{m_i\}^{\omega} = LP \mid i = max\left(M_1, M_2, M_3\right)^{amo}$$

$$\{m_i\}^{\omega} = PD \mid i = max\left(M_1, M_2, M_3\right)^{amo} > M_i \text{ and } M_i < min\left(M_1, M_2, M_3\right)^{amo}$$

end

8.2 Minimum-Distance Classification

The membership of a burst to a class is determined based on the minimum distance of the burst to the solution vectors. If m_1, m_2, ..., m_c are the mean vectors (templates) for the k classes, D_1, D_2, ..., D_k, the feature vector x of a burst is classified by measuring the distance from x to each of the means and assigning x to the class for which the distance is minimum. The decision boundary is not perfectly linear. Therefore, ambiguities will certainly arise. Ties are resolved by probabilistically determining the classification of the ambiguous bursts.

8.3 Minimum-Error Classification

Template matching can easily be expressed mathematically. Let x be the feature vector for the unknown input, and let m1, m2, ..., mc be templates (i.e., perfect, noise-free feature vectors) for the c classes. Then the error in matching x against mk is given by

|| x - mk || . Here || u || is called the norm of the vector u. A minimum-error classifier computes || x - mk || for k = 1 to c and chooses the class for which this error is minimum.

8.4 Phase Analysis

The phase information gives the relative fraction of burst in each cycle. This is calculated using the following formula:

PD End = $\frac{1}{PDs}\sum PD$ (End of PD – Start of PD / Period) / Total number of PDs

LP Start = (Start of next LP – Start of Current PD / Period) / Total number of LPs

LP End = (End of next LP – Start of Current PD / Period) / Total number of LPs

Starting and ending values above are plotted as LP and PD bars on a graph. The onset of PD is marked as the beginning of a cycle and the time courses are applied to the rhythms, normalizing to the period. Phase diagrams show the relative fraction of each cycle with respect to one reference point. This information is useful in comparing rhythms of particular spike train with other data, correspondence between in-vitro and in-vivo recordings, etc.

This idea of temporal coding is supported by experimental evidence from sensory systems such as locust olfaction, electro-sensory system of electric fish, monkey vision and audition etc. The evidence points towards the role of spike timing particularly across an ensemble of cells (instead of single cells), in encoding various aspects of the stimulus.

Sometimes it is desirable to scale the data so that the resulting standard deviation is unity. This is easily done: just divide x by the standard deviation s. Similarly, in measuring the distance from x to m, it often makes sense to measure it relative to the standard deviation. The so-called standardized distance from x to m is given by the linear boundaries produced by a minimum-Euclidean-distance classifier may not be flexible enough. This will warp the feature space and prevent a linear discriminant function from performing well. Note that r is

CHAPTER 9

SUMMARY AND DISCUSSION

The proposed algorithm has four distinguishing features. The separation of multi-unit bursting activity is achieved using just single channel recordings. While *a priori* knowledge of the form of burst variability is assumed, no assumptions about the rhythmicity of the pattern or even the number of distinct patterns are made. Therefore, idiosyncrasies of the rhythm such as single-spike bursts and missing bursts are handled well. The entire dataset is used as training sample and accounts for all variability of spikes in the recording. Unlike earlier techniques (discussed in Chapter 4) that characterized variability based on the observed changes in the waveform of individual action potentials, the algorithm presented in this study performs classification based on variability in burst features.

9.1 Evaluation of the Approach

This section will critically evaluate the various techniques employed by the algorithm.

9.1.1 Empirically Determined Parameters

Several threshold values are determined empirically in this approach - θ_w, θ_{Sa}, θ_{Sisi} and θ_{Sirreg}. The values were chosen intuitively and only after carefully examining the performance of the algorithm over several recordings. The window threshold θ_w of 2 milliseconds is determined based on an *a priori* knowledge of the average time for an action potential to peak. The spike-cleaning threshold θ_{Sirreg} determines whether or not to repeat the preprocessing steps after removing the irregular spikes. The actual thresholds are a function of these θ values and adaptively calculated mean values (of amplitude, ISI etc.).

