ENGINEERING IN BRAIN RESEARCH

Processing Electroencephalograms and Chaos in Neural Networks

A thesis presented for the degree of Doctor of Philosophy in Electrical and Electronic Engineering at the University of Canterbury, Christchurch, New Zealand.


Alison A. Dingle
B. E. (Hons 1)
ABSTRACT

The structure and function of the brain, as it is presently understood, is outlined. The importance of technology in acquiring this knowledge is illustrated by tracing the history of brain research, and the contributions that engineering is currently making to brain research are discussed.

The manifestation of epilepsy in recordings of the electrical activity of the brain — electroencephalograms (EEGs) — is outlined. A new PC-based system for the automated detection of this epileptiform activity is presented. The system consists of three stages: data collection, feature extraction and event detection. The feature extraction stage detects candidate epileptiform transients on individual channels, while an expert system is used to detect focal and non-focal epileptiform events. Considerable use of spatial and temporal contextual information present in the EEG aids both in the detection of epileptiform events and in the rejection of artifacts and background activity as events. Classification of events as definite or probable overcomes, to some extent, the problem of maintaining satisfactory detection rates while eliminating false detections. Test results are presented which indicate that this system should be capable of performing reliably in routine clinical EEG screening.

Neural networks are introduced and their application to real-world problems examined. In particular, the application of back-propagation neural networks to the detection of epileptiform transients is discussed. A number of modifications to the back-propagation learning algorithm are proposed which should enable the desired network performance to be achieved.

The phenomenon of deterministic chaos is reviewed and evidence for chaos in the brain presented. A new model of a neuron, termed the versatile neural unit and based on a chaotic system is proposed. This neural unit is relatively simple, yet produces a wide range of activity reminiscent of that observed in neurons. The versatile neural unit provides a means for introducing chaos into neural networks to enhance their performance and, at the same time, provide insights into the roles of chaos in the brain. Networks of versatile neural units are shown to produce activity similar to that observed in EEGs and the introduction of chaos into the self-organizing map is shown to improve its ability to cluster input patterns and model their probability density function.
ACKNOWLEDGEMENTS

I am indebted to many people for the assistance, advice, support and encouragement I have received during the course of my research and thesis preparation.

In particular, I would like to thank my supervisors, the late Professor Richard Bates, Dr Richard Jones and Dr John Andreae for their guidance and encouragement. I also wish to thank Grant Carroll and Dr Richard Fright from Christchurch Hospital for the many valuable discussions I had with them.

I gratefully acknowledge receipt of New Zealand University Grants Committee Postgraduate Scholarship.

Finally, I wish to thank my family, friends and flatmates for their support, understanding and encouragement.
CONTENTS

PREFACE ix

I INTRODUCTION 1

CHAPTER 1 THE BRAIN 3
  1.1 Introduction 3
  1.2 Cells of the brain 3
    1.2.1 Neurons 4
    1.2.2 Glial Cells 5
  1.3 Organization of the Brain 5
    1.3.1 The Cerebral Cortex 5
    1.3.2 Subcortical Structures 7
  1.4 Protection of the Brain 7

CHAPTER 2 HISTORY OF BRAIN RESEARCH 11
  2.1 Introduction 11
  2.2 From Antiquity to the Middle Ages 11
    2.2.1 Source of Mental Processes 11
    2.2.2 Anatomy of the Nervous System 12
    2.2.3 Operation of the Nervous System 14
  2.3 The Middle Ages to the Nineteenth Century 15
    2.3.1 Localization of Brain Function 15
    2.3.2 Nerve Function 17
    2.3.3 The Neuron 20
  2.4 The Twentieth Century and Beyond 22
    2.4.1 The Action Potential 23
    2.4.2 Neural Communication 23
    2.4.3 Neuroanatomy 25
    2.4.4 Disorders of the Brain 25
    2.4.5 Higher Functions of the Brain 26

CHAPTER 3 ENGINEERS IN BRAIN RESEARCH 29
  3.1 Introduction 29
  3.2 Recording Techniques 29
  3.3 Stimulation Techniques 30
  3.4 Signal Processing Techniques 31
  3.5 Imaging the Brain 32
  3.6 Quantification of Neurologic Function 33
  3.7 Modelling Brain Function 33
7.4.3 Learning Algorithms
7.5 The Hopfield Network
7.6 The Self-Organizing Map
7.7 Back-Propagation Networks
7.8 Hardware Implementations
7.9 Comparison with the Brain
7.10 Summary

CHAPTER 8 APPLICATIONS OF NEURAL NETWORKS 101
8.1 Introduction 101
8.2 General Considerations 101
8.3 Application Examples 104
  8.3.1 Memory 104
  8.3.2 Optimization Problems 104
  8.3.3 Categorization 105
  8.3.4 Control Problems 105
  8.3.5 Data Processing 107
  8.3.6 Predictive Models 107
  8.3.7 Pattern Recognition 108
8.4 Modifying Back-propagation 108
  8.4.1 Learning Steepness Parameters 108
  8.4.2 Tuning Network Performance 110
8.5 Detection of Epileptiform Transients 111
  8.5.1 Neural Network Approaches 111
  8.5.2 Preliminary Investigation 112
8.6 Limitations of Neural Networks 113

IV CHAOS IN NEURAL NETWORKS 115

CHAPTER 9 DETERMINISTIC CHAOS 117
9.1 Introduction 117
9.2 Theoretical Background 117
  9.2.1 Dynamical Systems 118
  9.2.2 Phase Space 120
  9.2.3 Attractors 121
9.3 The Discovery of Chaos 122
9.4 Chaotic Systems 123
  9.4.1 Lorenz Model 123
  9.4.2 Rössler System 127
  9.4.3 Hénon Map 127
  9.4.4 Logistic Map 127
9.5 Strange Attractors 131
  9.5.1 Attractor Dimension 131
  9.5.2 Entropy 133
  9.5.3 Lyapunov Exponents 134
9.6 Distinguishing Chaotic from Random Signals 135
  9.6.1 Reconstruction of Attractors 136
  9.6.2 Calculation of Lyapunov Exponents 137
  9.6.3 Calculation of Attractor Dimension 138
  9.6.4 Calculation of Entropy 139
  9.6.5 Accuracy of Algorithms 140
"If the human brain were so simple that we could understand it, we would be so simple that we couldn’t."

Emerson Pugh

From time immemorial humans have been interested in understanding the brain. However, despite the long history of brain research, the operation of the brain is still relatively poorly understood.

Over the years the primary role of physical scientists in brain research has been to provide the necessary tools and techniques (e.g., optical and electron microscopes, galvonometer). With the advent of powerful computers and the development of new mathematical techniques, engineers have become more directly involved in brain research — recording, analysing, and modelling brain activity and function.

The purpose of this thesis is to examine, and add to, the contributions of engineering to brain research. Three areas in particular are considered — signal processing, neural networks and deterministic chaos.

Signal processing techniques can aid in the analysis and quantification of brain activity. Frequency analysis is used to analyse changes in the frequency content of brain activity over time and to examine the distribution of activity over the scalp. On the other hand, time series analysis is employed to detect certain types of brain activity, for example, waveforms characteristic of epilepsy. Automated analysis and quantification of brain activity would not only be a useful clinical tool aiding in the diagnosis of various brain disorders, but would also be a valuable research tool enabling studies of the effects of medication on brain activity.

Neural networks are adaptive parallel processing systems inspired by the anatomy and physiology of the brain. They provide an alternative means for processing information and, because of their ability to learn from examples, are applicable to a wide variety of tasks and may prove useful in the analysis of brain activity. Despite the emphasis of neural networks on technological applications rather than modelling the brain, these systems still provide insights into the operation of the brain and have led to cognitive scientists developing new theories of brain function.

Deterministic chaos, or simply chaos, is the phenomenon whereby unpredictable and seemingly random behaviour is produced by a completely deterministic, and often very simple, system. There is a growing body of evidence that chaos is present in the brain but, although the existence of chaos in the brain is becoming more widely accepted, many questions regarding its purpose remain unanswered. Better understanding of the roles of chaos in the brain are needed and may be provided by mathematical models.

This thesis is divided into five parts, the first of which provides an introduction to brain research and the roles of engineering in this field. Three aspects of engineering in brain
research are then considered in detail. Part 2 examines a number of signal processing techniques and their application to analysis of the electrical activity of the brain as recorded by electroencephalograms (EEGs). In particular, a system is presented that uses time series analysis techniques to detect waveforms in EEGs that are characteristic of epilepsy (i.e., epileptiform activity). This system has been developed over several years as a collaborative effort between the Departments of Neurology, and Medical Physics and Bioengineering at Christchurch Hospital and the Department of Electrical and Electronic Engineering at the University of Canterbury. Initial development was undertaken by Bruce Davey as part of his PhD research (1). However, the system developed by Davey et al. (1989) had only been tested on a single EEG. Consequently, substantial modifications and additions were needed to obtain a system capable of detecting epileptiform activity in its many forms, while eliminating false detections. Over the past four years, I have developed a new system based on a wide range of EEGs. The performance of this system has been evaluated on 21 EEGs and is presently being further investigated through a blind clinical trial.

Part 3 introduces the field of neural networks and illustrates the wide range of application domains. Particular attention is paid to back-propagation neural networks and their application to the detection of epileptiform activity in EEGs. Part 4 is concerned with deterministic chaos and its presence in the brain. A model of neuron behaviour is proposed that enables deterministic chaos to be introduced into neural networks which, consequently, display more realistic activity and improved performance at given tasks. The thesis concludes with a discussion of the implications of these engineering areas to brain research (Part 5).

At this point, it is worthwhile introducing the notation used throughout the thesis. A vector is represented as a bold lowercase character (e.g., $x, \xi$). Subscripts are used to denote the elements of a vector. For example, the elements of a vector $x$ are $x_1, x_2, ..., x_M$. A matrix is represented by a bold uppercase character (e.g., $A$) and its elements are again denoted with subscripts (e.g., $a_{11}, ..., a_{1M}; a_{21}, ..., a_{2M}; ..., a_{N1}, ..., a_{NM}$). Constants tend to be uppercase characters (e.g., $N, M$), while variables tend to be lowercase (e.g., $x, y, z$). Variables are a function of time. Continuous time is represented by $t$ while discrete time is represented by $n$. Thus, a variable of continuous time is written as $x(t)$ while a discrete sequence is represented as $x(n)$.

During the course of the work reported in this thesis the following papers and presentations have been prepared:


A. A. Dingle, R. D. Jones, G. J. Carroll and W. R. Fright (1993) A multi-stage system to detect epileptiform activity in the EEG. Accepted for publication in *IEEE Transactions on Biomedical Engineering*.

Part I

INTRODUCTION

From time immemorial humans have been interested in understanding the brain. However, despite the long history of brain research, the operation of the brain is still relatively poorly understood.

This Part provides an introduction to brain research and, in particular, to the roles of engineering in this field. Chapter 1 reviews the structure and function of the brain as it is presently understood, while Chapter 2 traces the history of brain research from the Ancient Greek philosophers (c. 500 B.C.) through to present day neuroscientists. Much of the knowledge about the brain has been hard won, with conflicting hypotheses often being debated for many years before finally being resolved through the advent of new technologies. Over the years, the place of physical scientists in brain research has been to provide the necessary tools and techniques. However, with the development of powerful computers, engineers have become more directly involved in brain research. Chapter 3 provides a brief overview of the roles of engineering in brain research, from the development of instrumentation, to the analysis of recorded signals, to the modelling of brain function.
CHAPTER 1

THE BRAIN

Brain, n. An apparatus with which we think that we think.
Ambrose Bierce, The Devil’s Dictionary, 1906

1.1 INTRODUCTION

The brain is the organ that enables an animal to survive in an ever-changing environment. It receives signals from the environment via sense organs (e.g., eyes, nose, skin), stores information that is likely to be relevant to future needs and determines appropriate responses to be performed by effector organs (e.g., muscles).

Despite the long history of brain research, which can be traced back as far as the Ancient Greek physicians and philosophers (c. 500 B.C.), the operation of the brain is still relatively poorly understood. In the words of John Eccles (1969):

“At the present time we can think of the total problem of brain operation as a pile of disconnected fragments of a jig-saw puzzle with only a relatively few fragments pieced together to give a meaningful understanding.”

As brain research and technology have advanced, knowledge has become scattered among an increasingly large number of disciplines, with each discipline addressing a different aspect of the problem. For example, neuroanatomists study the components of the brain and the ways in which they are connected, whereas neurophysiologists examine the operation of these components and how they work together. The problem of piecing together the fragmentary information available is compounded by the different levels of analysis employed by each discipline. While neurophysiologists and neuroanatomists concern themselves with the components of the brain, psychologists and philosophers are more interested in the higher functions of animals (e.g., language, memory) and neuropsychologists attempt to relate these functions to the underlying structures.

This Chapter provides a brief overview of the structure and operation of the brain. Although the discussion is restricted to the human brain, many of its features are common to the brains of other animals and, in particular, to those of mammals.

1.2 CELLS OF THE BRAIN

The brain is a tissue composed of many cells, the most important of which are the neurons. The neurons are the information processing units of the brain and are surrounded, supported and nourished by glial cells.
1.2.1 Neurons

Neurons are electrochemical cells specialized for the task of processing and communicating information. Although neurons come in many shapes and varieties, they usually consist of three structures (Fig. 1.1): (1) dendrites through which most input signals are received, (2) a cell body or soma which provides the life processes for the cell, and (3) an axon along which the output is transmitted to other neurons.

![Figure 1.1. Structure of a neuron.](image)

Each neuron is bounded by a cell membrane, which plays an important role in its operation. The intracellular and extracellular fluids contain ions (e.g., Na\(^+\), K\(^+\), Cl\(^-\)) and the differences in ion concentrations give rise to a membrane potential (typically \(-70\) mV at rest). The membrane potential at the origin of the axon, the axon hillock, determines whether an electrical output pulse, known as an action potential, is generated and transmitted along the axon to other neurons.

Neurons communicate with each other by way of uni-directional chemical interfaces termed synapses. Typically, synapses form between the axon of one neuron and a dendrite of another neuron (axodendritic synapses) or between the axon of one neuron and the soma of another neuron (axosomatic synapses). When an action potential reaches a synapse a chemical substance, known as a neurotransmitter, is released. The neurotransmitter crosses the synapse and activates special molecules, termed receptors, which are attached to the membrane of the receiving or post-synaptic neuron. The activation of the receptors causes a localized change in the membrane potential of the post-synaptic neuron. This induced potential is referred to as the post-synaptic potential and may be either excitatory (positive) or inhibitory (negative). The strength of synapse coupling is variable and alters with chemical composition, size and form. In fact, the most important learning and memory effects in the brain seem to be associated with synaptic connections.
Not all synapses in the brain are axosomatic or axodendritic, although these are the most common and also appear to be the most important. Axoaxonic synapses form between axons of neurons and modify the amount of neurotransmitter released by the post-synaptic neuron. Dendrodendritic synapses (between dendrites) also occur but their function is not presently understood (Carlson, 1986).

A neuron may receive thousands or tens of thousands of inputs by way of synapses. The resultant post-synaptic potentials are integrated, both spatially and temporally, to produce a change in the membrane potential. When the membrane potential at the axon hillock exceeds a threshold value (typically $-20 \text{ mV}$), an action potential is generated and transmitted along the axon to other neurons. After the emission of an action potential the neuron needs a finite time (1–2 ms) to recover. This is the absolute refractory period during which the neuron is unable to produce an action potential, regardless of the strength of the inputs. A relative refractory period of 5–7 ms follows this, during which the neuron operates with an elevated threshold, before resuming normal activity (Amit, 1989).

Neuroscientists often measure the activity of a neuron by its firing frequency. A nerve signal is a train of equal amplitude pulses and, therefore, the only significant expression of information appears to be the frequency of action potentials. Thus, a neuron can be likened to a voltage controlled oscillator with its membrane potential determining its firing frequency.

1.2.2 Glial Cells

Neurons only constitute about 10% of the cells of the brain (Thompson, 1967). Most of the remaining cells surround, support and nourish the neurons, and are collectively termed glial cells. The astrocytes provide neurons with physical and nutritional support, the microglia, along with some astrocytes, remove neurons which have died as a result of injury or from ‘old age’ and the oligodendroglia produce the myelin sheath which insulates the axons and, thus, prevents messages from interfering with each other (Carlson, 1986).

1.3 ORGANIZATION OF THE BRAIN

The $10^{11}$ or so neurons which make up the human brain are grouped together in an organized fashion. There are three major regions of the brain (Fig. 1.2): the brain stem, the cerebellum and the cerebrum.

The brain stem contains structures that control arousal and sleep as well as vital functions, such as respiration and the cardiovascular system. The cerebellum receives a large amount of sensory information which it uses to exert a coordinating and smoothing effect on movements. The cerebrum is the largest of the three regions and is divided into two almost equal halves known as the left and right cerebral hemispheres. The two hemispheres are interconnected by several commissures, the largest and most important of which is the corpus callosum.

1.3.1 The Cerebral Cortex

The outer portion of the cerebrum is 3–4 mm thick and is known as the cerebral cortex. The cerebral cortex is highly convoluted consisting of sulci (small grooves), fissures (larger grooves) and gyri (bulges between adjacent grooves). These convolutions mean that the cerebral cortex has a large surface area. In fact, there is a general correlation between the surface area of the cerebral cortex of an animal and the complexity of its behaviour (Hilgard et al., 1979).
The surface of each cerebral hemisphere is divided into four lobes: the frontal lobe, the temporal lobe, the parietal lobe and the occipital lobe (Fig. 1.2). Many regions of the cerebral cortex appear to be specialized for particular functions (Fig. 1.3).

The frontal lobes are specialized for the planning, control and execution of movements. In particular, the precentral gyrus, known as the primary motor cortex, contains neurons that participate in the control of movement. Stimulation studies have revealed that particular parts of the motor cortex cause movements of particular parts of the body. In fact, the body is effectively mapped onto the motor cortex, with disproportionately large areas dedicated to movement of the muscles involved in speech and to the fingers. Damage to the frontal lobes
may cause changes in personality (e.g., lack of inhibition, impulsivity, lack of concern) and
in intellectual abilities (e.g., inability to plan and follow through a course of action, inability
to take into account future consequences of actions) (Walsh, 1978).

The posterior lobes (temporal, parietal and occipital) are specialized for sensation, percep­tion and cognition. The temporal lobe contains the primary auditory cortex which receives
sensory information from the receptors of the inner ear. Auditory signals are mapped onto
the auditory cortex according to their frequency, which is known as a tonotopic mapping.
The temporal lobes are also concerned with olfaction and the integration of visual experience
with other forms of sensory information. Studies of anterograde amnesia — the inability to
learn new information — have revealed that the temporal lobes also play an important part
in memory (Geschwind, 1979).

The parietal lobes are concerned with somatic sensation and perception, as well as in­tersensory or cross-modal association. The post-central gyrus is referred to as the primary
somatosensory cortex because it receives information from the somatosenses (touch, pressure,
pain, temperature, body position and vibration). The body is mapped onto the somatosen­sory cortex in a similar way as onto the motor cortex.

The occipital lobes are specialized for vision, containing the primary visual cortex, which
receives sensory signals from the retina, and being involved with visual perception and visual
memory.

Despite the appearance of bilateral symmetry, each hemisphere has its own specialized
talents. The somatosensory and motor regions of each hemisphere are concerned mainly with
the opposite or contralateral side of the body. In the visual system a similar phenomenon
is found — the left hemisphere tends to process the right half of the visual field while the
right hemisphere processes the left half of the visual field. More specialized functions are
often represented asymmetrically in the brain. For example, linguistic abilities tend to reside
mainly in the left hemisphere, while the right hemisphere appears to be important for the
recognition of emotion, perception of melodies and the recognition of complex non-verbal
visual patterns (Walsh, 1978).

1.3.2 Subcortical Structures

The cerebral cortex consists mostly of glia and the cell bodies, dendrites and interconnecting
axons of neurons. Because the predominance of cells gives the cortex a greyish brown appear­ance, it is often referred to as grey matter. Beneath the cerebral cortex run millions of axons
which connect the neurons of the cerebral cortex with those of the subcortical structures.
The large concentration of myelin gives this tissue an opaque appearance and, hence, the
name white matter.

The subcortical structures include the basal ganglia, which form an important part of
the motor system; the thalamus, which relays sensory information to the cerebral cortex; the
hypothalamus, which organizes behaviours related to survival (e.g., fleeing, feeding, fighting
and mating) and the limbic system, which is involved in emotional behaviour, motivation and
learning. An important structure of the limbic system is the hippocampus which appears to
play a major role in learning and memory (Carlson, 1986).

1.4 PROTECTION OF THE BRAIN

The brain is the most protected organ of the body, being encased in a tough bony skull.
Furthermore, there are protective sheaths around the brain which are referred to as meninges. The meninges consist of three layers (Fig. 1.4). The outer layer, known as the dura mater, is thick, tough and flexible, while the middle layer or arachnoid membrane is soft and spongy. The innermost layer, known as the pia mater, is closely attached to the brain and follows every convolution of its surface. Between the arachnoid membrane and the pia mater is a gap, known as the subarachnoid space, which is filled with cerebrospinal fluid. This pool of cerebrospinal fluid surrounding the brain absorbs much of the shock caused by sudden head movements.

Cerebrospinal fluid is produced in the ventricles of the brain (Fig. 1.5). These are a series of hollow interconnected chambers, which are connected to the subarachnoid space through a number of small openings or foramen. The two lateral ventricles are the largest of these chambers and lie one in each hemisphere in the anterior region of the brain. They are
connected to the *third ventricle* which is located at the midline of the brain. This chamber in turn connects to the *fourth ventricle* which lies in the posterior region of the brain.
CHAPTER 2

HISTORY OF BRAIN RESEARCH

"The great tragedy of Science — the slaying of a beautiful hypothesis by an ugly fact."

T. H. Huxley (1825–1895)

2.1 INTRODUCTION

The history of brain research stretches back to the Ancient Greek physicians and philosophers (c. 500 B.C.). Despite more than 2000 years of investigation into the brain, relatively little is known of its operation. The knowledge that has been acquired has been hard won. Often conflicting hypotheses have been debated for many years, and in some cases for several centuries, before finally being resolved by the advent of new technology or through improved experimental procedure.

This Chapter reviews some of the major conflicts that have arisen throughout the long history of brain research and the tools and techniques that have enabled them to be resolved. A number of extracts from the extensive collection of writings “The human brain and spinal cord” (Clarke and O’Malley, 1968) are used to illustrate the thinking of the time.

2.2 FROM ANTIQUITY TO THE MIDDLE AGES

The most important idea that arose during this period was that the brain, rather than the heart, was the source of mental processes. Much of the visual or gross anatomy of the nervous system was uncovered during this time and several theories regarding its operation were proposed.

2.2.1 Source of Mental Processes

The controversy regarding the location of mental processes can be traced to the Ancient Greeks (c. 500 B.C.). The prevailing opinion of this time was the cardiac hypothesis, which maintained that the heart was the central organ of the body. The foremost proponent of this view was Aristotle (384–322 B.C.), who argued that because the heart is warm and active, it must be the source of mental processes. Despite its secondary status, Aristotle did recognize the importance of the brain:

"The purpose of its presence in animals is no less than the preservation of the whole body." (Clarke and O’Malley, 1968, p. 9)
But regarding its function he wrote:

"... nature has contrived the brain as a counterpoise to the region of the heart with its contained heat, and has given it to animals to moderate the latter ... . The brain, then, tempers the heat and seething of the heart." (Clarke and O'Malley, 1968, p. 10)

Aristotle, thus, believed that the purpose of the brain was to cool the blood and went on to explain the difference in brain size between humans and animals as follows:

"Of all animals, man has the largest brain in proportion to his size; and it is larger in men than in women. This is because the region of the heart and of the lung is hotter and richer in man than in any other animal; and in men than in women." (Clarke and O'Malley, 1968, p. 10)

The opposing opinion or brain hypothesis appears to have originated with Alcmaeon of Croton (c. 500 B.C.), who asserted that the brain was the centre of sensation and thought. His followers included the Hippocratic physicians (c. 430–350 B.C.), who described the brain as "the most powerful organ of man's body", Plato (c. 429–347 B.C.), who argued in favour of the brain hypothesis because the brain is located closest to the heavens, Herophilus (c. 300 B.C.) and Erasistratus (c. 260 B.C.).