Such predefined θ values may restrict the applicability of the algorithm to different data sets. However, the accuracy of detection and classification performance appears insensitive to small changes in the values. Also, with a good user interface that provides appropriate quantitative and visual feedback during the classification process, the user can adjust these values to improve performance.

9.1.2 Amplitude Thresholds for Noise Removal and Spike Detection

Spectral subtraction and other low-pass filtering techniques that have been proposed to automatically remove noise, assume that noise is stationary. The noise spectrum is automatically estimated and subtracted from the signal. The drawback is that when one estimate is applied to the whole signal, it tends to distort the waveform shape of valid spikes as well.

This study estimates amplitude thresholds based on secondary statistics of the entire signal, thereby also running the risk of assuming signal stationarity. The technique is sensitive to the signal-to-noise ratio and also the frequency of burst and noise activity in the recording. For example, if there are sequences where there is no bursting activity and only electrical noise, the estimated threshold is unduly lowered and noisy low-amplitude spikes will not be removed.

A computationally efficient moving-average technique can be applied instead, to accommodate transience. The statistics could be calculated *in-context*, within pre-defined time windows. For instance, group means could be calculated as $\bar{x}_g = \frac{1}{w} \sum_{i=w*g}^{w(1+g)} v_i$ where w is the predefined time window for each group g. "Moving" statistics thus obtained, accommodate non-stationarity of the signal by using only the most current values. This can

be viewed as being similar to a primitive resampling technique that yields a reasonable estimate of general statistics such as mean and variance, when no underlying distribution is assumed.

In this study, the bootstrap resampling method was tested to adaptively estimate the mean and variance. The noise threshold was calculated over non-overlapping time windows, during which the secondary statistics were derived. (For recordings of the pyloric rhythm, this yielded no significant improvement in the estimates.) With such an approach, the choice of window length becomes very critical; a poorly selected window threshold can yield inaccurate estimates of variance.

Erroneous inclusion of noise peak or missing a valid neural spike is avoided by using two thresholds (a positive and negative threshold) instead of one.

9.1.3 Outlier Detection and Removal

With the proposed spike detection algorithm, significant changes in the waveform of individual spikes do not degrade the accuracy of peak detection. Spikes of uncharacteristically high amplitude are detected as outliers and removed. This step is especially useful when the interference from neighboring neurons is occluding foreground activity. Only the high-amplitude interference is removed, and the pyloric burst is left intact. However, if the interference does not have higher amplitude, then the threshold method will not be effective. The result is a burst with uncharacteristically long duration or high frequency. This is then detected and removed in the burst cleaning stage. The algorithm thus improves on earlier approaches by explicitly handling outliers in the early stages.

9.1.4 Resolving Spike Overlap

The proposed algorithm does not resolve overlaps. The main argument in favor of elaborate overlap resolution is that accurate spike firing times cannot be determined in the presence of overlaps. With the burst extraction approach to multi-unit separation, it may not be necessary or even desirable to resolve spike superposition based on the template subtraction techniques that have been proposed. The time series is governed by well-understood deterministic processes. If the general firing pattern and the number distinct patterns are known beforehand, several heuristics can be used to resolve overlaps sensibly and efficiently. For instance, the peak-to-peak amplitude of one spike in a burst could be uncharacteristically high because of overlap with spikes from another neuron. The average amplitude of the spikes in the burst can be estimated without taking into account this overlapped spike, and the burst frequency can be adjusted to reflect overlap. Prior knowledge of the rhythm can thus be used to construct or refute a reasonable model for overlaps.

9.1.5 Burst Edge Detection

The performance of the automatic burst edge detection technique can degrade in noisy conditions or when there is high rate of correlation. When several neurons burst at roughly the same time with similar amplitudes, the burst detection approach will fail since it uses only time-domain parameters. Biophysical parameters (like the refractory period and ISI distribution used by (Fee, Mitra et al. 1996)) could be used to supplement the algorithm.