The relative merits of the brain and cardiac hypotheses were debated for many centuries with evidence and logic being presented in support of each. Some of the most notable arguments were presented by Galen of Pergamum (129–199 A.D.), who used experiment, analogy and reasoning to refute the cardiac hypothesis and to advocate the brain as the seat of intelligence, sensation and motion. He demonstrated that when pressure was applied to the brain it rendered an animal senseless and immobile, while pressure applied to the heart caused only pain. Furthermore, he pointed out that the sense organs were connected to the brain rather than to the heart and that nerves originated from the brain. Even today many of Galen's arguments are impressive and it is difficult to understand how the cardiac hypothesis survived as long as it did. In fact, supporters of the cardiac hypothesis can be found as late as the sixteenth century.

2.2.2 Anatomy of the Nervous System

Although Aristotle argued that the heart was the centre for thought and sensation, it is in his writings that the first account of any detail is found of the structures of the nervous system:

"But, taking size for size of animal, the largest brain, and the moistest, is that of man. Two membranes enclose it: the stronger one nearer the bone of the skull [dura mater]; the inner one [pia mater], round the brain itself, is finer. The brain in all cases is bilateral. Behind this, right at the back, comes what is termed the 'cerebellum', differing in form from the brain as we may both feel and see." (Clarke and O'Malley, 1968, p. 8)

Herophilus and Erasistratus, renowned for their knowledge of anatomy, differentiated between motor and sensory nerves and their suggestion that nerves were hollow had repercussions for many centuries. Erasistratus appears to be the first to describe the four cavities or ventricles of the brain, which subsequently formed the basis for many theories of brain operation:
"I investigated the nature of the [human] brain; and the brain was indeed bipartite as in other animals. It had a ventricle placed longitudinally on each side [the lateral ventricles], and these were pierced through into [another] one at the junction of the [two] parts. This one [the third ventricle] extended to the so-called cerebellum, where there was another, smaller ventricle [the fourth ventricle], each side walled off by membranes; for the cerebellum was set off by itself, as well as the cerebrum, and was like the jujenum and very much folded. The cerebrum was constructed from even more and differing folds." (Clarke and O'Malley, 1968, p. 12)

Erasistratus, after describing the convolutions of the surface of the brain, went further suggesting that they were related to intelligence:

"... since man greatly surpasses other beings in intelligence, his brain was greatly convoluted." (Clarke and O'Malley, 1968, p. 12)

However, this idea was later rejected by Galen. In this excerpt, although noting the lack of convolutions in the cerebrum of an ass, Galen relates the superior intelligence of humans to the quality of their spirit:

"Erasistratus demonstrates very well that the epenkranis (as he calls the cerebellum) is of a more varied composition than the cerebrum; but when he claims that the cerebellum, and with it the cerebrum, is more complex in man than in other animals, because these latter do not have intelligence like that of man, it does not appear to me that he is reasoning correctly, since even asses have a very complicated cerebellum although their imbecile character demands a very simple and unvariegated cerebrum. It is better to believe that intelligence results from the good temperament of the body charged with thinking, whatever this body, and not the variety of its composition. Indeed, it seems to me that it is less the quantity than the quality of the psychic pneuma that is necessary to produce perfection of thought." (Clarke and O'Malley, 1968, p. 631)

Some of the best descriptions of brain anatomy are found in the writings of Rufus of Ephesus (fl. 98–117 A.D.), who seems to be the first to consider the brain, spinal cord and nerves as a connected system:

"The brain is located inside the skull; it is covered by meninges; one, denser and more resistant, is attached to the bone [dura mater]; the other, thinner but also resistant, although to a lesser extent, stretches over the brain [pia mater]. The upper surface of the brain is called varicose [convoluted], its inner and posterior surface is called the base; the extension from the base is the parencephalon [cerebellum]; the cavities of the brain have been designated hollows [ventricles]. ... The processes springing from the brain are the sensory and the motor nerves with the help of which we are able to feel and to move voluntarily and which are responsible for all activities of the body." (Clarke and O'Malley, 1968, p. 13)

By the early centuries A.D. much of the visual anatomy of the nervous system was known. It was realized that nerves were responsible for voluntary movement and sensation but the cardiac-brain debate continued, as can be seen by Galen's writings:

"It is agreed by all physicians that none of the parts of the animal possesses either what we term voluntary motion or sensation without a nerve; and that, if
the nerve should be cut, the part immediately becomes motionless and insensitive. But it is not yet realized by all that the brain is the beginning of the nerves, just as it is of the spinal cord; and yet in dissections the fact is thus.” (Clarke and O’Malley, 1968, p. 20)

2.2.3 Operation of the Nervous System

Speculations regarding the operation of the nervous system began even before much of its anatomy had been uncovered. Diogenes (c. 440 B.C.) thought that air was responsible for the movement of limbs and that it found its way to them by means of blood vessels. The Hippocratic physicians believed that the body acted “according to the brain’s understanding” and that this understanding was provided by inhaled air:

“When man draws breath into himself, this breath first arrives at the brain, and it is in this way that air disperses itself through the rest of the body, leaving in the brain its most active part, the part that has intelligence and understanding. If the air were to proceed first of all into the body, and from there to the brain, it would leave the intelligence in the flesh and veins, and this would reach the brain overheated, not pure but mixed with the humors originated from the flesh and blood, so that it would no longer have its perfect qualities.” (Clarke and O’Malley, 1968, p. 5)

Plato also had little knowledge of anatomy and tended to favour blood as the agent by which sensory impressions were conveyed. On the other hand, Aristotle considered this task to be performed by a spirit or ‘pneuma’. While Herophilus introduced the idea that nerves were hollow and contained ‘psychic pneuma’ or ‘animal spirits’, it was Erasistratus who extended this concept to muscle contraction, explaining that when the psychic pneuma reached the muscle (via a motor nerve) the muscle inflated. The consequent increase in breadth of the muscle was accompanied by a corresponding decrease in its length and the muscle contracted.

Of all the early Greek physicians, Galen was probably the most influential. Building on the ideas of his predecessors, he formulated a ‘complete’ theory of nervous system operation. He adopted the doctrine of the hollow nerve which either transmitted psychic pneuma to the muscles or received sensory impressions:

“The nerves which in consequence enjoy the role of conduits, carry to the muscle the forces that they draw from the brain as from a source ...” (Clarke and O’Malley, 1968, p. 150)

Galen considered the ventricles to be a reservoir for the psychic pneuma which was manufactured by the solid parts of the brain. He considered that the ‘vital spirit’ in blood from the heart was changed into psychic pneuma by the brain which were then stored in the ventricles.

Although Galen located mental faculties in the substance of the brain, the theory that was to be universally adopted placed them in the ventricles. Nemesius of Emesa (c. 390 A.D.) developed the idea of ventricular localization, placing imagination in the lateral ventricles, intellect in the third ventricle and memory in the fourth ventricle. Although the basic form of this idea remained unaltered, there were several variations. The accepted ventricular localization doctrine that was taught well into the sixteenth century stated that the lateral ventricles receive information from the five senses and transmit it to the third ventricle which is able to imagine, reason and cogitate, while the fourth ventricle is dedicated to memory.
2.3 THE MIDDLE AGES TO THE NINETEENTH CENTURY

Throughout the Middle Ages there was continued controversy over the relative merits of the heart and the brain as the source of mental processes. In 1588, Andrea Cesalpino still believed that "the heart is not only the origin of all the veins but also of the nerves" and René Descartes, writing in 1649, noted that some would place the seat of the soul in the heart, although he himself located it in the brain (Clarke and O'Malley, 1968, p. 25).

During the Middle Ages, the Moslems were active in preserving and, in some cases, extending Greek medical learning. In particular, the works of Galen became a body of dogmata for the medical world, and little new was added until the sixteenth century.

From the sixteenth to the nineteenth centuries, several major discoveries about the nervous system were made, the most important of which was the discovery of the neuron and the idea that it was the unit of brain structure and function.

2.3.1 Localization of Brain Function

Ventricular localization was still being taught well into the sixteenth century when Andreas Vesalius began to oppose this doctrine. Although Vesalius still believed the ventricles played a part in the production of animal spirits, he denounced localization of mental functions:

"I have decided for the present to say nothing about the ventricles other than that they are hollows or cavities in which air, drawn in during inspiration, and vital spirit, transmitted to them by the heart, are altered by the peculiar substance of the brain into animal spirit ... . Hence I do not hesitate to ascribe to the ventricles a share in the production of animal spirit, but I believe nothing ought to be said of the locations of the faculties of the principal soul in the brain." (Clarke and O'Malley, 1968, p. 718)

Vesalius went on to show that the difference between the brains of humans and of animals is not in the ventricles but in brain mass:

"Nevertheless in dissection we see that men do not differ from animals by possession of any special ventricle; not only is the number of ventricles the same, but all other things are very much alike in man and animal except in respect to the mass of the brain and a temperamental urge towards upright conduct." (Clarke and O'Malley, 1968, p. 718)

Gradually the ventricles were abandoned as the site of mental functions and several theories were proposed localizing functions in the substance of the brain. Thomas Willis (1621-1675) suggested that the cerebellum controlled involuntary movements such as those of the heart and the lungs, while the cerebrum was the centre of voluntary movements and sensation. Within the cerebrum he localized sensation in the corpus striatum, imagination in the white matter and memory in the cerebral cortex.

Franz Gall (1758-1828) and Johann Spurzheim (1776-1832) drew widespread attention to the cerebral cortex which they suggested was made up of many organs, one for each psychological faculty. They believed that superlative development of a particular trait required superlative development of the relevant brain centre. Because these centres were located on the surface of the brain, their atrophy or supertrophy would be reflected by the shape of the skull. Gall and Spurzheim examined the external features of the skull and correlated the
bumps and depressions with what they considered to be important aspects of behaviour (e.g., secretiveness, cautiousness, destructiveness). The study of the shape of the skull in relation to psychological traits was termed phrenology. From examination of skull shapes of criminals and mentally ill patients, as well as busts of famous men, Gall and Spurzheim established maps of functional localization. Of most significance was the localization of language in the frontal lobes, which resulted from Gall's study of a penetrating brain injury that caused language deficits.

Pierre Flourens (1794–1867) opposed both phrenology and the concept of localization of function. Flourens developed a crude technique for lesioning an animal's brain and studying any resultant behavioural changes. From studies of chickens and pigeons, he concluded that the cerebellum controlled locomotion, that the medulla oblongata (bottom of brain stem) was the seat of the principle of life because its damage arrested respiration, and that the cerebrum was the seat of intelligence. He believed that there was no localization within the cerebrum and that all faculties existed there coextensively. Thus, loss of function was correlated with the extent of the lesion:

"... a large section of the cerebral lobes can be removed without loss of function. As more is removed all functions weaken and gradually disappear. Thus, the cerebral lobes operate in unison for the full exercise of their functions."
(Changeux, 1985, p. 16-17)

Flourens' conclusions were accepted universally for many decades, despite the fact that most of his experiments were carried out on birds, which have very little cerebral cortex.

The first irrefutable evidence supporting cortical localization was provided by Jean Baptiste Bouillard (1796–1881) who, from his experience of patients with brain disease, agreed with Gall that language function was localized in the frontal lobe. In 1836, Marc Dax presented a series of clinical cases which demonstrated that speech disorders were constantly associated with lesions of the left cerebral hemisphere. Ernest Auburtin, in 1861, reported a case where a patient ceased to speak when pressure was applied to his exposed anterior lobes. By 1863, Pierre-Paul Broca had studied eight patients, who were unable to speak despite having perfect comprehension and intact vocal muscles:

"Here are eight instances in which the lesion was in the posterior third of the third frontal convolution. This number seems to me to be sufficient to give strong presumptions. And the most remarkable thing is that in all the patients the lesion was on the left side. I do not dare draw conclusions from this. I await new facts"
(Joynt, 1964).

By establishing correlations between anatomical and behavioural features, Broca gave the first demonstration of the discrete cortical localization of well-defined faculties and, furthermore, revealed an asymmetry between the hemispheres. In 1874, Carl Wernicke claimed that there was a second centre for language located on the first and second temporal convolutions. Lesions of this area resulted in a deficit in comprehension but not in expression; patients could speak fluently but what they said made little sense and, although they could hear, they were unable to understand or repeat what was said. Wernicke theorized that this region of the temporal lobe stored sound images of objects and these were sent over a pathway to Broca's area where representations of speech movements were retained. Furthermore, he suggested that if the fibres connecting the two language regions were damaged, a speech deficit would result where expression would be retained, as would comprehension, but speech would still be nonsense. The work of clinical neurologists, such as Broca and Wernicke, seemed to indicate
that functions were somehow localized in the cerebral cortex. Although many were excited by this idea others maintained strong objections.

In 1870, Eduard Hitzig and Gustav Fritsch introduced a new technique to study localization of function and provided further evidence in support of the theory of cortical localization. From experiments stimulating the exposed cortex of dogs with electric currents, they showed not only that the cortex was excitable but that it was selectively so. Direct application of electric current to portions of the anterior cortex caused movements on the opposite side of the body. Furthermore, stimulation of restricted portions of the anterior cortex elicited movements of particular body parts, suggesting that there were topographic representations of the different parts of the motor system.

Over the next few years David Ferrier refined the stimulation techniques and confirmed the results of Hitzig and Fritsch on monkeys, cats, rabbits, guinea pigs, fish, pigeons and frogs. He was able to delineate the motor cortex and concluded that “there are certain regions in the cortex to which definite functions can be assigned” (Clarke and O’Malley, 1968, p. 518).

Roberts Bartholow, in 1874, was the first to report electrical stimulation of the human brain. However, it was Fedor Krause, in the early 1900s, who applied this technique to conscious patients undergoing neurosurgery and was, consequently, able to map the human motor region.

Despite the evidence supporting functional localization, opposition to the hypothesis continued. In 1892, Friedrich Goltz performed a series of experiments that were intended to criticize cortical localization. He reasoned that if a portion of the cerebral cortex had a function, as maintained by the localizationists, then its removal should lead to a loss of that function. Having removed the whole cortex of three dogs, he observed that the dogs’ movements and sensations were not grossly impaired as predicted. The dogs still panted when warm, shivered when cold, could catch their balance when they slipped, and were able to orient to pinches, light and sound, although their response thresholds were elevated. Goltz interpreted his findings as indicating a general lowering of will and intellect but argued that they did not support the localization hypothesis.

John Hughlings Jackson (1835–1911) proposed a theory of brain operation that resolved the differences between Goltz’s findings and the views held by localizationists. He viewed the brain as being organized into a number of layers arranged in a functional hierarchy. Each successively higher level was thought to control more complex aspects of behaviour but to do so through the lower levels. He adopted the theory of hierarchy from Herbert Spencer’s argument that the brain evolved in a series of steps, each of which gave animals the capacity to engage in a constellation of new behaviours. Jackson suggested that destruction of high level structures (e.g., the cortex) would impair more complicated versions of behaviour but, so long as lower level structures were intact, simpler low-level versions would remain. Therefore, Goltz’s dogs were effectively ‘low-level’ dogs.

The relevant question then is not where is a particular function localized but what unique contribution is made by each part of the cortex. Jackson was remarkably modern in his ideas which are receiving more serious consideration today than they did in his own time. On the whole, the criticisms of the anti-localizationists denounced the excesses of strict localizationists and have led to a better understanding of the complexity of brain operation.

2.3.2 Nerve Function

Although experiments and subsequent theories regarding functional localization appear to have carried on independently from other aspects of brain research, many discoveries re-
garding nerve function had a profound influence of the type of experiments performed. For example, the discovery that nerve messages were electrical in nature led to experiments involving electrical stimulation of the brain. However, this revelation only came about after many years of research.

Although the purpose of nerves was understood in the sixteenth century, it was still generally accepted that nerves were hollow and that through them flowed animal spirits. Andreas Vesalius described the function of nerves:

"The nerves may be considered as the diligent servants and messengers of the brain as they deliver the spirit prepared by the brain to those instruments to which it must be sent." (Clarke and O'Malley, 1968, p. 154)

Attention, at this time, centred on the transmission of animal spirits through the nerves, but Vesalius' refused to discuss this subject, denying that nerves were hollow:

"I scarcely dare to deny the hollowness of the nerve although I have never seen a channel even in the optic nerve; this despite the fact that the nerves, like the veins and arteries, are called vessels and that I am aware that professors of anatomy declare the optic nerves to be hollow." (Clarke and O'Malley, 1968, p. 155)

René Descartes was among one of the first to suggest that animal spirits were material rather than spiritual bodies. He espoused a balloonist theory that required a fluid to flow through the nerves, with movement being caused by the filling and emptying of muscles. He illustrated his theory with the following example. Suppose the skin of the foot is touched by a burning ember. This displaces the skin and pulls a tiny thread stretching from the foot to the brain. This in turn pulls open a pore in the brain, permitting animal spirits to flow through the nerves and inflate the muscles causing movement (Churchland, 1986, p. 16).

Whereas Descartes and other French and Italian scientists applied the laws of the physical sciences to nerve function, another group of investigators, mainly in England and Holland, preferred to relate chemical data to the problem. These were the iatrochemists and one of the most important among them was Thomas Willis. Willis believed that when the vital spirit of the blood and the animal spirits of the nerves met in the muscle, the resulting reaction produced contraction.

Giovanni Borelli (1608–1679) showed that animal spirits could not be gaseous by submerging a struggling animal in water. When the animal's muscles were slit no bubbles appeared. Subsequently, Borelli suggested that nerves must use a liquid medium for transmission of messages:

"... the tubes of the nerve fibres containing a spongy pith can be saturated and swollen by the spirituous juice of the brain, and then if one end is compressed, set in motion, struck, or twitched, the vibration, shaking, or surge must immediately be communicated to the other end". (Clarke and O'Malley, 1968, p. 164)

At the muscle Borelli suggested that a chemical reaction caused contraction:

"... some spirituous droplets can be squeezed out from their terminal orifices into the corresponding muscle, whence a boiling up and dilatation follow and the muscle is thereby contracted and dilated." (Clarke and O'Malley, 1968, p. 164)
Francis Glisson (1597–1677) disproved balloonist theories of nerve function by submerging a man’s arm in water and measuring the displacement of water when the muscles were contracted. Because no change in water level could be observed, Glisson concluded that muscle contraction was not the result of fluid flowing into the muscle as was commonly thought. Jan Swammerdam, in Holland, reached the same conclusion from similar experiments on frogs.

The results of Glisson’s experiments on muscle contraction led him to reject materialistic views of nerve function. In their place he proposed the vitalistic concept of irritability, which, in the hands of Albrecht von Haller (1708–1777), became accepted as a fundamental property of living matter. When irritated by touch or by chemicals, the nerves would make the muscles contract. Although Haller realized that electricity was a candidate for nerve transmission, he did not consider it to be the natural transmission medium, believing nerves to lack the insulation required for conduction.

Interest in theories of electrical conduction arose in 1731 when Stephen Gray demonstrated that the human body could be electrified. In 1780 Luigi Galvani observed that discharges of static electricity from Leyden jars could cause muscles to contract. Then in 1786, while trying to observe the effect of atmospheric electricity on the contraction of a frog’s leg, Galvani saw the limb contract spontaneously. Because this took place on a calm day with no sign of an electrical storm, he concluded that the frog itself had produced ‘animal electricity’, which caused the muscle to contract:

“From what has hitherto been learned and investigated I consider it sufficiently well established that there is an electricity in animals”. (Clarke and O’Malley, 1968, p. 184)

However, in this experiment the preparation was suspended on an iron bar with a copper hook implanted in the spinal cord. Consequently, Galvani’s conclusion became the target of criticism from Alessandro Volta, who correctly claimed that Galvani had not proved the existence of animal electricity but had simply shown that contact between iron and copper caused muscle contraction. Galvani, in collaboration with his nephew, Aldini, quickly replied to this criticism with an experiment involving no metal; the exposed muscle of the frog’s leg was shown to contract when placed in contact with its spinal cord.

The controversy over ‘metallic’ versus ‘animal’ electricity only came to an end with the development of adequate measuring devices. Hans Oersted and Johann Schweigger invented the galvonometer, which was able to detect the presence and strength of electric currents. In 1825, Leopoldo Nobili improved this instrument and produced the astatic galvonometer. Using this galvonometer, Carlo Matteucci was able to record electric currents in muscles and the field of electrophysiology was born.

In 1848, Emil du Bois-Reymond was the first to demonstrate that the nervous effect was an electrical phenomenon and that a wave of electrical negativity, an action potential, passes down the nerve:

“It is well known that I have since extended those experiments successfully to the living body of the frog, of the rabbit, and even of man and that I have detected a negative variation of a similar description which the nerve current undergoes while the nerve is conveying to the brain or to the muscle, those material changes which give rise to sensation and motion.” (Clarke and O’Malley, 1968, p. 196)

In addition, du Bois-Reymond proposed a universal law of nerve stimulation by electric current:
"It is not the absolute value of the current density at any given moment to which the motor nerve responds by the contraction of its appropriate muscle, but the change of this value from one moment to the next. This means that the stimulation to move, which results from these changes, is the more powerful the more rapidly these changes of the same magnitude occur or the greater they are per unit time." (Clarke and O'Malley, 1968, p. 197)

It was well known that nerves could be excited by electricity, but establishing that electricity was the essential feature of nerve function was of great significance. Du Bois-Reymond realized the importance of his discovery stating:

"If I have not completely deluded myself, I have succeeded in restoring to life in full reality the hundred-year-dream of the physicist and physiologist, the identity of the nerve principle with electricity." (Clarke and O'Malley, 1968, p. 193)

The study of nervous transmission had finally resulted in the animal spirits of the Ancient Greeks being identified as electrical impulses.

Richard Caton reasoned that if activity in the peripheral nerve was accompanied by electrical activity a similar phenomenon may occur in the brain. By placing electrodes in contact with the exposed brains of rabbits and monkeys, he was able to observe electrical fluctuations:

"In every brain hitherto examined, the galvonometer has indicated the existence of electrical currents." (Caton, 1875)

Caton was convinced that these potential fluctuations were related to the functional activity of the brain:

"The electric currents of the grey matter appear to have a relation to its function. When any part of the grey matter is in a state of functional activity, its electrical current usually exhibits negative variation." (Caton, 1875)

Caton's work established that in addition to being sensitive to electrical stimulation the brain itself produces electrical activity. This discovery formed the basis for the field of electroencephalography (Chapter 4).

### 2.3.3 The Neuron

Early views of the brain showed little concern for its structure, probably because to the naked eye it appeared to be a relatively homogeneous mass. Deciphering the fundamental structure of the nervous substance depended directly on the development of the optical microscope.