9.1.6 Selection of the Feature Set

The performance of any clustering algorithm relies on the original differences among the groups in the dataset. One major challenge is a large degree of variability in the bursts belonging to the same class, relative to the differences between patterns in different classes. The efficiency of the feature extraction method in representing those differences is crucial for the success of clustering.

The algorithm attempts to cope with this problem by first manually defining variability (in the burst amplitude, duration, frequency and IBI) based on *a priori* knowledge and then automatically looking for characterizing features in this reduced feature set. In general, the manually selected features will effectively lower the dimension of feature space and signify differences among burst patterns in recordings from other systems as well.

Unlike automatic feature extraction algorithms, this procedure models the method used by human observers in detecting and classifying bursts, without introducing any unnecessary complexity into a primitive feature space. The feature values are standardized and invariant to differences in scale.

9.1.7 Bin-Estimation Procedure

Initial clustering of the bursts is achieved by applying a single-pass bin-estimation procedure to a sorted list of feature vectors. Each bin corresponds to closely related groups of bursts. Neither the number of bins nor the length is defined *a priori.* Thus, the algorithm determines the underlying characteristics of the data without underestimating or over-representing the variations in the data. Distinct subclasses in the data (possibly due to undetected outliers) are handled well by the bin-estimation approach.

9.1.8 Redundant Solution Vectors

Typically, only one mean vector m_i is used to represent the points in a cluster D_i. The solution vectors minimize the sum of the squared lengths of the error vectors $x-m_i$ in D_i. When the clusters are not perfectly spherical, several means can be estimated for a cluster to cater to the minor variations within the cluster structure. The performance of the classifier depends upon how closely the feature vectors to be classified, resemble the solution vectors. Since the entire dataset is used as a training sample, the variability in the entire signal can be captured.

9.1.9 Efficiency and Time Complexity

The initial clustering procedure is a single-pass algorithm. The classifier has the flexibility to suggest categories based on the degree of membership of each feature vector to the different clusters. This approach is therefore flexible for real time processing when the algorithm is presented with novel patterns. The decision boundaries produced by the L_1 distance classifier are linear. If the feature space is warped because of a noisy signal, the linear discriminant function may not perform well.

While the algorithm is suited for bursting neurons that occasionally switch to spiking mode, it is not very effective for spiking neurons. Some of the features (such as frequency and duration) and detecting "edges" do not make sense for spiking neurons. It would be useful to enhance the algorithm to cater to both bursting and spiking neurons.

The algorithm is very practical because the procedures for variability representation, detection and classification are computationally feasible.

CHAPTER 10

CONCLUSION

This study has proposed and implemented an algorithm that classifies multi-unit neuronal bursts into single-unit activity based on *a priori* assumptions about the form of variability in burst patterns. A single-pass clustering procedure is implemented that efficiently separates real multi-unit recordings in the presence of noise. A unified procedure to extract and classify neuronal bursts has been presented. There are four distinguishing features of the algorithm:

Multi-unit recordings are classified into single-unit activity in the presence of realistic distributions of neuronal noise. The decomposition of spikes that result from overlap is not addressed here. This omission is justified by the following:

In this paper, a "smart" feature extraction method has been discussed, that efficiently determines the optimum subset of features for signifying the differences among burst types. A technique to automatically cluster the data using a bin-estimation algorithm has been presented. Redundant solution vectors that completely represent the clusters have been derived. A template-matching classifier has been designed based on the structure of the data as revealed by the natural clusters in the data. A fuzzy classification approach has been taken to determine the membership of ambiguous instances. The instance is assigned the class that is closest to it, based on the L1 norm. The algorithm has been applied to real multi-neuronal recordings from the crustacean gastric system and evaluated against manual extraction and classification. Despite the employment of potentially restrictive heuristics and empirically determined parameters,

the algorithm can be applied to several types of neuronal recordings in a computationally efficient manner.