One of the first cellular anatomists, Antony van Leeuwenhoek (1632–1723), searched for cavities in nerves with a primitive microscope. But his results were inconclusive and even a hundred years later the hollowness of nerves was still being debated. No significant improvement was made to the microscope until the 1830s, by which time the instrument had, on the whole, fallen into disrepute because of artifacts which frequently misled investigators. Using the new achromatic microscope, Christian Ehrenberg was able to describe the nerve cell of the cerebral cortex. Jan Purkyně observed cell bodies through the microscope and on other slides saw the long skinny fibres that are now termed axons. However, it was not until
1838, when Robert Remak ventured the opinion that the nerve fibres were actually extensions of the nerve cells, that anyone suspected nerve cells were radically different in shape from red blood cells. Adolph Hannover substantiated the opinion that nerve fibres are connected to nerve cells, after studying slices of the cerebral cortex:

"... it now seems to me more than probable that the brain fibres originate from the brain cells and retain a lifelong, permanent connection with these central structures". (Clarke and O'Malley, 1968, p. 61)

However, it was Rudolf Albert von Koelliker who, in 1849, demonstrated conclusively that nerve fibres originated from nerve cells. He then made reference to the problem of the relationship between cells, which was later to engage the attention of many researchers:

"How these ultimate [fibre] processes finally end, whether freely or in connection with nerve fibres, or whether they anastomose with similar processes, is not yet known, although it seems not at all improbable that the three possibilities mentioned according to the location may be found." (Clarke and O'Malley, 1968, p. 62)

Understanding the cellular structure of the brain was advanced chiefly by improvements in technology. The early anatomists, who tried to examine the microscopic structure of the brain, were faced with a gelatinous substance. In 1809, Johann Reil developed the technique of fixing and hardening tissue by placing it in alcohol. In 1840, Adolph Hannover found that chromic acid was a more effective fixer. The formaldehyde technique, which is the preferred modern fixative procedure, was discovered by Ferdinand Blum in 1893. During the 1830s Luigi Rolando devised techniques for cutting thin sections of tissue and Benedikt Stilling found, in 1842, that by first freezing the tissue extremely thin sections could be obtained. The most exciting development in the field of neuroanatomy was the technique of staining, which allowed different portions of cells to be visualized. In 1858, Joseph von Gerlach devised the carmine stain which turns cells and their fibres red. Franz Nissl, in 1894, discovered the methylene blue technique which stains cell bodies and reveals many of their internal structures. The most amazing staining procedure was developed by Camillo Golgi in 1875. He found that when the tissue was impregnated with silver nitrate only a few cells (1-10%) would take up the stain and each of these cells would do so in its entirety. Thus, the whole neuron (its dendrites, soma and axon) could be visualized at one time.

Using carmine staining techniques, Otto Deiters was, in 1865, able to observe, for the first time, a network of very thin and delicate dendrites extending from the cell body. On the basis of observed structural differences, he drew a distinction between 'protoplasmic processes' (dendrites) and 'nervous processes' (axons). Deiters then began to explore the problem of contact between neighbouring nerve cells, which was to become the most important issue in the investigation of nervous tissue.

Two schools of thought developed regarding the connection of nerve cells. The reticular theory, which maintained that neurons fused together, originated in 1855 with Franz von Leydig, who observed numerous interlacing fibrils in the nervous systems of spiders. Later, Joseph von Gerlach, while studying dendrites in the human cortex, found two distinct fibre networks: a fine network formed of dendrites and a coarser axonal network. In 1883, Golgi dismissed Gerlach's hypothesis of an interlinking network of dendrites but maintained that axons were connected to form an axonal network. He believed that nerve cells must fuse or 'anastomose' to form a continuous nerve net or reticulum.

The neuron theory maintained that neurons were independent cells. Experiments with neuronal growth and degeneration provided the first evidence supporting this hypothesis.
Wilhelm His, in 1888, showed through a series of experiments that foetal neurons definitely start out as independent entities and then proceed to extend their axonal and dendritic processes. There seemed no evidence that neurons subsequently fused. In the mirror image of His's tests, August-Henri Forel found that when a cell body was damaged only the axon attached to it degenerated and, conversely, when the axon was injured only its cell body showed typical degenerative signs. Meanwhile in Spain, Santiago Ramón y Cajal was working on the same problem using the Golgi stain. From his preparations it appeared that axons had terminal bulbs which came very close to membranes of other cells but did not actually fuse with them. Consequently, he rejected the reticular theory:

“We have made careful investigation of the course and connections of nerve fibres in the cerebral and cerebellar convolutions of man, monkey, dog, etc., and we have never seen an anastomosis between the ramifications of two different protoplasmic expansions, nor have we observed them between the filaments emanating from the expansions of Deiters [i.e., from axons]; the fibres intermingle in a very complex manner, producing a thick and intricate plexus but never a net. ... This is not to deny indirect anastomosis (by branches of the filaments of Deiters) but to affirm simply that never having seen them, we dismiss them from our opinion.” (Clarke and O'Malley, 1968, p. 112)

Apparently what hardened the resolve of Golgi was that unless neurons formed a reticulum the manner of their communication would be unexplainable and, as a matter of consequence, old theories of animal spirits would be revived:

“I cannot abandon the idea of a unitarian action of the nervous system without being uneasy that by so doing I shall become reconciled to the old beliefs.” (Clarke and O'Malley, 1968, p. 96)

Cajal was not insensitive to Golgi's concerns, agreeing that the reticular theory, if true, would make communication easier but concluded that this hypothesis was unsupported by the evidence:

“From the analytic point of view it would be very convenient and economical if all the nerve centers formed a continuous network intermediate between motor nerves and sensory nerves. Unfortunately, nature seems to ignore our intellectual need for convenience and unity, and is very often pleased with complexity and diversity.” (Clarke and O'Malley, 1968, p. 128)

Cajal thought it reasonable to conjecture that electrical induction might account for all interneuronal communication.

In 1891, Wilhelm Waldeyer critically analysed the evidence available and concluded that it favoured the independence of nerve cells which he termed neurons and, by the turn of the century, opposition to the neuron theory had considerably declined.

2.4 THE TWENTIETH CENTURY AND BEYOND

By the beginning of the twentieth century it was well established that the neuron was the fundamental unit of nervous system structure and the independence of neurons was becoming more widely accepted. The remaining questions concerned their operation, how they
communicate with each other and how they relate to higher functions (such as memory and reasoning).

The advent of new and powerful technologies has enabled brain research to continue to advance. The availability of powerful computers has led engineers to become directly involved in brain research. In fact, computer simulations may provide the key to understanding the operation of the brain.

2.4.1 The Action Potential

Although the action potential had been discovered in the nineteenth century, Keith Lucas was the first to demonstrate the 'all-or-nothing' property of nervous transmission. Following this discovery, Alexander Forbes showed that the nature of nerve impulses was the same regardless of the type of stimulus. Finally, in 1926, Edgar Adrian conclusively demonstrated that nerve impulses were, in fact, the nerve message and showed that an increase in stimulus strength increased their frequency.

The development of the cathode ray oscilloscope provided a means for easy and accurate recording of action potentials. Joseph Erlanger and Herbert Gasser introduced this technology to nerve physiology in 1921 and were subsequently able to give detailed descriptions of nerve impulses. By 1886 Julius Bernstein had developed a membrane theory of nerve conduction based on the observation that all living animal tissues contain electrically charged ions of salt solutions (e.g Na+, K+, Cl−). Bernstein argued that these ions were the source of the electrical charge in the nerve. However, it was not until the development of microelectrodes in 1939 that the basic mechanism of nerve impulse transmission was demonstrated to be ionic exchange between intracellular and extracellular fluids. From microelectrode studies of the giant axon of squid, Alan Hodgkin and A. Huxley were able to describe the nature of nerve impulses:

“In such experiments it is found that the inside of a resting nerve fibre is 50 to 70 mV negative to the external solution; this standing difference is known as the resting potential. When an impulse travels along the fibre, the inside swings momentarily positive giving a transient action potential with an amplitude of 100–120 mV. At the crest of the action potential the inside of the fibre is 40–50 mV positive to the external solution.” (Clarke and O'Malley, 1968, p. 235–236)

Based on their experimental observations, Hodgkin and Huxley went on to develop a detailed mathematical model of nervous transmission, which today provides the basis for many simulations of neuron operation.

2.4.2 Neural Communication

The next question to be asked was how the electrical impulse travelling along the axon crosses the neuromuscular junction and causes muscle contraction. Already du Bois-Reymond had made mention of a chemical mechanism, envisaging “that there must be either a stimulating secretion, perhaps in the form of a thin layer of ammonia or lactic acid or some other substance on the surface of the contractile tissue, in such a way that a vigorous excitation of the muscle follows, or an electrical influence” (Changeux, 1985, p. 33).

However, the theory of chemical transmission was only to take shape through discoveries made in the field of pharmacology, a very old discipline concerned with medicines, drugs and
poisons. Analysis of the action of these substance only began in 1857 with Claude Bernard, who attempted to understand the effect of the South American arrow poison curare. He found that messages were somehow prevented from reaching the muscles and suggested that this was due to "the natural death of the motor nerve". Later it was shown that the nerve itself was uninfluenced by the poison and, in 1866, Edmond Vulpian concluded that the site of action of curare was not in the nerve, as Bernard thought, but at the neuromuscular junction.

In 1906, Charles Sherrington published a landmark book *The Integrative Action of the Nervous System* in which he introduced the term *synapse* to describe the communication structures between neurons. Sherrington's claim that the nervous system contained synapses was not based on direct observation but rather on inferences drawn from careful studies of simple reflexes in dogs. Sherrington concluded that conduction along nerve fibres is not the only mode of signal transmission and that, in order to account for the observed reflex delay, the signal must be transmitted across a gap between sensory and motor neurons by a slower process. With the development of the electron microscope in the 1950s, it became possible to observe synapses and the gap between membranes was measured to be approximately 200 Å.

In 1904, Thomas Elliott was the first to suggest that a chemical transmitter might be liberated at the nerve endings. He demonstrated the similarity between the effects of adrenalin and the electrical stimulation of sympathetic nerves and argued that perhaps adrenalin was the chemical stimulant released each time a nerve impulse arrived at the neuromuscular junction.

Otto Loewi, in 1921, provided direct experimental evidence of chemical transmission at the neuromuscular junction. He unequivocally proved that nerves do not influence muscles directly but that a chemical mediator is responsible for the observed effects of nerve stimulation. The chemical substance released by motor nerves was identified to be acetylcholine. Acetylcholine was, thus, one of the first neurotransmitters to be discovered and it was found that curare blocked its action.

Although chemical transmission had been discovered in the peripheral nervous system, it was to be several years before it was accepted that a similar process occurred between neurons in the brain. During the 1940s, pharmacologists found that acetylcholine was present in various regions of the brain and particularly in the cortex. In 1954, Marthe Vogt showed that noradrenalin (closely related to adrenalin) was present in the hypothalamus.

Since then it has been found that the nervous system contains a variety of neurotransmitters (e.g., acetylcholine, noradrenalin, serotonin, dopamine, glutamic acid, GABA, glycine), each of which interacts with a specialized receptor. Although a single neuron typically responds to a number of neurotransmitters, it produces and releases only one. While some neurotransmitters appear to be exclusively excitatory or inhibitory, others can produce either excitation or inhibition, depending on the nature of the post-synaptic receptors (Carlson, 1986).

Studies continue to identify further neurotransmitters and examine the steps necessary for synaptic transmission. Investigators have discovered many drugs that selectively affect the synthesis, storage, release, deactivation and re-uptake of neurotransmitters or that stimulate or block post-synaptic receptors. Drugs that facilitate post-synaptic effects are termed *agonists* while those that inhibit them are known as *antagonists*. These drugs serve as very important tools for investigating the functions of various classes of neurons and are used to treat neurological and psychological disorders.

Some neurons produce and secrete substances, called *neuromodulators*, that have a widespread effect. These alter the excitability of circuits of neurons in distant parts of the brain and affect mood, behaviour and appetite.
2.4.3 Neuroanatomy

Having established that neurons are the structural unit of the nervous system, the attention of neuroanatomists turned to neural circuitry. Many of the advances made in this field have been a direct result of new and powerful techniques.

In the early 1950s, Walle Nauta developed a staining technique that enabled the mapping of connections between different brain structures. When a neuron is destroyed its axon degenerates and, before it disappears altogether, it can be stained differently from its neighbours. If a particular region of the brain is destroyed and, a few days later, stained and examined under a microscope, the presence of selectively stained fibres in another (possibly quite distant) part means that this region received fibres from the destroyed structure.

The origin and destination of axons can be traced by injecting certain radioactive chemicals into brain structures. These chemicals are absorbed by neurons and can be located by photographic emulsion. Many neural pathways in the brain have been traced and reveal ordered topographic maps in the cerebral cortex (Fischbach, 1992).

A useful procedure for locating areas involved in a particular task is the deoxyglucose technique developed by Louis Sokoloff and his colleagues (Iversen, 1979). Glucose provides neurons with energy and, consequently, more active neurons absorb more glucose. Radioactive deoxyglucose is absorbed as though it were glucose but cannot be completely metabolized and, therefore, accumulates in active neurons. In a typical experiment, radioactive glucose is administered to the animal which is then stimulated with, for example, sound. The areas involved in hearing are those displaying the highest levels of radioactivity. Before the development of positron emission tomography (PET), radioactivity could only be detected after microscopic dissection. PET, however, enables the active brain structures in living animals to be mapped. PET and magnetic resonance imaging (MRI) are also being used to detect changes in cerebral blood flow. Increased blood flow in localized regions is associated with increased neural activity and, thus, changes in blood flow can help locate areas involved in a particular task (Fischbach, 1992).

With the help of technology, neuroscientists are making great advances in detailing the anatomy of the brain and in uncovering its relationships with cognitive processes (Goldman-Rakic, 1992).

2.4.4 Disorders of the Brain

Although for most people the brain functions effectively and unceasingly for more than 60 years, defects of the brain do sometimes arise in its structure or in its electrical and chemical processes. Adequate treatment of brain malfunction depends on an understanding of its mode of operation. Parkinson's disease is a progressive chronic condition characterized by, for example, muscle rigidity and involuntary tremor, which reduce walking to a shuffle. The work of Arvid Carlsson revealed that the disease is caused by progressive destruction of dopamine pathways in the brain. This information enabled an effective treatment to be developed with patients now being treated with L-dopa, the precursor of dopamine (Kety, 1979).

Because present understanding of brain malfunction is very limited, it should be expected that many disorders cannot be satisfactorily treated. However, empirical treatments are often surprisingly successful. Epilepsy is a condition characterized by recurrent seizures associated with a disturbance of consciousness (Kolb and Whishaw, 1980). In many cases, the causes and mechanisms of seizure activity are not understood but the condition can often be effectively treated with anticonvulsant drugs such as Dilantin. If medication fails, surgery
can be performed to remove the source of seizure activity (if it can be accurately localized), or to sever the corpus callosum, thus preventing seizures from spreading to both hemispheres. Treatments for schizophrenia and depression were discovered accidentally, without knowledge of their effects on the brain. For example, chlorpromazine was developed as an anaesthetic but proved to alleviate the symptoms of schizophrenia and mania. Studies of the effects of such drugs on the brain are aiding in understanding the chemistry of the brain, which will result in more effective treatments for many brain disorders.

Disorders of the brain do, however, provide a valuable source of information about operation of the brain. The oldest and still most widely used approach to the problem of brain function is to analyse the effects of lesions to areas of the brain. In humans this is restricted to naturally occurring lesions or to patients who have an epileptic focus surgically removed. However, results from lesion studies must be interpreted carefully to avoid incorrect conclusions being drawn. Consider the following analogy. If a resistor is removed from a radio that subsequently buzzes, the resistor could be assigned the role of buzz inhibition. In a similar way, misleading conclusions can be drawn from lesion data.

Imaging techniques, such as MRI and PET, are revealing both structural and physiological differences between the brains of certain patients and those of the normal population. For example, MRI has revealed that the cerebral ventricles are often much larger in patients with schizophrenia than in normals, while the hippocampus may be considerably smaller. Patients with schizophrenia also tend to have reduced blood flow to the prefrontal cortex indicating decreased neural activity in this area (Gershon and Reider, 1992).

2.4.5 Higher Functions of the Brain

Joseph Weizenbaum, a Professor of computer science, while talking about brain research said “prediction doesn't imply knowledge about what is happening at a given point in time” (Restak, 1979). Even if future states of the brain could be predicted from a physical point of view (which they cannot), it would be impossible to predict reliably what a person is thinking and why. Consider the analogy which is often drawn between the brain and the computer. Even if one knew the state of every flip-flop, every storage cell and every input, and the state transition map of the computer so that its state at any time in the future could be predicted, it would still be impossible to predict reliably what application the computer was working on. Similarly, describing brain function in terms of physics and chemistry does not imply that higher mental functions are understood.

There are many different levels of brain research. On the lowest levels, physics and bichemistry are employed to explain membrane function, cellular interaction and electrical conductivity, while the higher levels call upon psychology to study and characterize external behaviour with minimal regard to internal structure. Between these extremes lie other levels of study. For example, the field of neuropsychology tries to correlate behaviour with underlying structure. However, the many levels of brain research are ill-defined and scientists working on any one level need occasional forays both downward to find mechanisms for the functions studied, and upward to understand their role in the overall scheme.

In contrast to the general consensus reached by the 1950s concerning the fundamental structure and function of neurons, many conflicts and puzzles concerning higher functions and their implementation have remained unresolved (e.g., motor control, learning, memory, perception, cognition).

One of the most notable features of the brain is its ability to learn new tasks and adapt to environmental changes. However, learning cannot be achieved without memory. Learning
is the process by which knowledge is acquired, while memory is the process by which this knowledge is retained over time. Different types of learning have been uncovered from studies of patients with temporal lobe lesions. Although such lesions greatly impair learning that requires conscious participation (i.e., explicit learning), those forms of learning that do not require conscious effort (i.e., implicit learning which takes place slowly through repetition) remain intact. Evidence now exists that experiences can produce physiological changes in the brain, in particular to the synapses (Carlson, 1986). The details of the cellular mechanisms of learning are presently being uncovered (Kandel and Hawkins, 1992).

Since the 1950s, neuroscientists have become increasingly aware that memory is distributed around a networks of neurons rather than being localized to a single structure. Present thinking is that long-term storage takes the form of an associative memory with stored information able to be accessed when required. Working memory is another type of memory that provides short-term storage of information relevant to the task at hand and enables manipulation of this information. For example, working memory is required when performing mental arithmetic and for planning chess moves. Evidence indicates that working memory is in the prefrontal lobes of the cerebral cortex and studies of patients with damaged prefrontal lobes are helping to uncover its operation. Furthermore, these studies are providing insights into schizophrenia because the abnormal mental attributes associated with schizophrenia are remarkably similar to those caused by damage to the prefrontal cortex (Goldman-Rakic, 1992).

Another important aspect of brain research is perception or the interpretation of sensations. Perception has often been regarded as a learned phenomenon because of the significant influence of experience. For example, a subject manipulating objects while viewing them through distorting glasses will rapidly display perceptual adaptation. However, studies of young infants have shown that certain perceptions are innate. Young children will not crawl out on a glass plate extended over the edge of a table, indicating that they have some perception of depth (Restak, 1979). It is hoped that a fuller understanding of perception may be obtained from studies of optical illusions, which result from misinterpretation of visual sensations, and from experiments where very different physical stimuli produce the same perception (Restak, 1979).

At a higher level cognitive psychologists attempt to understand the nature of human intelligence and how people think. Presently this field is dominated by the information processing approach, which analyses mental processes into a sequence of ordered stages with each stage reflecting an important step in the processing of cognitive information. The study of cognitive development has provided many interesting results and theories. Young children have particular difficulty with setting and remembering the subgoals involved in solving a problem (Anderson, 1985). With age, children appear to pass through a series of stages in which they develop better representations of the world and improved reasoning processes. Often cognitive development is modelled by the addition or modification of rules which enable the problem to be solved.

Mathematical models, such as neural networks (Chapter 7), are providing ways of studying learning, memory and representation (Hanson and Burr, 1990; Hinton, 1992), and are helping to direct experimental research.
CHAPTER 3

ENGINEERS IN BRAIN RESEARCH

"Progress in science depends on new techniques, new discoveries and new ideas, probably in that order."
Sydney Brenner, 1990

3.1 INTRODUCTION

Advances in brain research have often relied on the discovery of new tools and techniques. For example, study of the microstructure of the brain had to wait until the development of the optical microscope and then the electron microscope, while measurement of the electrical activity of neurons depended on the invention of accurate recording devices.

Over the years, the primary role of physical scientists in brain research has been to provide the necessary tools and techniques. With the advent of powerful computers and the development of new mathematical techniques, engineers have become more directly involved in brain research — measuring, analysing, and modelling brain activity and function.

The overall objective of this Chapter is to illustrate the wide range of expertise that engineers contribute to brain research. However, rather than being an exhaustive review, this Chapter briefly discusses a number of the contributions of engineering to brain research that are relevant to the work reported in this thesis. Thus, although there are no doubt may omissions and shortcomings, the diversity of roles that engineers play in brain research is demonstrated.

3.2 RECORDING TECHNIQUES

The electrical activity of the brain can be detected, amplified, displayed and stored through the use of appropriate instrumentation.

The membrane potential of an individual neuron can be detected using a microelectrode (∼0.05 μm in diameter, ∼1.0 μm in length) constructed from fine wire or from glass filled with an electrolyte. The microelectrode is inserted into the intracellular fluid and the potential measured with respect to an extracellular electrode. The signal is then amplified using a differential amplifier with a high input impedance (∼10^{10}Ω), a voltage gain of 50–10 000, a flat frequency response up to 10 kHz and a high common mode rejection ratio to reduce interference from other electrical sources (Bronzino, 1986).

Electrical potentials generated by the brain can be recorded by using depth electrodes implanted in the brain, electrodes in contact with the exposed cortex, needle electrodes inserted subcutaneously and even by surface electrodes secured to the scalp. Depth electrodes
consist of a bundle of insulated metal wires and are inserted into the brain to record the activity of subcortical neurons. Needle and surface electrodes are employed to measure the potentials generated by neurons in the cerebral cortex. The potentials recorded represent the activity of a large number of neurons.

Recordings of the electrical activity of the brain are known as *electroencephalograms* (EEGs) and are usually made with non-invasive surface electrodes (Chapter 4). Typically, these electrodes are disks of up to 10 mm in diameter made from or plated with gold or silver. A conductive gel is placed between the electrodes and the scalp to help secure them in place and to reduce contact impedance. From carefully applied electrodes, signals of up to several hundred microvolts can be obtained. These measured potentials are amplified by a factor of $\sim 10^6$ again using differential amplifiers and are recorded on FM analog tape, polygraph chart or by a computer data acquisition system. Techniques whereby a patient’s EEG and behaviour are simultaneously recorded were first proposed by Hunter and Jasper (1949) and have gradually become more sophisticated allowing long-term monitoring of EEGs. The development of portable recording devices (Koffler and Gotman, 1985; Sato *et al.*, 1976; Smith *et al.*, 1979) has enabled EEGs to be recorded while patients perform many of their normal daily activities outside the hospital environment.

The electrical impedance of the brain can also be recorded from electrodes attached to the scalp. Such a recording is known as a *rheoencephalogram* and can be used to assess cerebral blood flow (Geddes and Baker, 1989).

An effective stimulus delivered to a sense organ produces an electrical potential that is detectable on the cerebral cortex. However, the *evoked potential* (EP) appears against a background of spontaneous electrical activity (noise). In 1947 Dawson employed a signal averaging technique to record evoked potentials. This method assumes that the evoked response is time-locked to the stimulus while the background activity is assumed random. Evoked potentials can be detected in response to somatosensory stimulation (e.g., electrical shock to a peripheral nerve), visual stimulation (e.g., a flash or pattern reversal) or auditory stimulation (e.g., a click) by recording and averaging EEG activity over the appropriate area of the cortex. Averaging EEG activity over $N$ trials increases the signal to noise ratio by $\sqrt{N}$. The latency time from stimulation to the peak of the EP is, for a given modality, very consistent in normals and, therefore, can be used to report on the integrity of the sense organ and its neural pathways (Geddes and Baker, 1989).

Recent advances in sensor technology have allowed the magnetic fields of the brain to be routinely recorded. The *magnetoencephalogram* (MEG) is a recording of the extracranial magnetic fields produced by electric currents within the brain. These magnetic fields are very weak, being about one billionth the strength of the earth’s magnetic field, and have only been able to be measured through the development of superconducting quantum interference devices (SQUIDs) and gradiometers which discriminate against distant noise sources (Nunez, 1981). The theoretical advantage of the MEG is that the magnetic fields of the brain are not greatly affected by the medium through which they pass, whereas the scalp and skull tend to cause spatial averaging and smearing of electrical potentials (Rose *et al.*, 1987a).

### 3.3 STIMULATION TECHNIQUES

The brain can be stimulated by application of electric currents — *electrical stimulation* — or by varying magnetic fields which induce electric currents in the brain — *magnetic stimulation*.

Electrical stimulation of the human brain began in the last century when Roberts Bartholow had under his care a patient with a cranial defect that exposed the posterior portions
of her cerebral hemispheres. The development of neurosurgical techniques, early this century, enabled further studies of electrical stimulation in conscious patients. For example, during neurosurgery Penfield (1958) used electrical stimulation to study the localization of brain functions. Electrical brain stimulation is also used to evaluate the likely effects of neurosurgery. For example, patients with intractable epilepsy may have the region of their brain removed that is responsible for causing the seizures. In these cases, arrays of electrodes are implanted in the brain and electrical stimulation is used to examine the functional deficits that may result from neurosurgery and, thus, ensure that the cure will not be worse than the problem.

The cerebral cortex can also be stimulated by extracranial electrodes. Merton et al. (1982) reported that it was possible to stimulate the motor cortex using relatively large electrodes (∼4 cm²) and a short duration pulse (50 μs time constant) of a few hundred volts. They also found that the visual cortex could be stimulated but required a larger voltage and was consequently quite painful.

More recently, magnetic fields have been employed to stimulate the cortex. The current induced in the brain tissue is proportional to the rate of change of the magnetic field. In contrast to electrical stimulation, magnetic stimulation causes negligible patient discomfort and is not attenuated by passing through the skull (Barker et al., 1985). Hess et al. (1986) used magnetic stimulation to measure the central motor conduction time in patients with multiple sclerosis and concluded that the procedure could reveal the presence of subclinical lesions and confirm signs of upper-motor-neuron dysfunction.

One of the disadvantages of magnetic stimulation is that the site of stimulation is not well-defined. Consequently, Balducci et al. (1990) proposed the use of a combination of magnetic and electrical stimulation to reduce patient discomfort but maintain the ability to focus the current in the area of interest. Electrical stimulation produces a current perpendicular to the skin surface while magnetic stimulation induces a current parallel to the skin surface and, therefore, the two techniques tend to stimulate structures with different orientations (Barker et al., 1987).

3.4 SIGNAL PROCESSING TECHNIQUES

A number of signal processing techniques can aid in analysis and quantification of brain activity. For example, signal averaging is used to detect evoked potentials (§3.2). Having recorded the evoked potential, it is necessary to extract clinically useful information, which resides in the latencies and amplitudes of certain components of the waveform. A variety of techniques have been devised to identify certain components of the waveforms which then allow the latencies and amplitudes to be easily measured (Duffy et al., 1989).

EEG activity can be examined by both frequency and time series analysis (Chapter 5). The background EEG activity is typically quantified by calculating its frequency content using either amplitude-period analysis or the Fourier transform. Frequency information can be displayed as a compressed spectral array to show changes in frequency content over time. Alternatively, the signal power over certain frequency ranges (e.g., delta activity of 0.5–4 Hz) can be displayed as a topographic brain map, which indicates the distribution of activity over the scalp (Duffy et al., 1989). Frequency analysis can be used to detect lesions (Gotman et al., 1973; Gotman et al., 1975) and sleep disorders (Jansen, 1989; Principe et al., 1989), and to evaluate the effects of certain drugs (Künkel et al., 1976; Matějček and Devos, 1976).

Time series analysis of the EEG is usually concerned with the detection of individual waves with certain characteristics (e.g., sleep spindles, vertex waves, spikes, sharp waves). Tech-
niques such as matched filtering (Barlow, 1980), template matching (Fischer et al., 1980; Lloyd et al., 1972), and parameter calculation (Gotman and Gloor, 1976; Glover, Jr. et al., 1986) are employed to detect such waves. One such system for detecting waveforms characteristic of epilepsy is presented in Chapter 6. In another approach, the EEG is considered to be generated by a random process and short segments of EEG are assumed to be stationary. Autoregressive filtering techniques are employed to detect non-stationarities (Arakawa et al., 1986; Johnson et al., 1990; Lopes da Silva et al., 1975) and can even be used to segment the EEG (Praetorius et al., 1977).

The origins and significance of EEG and MEG signals are not well understood. Considerable attention has focussed on deducing current distributions within the brain from EEGs (Nunez, 1981) and MEGs (Rose et al., 1987b; Sarvas, 1987), in the hope of gaining a better understanding of the electrical events of the brain. Because magnetic fields are not significantly affected by the scalp and skull, it has been suggested that the MEG may be superior to the EEG for spatial localization (Modena et al., 1982; Rose et al., 1987a). However, the EEG and MEG tend to be related to different aspects of current flow (Cohen and Cuffin, 1987; Nunez, 1981) and, therefore, use of both techniques when attempting to determine the current source distribution should be advantageous (Cohen and Cuffin, 1987).

Recently the discovery of deterministic chaos (Chapter 9) — the phenomenon whereby deterministic systems generate apparently random activity — has enabled the EEG to be considered as the output of a deterministic system. At present, techniques developed for chaotic systems analysis are being applied to the EEG in an attempt to understand the nature of the underlying system (Chapter 10).

3.5 IMAGING THE BRAIN

The development of imaging techniques has enabled the living brain to be studied. The discovery of x-rays at the end of the nineteenth century provided the first method for viewing the brain. By replacing the cerebrospinal fluid (CSF) with air, in a technique known as pneumoencephalography, the shape and position of the ventricles can be observed. The blood vessels supplying the brain can also be imaged by cerebral angiography, involving the injection of an x-ray opaque dye into the bloodstream. However, these x-ray projection techniques do not take into account depth information and, hence, structures may overlap and become indistinguishable in the final image (Kety, 1979).

In the 1950s, Leksell developed echoencephalography which is a one-dimensional ultrasound imaging technique. It uses pulse-echo information to identify the location of midline structures, the displacement of which indicate the presence of a space-occupying lesion in one hemisphere (Geddes and Baker, 1989).

The development of computed tomography (CT) has enabled the internal structures of the brain to be imaged (Hounsfield, 1972). CT is the reconstruction of an image from measurements of emanations that have passed through the body (Bates et al., 1983). Although CT is a general term, it usually refers to x-ray transmission CT where one-dimensional x-ray projections are taken from many angles and then combined to produce a two-dimensional cross-sectional image. The development of CT has decreased the need for the dangerous and uncomfortable procedures that were previously required to image the brain. CT has become an important diagnostic tool, particularly for the localization of tumours (Axel et al., 1983).

More recently magnetic resonance imaging (MRI) techniques have emerged which produce higher quality images than x-ray CT. The patient is placed in a strong magnetic field which causes atomic nuclei of the tissues to align themselves in the direction of the magnetic field.
These nuclei can be made to resonate and change their alignment by irradiation with radio-frequency waves. When the radio-frequency pulse is turned off, the nuclei return to their original orientations and, in so doing, release energy at particular frequencies. An image of the tissue is constructed by computer analysis of the radio-frequency energy released (Duffy et al., 1989). MRI is diagnostically superior to CT; for example, MRI is able to image plaques found in the brains of multiple sclerosis patients (Young et al., 1987).

MRI and x-ray CT are used to produce anatomical images of the brain, which are used, for example, to plan neurosurgery (Peters et al., 1989). However, emission computed tomography (ECT) enables the physical state of the brain to be imaged (Jaszczak, 1988). In ECT, a radiopharmaceutical is administered to the patient and the distribution of this substance within the body determined from measurements of the radiation released. There are two types of ECT depending on the type of radiation emitted. Single photon emission computed tomography (SPECT) detects gamma photons while positron emission tomography (PET) detects pairs of coincident photons which are emitted during the annihilation of positrons. These techniques can be used to image cerebral blood flow by radioactively labelling red blood cells, or brain metabolism by administering radioactive deoxyglucose (Carlson, 1986).

Three-dimensional images of the brain can be displayed, by techniques such as volume and surface rendering, from a database consisting of two-dimensional cross-sectional images (Kennedy and Nelson, 1987; Robb and Barillot, 1989). Furthermore, anatomical information obtained from CT or MRI can be integrated with physiological data from SPECT or PET (Robb and Barillot, 1989).

3.6 QUANTIFICATION OF NEUROLOGIC FUNCTION

In many instances it is useful to quantify neurologic function, for example, to assess recovery following acute brain damage such as head injury and stroke, to study the effects of brain disorders such as Parkinson's disease, Alzheimer's disease and multiple sclerosis, to evaluate the efficacy of various treatment strategies (drugs, physiotherapy) or to provide information for understanding and modelling brain function.

A number of tests have been developed to measure specific dimensions of neurologic function. The major functional categories include mental state, vision, audition, tactile sensation, strength, range of motion, speed of movement, response time, co-ordination and steadiness (Kondraske et al., 1984; Kondraske et al., 1988).

Pursuit tracking tasks are often used to measure integrated function (Jones et al., 1986; Lynn et al., 1977), while a number of other tasks can be employed to break tracking into its various sensory, perceptual and motor component functions (e.g., visual resolution, visuospatial perception, reaction time, range of movement). These tests can be used to quantify sensory-motor function, to examine the effects of sex, age and handedness (Jones et al., 1986) and to study impairments due to stroke (Jones et al., 1987; Jones et al., 1989b) and Parkinson's disease (Jones and Donaldson, 1989).

3.7 MODELLING BRAIN FUNCTION

The brain can be modelled on a number of levels. At the lowest levels, sets of differential equations are employed to simulate the operation and interaction of neurons (Traub et al., 1988). On the next level, arrays of neuron-like elements are constructed to function in ways akin to certain subsystems of the brain (e.g., visual system), while on still higher levels
rule-based programs are developed to produce intelligent behaviour (e.g., playing checkers). To begin to comprehend the brain one needs an understanding of its function on each of these levels.

Based on results obtained from tracking tasks and their analysis by engineering control theories, several models of sensory-motor function have been proposed (Bösser, 1984; Lynn et al., 1977; Neilson et al., 1992). These models tend to be of two types: (1) those based on control theory which assume the existence of continuous negative feedback and (2) those that view tracking as a sequence of discrete motor programmes of varying length. Neilson et al. (1988) proposed adaptive model theory to bridge the gap between these two approaches. The basic contention of adaptive model theory is that the brain contains neural circuitry to compute, store and adaptively maintain the accuracy of inverse internal models describing the relationships between multiple sensory and motor signals. This theory has been used to gain insights into disorders of motor control such as cerebral palsy and stuttering (Neilson et al., 1992).

The field of artificial intelligence (AI) attempts to construct machines that behave intelligently and its emergence was closely related to the development of the information processing approach to cognitive psychology (§2.4.5). AI tries to capture intelligent behaviour without regard to the underlying mechanisms producing the behaviour. This usually involves manipulating symbols according to a sequence of formal rules, known as symbolic processing. AI techniques have proven to be effective in forming some kinds of expert systems (Chapter 6). Expert systems attempt to reproduce the abilities of a human expert in some field, and typically assume that the actions and knowledge of the expert can be embodied in a set of facts and an accompanying hierarchy of rules (Parsaye and Chignell, 1988). Fuzzy logic (Zadeh, 1989) and probability are used to provide expert systems with a way to handle imprecision and uncertainty (e.g., "The man is fairly tall", "If the man is tall, he might play basketball").

AI is also concerned with learning — teaching robots to perform certain tasks. Consequently, a variety of techniques for machine learning have been developed (Michalski et al., 1983). Following the theories of cognitive psychology, many machine learning strategies are based on rule generation and modification, with rules being formed based on new experiences, generalized when similar situations are encountered and altered when contradictions arise (Winston, 1984).

Neural networks (Chapter 7) provide an alternative approach to artificial intelligence based on modelling neurons, their interconnections and their ability to learn. Thus, neural networks are arrays of simple neuron-like elements which interact with each other through weighted connections. By adjusting the strengths of these connections, neural networks can be trained to perform specific tasks. Rather than forming explicit rules, neural networks acquire connection strengths which enable them to behave as though they know the rules. They exhibit rule-like behaviour by responding to regularities in the input patterns.

Neural networks are based on the present understanding of the structure and function of neurons. However, much of the operation of the brain is still not understood. Rather than waiting until knowledge of brain function is complete, engineers are developing more brain-like computers from neuron-like elements. These neural networks are being employed to solve real-world problems (Chapter 8) and, at the same time, are providing insights into the operation of the brain (§7.9).

Although the ultimate aim of neural networks is to produce artificial intelligence, the development of neural networks is usually considered as a separate and independent field. Whereas AI can be thought of as a 'top-down' approach to modelling the brain, neural
networks are considered to be a 'bottom-up' approach. Therefore, investigators in the field of neural networks attempt to solve similar problems to those in AI but using a different level of implementation.

Understanding the operation of the brain, particularly at higher levels, is still a long way off but collaboration between neuroscientists, engineers, psychologists and computer scientists has enabled theories of brain function to be formulated and tested. In the end, computer simulations are likely to provide an important tool and perspective in understanding the brain.
Part II

PROCESSING ELECTROENCEPHALOGRAMS

The electroencephalogram (EEG) is a recording of the electrical activity of the brain. It provides information pertinent to the diagnosis of a number of brain disorders and is of particular value in the investigation of epilepsy.

Between seizures the EEG of a person with epilepsy may be characterized by occasional epileptiform transients — spikes and sharp waves. Consequently, a routine clinical recording of 20–30 minutes can be useful in the diagnosis of epilepsy. An electroencephalographer (EEGer) detects epileptiform transients by visual inspection of the recording, which requires considerable skill and is time consuming. Automation of this process could save time, increase objectivity and uniformity, and enable quantification for research studies.

Despite numerous attempts to automate the detection of epileptiform transients in EEGs, very few methods have been introduced into the clinical setting and, even then, with only limited success. The central problem is the high number of false detections, which reflects the difficulty in distinguishing epileptiform transients from a wide variety of background activities and artifacts. It is now generally accepted that the only way to separate epileptiform transients from non-epileptiform waves is to make use of spatial and temporal contextual cues present in the EEG. This Part presents a system to detect epileptiform activity which makes considerable use of both spatial and temporal contextual information and has proven particularly successful at rejecting non-epileptiform waves in routine clinical recordings.

This Part begins by introducing the EEG — its recording and clinical application — in Chapter 4, before reviewing a number of techniques employed for automated EEG analysis in Chapter 5. The system for automated detection of epileptiform activity in the EEG is presented in Chapter 6.
CHAPTER 4

THE ELECTROENCEPHALOGRAM

So easy it seemed
Once found, which yet unfound most would have thought impossible.
John Milton (1608–1674)

4.1 INTRODUCTION

The electroencephalogram (EEG) is a recording of the electrical potentials generated by the brain. The potential difference between a pair of electrodes, typically placed on the scalp, is measured, and represents the electrical activity of many millions of neurons. Although the origin and significance of the EEG is not fully understood, correlations between clinical findings and EEG activity have led to useful diagnostic criteria. Today, EEG recording is a routine clinical procedure that provides information pertinent to the diagnosis of a number of brain disorders and is particularly helpful in the investigation of epilepsy.

This Chapter looks briefly at the discovery of the EEG before reviewing standard recording techniques and typical EEG activity. Finally, the clinical application of the EEG as a diagnostic tool is considered, with particular reference to the diagnosis of epilepsy.

4.2 DISCOVERY OF THE EEG

In 1875, Richard Caton observed electrical fluctuations when electrodes were placed in contact with the exposed cortex of rabbits and monkeys. Adolf Beck (1890) unknowingly repeated much of Caton’s work, recording oscillations in electrical potential from the occipital cortex of dogs. Similar experiments were performed by Vasili Danilevsky, who produced a thesis in 1876 but did not publish, and by Fleischl von Marxow, whose findings were finally published in 1890 after being sealed in an envelope seven years earlier. Due to the very crude measurement techniques employed, scientists had reservations about the significance of experimental findings, until they were confirmed using the Einthoven string galvonometer in the early 1900s.

However, it was not until the 1920s, when a German psychiatrist, Hans Berger, found electrical potentials could be measured through the intact skull, that attention turned to recording and studying human brain potentials. After years of careful research to establish the cerebral origin of the measured potentials, Berger published the first electroencephalogram (EEG) in 1929. Berger was the first to observe and accurately describe many features of the EEG and noted that it had different characteristics in neurological disorders such as epilepsy. Berger hoped that the EEG would provide a “window on the mind” which would be useful for treating psychiatric disorders.
Berger did not get the credit he deserved for laying the foundations of human electroencephalography until his work was replicated by Adrian and Matthews (1934). A number of studies were then performed in an attempt to correlate EEG activity with physiological states and clinical findings. Gibbs et al. (1935) carried out one such study, observing EEG activity during sleep and in patients with epilepsy. The results of these studies led to the EEG being used as a diagnostic tool which, in turn, necessitated the standardization of recording techniques.

4.3 EEG RECORDING

The electrical activity of the brain can be detected by electrodes either implanted in the brain (depth electrodes), inserted under the dural covering (subdural electrodes), placed on dural covering (epidural electrodes), inserted subcutaneously (needle electrodes) or secured to the scalp (surface electrodes). Most clinical EEGs are recorded using surface electrodes because of their non-invasive nature. Surface electrodes are typically metal disks of 4–10 mm in diameter which are held in place by a cap, or more securely by an adhesive such as collodion. Conductive gel is placed between the electrodes and the scalp to reduce contact impedance.

![Figure 4.1. Placement of 21 electrodes according to the 10–20 International System.](image)

The location of electrodes on the scalp was standardized by Jasper (1958) to the 10–20 International System (Fig. 4.1). This system affords adequate coverage of all parts of the head with electrode positions being designated in terms of the underlying brain regions (e.g., frontal, central, parietal, temporal, occipital). Electrode locations are specified in terms of measurements from standard landmarks on the skull — the inion, nasion and left and right preauricular points. The measurements are 10% or 20% of the total distance between the inion and the nasion, and between preauricular points. Thus, the placement of electrodes is proportional to skull size and shape (Appendix A).
The electrical activity differs in amplitude and frequency over different regions of the brain. Therefore, recordings need to be made simultaneously from several locations on the scalp. This allows features of the normal EEG to be adequately described and any abnormalities to be localized. An EEG channel measures the potential difference between a pair of electrodes. The particular arrangement whereby a number of channels are recorded simultaneously is known as a montage. There are two basic types of montage — referential and bipolar — and it is advantageous to use both during an EEG recording (Duffy et al., 1989).

On a referential montage, electrode potentials are measured with respect to a common reference which, ideally, is unaffected by cerebral electrical activity. The advantage of a referential montage is that valid comparisons can be made between amplitude measurements on different channels. However, the main problem is finding an ideal reference. Commonly chosen references — the ear lobes and the vertex — pick up a considerable amount of cerebral activity, while extracerebral references tend to be contaminated by the electrical activity of the heart (ECG) and muscles (EMG). The average potential over a number of channels is also used as a reference on the assumption that EEG signals are random and average to zero. Although the average potential can be affected by large potentials on one or more electrodes, recordings are often superior to those obtained using an ear lobe or vertex reference (Duffy et al., 1989).

On a bipolar montage, the potential difference between pairs of adjacent electrodes is measured. Electrodes are connected in a sequential manner to form electrode chains, which may run longitudinally, laterally or circumferentially. The advantage of bipolar montages is that electrical discharges can be easily localized; an electrical discharge originating at the common electrode of two adjacent channels shows the phenomenon of phase reversal.

Not all electrical potentials recorded on the scalp are generated by the brain. Extracerebral potentials are collectively termed artifacts and may arise from eye and tongue movements, muscle contraction, ECG, electrode movement or electrical equipment. Artifacts can be minimized through careful recording, amplification and filtering procedures.

The voltages detected on each EEG channel are very small, being a maximum of several hundred microvolts. Before these signals can be displayed on a standard output device, such as a chart recorder, they must be amplified by a factor of $10^4 - 10^6$, which is achieved in a number of stages. The first stage of amplification employs differential amplifiers with high common mode rejection ratios which reduce interference from other electrical sources. The amplified signals are then band-pass filtered between 0.5–70 Hz to eliminate much of the electrical activity produced by the skin and muscles (Duffy et al., 1989). The resulting signals are displayed by a multi-channel chart recorder which, by convention, produces an upward deflection on the chart in response to a negative change in potential.

### 4.4 EEG ACTIVITY

The EEG is a fluctuating signal produced by the brain. It derives from graded synaptic potentials (rather than action potentials) generated by neurons in the cerebral cortex (Gevins et al., 1975; Li and Jasper, 1953). These potentials are diffused and attenuated as they pass through the cerebrospinal fluid, skull and scalp. The signals recorded on the scalp typically range in amplitude from 10–200 $\mu$V and in frequency from 1–70 Hz (Duffy et al., 1989).

As first noted by Gibbs et al. (1935), the form of the EEG differs from person to person:

"The pattern may change from moment to moment, but its general appearance is characteristic of the individual subject."
Despite the individuality of one's EEG, all records have certain basic features in common, which can be described in terms of their frequency, amplitude, morphology, symmetry, synchrony, location and reactivity.

The *amplitude* of EEG activity depends on the age of the patient — EEGs of children are usually of much higher amplitude than those of adults — and on the recording montage — bipolar recordings are of lower amplitude than referential recordings, because of the shorter inter-electrode distance.

On-going EEG activity is referred to as the *background activity*. Although it is essentially a mixture of waves of multiple frequencies, under particular conditions waves of a certain frequency may dominate. Depending on frequency, rhythmic activity is classified as delta (< 4 Hz), theta (4–8 Hz), alpha (8–13 Hz) or beta (> 13 Hz) (Fig. 4.2).

![Figure 4.2. Four common EEG rhythms: (a) Delta < 4 Hz, (b) Theta 4–8 Hz, (c) Alpha 8–13 Hz, and (d) Beta > 13 Hz.](image)

The background activity varies in amplitude and frequency over different regions of the head, but is normally of similar amplitude and frequency over homologous areas of the two hemispheres — that is, *symmetrical*.

The type of background activity observed also depends on the *age* and *state* of the patient. The most common EEG activity, particularly over occipital regions, is the posterior dominant or *alpha* rhythm, which arises in awake individuals with eyes closed and under conditions of physical relaxation and mental inactivity. The frequency of the posterior dominant rhythm changes with age, increasing in children to reach the alpha range (8–13 Hz) by about 8 years and often decreasing again in old age. *Beta* activity is also common in the awake EEG but tends to be of low amplitude (<20 μV). When eyes are closed, beta activity may be masked by alpha activity in posterior regions and, therefore, is best observed with eyes open. In
adults, *theta* waves normally only accompany drowsiness, but are quite common in awake EEGs of children. *Delta* waves arise during deeper stages of sleep (stage III and IV) and tend to be of high amplitude.