This study is motivated by the need to formalize the temporal dynamics of a group of neurons to further an understanding of how the neuronal population generates patterns of rhythmic activity and how the patterns are modified and regulated. By observing the activity of the neurons in concert, it becomes possible to quantify the collective computation of the sub-system. A comprehensive framework for automatic burst extraction and classification has been proposed. An efficient single-pass clustering procedure classifies bursts using only single-channel recordings.

Recordings of the pyloric rhythm are typically characterized by high signal-to-noise ratio (SNR), which allows reliable detection of spikes in the presence of electrical noise. However, when in circumstances where the SNR is low, amplitude thresholds methods may fail.

In this approach we explicitly remove outliers such as neuronal noise and random patterns before clustering or classification. This minimizes non-gaussian variance in the single-unit firing. Since the variability among burst patterns is characterize by a reduced feature set, the algorithm is able to cater to irregularities in in vivo recordings as well. Several threshold values are determined empirically in this approach. The values were chosen intuitively and only after carefully examining the performance of the algorithm over several recordings. Such predefined values may restrict the applicability of the algorithm to different data sets. However, the accuracy of detection and classification performance appears insensitive to small changes in the values.

With this approach, sophisticated methods for resolution of spike overlap may be unnecessary. If the general firing pattern and the number distinct patterns are known beforehand, several heuristics can be used to resolve overlaps sensibly and efficiently.

REFERENCES

Azouz, R. and C. M. Gray (1999). "Cellular mechanisms contributing to response variability of cortical neurons in vivo." Journal of Neuroscience **19**(6): 2209-2223.

Bankman, I. N. and S. J. Janselewitz (1995). "Neural waveform detector for prosthesis control." Proceedings of 17th Annual Conf. IEEE EMBS: 963–964.

Barbieri, R., M. C. Quirk, et al. (2001). "Construction and analysis of non-Poisson stimulus-response models of neural spiking activity." Journal of Neuroscience Methods **105**(1): 25-37.

Bergman, H. and M. R. DeLong (1992). " A personal computer based spike detector and sorter: implementation and evaluation." Journal of Neuroscience Methods **41**: 187–197.

Bohté, S. M., H. L. Poutré, et al. (2002). "Unsupervised Clustering with Spiking Neurons by Sparse Temporal Coding and Multi-Layer RBF Networks." IEEE Transactions on Neural Networks **13**(2): 1-10.

Borst, A. and F. E. Theunissen (1999). "Information theory and neural coding." Nature **2**(11): 947 - 957.

Cajal, S. R. (1909). "Histologie du système nerveux de l'homme et des vertébré." A. Maloine, Paris.

Chandra, R. and L. M. Optican (1997). "Detection, classification, and superposition resolution of action potentials in multiunit single-channel recordings by an on-line real-time neural network." IEEE Trans. Biomed. Eng. **44**(5): 403-412.

Cooper, D. C. (2002). "The significance of action potential bursting in the brain reward circuit." Neurochemistry International **41**(5): 333-340.

Cowan, W. M., T. C. Südhof, et al. (2000). Synapses. Baltimore, Maryland, Johns Hopkins Univ. Press.

Duda, R. O., P. E. Hart, et al. (2000). Pattern Classification, Wiley, New York.

Fee, M. S., P. P. Mitra, et al. (1996). "Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability." Journal of Neuroscience Methods **69**(2): 175-188.

Gerstner, W. and W. M. Kistler (2002). Spiking Neuron Models. Single Neurons, Populations, Plasticity. Cambridge, UK, Cambridge University Press.

Glaser, E. M. and W. B. Marks (1968). "Online separation of interleaved neuronal pulse sequences." Data Acquisition Process Biol. Med. **5**: 137-156.

Goense, J. B. M., R. Ratnam, et al. (2003). "Burst firing improves the detection of weak signals in spike trains." Neurocomputing **In Press, Corrected Proof, Available online 22 January 2003**.

Goren, Y. (2001). Clustering with Spiking Neurons, Technion - Israel Institute of Technology, Computer Science Department.