Background rhythms may also change in response to a stimulus. This phenomenon is referred to as their *reactivity*. For example, alpha rhythms are suppressed or attenuated by attention, in particular by visual stimuli and mental effort.

In addition to background activity, a number of other patterns may be present in the EEG (e.g., paroxysmal activity). Such features may arise simultaneously on a number of channels, in which case they are referred to as *synchronous*, and may be *localized* to a specific region of the head, *lateralized* to a single hemisphere or *generalized*.

*Paroxysmal activity* is a term used to describe patterns that occur with sudden onset and termination and stand out distinctly from the on-going background activity. Paroxysmal activity includes bursts of activity of much higher amplitude than the background and *transients* superimposed on the background activity. Transients are isolated waves that stand out from the background activity, and tend to be either *epileptiform transients* (spikes and sharp waves, Fig. 4.3) or *sleep transients* (e.g., vertex waves, Fig. 4.4).

![Figure 4.3](image)

**Figure 4.3.** Epileptiform transients: (a) spike and (b) sharp wave.

Despite careful recording procedures, many artifacts still appear in the EEG (e.g., eye blink, eye movement, ECG, muscle, electrode movement). These artifacts tend to have distinctive appearance (Fig. 4.5), but can be confused with other EEG activity.
Figure 4.4. Vertex waves.

Figure 4.5. Typical appearance of artifacts in the EEG: (a) Eye blink, (b) Muscle, (c) ECG, and (d) Electrode artifact.
4.5 EEG INTERPRETATION

When Berger discovered the EEG, he hoped that it would provide a "window on the mind", but this has not yet eventuated:

"The study of the relation between the EEG and higher mental functions has been singularly disappointing though increasing sophistication in recording and analysing equipment, including computer analysis, may yet prove of value." (Walsh, 1978)

However, correlations between EEG activity and clinical findings have led to the EEG becoming a routine clinical procedure of considerable diagnostic value.

EEG interpretation essentially involves determining whether the tracing is within normal limits for the age and state of the patient. This task requires the detection of features and patterns characteristic of abnormal conditions and is performed by an electroencephalographer (EEGer). The EEGer employs complex and subjective pattern recognition techniques to detect EEG abnormalities. Although agreement among EEGers is relatively high on the overall classification of EEGs, there are often considerable differences on individual patterns (Barlow, 1979; Gevins et al., 1975).

EEG abnormalities tend to be of three types: (1) abnormal background activity, (2) abnormal sleep patterns and (3) the presence of epileptiform activity. Abnormalities may arise in the amplitude, frequency, symmetry, or reactivity of background activity. For example, in an adult, a posterior dominant rhythm of less than 8 Hz is abnormal and may be a sign of a metabolic, toxic or infectious encephalopathy, or of dementia. Alpha rhythms are normally confined to posterior regions; widespread or generalized alpha activity may be a drug effect or may be alpha coma, which is caused by brain stem lesions or cerebral anoxia. Lack of reactivity of alpha rhythms to eye opening is also abnormal and may again be due to alpha coma or, if lateralized to one hemisphere, to an occipital lobe lesion. The presence of a lesion is also indicated by hemispheric asymmetries in the amplitude or frequency of background activity and by slow wave or delta activity in the adult awake EEG.

Sleep has been divided into four stages (I–IV) according to the type of EEG activity present. The deeper stages of sleep (III and IV) are characterized by increasing amounts of delta activity and, hence, are termed slow-wave sleep. REM (rapid eye movement) or paradoxical sleep produces EEG activity similar to that of an awake subject. The first period of REM sleep normally arises approximately 90 minutes after sleep onset. Sleep patterns then alternate between slow-wave and REM sleep with each cycle lasting about 90 minutes. Abnormalities may arise in this sleep architecture; for example, the immediate onset of REM sleep, before stage I and II sleep, is characteristic of narcolepsy. Sleep transients tend to arise during stage II and III sleep and asymmetries in their amplitude and distribution may indicate the presence of a lesion.

The presence of epileptiform activity in the EEG supports a clinical diagnosis of epilepsy. Epileptiform activity includes epileptiform transients (spikes and sharp waves), which arise between seizures (i.e., in the inter-ictal period), and ictal or seizure patterns, which usually consist of "repetitive EEG discharges with relatively abrupt onset and termination and characteristic pattern of evolution, lasting at least several seconds" (Chatrian et al., 1974). Seizure patterns that are not accompanied by clinical manifestations are referred to as subclinical seizures.
4.6 EEG AND EPILEPSY

The EEG is of particular value in the investigation of epilepsy, which is essentially a sudden disruption of mental, motor, sensory or autonomic activity. Epilepsy is a disorder characterized by recurrent seizures which may take a variety of forms. Generalized seizures involve both cerebral hemispheres and cause a loss or impairment of consciousness. They tend to be of two types: (1) grand mal or tonic-clonic seizures which involve major convulsions of musculature and (2) petit mal or absence seizures which cause an abrupt loss of mental functions for 5–15 seconds. Unilateral seizures are confined to one hemisphere and manifest themselves through clinical phenomena on the opposite (contralateral) side of the body. Partial or focal seizures arise from a localized region of the brain but may spread and produce generalized seizures. Partial seizures are classified as simple if they produce sensory, motor or autonomic symptoms and as complex if they involve a disruption of higher cortical functions. For example, complex partial seizures, also termed temporal lobe seizures, may produce hallucinations or sensory illusions (Walsh, 1978).

Epileptic seizures arise from the anomalous behaviour of a large number of neurons. The causes of this abnormal neural activity are not well understood, but include head injury and chemical imbalances. Recording the EEG during a seizure is particularly helpful in determining whether a patient has epilepsy. Absence seizures are characterized by very large amplitude, approximately sinusoidal waves of about 3 Hz with a sharp negative spike near the positive crest of the wave (Fig. 4.6). Tonic-clonic seizures cause a gradual alteration in EEG activity 15–20 seconds before any clinical signs develop. During this time the predominant waves become faster and of higher amplitude. Following the seizure, the EEG is abnormally flat but resumes its usual characteristics in about 20 minutes (Gibbs et al., 1935). Partial or focal seizures are characterized by a high frequency discharge over the focus, followed by a slower rhythmic discharge which may take the form of spikes or sharp waves, followed in turn by post-ictal slowing (Duffy et al., 1989). However, seizures tend to occur infrequently and unpredictably so obtaining a recording during a seizure might require an EEG extending over several days (i.e., long-term EEG). Such recordings can be made using seizure monitoring techniques, where the patient's EEG and behaviour are simultaneously displayed on a video screen, and stored whenever the patient presses an alarm button or the technician observes abnormal behaviour (e.g., seizure) or an automated detection system (§5.3.1) signals a seizure.

Between seizures the EEG of a person with epilepsy is “essentially normal, except that scattered here and there are brief groups of waves of pattern similar to those seen at the beginning of a seizure” (Gibbs et al., 1935). An inter-ictal EEG may be characterized by occasional epileptiform transients (spikes and sharp waves) and, consequently, relatively short recordings can still be useful in the diagnosis of epilepsy.

In the routine clinical procedure, an EEG of 20–30 minutes is obtained, which includes recordings on a number of bipolar and referential montages during periods of eyes open and eyes closed. Activation procedures, such as hyperventilation and photic stimulation, are often employed to bring out latent abnormalities and may be effective in eliciting epileptic discharges. The EEGer visually inspects the recording for epileptiform transients, the presence of which supports a diagnosis of epilepsy, but does not establish the diagnosis beyond doubt. Although epileptiform transients are the signature of epilepsy, a patient displaying these features may not have clinical epilepsy because clinical manifestations (i.e., seizures) depend on the location and spread of abnormal neural activity to other parts of the brain (Mac Gillivray, 1977). Similarly, a normal EEG does not exclude the possibility of epilepsy, because epileptiform transients arise intermittently and the routine EEG recording is just a short sample of the brain's continuous electrical activity.
4.7 EPILEPTIFORM TRANSIENTS

Epileptiform transients cannot be precisely defined and require years of accumulated experience to recognize reliably (Davey et al., 1989). Detection of epileptiform transients is highly subjective and, consequently, there are often considerable differences among EEGers (Barlow, 1979; Gevins et al., 1975).

Although definitions of epileptiform transients are subjective and largely acquired by experience, a few objective criteria have been proposed. Epileptiform transients are defined to be isolated waves that are clearly distinguished from background activity with pointed peaks at conventional paper speeds (i.e., 30 mm/s). They are often classified as spikes or
sharp waves, depending on their duration — spikes have durations of 20–70 ms, while sharp waves have durations of 70–200 ms (Chatrian et al., 1974). However, there is no evidence that spikes and sharp waves have different pathophysiological origins or different clinical significance (Gotman, 1985) and, hence, spikes and sharp waves are often jointly referred to as spikes, SSWs or epileptiform transients. Throughout this thesis no distinction is made between spikes and sharp waves and, therefore, they are referred to collectively as epileptiform transients.

Epileptiform transients are usually surface negative, that is, due to a negative change in potential at the surface of the brain. Surface negative spikes and sharp waves are of considerably greater epileptogenic significance than surface positive ones. The polarity of EEG waves is either positive (conventionally downward-going) or negative (upward-going). On a referential montage, a negative change in potential at an electrode produces a negative wave in the tracing, while on bipolar montages both positive and negative waves are recorded (Fig. 4.7).

Epileptiform transients arise with a wide variety of morphologies (Fig. 4.8). They tend to be either monophasic or biphasic and may be accompanied by a slow wave. The term polyspike is used when several spikes comprise a single EEG pattern.

Figure 4.7. Polarity of EEG transients: (a) negative spikes produced by a negative change in potential on a referential montage and (b) positive and negative spikes resulting from a negative potential change in a bipolar montage.
Figure 4.8. Epileptiform transients arise with a wide variety of morphologies: (a) monophasic spike, (b) biphasic spike and (c) polyspike.

Figure 4.9. Decreasing amplitude of spikes with distance from focus.

Epileptiform transients often arise simultaneously on several EEG channels, in which case they are collectively termed an epileptiform event (Davey et al., 1989). Epileptiform events can be classified as focal or non-focal in reference to the spatial distribution of constituent epileptiform transients.
Focal events arise from a localized centre of negativity known as the focus. Epileptiform transients decrease in amplitude and possibly in sharpness with distance from the focus (Fig. 4.9). On bipolar montages, focal events are usually characterized by phase reversal on adjacent channels (Fig. 4.10a), although there are several other possibilities. For example, the phase reversal may be separated by a null channel (Fig. 4.10b); no phase reversal is observed when the focus is at the beginning (Fig. 4.10c) or the end of an electrode chain (Fig. 4.10d).

In non-focal events, epileptiform transients tend to arise on several channels with similar amplitude, sharpness and polarity (Fig. 4.11).

4.8 SUMMARY

The EEG is a recording of the electrical activity of the brain which provides information pertinent to the diagnosis of a number of brain disorders, in particular, epilepsy. Between seizures the EEG of a person with epilepsy may be characterized by occasional epileptiform transients and, therefore, relatively short recordings can be useful in the diagnosis of epilepsy.

A routine clinical recording typically lasts 20–30 minutes and includes recordings on a number of bipolar and referential montages with both eyes open and eyes closed. Often recordings are also made during periods of hyperventilation and photic stimulation which are activation procedures that may elicit epileptic discharges.

An EEGer visually inspects the recording for epileptiform events, which are classified as focal or non-focal depending on the spatial relationships between constituent epileptiform transients. Because the detection of epileptiform activity is a highly subjective process, there
are often significant differences between EEGers, particularly on individual patterns.
CHAPTER 5

ANALYSIS OF THE ELECTROENCEPHALOGRAM

"With experience one can learn to detect certain gross features which have significant clinical and physiological correlates, but in this there is more art than science."

Grass and Gibbs (1938)

5.1 INTRODUCTION

EEG interpretation involves the detection of patterns and features characteristic of abnormal conditions. It is usually performed by a highly trained electroencephalographer (EEGer), who reduces the tracing to a few interpretive statements. This process requires considerable skill and is time consuming. Hence, automated EEG analysis could save time, increase uniformity and objectivity, and enable quantification for research studies.

Despite its widespread use, the EEG is one of the last routine clinical procedures to be fully automated (Gevins, 1984). This may be due to the complexity of the signal, the lack of basic knowledge about its origin and significance, and the subjective nature of EEG interpretation.

Many quantitative approaches have been developed to assist in EEG interpretation, but progress towards fully automated EEG analysis has been slow. This Chapter reviews current analysis techniques with particular emphasis on the detection of epileptiform transients.

5.2 ANALYSIS OF BACKGROUND ACTIVITY

The background activity of the EEG is most commonly analysed in terms of its spectral properties. Frequency information can be extracted from the signal in a number of ways including the Fourier transform, period analysis and autoregressive modelling.

In the earliest analyses of EEGs, power spectra were obtained by means of the Fourier transform. In 1938, Grass and Gibbs reported that:

"After having made [Fourier] transforms of 300 electroencephalograms, we are convinced that the system not only expresses data in a manner more useful and concise than is possible by present methods, but that in many cases it indicates important changes in the electroencephalogram which would otherwise remain hidden."

With the development of spectral analysers and, later, the Fast Fourier Transform (FFT)
and digital computers, application of frequency analysis to the EEG has become common
place.

The amplitude and frequency of EEG activity changes with time and varies over different
regions of the scalp. By plotting the power spectra of successive short segments of EEG
activity as a compressed spectral array, the temporal distribution of EEG activity can be
examined (Bickford et al., 1972; Chiappa et al., 1976; Etevenon et al., 1976).

Alternatively, the spatial distribution of EEG activity over the scalp can be displayed
as a topographic map. In contrast to the compressed spectral array, which displays changes
in frequency content of the signal with time, topographic mapping assumes EEG activity
is stationary. A consistent estimate of the frequency content of the signal is obtained by
averaging spectra obtained from a number of consecutive EEG segments. The resulting
EEG spectrum is usually divided into four frequency bands corresponding to the four major
rhythms of the EEG — delta (<4 Hz), theta (4–8 Hz), alpha (8–13 Hz) and beta (>13 Hz) —
and the average level of activity in each band computed. The spatial distribution of activity
in each band can then be displayed as a series of topographic maps.

Gotman et al. (1973) employed topographic mapping to localize EEG abnormalities. Localized
slow waves are the most common indicator of a destructive brain lesion and, conse-
quently, a topographic map of the delta frequency band and, to a lesser extent, the theta
band may be useful in detecting lesions. The EEG from areas nearest the lesion is not only
characterized by the presence of slow waves but also by the absence of normal background
activity. Hence, displaying the ratio of slow-wave activity (i.e., delta and theta) to normal
background activity (i.e., alpha and beta) can provide improved localization of lesions.

Period analysis provides a simplified approach to EEG spectral estimation. The EEG
is segmented into individual waves, on the basis of the zero crossings of the signal and
its derivatives, and the amplitude and period of each wave measured (Leader et al., 1967;
Salb, 1980; Schenk, 1976). Statistical information about the signal, such as the probability
density functions of wave amplitudes and periods, can then be derived (Goldberg et al., 1973).

An alternative approach to the determination of power spectra is autoregressive modelling
which, as with other spectral analysis techniques, assumes EEG activity is stationary. Each
EEG sample is expressed as a weighted sum of the previous sample values plus an error term.
The weights are determined from the autocorrelation function of an EEG segment and can
be used to estimate the power spectrum. For the same sample length, the autoregressive
model tends to provide better estimates of the true EEG power spectrum than does the FFT
(Barlow, 1979). Praetorius et al. (1977) used autoregressive modelling to partition the EEG
into quasi-stationary segments by sensing significant changes in the model parameters and
characterized each segment by its power spectrum.

Considerable data reduction can be achieved by summarizing the power spectrum of the
background EEG by a few parameters such as the mean frequency, the bandwidth and the
signal power (Wennberg and Zetterberg, 1971).

Quantification of background activity is used in conjunction with either statistical analysis
or artificial intelligence techniques to examine the effects of age (Matějček and Devos, 1976),
drugs (Künkel et al., 1976; Matějček and Devos, 1976) and mental activity (Koles and Flor-
Henry, 1981) on the EEG, to categorize background activity (Jagannathan et al., 1982; Leader
et al., 1967; Serafini, 1973; Walter et al., 1967) and sleep stages (Jansen, 1989; Jestico
et al., 1977; Principe et al., 1989), to distinguish between normal and abnormal background EEG
activity (Friberg et al., 1976; Gotman et al., 1973; Jansen, 1989), and to reveal differences
between certain patient groups, for example, between dyslexic and normal children (Sklar et
al., 1973).
5.3 DETECTION OF EPILEPTIFORM ACTIVITY

Over half the EEG referrals relate to epilepsy, with the EEG being the most useful procedure in its diagnosis. The EEG of a patient with epilepsy may be characterized by epileptiform activity, which includes both seizure patterns and epileptiform transients.

5.3.1 Seizure Patterns

Recording the EEG during a seizure is particularly helpful in determining whether a patient has epilepsy. Because seizures often occur infrequently and unpredictably, obtaining a recording during a seizure may require an EEG extending over several days (i.e., long-term monitoring). To identify seizures in a long-term EEG, it is necessary either to inspect the entire recording visually which is extremely time consuming, or to review the sections during which seizures are known to have occurred. Although the latter approach is more time efficient, it relies on the patient being aware of the seizure or on an observer noticing the seizure. Automatic detection of seizures enables the EEGer to review a small proportion of the entire record and does not rely on the patient or an observer being aware of the seizure (Gotman, 1985). Thus, automatic seizure detection is employed as a data reduction process. Although emphasis tends to be on detecting a high proportion of seizures rather than on eliminating false detections, the number of false detections must be small enough so that significant data reduction is achieved.

Seizures are not primarily electrographic patterns but rather behavioural events. Therefore, an automatic detection system can only be expected to detect those seizures with clear EEG manifestations. Fortunately, many seizures are characterized by distinctive electrographic patterns. For example, absence seizures are associated with spike-and-wave activity of approximately 3 Hz and a number of techniques have been developed to detect this characteristic pattern. Bursts of spike-and-wave can be detected based on the presence of slow waves alone. Jestico et al. (1977) band-pass filtered the EEG between 2.2–4 Hz, making a detection when the filtered signal exceeded a threshold. Smith et al. (1979) also relied solely on the presence of slow waves but required a minimum number of such waves to occur within a one second epoch. Most other such detection systems require slow waves to be accompanied by spikes. Vries et al. (1981) passed the EEG through two parallel band-pass filters, one filtering the signal between 1–4 Hz and the other between 15–30 Hz. The filtered signals were thresholded to recognize slow waves and spikes respectively and a seizure was detected when a spike and a slow wave occurred within 0.3 seconds. Quy et al. (1980) also employed two parallel band-pass filters to recognize spikes and slow waves but required a minimum number of spike-and-wave patterns before making a detection. Similarly, Principe and Smith (1982) required spike-and-wave patterns to repeat at a frequency of 1.8–4.5 Hz. Whisler et al. (1982) detected slow waves by filtering and thresholding the EEG but recognized spikes on the basis of their amplitude, duration and slope. Carrie and Frost, Jr (1977) calculated the second derivative of the EEG and compared it with a running average to detect spikes and recognized slow waves by calculating their duration and comparing their amplitude with that of the preceding waves. Koffler and Gotman (1985) relied more heavily on the presence and amplitude of spikes, only requiring an accompanying slow wave when the spikes were of moderate amplitude and, thus, enabling the detection of irregular spike-and-wave activity.

Although many seizures do not display a stereotypical EEG pattern, most include some rhythmic discharge of large amplitude (Gotman, 1985). Because the pattern to be recognized is poorly defined, few detection systems have been developed. The most notable seizure detection system was published by Gotman (1982). The EEG is divided into two second
epochs and, for each epoch, a number of parameters calculated — the average halfwave amplitude, average frequency and its coefficient of variation (i.e., mean/standard deviation). A detection is made when the average frequency lies in the range 3–20 Hz, its coefficient of variation is less than 60% and the average halfwave amplitude is at least three times the background amplitude. The background amplitude is computed as the average halfwave amplitude over 16 seconds, ending 12 seconds prior to the current epoch. A number of modifications to this system have improved its performance (Gotman, 1990). The background amplitude is now calculated 20 seconds prior to the current epoch to avoid problems detecting seizures with slow onset. To reduce false detections due to short bursts of artifacts, the average halfwave amplitude of 8 seconds following the current epoch must be at least 1.6 times the background. A detection is also made if the amplitude of the current epoch is greater than the background and its average duration is less than 1/3 that of the background. This additional detection criterion enables seizures consisting of rapid low amplitude activity to be detected.

5.3.2 Epileptiform Transients

Automated detection of epileptiform transients has two primary areas of clinical application. The first is in long-term EEG monitoring where it acts essentially as a data reduction process (Gotman, 1979; Gotman, 1981). A segment of EEG is recorded only when a transient is detected and all segments are reviewed by an EEGer. Consequently, the goal is to detect a high proportion of epileptiform transients while minimizing false detections. The second area is in routine clinical recordings where a major objective is to minimize the visual inspection process. In this case, it is important not to precipitate a misdiagnosis of epilepsy and, therefore, the aim is to eliminate false detections while detecting a satisfactory proportion of epileptiform transients.

Despite numerous attempts to automate the detection of epileptiform transients, very few systems have been introduced into routine clinical use and, even then, only with limited success. Due to the high number of false detections, these systems are restricted to long-term monitoring with all detections being reviewed by an EEGer.

5.3.2.1 Mimetic Methods

The most commonly employed approach to the detection of epileptiform transients is to mimic the pattern recognition techniques believed to be used by EEGers. The EEG is divided into individual waves consisting of two contiguous halfwaves, and those waves that fit the definitions of epileptiform transients identified. The major difficulty with this approach is the lack of precise and objective definitions of epileptiform transients.

The simplest mimetic techniques examine a single variable of the signal. For example, Walter et al. (1973) monitored the second derivative or sharpness of the EEG making a detection when it exceeded a threshold, while Smith (1974) examined the first derivative or slope of the EEG, requiring two slopes of sufficient amplitude and of opposite polarity within a brief time interval.

Carrie (1972) noted that “the content of the abnormal part of an EEG is not definable in absolute terms but is exceptional in relation to the characteristics of the background signal” and, therefore, wave parameters need to be normalized with respect to the background activity. For example, Carrie (1972) compared parameters of individual waves with average measurements of the preceding 128 waves. In a similar approach, Gevins et al. (1975) measured the curvature at the apex of each wave and compared it with the average curvature.
calculated over 4 seconds at the beginning of the recording, while Carrie (1976) compared the sharpness of each wave with the average sharpness of the preceding 256 waves.

In order to distinguish epileptiform transients from non-epileptiform waves, many groups have found it necessary to calculate more than one parameter of waves. Saltzberg et al. (1967) recognized that a detection scheme based on sharpness alone may give rise to an excessive number of false detections, and that slope and amplitude may also be important distinguishing features. Ma et al. (1977) calculated the amplitude and duration of waves and found considerable overlap between clusters of epileptiform transients and non-epileptiform waves. Guedes de Oliveira et al. (1983) determined the amplitude, slope and sharpness of waves relative to a running average of the background as well as the wave duration. Using discriminant analysis, they found that total wave duration, relative slope and relative sharpness were the best parameters for distinguishing between epileptiform and non-epileptiform activity. Gotman and Gloor (1976) calculated the sharpness and amplitude of waves with respect to a measure of the background amplitude, making a detection when the wave parameters exceeded a set of thresholds. The sharpness threshold depended on the wave amplitude, while the amplitude threshold, in turn, was a function of the wave duration. Frost, Jr (1979) computed the sharpness of waves relative to the background activity and when this exceeded a threshold also measured the halfwave amplitudes and the wave duration. Ktonas and Smith (1974) proposed a set of six parameters to describe the morphology of epileptiform transients — the slope at the inflection point of each halfwave, the duration from inflection point to peak for each halfwave, duration between inflection points and the total wave duration. In addition to these six parameters, Glover, Jr et al. (1986) also calculated the amplitude to baseline, halfwave amplitudes and sharpness at the peak.

Detection systems based on mimetic techniques often also include a number of artifact rejection procedures. For example, the system developed by Gotman and Gloor (1976) rejects waves (a) as muscle if there are a minimum number of high amplitude waves within 1/3 second and (b) as eye blink if they arise on frontal channels with positive polarity and a duration of at least 150 ms, and there are similar waves on the homologous contralateral channels.

5.3.2.2 Template Matching

A particular waveform can be identified by template matching or matched filtering. A well-defined pattern or template is chosen and a detection made when the cross-correlation of the template with an EEG segment exceeds a predetermined threshold. However, detections cease as soon as the EEG differs slightly from the template (Gotman, 1985) and, therefore, difficulties arise in choosing an appropriate template. Pola and Romagnoli (1979) recognized this problem and used template matching to study the temporal distribution of epileptiform transients of a particular morphology, while Saltzberg et al. (1971) employed template matching to detect activity in the scalp EEG which corresponded to subcortical spikes. In order to enhance the performance of template matching techniques, Lloyd et al. (1972) included a second stage in the detection process. After passing the EEG through a matched filter, they calculated its second derivative, making a detection when a threshold was exceeded.

5.3.2.3 Parametric Methods

Parametric methods assume that the background EEG activity is generated by a stationary random process and represent its statistical properties by a small number of parameters. These parameters, in conjunction with previous EEG samples, are used to predict the next sample value. When the difference between the predicted and actual EEG values exceeds a
threshold, a non-stationarity is detected. Thus, epileptiform transients are considered to be an unexpected or statistically improbable occurrence.

Lopes da Silva et al. (1975) employed an autoregressive model to detect unexpected features in the EEG, in particular, epileptiform transients. Arakawa et al. (1986) used autoregressive modelling and a nonlinear function to separate non-stationary waves from the background activity, while Johnson et al. (1990) used an autoregressive EEG model followed by an adaptive binomial decision rule to detect epileptic discharges. Barlow (1980) developed an inverse filter based on the frequency spectrum of the EEG to detect non-stationarities.

Because these parametric techniques detect a large number of non-stationarities which are not epileptiform transients, a number of groups have added a second detection stage to improve their performance. After autoregressive filtering the EEG, Birkemeier et al. (1978) calculated and thresholded its second derivative, while Pfurtscheller and Fischer (1978) and Fischer et al. (1980) followed the autoregressive filter with template matching.

5.3.2.4 Syntactic Methods

Syntactic systems classify a pattern based on a structural combination of features and, in this way, are able to take into account the immediate temporal context of a waveform. Walters et al. (1989) converted the EEG signal into a symbolic representation based on the slope and duration of halfwaves. A set of 9 symbols was used to describe each wave and its context, and a detection was made when this symbolic string matched a predetermined syntax.

5.3.2.5 Expert Systems

An expert system attempts to replicate the knowledge and reasoning of an expert, in this case, an EEGer. Expert systems have been employed to enhance the performance of mimetic approaches and are essentially rule-based systems which incorporate spatial and temporal contextual information into the detection process.

The distinction between mimetic techniques and expert systems is not clear, but probably begins with the system developed by Glover, Jr et al. (1986), which uses information about activity arising simultaneously on other channels both to detect epileptiform transients and to reject artifacts. For example, all waves that occur synchronously with an artifactual burst are rejected, while candidate epileptiform transients must be accompanied by synchronous epileptiform transients on other channels. Increasing emphasis on spatial and temporal contextual cues present in the EEG has continued to decrease the number of false detection of these systems (Glover, Jr. et al., 1989).

5.3.2.6 Neural Networks

Neural networks (Chapter 7) are adaptive, parallel processing systems inspired by the architecture of the brain. Because neural networks are able to learn from experience, they can be trained to perform certain tasks. In particular, back-propagation neural networks (§7.7) are able to learn arbitrary associations between input and output patterns and, therefore, can be trained to recognize certain patterns. Eberhart et al. (1989a) have trained back-propagation neural networks to detect epileptiform transients (§8.5).
5.4 SUMMARY

EEG interpretation is a subjective and time consuming task which is performed by an EEGer. Automation of this process could save time, increase uniformity and objectivity, and enable quantification for research studies.

Although many attempts have been made to automate the detection of epileptiform transients in the EEG, these have met with limited success. The main problem is the high number of false detections, which reflects the difficulty in distinguishing epileptiform transients of various morphologies from a wide range of background activities and artifacts. Consequently, the systems developed to date are restricted to long-term EEG monitoring, where they act as data reduction systems and all detections are reviewed by an EEGer.
CHAPTER 6

A SYSTEM TO DETECT EPILEPTIFORM ACTIVITY

"The classification of an EEG waveform as normal or abnormal is probably one of the most complex problems encountered in computerized analysis of the EEG".

Ktonas and Smith (1974)

6.1 INTRODUCTION

To date, the systems developed to automatically detect epileptiform activity in EEGs are restricted to long-term monitoring where they act as data reduction systems. A segment of EEG is recorded when a detection is made and all such segments are reviewed by an EEGer. Due to the high number of false detections, these systems cannot perform satisfactorily in the routine clinical setting, where it is important not to precipitate a misdiagnosis of epilepsy.

It is becoming generally accepted that the only way to separate epileptiform from non-epileptiform waves is to make use of a wide spatial and temporal context (Glover, Jr. et al., 1989; Gotman and Wang, 1991). Several groups are implementing this approach in an effort to minimize false detections. Glover, Jr. et al. (1989) have developed a system that relies on a wide spatial context with 12 EEG channels being analysed together with additional contextual information provided by EKG, EOG, and EMG channels. Conversely, the system developed by Gotman and Wang (1991) implements a wide temporal context, where sections of EEG are classified into one of five states (active wakefulness, quiet wakefulness, desynchronized EEG, phasic EEG or slow wave EEG) before state dependent rules are applied to reject non-epileptiform activity.

This Chapter describes a system that makes considerable use of both spatial and temporal contextual information. This system has proven particularly successful at rejecting non-epileptiform activity in routine clinical EEG recordings (Davey et al., 1989; Jones et al., 1989a; Dingle et al., 1990; Dingle et al., 1992; Jones et al., 1992). It uses a mimetic approach to detect candidate transients which are subsequently confirmed or rejected as epileptiform by an expert system. The expert system integrates both spatial and temporal contextual information to detect epileptiform activity and to reject non-epileptiform waves. Preliminary results indicate that this system should be capable of performing reliably in the routine clinical EEG setting.

6.2 SYSTEM OBJECTIVES AND PHILOSOPHY

The system was designed to detect epileptiform activity specifically in routine clinical EEG recordings. In order for such a system to be of real value in assisting the EEGer, it must have no false detections. Otherwise, the EEGer must inspect all activity reported to prevent
a misdiagnosis of epilepsy, which could well take longer than reviewing the entire record independently. Therefore, it is essential to eliminate false detections while, at the same time, maintaining reasonable detection rates.

The approach taken was to replicate, as far as possible, the procedures employed by the EEGer. An expert system provides a means for implementing such an approach.

6.2.1 Expert Systems

An expert is someone who has exceptional knowledge or skill in a limited field, in this case an EEGer. An expert system provides a way of automating the skills of a human expert, by assuming that experts apply their knowledge in a highly rational manner. Thus, an expert system reasons and arrives at conclusions based on the knowledge it possesses.

An expert system consists of a knowledge base, a rule base and an inference engine. The knowledge base contains the information that allows an expert to make decisions. In order for the system to reason and derive new knowledge, knowledge and data must be combined with rules which can be expressed in the general form:

\[
\text{IF } \text{premise} \quad \text{THEN } \text{conclusion}
\]

For example,

\[
\text{IF } \text{Wave duration} \leq 40 \text{ ms} \\
\text{AND } \text{Background frequency} \geq 25 \text{ Hz} \\
\text{THEN } \text{Wave is due to muscle activity.}
\]

The inference engine attempts to establish the truth or falsity of a hypothesis or goal. An example of a goal may be:

*The EEG contains epileptiform activity.*

At each step in the reasoning process, the inference engine determines which rule should be applied to the available knowledge and data. There are two major methods of controlling inference — forward chaining and backward chaining. In forward chaining, the system starts with the available facts and begins to reason and deduce new facts in the hope that it will eventually deduce the goal. While in backward chaining, the system starts with the goal and tries to prove a series of subgoals working backwards from the goal. The optimal inference strategy depends on the problem, but forward chaining works well when the system has to interpret a set of incoming data (Parsaye and Chignell, 1988) and, therefore, is the strategy employed for the detection of epileptiform activity.

6.2.2 The Human Expert

In order to design an expert system to detect epileptiform activity in EEGs, extensive collaboration with the human expert — the EEGer — was required.

When reading an EEG, the EEGer marks epileptiform events rather than individual epileptiform transients, and classifies these as focal or non-focal depending on the spatial distribution of constituent epileptiform transients (§4.7).
In distinguishing epileptiform events from artifacts and background activity, the EEGer first searches for waveforms fitting the definition of epileptiform transients, that is, isolated waves with pointed peaks, which are clearly distinguished from background activity.

The EEGer then considers the *spatial context* of these waves, that is, their location on the scalp and the presence of artifacts or epileptiform transients on neighbouring channels. Epileptiform transients always occupy a definable electric field on the scalp and, therefore, should be seen in at least two electrode sites (Maulsby, 1971). Thus, the presence of epileptiform transients arising synchronously on neighbouring channels suggests that the wave under consideration is also epileptiform, while the presence of artifacts on neighbouring channels suggests that it is not epileptiform.

The EEGer also makes use of *temporal information*, that is, the presence and distribution of abnormal activity throughout the EEG. When suspicious activity is observed, the EEGer carefully reviews the electrodes involved searching for similar activity arising at other times in the EEG. The presence of other epileptiform events with the same spatial distribution suggests that the suspicious activity is in fact epileptiform.

Thus, in detecting epileptiform activity, the EEGer not only relies on the sharpness, amplitude and duration of individual waves but also on their spatial and temporal context.

### 6.2.3 System Overview

The task of detecting epileptiform events has been implemented in three stages (Fig. 6.1):

1. **Data collection** — 16 channel EEGs are filtered, sampled and digitized.
2. **Feature extraction** — a mimetic approach is used to detect candidate epileptiform transients.
3. **Event detection** — an expert system uses spatial and temporal contextual information to detect and classify epileptiform events.

![Figure 6.1](image)

*Figure 6.1*. Block diagram of system for detecting epileptiform activity. EEG is amplified and filtered before being sampled at 200 Hz and digitized to 12 bits. The feature extraction stage detects candidate epileptiform transients using a mimetic approach, while the expert system detects focal and non-focal epileptiform events.

Because of the difficulty in reliably distinguishing epileptiform events from all background activities and artifacts, and because of sometimes conflicting clinical requirements or priorities, the system produces two types of output. The first is *definite* epileptiform events. For these it is essential that all artifacts and background activity are rejected and, therefore, the subsequent detection rate of epileptiform events may not be particularly high. The second
type of output is probable events. The aim here is to detect a higher proportion of epileptiform events but this is likely to be at the expense of several false detections.

This approach should prove satisfactory for most situations. When a data reduction process is required (e.g., in long-term EEG monitoring), probable events can be used to trigger data storage. In the routine EEG setting, it is intended that the EEGer need only look at the probable event results if there are few or no definite events detected.

6.3 DATA COLLECTION

Sixteen channels of EEG are recorded from a number of bipolar and referential montages. The four bipolar montages used are shown in Fig. 6.2 and are referred to as longitudinal, transverse, longitudinal-transverse, and circumferential. Recordings are also made on three referential montages which use either the ear lobes, the vertex or the average EEG potential as a reference.

Figure 6.2. The four bipolar montages used. For each channel, the solid line indicates the positive electrode while the dotted line indicates the negative electrode.
Recordings are made while the patient is awake but resting and include periods of eyes open, eyes closed, hyperventilation and photic stimulation. One of three recording protocols is employed depending on the age of the patient:

1. Adults — recordings of up to 5 minutes are made on each of the four bipolar and three referential montages and include both periods of eyes open and eyes closed. Recordings are then made during periods of hyperventilation and photic stimulation on the longitudinal montage.

2. Children up to 12 years old — recordings are made as for adults except that the ear electrodes $a1$ and $a2$ are not used. Consequently, only 14 channels of EEG are recorded on the transverse montage and the ear reference montage is not employed.

3. Babies — recordings are only made on the longitudinal montage and include photic stimulation.

The EEG is amplified by an EEG machine (Siemens Minograph Universal) before being high-pass filtered above 0.5 Hz (1-pole) by the EEG machine and low-pass filtered to 70 Hz using a 5-pole analog Butterworth filter. The band-pass filtered EEG is then sampled at 200 Hz and digitized to 12 bits. Presently, EEG data are stored for off-line processing.

6.4 FEATURE EXTRACTION

The first stage of analysis is feature extraction, the purpose of which is to detect candidate epileptiform transients (i.e., isolated waves with pointed peak clearly distinguished from background activity). Feature extraction is essentially a data reduction process that extracts pertinent information for use by the expert system. Therefore, it is necessary to detect a high proportion of epileptiform transients and provide information about their context, without detecting an unnecessarily large number of non-epileptiform waves.

A mimetic approach (§5.3.2.1) was adopted in an attempt to replicate the pattern recognition techniques employed by the EEGer. The EEG is divided into halfwaves by a simple peak detection algorithm, with a wave comprising two contiguous halfwaves. Parameters of each wave and its constituent halfwaves are calculated and thresholded. Because epileptiform transients are defined to be clearly distinguished from background activity, parameters of individual waves are compared with measures of the background activity. Waves whose parameters exceed all thresholds are put forward as candidate epileptiform transients. The following sections detail each step in this process.

6.4.1 Scaling EEG Amplitude

The overall amplitude of EEGs varies considerably between individuals and, in particular, with the age of the patient. For example, EEGs of infants are normally of much higher amplitude than those of adults. The EEGer relies on the recording technician to select an appropriate gain setting, which produces a similar deflection on the chart recorder for all EEGs. However, this process is subjective and also very coarse because of the limited number of gain settings available.

A means for objective scaling of EEGs has been implemented by estimating the global EEG amplitude. After automatic elimination of EMG and electrode artifacts, the average halfwave amplitude over 16 channels and over 60 seconds at the beginning of the recording
is calculated. The resulting value is used to scale the EEG halfwave amplitude to 8.4 μV (which was the average value calculated for 10 EEGs).

EMG and electrode artifacts are detected when the average halfwave amplitude over one second on any channel is more than 2.5 times the 60 second average on that channel. EEG data on all channels within one second of such a detection are eliminated. Having removed artifactual sections of EEG, the average halfwave amplitude is recalculated over all channels.

### 6.4.2 Peak Detection

Having scaled the overall EEG amplitude, a simple peak detection algorithm is used to divide the EEG into halfwaves. The system alternately searches for halfwaves of opposite polarity and with a minimum amplitude of 4.2 μV (see Algorithm 6.1).

```plaintext
Algorithm 6.1. Peak detection algorithm.
```

$$
\text{LastDifference} = \text{eeg[ThisSample]} - \text{eeg[ThisSample-1]}
$$

$$
\text{ThisDifference} = \text{eeg[ThisSample+1]} - \text{eeg[ThisSample]}
$$

$$
\text{NextDifference} = \text{eeg[ThisSample+2]} - \text{eeg[ThisSample+1]}
$$

IF $\text{LastDifference} \times \text{ThisDifference} < 0$ OR $\text{ThisDifference} = 0$

THEN (* Found peak *)

$$
\text{HalfwaveAmplitude} = \text{eeg[ThisSample]} - \text{eeg[LastHalfwave]}
$$

IF $\text{ABS( HalfwaveAmplitude )} > 4.2 \mu\text{V AND HalfwaveAmplitude} \times \text{LastAmplitude} < 0$

THEN (* Opposite polarity → Found new halfwave *)

Increment $\text{NumHalfwaves}$

$$
\text{Halfwave[NumHalfwaves]} = \text{ThisSample}
$$

$$
\text{LastHalfwave} = \text{ThisSample}
$$

$$
\text{LastAmplitude} = \text{HalfwaveAmplitude}
$$

ELSE

IF $\text{HalfwaveAmplitude} \times \text{LastAmplitude} > 0$

THEN (* Same polarity → Replace last halfwave *)

$$
\text{Halfwave[NumHalfwaves]} = \text{ThisSample}
$$

$$
\text{LastHalfwave} = \text{ThisSample}
$$

$$
\text{LastAmplitude} = \text{HalfwaveAmplitude}
$$

### 6.4.3 Parameters

Parameters of each wave and its constituent halfwaves are then calculated, thresholded and compared with measures of the background activity. It is generally accepted that the important parameters for distinguishing between epileptiform transients and non-epileptiform waves are duration, amplitude and sharpness (Gotman, 1980; Ktonas and Smith, 1974; Saltzberg et al., 1967). The following definitions of these parameters are used (Fig. 6.3):

1. **Duration** of a wave is the sum of the durations of its constituent halfwaves. The duration of each halfwave is calculated from the peak to the point where the slope of the halfwave (calculated over 2 samples, 10 ms) changes rapidly (i.e., changes direction, or more than a 60% drop in slope). This duration measurement ensures muscle spikes (even those superimposed on slow waves) have short halfwave durations.
2. Amplitude of a wave is the difference between the peak and the floating mean, which is the average EEG value over 75 ms centred on the peak. The amplitude measure is, therefore, dependent on wave duration.

3. Sharpness of a wave is the sum of the magnitudes of the peak halfwave slopes. The peak slope of each halfwave is (a) the peak to peak slope when the halfwave duration is less than 20 ms or (b) obtained by a least squares estimation based on 4 samples (excluding peak).
Epileptiform transients are defined to be clearly distinguished from background activity. Therefore, parameters of each wave need to be compared with those of the background activity. The following measures of the background activity are calculated:

1. **Background amplitude** is the average difference between the EEG and the floating mean. This measure of background amplitude means that slow activity (e.g., delta waves, slow waves of spike-and-wave activity) do not contribute significantly to the background activity.

2. **Background duration** is the average peak-to-peak duration of the halfwaves.

3. **Background rhythmicity** is defined by two parameters, the duration rhythmicity which is the coefficient of variation (standard deviation/mean) of halfwave durations and the amplitude rhythmicity which is the coefficient of variation of halfwave amplitudes.

Measures of the background activity are calculated over one second centred on the wave under consideration. A one second background is sufficiently long that an epileptiform transient cannot dominate the background measure but is sufficiently short that bursts of alpha waves or muscle activity have a substantial effect. Therefore, waves making up such a burst do not appear significantly larger than the background.

Parameters of each wave are compared with those of the background activity by dividing the amplitude and sharpness of the wave by the background amplitude. The resulting parameters are termed the **relative amplitude** and **relative sharpness** respectively. Because these are the most important parameters for the detection of epileptiform transients, accurate scaling of EEGs (§6.4.1) is not critical.

Optimal definitions for parameters of both the individual waves (amplitude, sharpness and duration) and the background activity (amplitude, duration and rhythmicity) were determined by statistical analysis (Appendix B).

**6.4.4 Thresholds**

Epileptiform transients do not usually occur in isolation but arise synchronously on several channels. To detect a high proportion of synchronous transients a two threshold system is employed. When a wave is detected whose parameters exceed the thresholds, waves on all channels within 50 ms are reconsidered using a lower set of thresholds. This ensures that most epileptiform transients are detected, while smaller non-epileptiform waves throughout the EEG are rejected. The two sets of thresholds used are given in Table 6.1. Ideally, the higher set of thresholds allows at least one epileptiform transient from every event to be detected, while the lower set of thresholds enables all transients constituting the event to be detected.

**6.4.5 Contextual Information**

In order to replicate the reasoning of the EEGer, the expert system requires information about the spatial context of candidate transients. The spatial context of a wave includes knowledge of activity on adjacent channels. Therefore, whenever a group of candidate epileptiform transients is detected, measures of the background activity (amplitude, duration and rhythmicity) on all channels are recorded. This information can be interpreted spatially by the expert system if the recording montage is known. Provided that the recording protocol is known, the montage can be determined by detecting the times when the montage was
Thresholds for parameters given that the global EEG amplitude has been scaled to 8.4 μV. When a wave exceeds the thresholds for Screen Mode, all waves within 50 ms are reconsidered using the Review Mode thresholds.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(Units)</th>
<th>Screen Mode</th>
<th>Review Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halfwave Duration</td>
<td>(ms)</td>
<td>10 150</td>
<td>10 150</td>
</tr>
<tr>
<td>Total Duration</td>
<td>(ms)</td>
<td>30 250</td>
<td>30 250</td>
</tr>
<tr>
<td>Amplitude</td>
<td>(µV)</td>
<td>16.8 –</td>
<td>10.1 –</td>
</tr>
<tr>
<td>Amplitude/Background</td>
<td></td>
<td>4.0 –</td>
<td>2.8 –</td>
</tr>
<tr>
<td>Sharpness</td>
<td>(µV/ms)</td>
<td>2.52 –</td>
<td>1.68 –</td>
</tr>
<tr>
<td>Sharpness/Background</td>
<td></td>
<td>2.5 –</td>
<td>2.0 –</td>
</tr>
</tbody>
</table>

Table 6.1.

changed (i.e., when no EEG is present). This is achieved by monitoring the background amplitude and noting when it drops below a threshold of 2.5 µV on all 16 channels. Waves arising within 0.5 seconds of a montage change appear disproportionately large because the background amplitude includes a period when no EEG was present. Therefore, there is a refractory period of 0.5 seconds on either side of a montage change during which no detections can be made.

Artifacts, such as eye blinks and electrode movement, are characterized by substantial prolonged deviations of the EEG from baseline. Therefore, whenever the offset of the floating mean from baseline exceeds a threshold of 50 µV, this fact is reported to the expert system.

6.5 EVENT DETECTION

The expert system is the final stage in the detection of epileptiform activity and rejection of artifacts and epileptiform-like background waves. The expert system is written in Prolog (a declarative artificial intelligence language) and attempts to replicate the knowledge and reasoning of the EEGer.