Gozani, S. N. and J. P. Miller (1994). "Optimal discrimination and classification of neuronal action potential waveforms from multi-unit, multi-channel recordings using software-based linear filters." IEEE Transactions on Biomedical Engineering **41**: 358-372.

Gray, C. M. (1999). "The temporal correlation hypothesis of visual feature integration: Still alive and well." Neuron **24**: 31.

Holt, G. R., W. R. Softky, et al. (1996). "Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons." Journal of Neurophysiology **75**(5): 1806-1814.

Hulata, E., R. Segev, et al. (2002). "A method for spike sorting and detection based on wavelet packets and Shannon's mutual information." Journal of Neuroscience Methods **117**(1): 1-12.

Jansen, R. F. and A. T. Maat (1992). "Automatic wave form classification of extracellular multineuron recordings." Journal of Neuroscience Methods **41**(2): 123-132.

Johnson, D. H., C. M. Gruner, et al. (2001). "Information-theoretic analysis of neural coding." Journal of Computational Neuroscience **10**: 47-69.

K.H.Kim and S.J.Kim (2000). "Neural spike sorting under nearly 0 dB signal-to-noise ratio using nonlinear energy operator and artificial neural network classifier." IEEE Transactions on Biomedical Engineering **47**(10): 1406 -1411.

Letelier, J. C. and P. P. Weber (2000). "Spike sorting based on discrete wavelet transform coefficients." Journal of Neuroscience Methods **101**(2): 93-106.

Lewicki, M. S. (1998). "A review of methods for spike sorting: the detection and classification of neural action potentials." Network: Computation in Neural Systems **9**(4): R53-R78.

Li, D., D. S. K. Magnuson, et al. (2000). "Non-stationary analysis of extracellular neural activity." Neurocomputing **32-33**: 1083-1093.

Lisman, J. E. (1997). "Bursts as a unit of neural information: making unreliable synapses reliable." Trends in Neuroscience **20**(1): 38-43.

Marder, E. and D. Bucher (2001). "Central pattern generators and the control of rhythmic movements." Current Biology **11**(23): R986-R996.

Nusbaum, M. P. and M. P. Beenhakker (2002). "A small-systems aapproach to motor pattern generation." Nature **417**: 343-350.

Rieke, F., D. Warland, et al. (1999). Spikes: Exploring the Neural Code, MIT Press.

Roberts, W. M. and D. K. Hartline (1975). "Separation of multiunit nerve impulse trains by a multichannel linear filter algorithm." Brain Research **94**: 141-149.

Schmidt, E. M. (1984). "Computer separation of multi-unit neuroelectric data: a review." Journal of Neuroscience Methods **Vol. 12**: 95-111.

Singer, W. (1999). "Time Coding as Space?" Current Opinion in Neurobiology **9**: 189-194.

Snider, R. K. and A. B. Bonds (1998). "Classification of non-stationary neural signals." Journal of Neuroscience Methods **84**(1-2): 155-166.

Strang, G. (1993). "Wavelet Transforms versus Fourier Transforms." Bulletin of the American Mathematical Society **28**(2): 288-305.

Tam, D. C. (2002). "An alternate burst analysis for detecting intra-burst rings based on inter-burst periods." Neurocomputing **44–46**: 1155 –1159.

Wheeler, B. C. and W. J. Heetderks (1982). "A comparison of techniques for classification of multiple neural signals." IEEE Transactions on Biomedical Engineering **29**(12): 752-759.

Yang, X. and S. A. Shamma (1988). "A totally automated system for the detection and classification of neural spikes." IEEE Transactions on Biomedical Engineering **35**(10): 806-816.

Zouridakis, G. and D. C. Tam (1997). "Multi-unit spike discrimination using wavelet transforms." Computers in Biology and Medicine **27**(1): 9-18.

Zouridakis, G. and D. C. Tam (2000). "Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering." Computer Methods and Programs in Biomedicine **61**(2): 91-98.