When reading an EEG, the EEGer tends to mark epileptiform events (rather than the individual epileptiform transients) and classifies these as focal (localized) or non-focal (generalized). In distinguishing epileptiform activity from artifacts and background activity, the EEGer makes use not only of the amplitude and sharpness of waves but also of contextual information. Thus, for automated detection of epileptiform activity, the parameters of the individual waves must be used in conjunction with (a) spatial information (i.e., location of electrodes on the scalp, channel derivations, presence of transients or artifacts on adjacent channels) and (b) temporal information (i.e., the presence and distribution of abnormal activity during the EEG). When EEGers observe suspicious activity, they carefully review the electrodes involved searching for similar activity. This approach can be implemented by detecting all likely epileptiform events (based purely on spatial information) and grading them as definite, probable or possible. Probable and possible events can then be upgraded or rejected based on the presence and distribution of events elsewhere in the recording.
6.5.1 Knowledge and Data

In general, an expert system applies rules to facts and data to infer new facts and arrive at conclusions. The expert system receives data from the feature extraction stage, which takes the form of candidate epileptiform transients and their parameters, the background measures on all channels and the times when the offset of the floating mean from baseline exceeded 50 $\mu$V. Candidate epileptiform transients are represented as :

\[
\text{candidate}(\text{montage, sample, channel, } D1, D2, D3, \text{ polarity, sharpness, relative sharpness, amplitude, relative amplitude})
\]

where the first three arguments (montage, sample, channel) identify the wave and the remaining arguments are the parameters of the wave illustrated in Fig. 6.3. The background activity on each channel is represented as :

\[
\text{background}(\text{montage, sample, channel, background amplitude, background duration, duration rhythmicity, amplitude rhythmicity, offset})
\]

while the times when the offset exceeds 50 $\mu$V are described by :

\[
\text{movement}(\text{montage, channel, start sample, end sample, maximum offset})
\]

The permanent knowledge of the expert system is embodied in a number of facts which detail the location of electrodes and channel derivations for each montage. At present only four bipolar montages (Fig. 6.2) are included in the analysis because of multichannel artifacts which arise on referential montages when an artifact occurs at the reference electrode. The channel derivations are represented as :

\[
\text{channel}(\text{montage, channel, electrode, electrode})
\]

For example, the fact :

\[
\text{channel}(\text{longitudinal, 1, fp2, f4})
\]

indicates that on the longitudinal montage channel one measures the voltage of $fp2$ with respect to $f4$.

Further knowledge is derived through application of rules to the data and the facts. These rules are divided into two groups : (1) those that rely on the spatial context of waves and (2) those that rely on their temporal context.

6.5.2 Use of Spatial Context

Knowledge of electrode locations and channel derivations allows the expert system to interpret information spatially. Measures of the background activity on individual channels can, therefore, be combined to describe the distribution of EEG activity over the scalp.

Spatial contextual information is employed to reject artifacts and background activity, and to detect epileptiform events and classify them as definite, probable or possible. As described in Chapter 4, an epileptiform event comprises one or more synchronous epileptiform transients. There are two types of epileptiform event — focal and non-focal. Focal or localized
events arise from a centre of negativity (i.e., a focus) at the surface of the brain. On the other hand, constituent transients of non-focal events do not have any fixed polarity relationships, although negative waves tend to hold greater significance, especially on referential montages.

All epileptiform transients arising within a window of 125 ms are considered to be related to a single epileptiform event. Therefore, a maximum of one event every 125 ms can be reported. This is to avoid multiple detections due to polyspikes and bipolar transients. Waves occurring within a window of 125 ms are grouped together and each group of waves is searched first for artifacts and then for epileptiform events.

Groups of waves are eliminated if they are due to muscle contraction, eye blinks and electrode movement. A series of rules, which take into account channel location and background activity on adjacent channels, eliminates most of these artifacts. These rules take the following form:

1. **Bursts of muscle activity** are detected when the background activity on any channel is of high frequency ($\geq 25$ Hz) and large amplitude (background amplitude $\geq 12.6 \ \mu V$).

2. **Eye blinks** are detected when the floating mean drops significantly below baseline ($\geq 80 \ \mu V$) on at least two frontal channels.

3. **Electrode movement** is detected when the offset of the floating mean from baseline exceeds 50 $\mu V$ for more than 400 ms and reaches a maximum offset of at least 100 $\mu V$.

These rules successfully eliminate waves due to overt movement, sustained muscle activity, electrode movement and eye blinks.

![Figure 6.4](image-url)  
**Figure 6.4.** Four possibilities for focal events: (a) phase reversal on adjacent channels (focus at $f_4$), (b) phase reversal separated by a null channel (focus between $o_1$ and $t_5$), (c) focus at beginning of an electrode chain ($f_8$) and (d) focus at end of an electrode chain ($t_5$).

Remaining groups of waves must satisfy a number of criteria in order to be put forward as epileptiform events. For a focal event to be detected a distinct focus must be found and all
waves in the group must support its presence. A focus is defined by two waves arising within
20 ms of each other (Fig. 6.4) which (a) are on adjacent channels with opposite polarity or
(b) have opposite polarity but are separated by a null channel or (c) are both negative and
at the beginning of an electrode chain or (d) are both positive and at the end of an electrode
chain. The phase reversal must be such that the waves arise from a centre of negativity and
not from a centre of positivity. Waves which appear to be due to a positive surface potential
are invariably electrode artifacts. Two synchronous foci can be detected if they arise over
opposite hemispheres or at non-adjacent electrodes.

A non-focal event is detected when there are at least two epileptiform transients arising
within 40 ms of each other. No fixed polarity relationships between waves are necessary but
one negative epileptiform transient is required.

Although events usually require at least two synchronous transients, a single epileptiform
transient can be detected if it is particularly large and sharp. Single negative waves are
detected both as focal and non-focal events, while single positive waves are detected only as
focal. Allowing a single wave to constitute an event enables the detection of (a) non-focal
events that consist of only one epileptiform transient and (b) focal events where one of the
waves defining the focus has escaped detection by the feature extractor (either because it is
of very low amplitude, is insufficiently sharp, or occurs during a burst of muscle activity).

Figure 6.5. Examples of non-focal events from one EEG in which grading on spatial grounds led to (a) a
definite followed by a possible event and (b) a probable event.

Focal and non-focal events are graded as definite, probable or possible corresponding to
their level of certainty. The grading of events is determined from the amplitude and sharpness
of constituent epileptiform transients. For focal events the grading is based on the two waves
defining the focus, while for non-focal events it is based on the two largest waves. An example of the grading of non-focal events is shown in Fig. 6.5. Classification of focal events is less stringent than for non-focal events because of the well-defined polarity relationships which must exist between constituent waves. Events consisting of a single wave are classified more stringently and can only be possible.

A major problem has been distinguishing between isolated spikes due to muscle activity and epileptiform transients. Such muscle spikes tend to be characterized by short halfwave durations or a high frequency background. However, many epileptiform transients share these characteristics. To overcome this problem, the grading of events as definite, probable or possible is based on waves which do not have characteristics of muscle spikes. Therefore, events consisting of transients with similarities to muscle spikes may be rejected or have a lower level of certainty. Similarly, waves which appear to be part of the background activity are not included in the grading of events. For example, a wave is considered to be part of an alpha rhythm if its duration and the background duration fall within the alpha frequency range (8–13 Hz) and there are rhythmical waves of alpha frequency on occipital channels.

The classification of events is followed by a further artifact rejection stage. Electrode artifacts may give rise to large waves of opposite polarity on adjacent channels (Fig. 6.6a), which may initially be detected as definite focal events. Events of this type are rejected on the basis that definite focal events should manifest themselves on more than two channels (Fig. 6.6b).

![Figure 6.6](image)

Figure 6.6. An example of artifact rejection in which (a) an electrode artifact was initially detected as a definite focal event (focus at t5) but was rejected due to lack of supporting evidence on adjacent channels, whereas (b) a definite focal event (focus between o1-t5) also manifested itself on nearby channels (p4-p3 and p3-c3).

Based purely on spatial contextual information, it is not always clear whether a group of waves constitutes a focal or non-focal event (the reasons for this become more evident in §6.5.3). Each group is searched for both focal and non-focal events and, hence, a group of waves may be detected as both focal and non-focal. More than one focus may arise during
an EEG. The expert system, in addition to being able to detect all foci occurring separately, can discriminate two synchronous foci but may be unable to determine whether they are related or independent (based purely on spatial information). These classification difficulties are resolved by consideration of temporal contextual information.

6.5.3 Use of Temporal Context

Temporal contextual information is defined here to be the presence of abnormal patterns occurring with the same spatial distribution during the EEG. This information is used to upgrade or reject probable and possible epileptiform events, and to resolve any classification problems encountered.

Probable and possible events are upgraded if there is temporal support for their existence, otherwise they are rejected. Therefore, in the final output there are only two categories of event — definite and probable. Focal events are upgraded if there is at least one definite or two probable focal events with the same focus. Non-focal events, and any remaining focal events, are then upgraded if they arise with a similar spatial distribution as definite non-focal events. For example, the probable event and possible event of Fig. 6.5 arise on the same channels as the definite event and so are upgraded to definite and probable events respectively.

In some EEGs there are only one or two epileptiform events and, therefore, there is no temporal support for their existence. An additional rule ensures this activity is not rejected — if no events have been detected then all probable events are kept as probables (rather than being rejected).

![Figure 6.7](image)

Figure 6.7. An example of resolving classification difficulties during temporal analysis (a) a definite event detected as focal (focus at f8) and non-focal (transverse montage), (b) another event in the same EEG on the longitudinal montage which is focal (focus at f8). Thus (a) is also classified as focal (f8).

When a group of waves has been classified as both focal and non-focal, temporal information is used to determine the most likely event type. For example, the definite event of Fig. 6.7a was initially detected as both focal (focus at f8) and non-focal because all transients are negative and arise at the beginning of an electrode chain (transverse montage). The temporal analysis is able to resolve this difficulty because events detected on the longitudinal montage exhibit phase reversal at f8 (e.g., Fig. 6.7b). Similarly, the events of Fig. 6.5 were initially classified as both non-focal and focal (focus at fp1), but as there is no phase reversal at this electrode at other times in the EEG these events are classified as non-focal. In cases where correct classification is still unclear, focal events take precedence unless all constituent transients are negative.

If two synchronous foci are detected, temporal information is used to determine whether
they are independent. For example, foci may be detected at \( f_3 \) and \( f_4 \) on the longitudinal montage. These may be independent or may be due to a single focus at \( f_2 \) (i.e., between \( f_3 \) and \( f_4 \)). By considering similar events detected on the transverse and longitudinal-transverse montages, which include \( f_2 \), or by determining whether the foci also arise independently, it is possible for the system to decide the number of foci present.

6.6 SYSTEM PERFORMANCE

A system to detect epileptiform activity in EEGs should be both sensitive (i.e., detect epileptiform activity) and selective (i.e., reject non-epileptiform activity). A measure of system reliability can be obtained by calculating the percentage of epileptiform activity detected (sensitivity) and the percentage of the total detections that are epileptiform (selectivity). However, a better measure of system reliability may be its sensitivity and the number of false detections per unit time, as this is appropriate for EEGs both with and without epileptiform activity.

To date, system performance has been evaluated only on data used for development (i.e., the evaluation and training data are the same). Although this does not provide a true measure of performance, the results are nevertheless impressive.

Data from 21 patients aged 4–64 years (mean 29 years) have been analysed, totalling 380 minutes of 16 channel bipolar EEGs. The data covered a wide range of EEGs — 4 normal, 8 with only focal epileptiform activity and 9 with predominantly non-focal epileptiform activity. A variety of background activities also occurred (e.g., alpha, delta) and most EEGs contained significant amounts of artifact (e.g., eye blinks, electrode movement and muscle), particularly during periods of hyperventilation and photic stimulation. All bipolar EEG data available were used and no EEGs (or segments of EEGs) were rejected because of excessive artifact or 'difficult' background activities (e.g., sharp alpha activity).

Results obtained from the system were compared with those of a single EEGer (GJC). Throughout the development process the EEGer was consulted regarding events detected by the system. Thirteen events not initially detected by the EEGer were subsequently confirmed as being unequivocally epileptiform. These events tended to arise on page boundaries or during sections of EEG that contained considerable artifact on other channels. Thus, although the EEGer was considered to have 100% selectivity and 100% sensitivity (i.e., was the 'gold standard'), this was in part facilitated by his being able to reconsider his 'independent' scoring on viewing the system's results. Events detected by the system that the EEGer agreed were related to other epileptiform activity in the EEG, but would not normally report, have been termed questionable events.

Results obtained from the system are detailed in Table 6.2. On average, the system detected 53% of epileptiform events as definite with no false detections (i.e., 100% selectivity) and 64% as either definite or probable but at the expense of 3.5 false detections per hour. In all cases the system was able to distinguish normal from epileptiform EEGs and was able to localize the focus where one existed. Only in EEG #12 did the system not detect any epileptiform events as definite. Because there was only a single epileptiform event, there was no temporal support for its existence and, therefore, rather than being upgraded it was kept as a probable event.

The highest overall detection rates were achieved on EEGs containing little epileptiform activity. When classifying an EEG as normal or epileptiform the number of events detected is only of importance when there are very few of them, but it is not so critical when the EEG contains a large amount of epileptiform activity. Therefore, it is of little concern that the
<table>
<thead>
<tr>
<th>EEG (Type)</th>
<th>Duration (mins)</th>
<th>EEGER Events</th>
<th>Events Detected</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Definite</td>
<td>Probable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True (%) Quest False (/hr)</td>
<td>True (%) Quest False (/hr)</td>
</tr>
<tr>
<td>1 (N)</td>
<td>20</td>
<td>0</td>
<td>0 (-) 0 0 (0)</td>
<td>0 (-) 0 0 (0)</td>
</tr>
<tr>
<td>2 (N)</td>
<td>20</td>
<td>0</td>
<td>0 (-) 0 0 (0)</td>
<td>0 (-) 0 0 (0)</td>
</tr>
<tr>
<td>3 (N)</td>
<td>20</td>
<td>0</td>
<td>0 (-) 0 0 (0)</td>
<td>0 (-) 0 0 (0)</td>
</tr>
<tr>
<td>4 (N)</td>
<td>20</td>
<td>0</td>
<td>0 (-) 0 0 (0)</td>
<td>0 (-) 0 0 (0)</td>
</tr>
<tr>
<td>5 (F)</td>
<td>20</td>
<td>8</td>
<td>3 (37) 0 0 (0)</td>
<td>4 (50) 1 0 (0)</td>
</tr>
<tr>
<td>6 (F)</td>
<td>20</td>
<td>3</td>
<td>2 (67) 0 0 (0)</td>
<td>1 (33) 0 0 (0)</td>
</tr>
<tr>
<td>7 (F)</td>
<td>20</td>
<td>5</td>
<td>3 (60) 0 0 (0)</td>
<td>1 (20) 1 0 (0)</td>
</tr>
<tr>
<td>8 (F)</td>
<td>15</td>
<td>267</td>
<td>120 (45) 4 0 (0)</td>
<td>30 (11) 3 0 (0)</td>
</tr>
<tr>
<td>9 (F)</td>
<td>20</td>
<td>2</td>
<td>1 (50) 0 0 (0)</td>
<td>0 (0) 0 0 (0)</td>
</tr>
<tr>
<td>10 (F)</td>
<td>20</td>
<td>2</td>
<td>2 (100) 0 0 (0)</td>
<td>0 (0) 0 1 (3)</td>
</tr>
<tr>
<td>11 (F)</td>
<td>20</td>
<td>9</td>
<td>2 (22) 0 0 (0)</td>
<td>2 (22) 0 1 (3)</td>
</tr>
<tr>
<td>12 (F)</td>
<td>20</td>
<td>1</td>
<td>0 (0) 0 0 (0)</td>
<td>1 (100) 0 0 (0)</td>
</tr>
<tr>
<td>13 (NF)</td>
<td>15</td>
<td>70</td>
<td>42 (60) 1 0 (0)</td>
<td>4 (6) 0 0 (0)</td>
</tr>
<tr>
<td>14 (NF)</td>
<td>5</td>
<td>62</td>
<td>32 (52) 0 0 (0)</td>
<td>5 (8) 0 0 (0)</td>
</tr>
<tr>
<td>15 (NF)</td>
<td>5</td>
<td>37</td>
<td>14 (38) 0 0 (0)</td>
<td>8 (22) 0 0 (0)</td>
</tr>
<tr>
<td>16 (NF)</td>
<td>20</td>
<td>14</td>
<td>7 (50) 0 0 (0)</td>
<td>4 (29) 0 1 (3)</td>
</tr>
<tr>
<td>17 (NF)</td>
<td>20</td>
<td>49</td>
<td>27 (55) 0 0 (0)</td>
<td>2 (4) 0 2 (6)</td>
</tr>
<tr>
<td>18 (NF)</td>
<td>20</td>
<td>21</td>
<td>11 (52) 0 0 (0)</td>
<td>2 (10) 0 4 (12)</td>
</tr>
<tr>
<td>19 (NF)</td>
<td>20</td>
<td>57</td>
<td>44 (77) 3 0 (0)</td>
<td>4 (7) 0 3 (9)</td>
</tr>
<tr>
<td>20 (NF)</td>
<td>20</td>
<td>23</td>
<td>9 (39) 8 0 (0)</td>
<td>5 (22) 14 5 (15)</td>
</tr>
<tr>
<td>21 (NF)</td>
<td>20</td>
<td>41</td>
<td>37 (90) 0 0 (0)</td>
<td>0 (0) 0 5 (15)</td>
</tr>
<tr>
<td>Totals</td>
<td>380</td>
<td>671</td>
<td>356 (53) 16 0 (0)</td>
<td>73 (11) 19 22 (3.5)</td>
</tr>
</tbody>
</table>

Table 6.2. System performance on 21 EEGs: 4 normal (N), 8 with focal epileptiform events (F) and 9 with predominantly non-focal epileptiform events (NF).
total detection rate is only 56% for EEG #8 in which there were over 250 focal events. Of more concern is that only 44% of events were detected for EEG #11 which contained only 9 events.

![EEG waveforms](image)

**Figure 6.8.** A generalized burst of spike-and-wave activity detected as 5 events.

In addition to detecting isolated epileptiform events, the system is able to detect bursts of epileptiform activity. Because slow wave activity does not contribute significantly to the background amplitude measure, bursts of spike-and-wave activity are detected as several events (Fig. 6.8).

### 6.6.1 False Detections

No events were reported for any of the normal EEGs. Thus, false detections only occurred in EEGs containing epileptiform activity and then only in the probable category. All false detections had the same spatial distribution as epileptiform events but were generally too low in amplitude or insufficiently sharp to be considered epileptiform by the EEGer.
6.6.2 Missed Detections

Missed detections fall into two categories: (1) those completely missed and (2) those falsely rejected as artifact or background activity. This system is unable to detect epileptiform events that arise within 0.5 seconds of a montage change or during sustained muscle activity and reports a maximum of one event every 125 ms. The feature extraction stage may not detect epileptiform transients if they are of very low amplitude, are insufficiently sharp or have muscle spikes superimposed on them. The expert system tends to miss events if only one epileptiform transient has been detected by the feature extractor; this is because thresholds for the detection of a single transient are very high. In addition to missing events completely, the expert system may also reject true epileptiform events if all constituent transients have characteristics of muscle spikes or appear to be part of the background activity. The expert system also rejects non-definite events if they do not arise with the same spatial distribution as other detected events.

6.6.3 Comparison with other systems

Comparisons between systems for detecting epileptiform activity are made difficult by the wide range of ‘gold standards’ used for evaluating their performance. In particular, different definitions are used for both false detections and missed detections. For example, false detections can be defined as detections made by the system that were (a) obviously artifact (Gotman et al., 1978) or (b) not marked by any of n EEGers (Eberhart et al., 1989b) or (c) marked by fewer than m of n EEGers (Fischer et al., 1980). Similarly, missed detections may be (a) those not detected by the system but marked by at least m of n EEGers (Eberhart et al., 1989b) or (b) those falsely rejected by the system as non-epileptiform (Gotman et al., 1978). Despite these difficulties, an effort has been made to compare the performance of this system with that of others.

Most systems for the detection of epileptiform activity are still in the developmental stage and, consequently, studies of their performance tend to be limited. The most extensive system evaluation has been performed by Gotman et al. (1978) who recorded 2–3 mins of EEG from 110 patients while they were relaxed with eyes closed. An attempt was made to obtain recordings as artifact-free as possible. During a total of 255 mins of 16 channel bipolar EEG an average of 7 false detections per hour were reported. No account was made of the number of transients entirely missed by the system, but 16% of events were rejected as non-epileptiform. This system, which adopts a mimetic approach, has since been modified for long-term EEG monitoring and is now in regular clinical use as a data reduction system. It is only used during typical hours of sleep (11 p.m. to 5 a.m.) in an attempt to reduce movement, muscle and eye blink artifacts. However, Gotman (1979) reported its performance to be “highly variable” and stated that “the variety of morphologies of artifacts appeared to preclude a total automatic elimination”. Recent modifications (Gotman and Wang, 1991), which implement a wide temporal context, have considerably reduced the number of false detections (by up to 90%) with a minimal further loss of true epileptiform transients (<5%). However, they still expect approximately 15 false detections per hour during periods of wakefulness (due mainly to eye blink artifacts).

Glover, Jr. et al. (1990) also used a mimetic stage followed by a rule-based expert system. Twelve channels of bipolar or referential EEG were analysed and additional information was available from EMG, EOG and EKG channels. The recordings analysed included periods of sleep and results were reported for 3 patients. For development data 56% of events were detected with an average of 15 false detections per hour, while for evaluation data from the same 3 patients 21–57% (average 40%) of events were detected with 9–34 (average 16) false
detections per hour. Walters et al. (1989) used a syntactic approach and reported that for 3 EEGs their system detected 70% of epileptiform transients but reported 30 false detections in the 30 mins of EEG processed. Fischer et al. (1980) used a parametric method to detect transients in the EEG and followed this with template matching. The system detected 73% of transients marked by at least 4 of 7 EEGers in both a training and test set with 7 false detections (i.e., marked by no EEGers) in 8.3 mins (50/hr) and 3 false detections in 8 mins (23/hr) on the training and test sets respectively.

The overall detection rate for our system of 64% with an average of 3.5 false detections per hour compares very favourably with the results of other systems. The outstanding feature of our system is, however, its ability to detect, on average, 53% of events in an EEG as definite with no false detections.

6.7 SUMMARY

A system has been developed that makes considerable use of spatial and temporal contextual information to detect epileptiform activity in EEGs. The system reports epileptiform events rather than individual epileptiform transients, and classifies these as either focal or non-focal based on the polarity relationships between their constituent transients. Two categories of event — definite and probable — are employed to overcome the problem of maintaining satisfactory detection rates while minimizing false detections and to enable the system to be applied in different clinical situations. It is the ability of this system to eliminate false detections that will make it applicable to routine clinical recordings. However, more data are required to evaluate its performance further and a blind clinical study is in progress. Further stages of development include processing referential montages, application to recordings with periods of sleep (long-term monitoring) and integration of the data collection, feature extraction and expert system stages to achieve a system capable of real-time on-line detection of epileptiform activity in EEGs.
Although the expert system approach to the detection of epileptiform activity in EEGs has proven very successful, development time has been excessive due, on the whole, to the complexity of the problem and the difficulty acquiring and encoding relevant knowledge.

Neural networks provide an alternative means for processing information which may prove useful for EEG analysis. Inspired by the anatomy and physiology of the brain, neural networks employ an adaptive parallel architecture, modelling neurons, their interconnections and their capacity to learn. The ability of neural networks to process information in parallel and to learn by example makes them applicable to a wide variety of tasks (e.g., speech recognition, image classification, image compression).

This Part introduces neural networks in Chapter 7 before considering their application to real-world problems and, in particular, to the detection of epileptiform activity in the EEG, in Chapter 8.
CHAPTER 7

NEURAL NETWORKS

Experience is the best teacher.

Proverb

7.1 INTRODUCTION

Neural networks are adaptive parallel processing systems inspired by the anatomy and physiology of the brain. In essence, they are highly interconnected networks of relatively simple units, which are able to learn to perform complex tasks. This Chapter introduces the fundamental concepts of neural networks before examining three different neural networks — the Hopfield network, the self-organizing map and the back-propagation network. Hardware implementations of neural networks are then briefly reviewed and the Chapter concludes with a discussion of the similarities and differences between neural networks and the brain.

7.2 THE DEVELOPMENT OF NEURAL NETWORKS

The human brain, with its adaptive parallel architecture, outperforms conventional digital serial processors (computers) on perceptual tasks such as image classification and speech recognition, while computers are far better than humans at arithmetic and formal logic. Therefore, a single computer architecture may not be enough to solve all problems. Neural networks have developed as information processing systems with architectures inspired by the brain. Consequently, they hold the promise of solving a diversity of tasks and, in so doing, may provide further insights into the operation of the brain.

The field of neural networks grew from collaboration between technologists attempting to build intelligent machines, and neuroscientists interested in understanding how intelligence emerges from the interaction of many neurons in the brain. As a result of greater emphasis being placed on either biological plausibility or technological achievement, two subfields of neural networks have emerged — biological modelling and technological application. The aim of biological modelling is to produce a simplified version of a neural system that retains the same general behaviour so that the system can be more easily understood. Conversely, the intention of technologists is to develop capable machines and hence, rather than trying to mimic biology, they endeavour to extract concepts that can be used in new computational methodologies. The neural networks discussed in this Chapter have, for the most part, been developed with technological applications in mind. Although developers emphasize the ability of these networks to perform certain tasks, they can also provide valuable insights into brain operation.

The study of neural networks really began with McCulloch and Pitts (1943) who at-
tempted to understand what the brain might actually be doing. By modelling neurons as binary threshold devices, they showed that networks of simple elements could have immense computational power; specifically, they demonstrated such networks could realize any finite logical expression.

McCulloch and Pitts had shown how networks of neuron-like elements could compute. The next problem facing researchers was to understand how such networks could learn. In 1949, Donald O. Hebb provided the first explicit statement of a physiological learning rule:

“When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.”

Despite the non-mathematical nature of this learning law, it has formed the basis of many learning algorithms that modify the strength of synaptic coupling between neuron-like elements as a function of pre-synaptic and post-synaptic activity.

By 1958 Frank Rosenblatt had developed a learning machine which he called the Perceptron. The Perceptron was the first true neural network. It consisted of a number of neuron-like elements which received inputs through weighted connections. By modifying the strengths of these connections, according to a given algorithm, the Perceptron could learn to classify input patterns (provided that the input classes were linearly separable).

Following the development of the Perceptron, expectations of neural networks became unreasonably high and, consequently, were not able to be fulfilled. When Minsky and Papert (1969) published the limitations of the Perceptron (e.g., its inability to solve the ‘exclusive or’ problem), almost all interest in neural networks was lost, and the direction of artificial intelligence shifted to sequential symbolic processing.

Although some neural network research did continue, it was at a considerably slowed pace and was primarily directed towards biological modelling. However, with the 1980s came a revival of interest in technological neural networks and many important discoveries were made. In 1982 John Hopfield published a neural network that behaved as a content-addressable memory, correctly recreating an entire memory from any subpart of sufficient size (§7.5). In the same year, Teuvo Kohonen (1982) introduced a network, the self-organizing map (§7.6), which was able to discover important features in a set of input patterns and spatially order them to form a topographically organized map. However, probably the most significant advance in the field of neural networks was the development of the back-propagation learning algorithm, which provided a method for training multi-layer neural networks (§7.7).

7.3 PROPERTIES OF NEURAL NETWORKS

Until recently, information processing has involved devising algorithms or rules to solve a problem and encoding them in software. Neural networks provide a radically different approach to information processing. Based upon modern neurophysiology, neural networks are made up of many neural units (models of neurons) which interact with each other through weighted connections. Neural units tend to be very simple and, in isolation, have extremely limited computational power. The information processing capability of neural networks is a collective phenomenon resulting from the interaction of many neural units. The collective properties of neural networks tend to be relatively insensitive to the detailed operation of neural units.
Neural networks are able to learn from experience by modifying the strengths of connections between neural units. In this way, knowledge of a particular pattern becomes distributed over the connections among a large number of neural units. The patterns themselves are not stored, rather the weights of the network become such that the patterns can be recreated when required. Consequently, the time for a neural network to respond to a given input pattern is independent of the number of memories it contains.

Neural networks do not need to be trained on all possible input patterns because they are able to generalize from a set of typical examples. In other words, after appropriate training, neural networks are able to respond correctly to input patterns not previously encountered. For example, a neural network trained to recognize characters of the alphabet is still able to classify input characters correctly when they are corrupted by significant amounts of noise. Neural networks are also able to cope with incomplete input data and even with partially incorrect data.

Because information is distributed throughout the neural network, failure of individual neural units or their connections is not catastrophic, rather the performance of the network deteriorates gradually as more components fail. This is an important property of neural networks and is known as graceful degradation.

### 7.4 COMPONENTS OF NEURAL NETWORKS

A neural unit is a highly simplified model of a neuron. Many neural units are interconnected to form a neural network. Neural units interact with each other and with the environment through uni-directional weighted connections. The architecture of a neural network describes which neural units are interconnected and how the network interacts with the environment. The strengths of the connections between neural units are termed weights and represent the synaptic coupling between neurons. In general, the weights may be positive or negative corresponding to excitatory and inhibitory synapses respectively. The knowledge of a neural network is embodied in the weights between neural units and, therefore, learning is a matter of searching for a set of weights that produces the desired network behaviour. A learning algorithm provides the mechanism for finding an appropriate set of weights.

A neural network can be specified by its architecture, its neural units and its learning algorithm. Although each is discussed separately here, they are not independent considerations, as is revealed by examination of three different neural networks in §7.5 – §7.7.

#### 7.4.1 Network Architecture

The architecture of a neural network describes the connections between neural units. The strength of the connection from unit $i$ to unit $j$ is represented by the weight $w_{ji}$; if no such connection exists then $w_{ji} = 0$. Therefore, the network architecture specifies which of the weights $w_{ji}$ can be non-zero.

A neural network can be thought of as one or more layers of neural units (Fig. 7.1). The input layer consists of a number of fan-out elements which distribute the external inputs to the neural units of the following layer. The final or output layer contains the neural units whose outputs are available to the environment. Between the input and output layers there may be any number of hidden layers, so named because they are effectively hidden from the environment. When describing the architecture of a neural network the number of layers is often quoted. However, in the literature there is some disagreement as to whether the input layer should be counted. In this thesis, the number of layers refers to the number of
layers of neural units. The input layer is not counted because it merely distributes the inputs rather than processing them. Therefore, the network of Fig. 7.1 is described as a three layer network.

![Neural Network Diagram](image)

**Figure 7.1.** A three layer neural network consists of an input layer of fan-out elements, two hidden layers and an output layer.

Having detailed the arrangement of neural units within a neural network, it is necessary to specify the connections between them. Connections are described as (a) feedforward if they link neural units of one layer with those of a following layer, (b) feedback if they join neural units in one layer to those of a previous layer, or (c) lateral if they connect neural units within a layer. Thus, Fig. 7.1 shows a three layer neural network with feedforward connections between neural units of consecutive layers, which is the typical architecture of back-propagation neural networks (§7.7).

Many other neural network architectures are possible. For example, the self-organizing map (§7.6) is a single layer network where each neural unit receives the entire input pattern. Similarly, the Hopfield network (§7.5) is a single layer network but with a full complement of lateral connections. The Hopfield net is also described as fully connected, because each neural unit interacts with all other neural units.
7.4.2 Neural Units

Neural units are models of neurons and, as such, should be as simple as possible while retaining the features necessary for information processing. In general, a neural unit $j$ receives several, say $M$, inputs $x_i$ through weighted connections $w_{ji}$ and from these derives a single output $y_j$ (Fig. 7.2), which then becomes the input $x_j$ to other neural units. A neural unit

\[
    y_j = f \left( \sum_{i=1}^{M} w_{ji} x_i - \theta_j \right)
\]

(7.1)

where $f[\cdot]$ is the transfer function of the neural unit. It is often convenient to represent the threshold as a weight from an extra input $x_0$:

\[
    y_j = f \left( \sum_{i=0}^{M} w_{ji} x_i \right)
\]

(7.2)

where $x_0 = 1.0$ and $w_{j0} = -\theta_j$. The advantage of this representation is that an appropriate threshold value can be learnt along with the strengths of connections between neural units.

The form of the transfer function $f[\cdot]$ depends on the particular features of a neuron that are modelled. A neuron can be thought of as being 'on' when it produces an action potential and 'off' at other times. From this perspective, a neuron can be modelled as a logic device. The two logic states may be 0 or 1, in which case the transfer function is the unit step function (Fig. 7.3a) or they may be $-1$ and $+1$ with the transfer function becoming the signum function (Fig. 7.3b). Alternatively, the activity of a neuron can be characterized by the frequency at which it generates action potentials and, thus, neurons can be viewed as analogue devices. The transfer function of these neural units can take several forms (Fig. 7.3c,d,e,f,g), but is essentially a monotonic non-decreasing function. Although a linear function (Fig. 7.3c) is sometimes employed, nonlinear functions are far more common because they significantly increase the collective computational power of the neural network.
7.4.3 Learning Algorithms

Neural networks are able to learn from experience by appropriately adjusting the strengths of connections or weights between neural units. A learning algorithm provides the mechanism for modifying the weights of a network based upon a number of typical examples, known as the training set. The training set and the learning algorithm determine what the network learns and how well it generalizes to patterns not previously encountered. Once trained, the performance of the network is evaluated by presenting it with a number of novel input patterns, known as the test set, and calculating a performance measure (e.g., number of correct responses, mean squared error).
Learning algorithms modify the weights between neural units in accordance with the input pattern and the network's response to it. These algorithms form a spectrum, at one end of which is learning with an error-correcting teacher and at the other is completely spontaneous, unsupervised discovery. In between is a continuum of rules including a number of graded learning schemes where the neural network is given an indication of its performance.

Learning with an error-correcting teacher or supervised learning enables a neural network to learn arbitrary associations between input and output patterns. The network is presented with an input pattern and the corresponding target output pattern. The weights between neural units are then modified to reduce the error between the target pattern and the output pattern produced by the network.

Graded learning algorithms are usually less capable and less generally applicable than supervised schemes. Their main advantage is that it is not necessary to know the correct output pattern. Instead of being given the target output for each input pattern, the network receives only an indication of its performance after several training patterns. This is usually a score or grade that represents the value of some performance measure or cost function. Graded learning is particularly applicable to control problems where there is no way of knowing the correct outputs, for example, balancing a broomstick on its end. In this case the performance measure may be a binary value corresponding to success or failure (Barto et al., 1983), or it may be the sum of the absolute angular deviation of the broomstick from vertical.

Unsupervised learning is also known as self-organization because the network receives no external guidance. Unsupervised learning algorithms enable a network to learn something about the statistical properties of the input patterns with neural units often behaving like feature detectors. There are two classes of unsupervised learning: (a) coincidence learning where the weights are modified in response to events that occur simultaneously and (b) competitive learning where neural units compete for the privilege to learn.

### 7.5 THE HOPFIELD NETWORK

In 1982, John Hopfield published a paper describing a neural network, the collective properties of which produced a content-addressable memory. In contrast to conventional computer memories where stored information is accessed by knowing its address, a content-addressable memory retrieves stored data on the basis of partial information. Thus, a content-addressable memory yields an entire memory item from any subpart of sufficient size. In general, a content-addressable memory is characterized by a number of locally stable states to which the system is attracted. The locally stable states correspond to the stored patterns. Thus, from a given initial state \( x = \xi_p + \delta \), which represents partial knowledge of the memory \( \xi_p \), the system should converge to the locally stable state \( \xi_p \).

#### Architecture

This neural network, which has since become known as the Hopfield network, consists of a single layer of fully connected neural units (Fig. 7.4). Each neural unit receives a single input from the environment and, therefore, the network does not need an input layer of fan-out elements. All inputs are applied simultaneously, after which the network passes through a series of states until it reaches a stable state, which is the network's output. To ensure that the network converges to a stable state, the weights between neural units must be symmetrical:

\[
    w_{ji} = \begin{cases} 
    w_{ij} & i \neq j \\
    0 & i = j 
    \end{cases} \tag{7.3}
\]
Neural Units
The neural units of the Hopfield network have two possible output values (0 and 1). Using the notation of §7.4, these binary neural units can be expressed as:

\[ y_j = \mu \left[ \prod_{i=1}^{M} w_{ji} x_i - \theta_j \right] \quad (7.4) \]

where \( \mu[^{[\cdot]}] \) is the unit step function (Fig. 7.3a) and the inputs \( x_i \) are binary-valued. Typically, the threshold \( \theta_j \) is zero and each neural unit randomly updates its output at an average rate \( R \). The asynchronous operation of the neural units represents a combination of the propagation delays, jitter and noise that are present in real neural systems.

Learning Algorithms
For the Hopfield network to operate as a content-addressable memory, the weights between neural units are set as follows:

\[ w_{ji} = \begin{cases} \frac{1}{M} \sum_{p=1}^{P} \xi_{pi} \xi_{pj} & i \neq j \\ 0 & i = j \end{cases} \quad (7.5) \]

where \( M \) is the number of neural units, \( P \) is the number of patterns to be stored and \( \xi_{pi} \) is the \( i \)th element of the pattern \( \xi_p \). The weights of the network are calculated after a single pass through the training patterns.

Collective Phenomena
The collective properties of the Hopfield network produce a content-addressable memory for binary patterns. Hopfield (1982) analysed the performance of the neural network as a content-addressable memory by considering its energy function \( E \):

\[ E = -\frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} w_{ji} y_i y_j \quad (7.6) \]
The energy function, also known as the \textit{energy landscape}, is characterized by a number of hollows or \textit{local minima}, which correspond to the stored patterns. When the output of a neural unit changes, the energy of the system decreases and, thus, the state of the network converges to a local minimum. Unfortunately, the Hopfield network does not always converge to the nearest local minimum and, consequently, does not always yield the correct memory. A further problem arises if one attempts to 'store' too many patterns in the network. In this case, spurious local minima are generated that do not correspond to any of the training patterns.

**Alternative Implementations**

Several variations of the Hopfield network exist, the most common of which is to use bipolar neural units whose two possible outputs are ±1 (Lippmann, 1987). The inputs to this network are also bipolar which allows the weights between neural units to become negative. Hopfield (1984) showed that the network still behaved as a content-addressable memory even if the neural units had a graded response (e.g., a sigmoidal transfer function) and were updated synchronously.

When stored patterns are not orthogonal, interference between patterns occurs and local minima are produced in the energy landscape which do not correspond to stored patterns. Often these local minima are not as 'deep' as those corresponding to stored patterns. Therefore, the network needs a way to escape from shallow local minima. Addition of thermal noise allows the network to move to higher energy states and escape from local minima. This is usually implemented through a probabilistic update rule, where the probability that the next output of a neural unit is 1 is given by:

\[
P_j = \frac{1}{1 + e^{-\Delta E_j/T}}
\]

\[
\Delta E_j = \sum_{i=1}^{M} w_{ji} x_i - \theta_j
\]  

(7.7)

where \(T\) is a parameter that acts like the temperature of a physical system. The temperature \(T\) decreases with time and, therefore, this process is known as \textit{simulated annealing} (Ackley \textit{et al.}, 1985; Kirkpatrick \textit{et al.}, 1983).

### 7.6 THE SELF-ORGANIZING MAP

In 1982, Teuvo Kohonen introduced a neural network that is able to discover important features of the training patterns. Furthermore, this network, which is known as the self-organizing map, forms spatially ordered representations of these features so that the location of the active neural unit is specific to a certain characteristic feature of the input pattern. For example, the location of a neural unit may correspond to the frequency of the input signal.

**Architecture**

The self-organizing map consists of a single planar array of neural units (Fig. 7.5). Each neural unit receives the entire input pattern through a set of weighted connections.

**Neural Units**

The neural units which make up the self-organizing map are linear and can be expressed as:

\[
y_j = \sum_{i=1}^{M} w_{ji} x_{pi}
\]

(7.8)

This neural unit can also be expressed in vector notation as:

\[
y_j = w_j \cdot x_p
\]

(7.9)
which is the scalar or dot product of the input pattern $x_p$ with the weight vector $w_j$ of neural unit $j$. The dot product of two vectors is a measure of their similarity. Therefore, a linear neural unit compares the input pattern with its weight vector and responds with a measure of the degree of matching. Typically, all vectors are normalized to unit length (i.e., $||w_j|| = 1$ and $||x_p|| = 1$) and, thus, the output of the neural unit becomes the cosine of the angle $\theta_{pj}$ between the vectors:

$$y_j = w_j \cdot x_p$$

$$= ||w_j|| \cdot ||x_p|| \cdot \cos \theta_{pj}$$

$$= \cos \theta_{pj}$$

(7.10)

**Learning Algorithm**

The self-organizing map employs an unsupervised competitive learning algorithm where neural units compete for the privilege to learn. The winning unit $j^*$ is the one with the largest output value:

$$y_{j^*} = \max_j (y_j)$$

$$= \max_j (w_j \cdot x_p)$$

$$= \max_j (\cos \theta_{pj})$$

(7.11)

Because the vectors have been normalized to unit length, the maximum output corresponds to the minimum angle $\theta_{pj}$ between the input pattern and the weight vector.

The weight vectors of the winning unit and its neighbours are modified according to the following rule:

$$w_j = \frac{w_j^{old} + \eta x_p}{||w_j^{old} + \eta x_p||}$$

(7.12)
where $\eta$ is the learning rate and is typically a decreasing function of time. This rule ensures that the weight vectors $w_j$ remain normalized to unit length and corresponds to a rotation around the unit circle towards the input pattern. The weight vectors of all neural units in the "neighbourhood" of the winning unit are adjusted and, thus, neural units do not learn independently of each other. This interaction of neural units during learning is crucial to the formation of globally ordered feature maps. In typical implementations of the self-organizing map, the size of the neighbourhood around the winning neural unit is initially very large and is slowly decreased with time until it finally includes only the winning unit.

Collective Phenomena
The competitive learning algorithm employed by the self-organizing map causes the neural units to develop into a set of feature-sensitive detectors. Each weight vector moves to the average of the cluster of input patterns for which the corresponding neural unit won the competition to learn. Therefore, the self-organizing map clusters the input patterns. Furthermore, the set of weight vectors tends to approximate the probability density function of the input patterns and, thus, the self-organizing map also performs vector quantization (Kohonen, 1990). In classical vector quantization techniques (Makhoul et al., 1985), a finite number of 'codebook' vectors are arranged in space such that their point density function approximates the continuous probability density function of the input patterns.

If only the weight vector of the winning unit is modified, all neural units operate independently and the order in which they are assigned to clusters is effectively random. However, Kohonen emphasized the importance of the global organization or ordering of feature maps. The fundamental principle of a topographically organized system is that nearby units must respond similarly and, thus, it is crucial to the formation of ordered maps that neural units do not learn independently. Because the weight vectors of all neural units in the neighbourhood of the winning unit are modified, changes in individual neural units are only reinforced if they result in a global order. As the learning rate and size of the neighbourhood decrease, the mapping moves from being very coarse to being finely tuned which corresponds to global ordering followed by increased selectivity of individual elements.

The self-organizing map clusters input patterns on the basis of their characteristic features and displays the overall similarity relations of the input data in a small number of dimensions, limited by our ability to display and perceive multi-dimensional data. For example, the self-organizing map is able to produce mappings that transform a signal pattern of arbitrary dimensionality onto a planar array. Thus, the self-organizing map can preserve topological relations while performing a dimensionality reduction of the input space.

The self-organizing map models the statistical properties of the input patterns. However, the model is only as accurate as the size of the network allows. The more neural units available the less area of input space each weight vector must cover and the more accurate the model. For a perfect model there would be one neural unit for each pattern, but this merely states that the ideal model of the input data set is the input data set itself. Therefore, for a network to be capable of discovering anything non-trivial about the data there must be fewer neural units than input patterns.

Alternative Implementations
Kohonen (1988) proposed using the Euclidean distance to measure the similarity between a neural unit's weight vector and the input pattern. This eliminates the need to normalize the vectors to unit length. A neural unit of this type can be expressed as:

$$y_j = \left( \sum_{i=1}^{M} (x_{pi} - w_{ji})^2 \right)^{1/2}$$

$$= \|x_p - w_j\|$$

(7.13)
The winning unit $j^*$ is then the neural unit with the minimum output value, which corresponds to the minimum distance between the input pattern and the weight vector:

$$y_{j^*} = \min_j (y_j)$$

$$= \min_j (\|x_p - w_j\|) \quad (7.14)$$

The similarity measure employed must be compatible with the learning law (Kohonen, 1988) and, therefore, the rule by which the winning neural unit and its neighbours modify their weight vectors becomes:

$$w_j = w_j^{old} + \eta (x_p - w_j^{old}) \quad (7.15)$$

The choice of metric (i.e., scalar product or Euclidean distance) determines how the input patterns are clustered. In fact, the Euclidean distance version is effectively a neural network implementation of the $k$-means clustering procedure (Hunt, 1975), which also spatially orders the clusters.

Several problems encountered when implementing the self-organizing map have been solved, at least to some extent, by relatively simple modifications. To ensure that all neural units participate in learning, the weight vector of every neural unit needs to be initialized so that it has the opportunity to win or, at least, to be the neighbour of a winner. This can be achieved in several ways all of which result in a considerable reduction in training speed (Hecht-Nielsen, 1990):

1. Begin training with very large neighbourhoods and decrease them slowly with time.

2. Initialize all weight vectors to a zero vector and multiply the inputs by a small number $\beta$ which slowly increases with time to a final value of 1.0. This solution, known as radial sprouting, works fairly well but tends to leave a few weight vectors behind and can only be applied if weight vectors are not being normalized.

3. Add random noise to the input vectors, beginning with a large amount of noise and slowly decreasing it with time. This enables the weight vectors to follow input patterns.

A further problem is that the weight vectors do not always accurately model the probability density function of the input patterns. Regions of input space where the probability density function is low tend to be under-represented while those where it is high are often over-represented. By providing neural units with a 'conscience', the probability density function of the inputs can be almost perfectly reproduced (Hecht-Nielsen, 1990). Each neural unit should have equal probability of winning and, therefore, if a neural unit wins considerably more often than $1/M$ (where $M$ is the total number of units) then its conscience should take it out of the competition for a while. This is implemented by keeping track of the fraction of time $f_j$ that unit $j$ wins:

$$f_j = f_j^{old} + \beta (z_j - f_j^{old}) \quad (7.16)$$

where $\beta$ is a small positive constant ($\sim 0.0001$) and $z_j$ is 1 when unit $j$ wins and 0 otherwise. A bias value is then calculated as follows:

$$b_j = \gamma \left( \frac{1}{M} - f_j \right) \quad (7.17)$$

where $\gamma$ is a positive constant ($\sim 10$). This bias value is incorporated into the competition, with the winner $j^*$ being chosen from minimum Euclidean distance or maximum scalar
7.7 BACK-PROPAGATION NETWORKS

Back-propagation neural networks are effectively multi-layer Perceptrons that employ the supervised learning algorithm known as back-propagation. Back-propagation networks are applicable to a wide variety of problems because they are able to learn arbitrary associations between input and output patterns. Consequently, these networks are also the most common.

Architecture
A typical back-propagation neural network is made up of several layers with feedforward connections between neural units of consecutive layers (Fig. 7.1). Each neural unit receives inputs from all neural units of the previous layer and projects its output to all those of the following layer.

Multi-layer networks are able to solve very complex problems. Whereas single layer networks can only distinguish classes that are linearly separable (Minsky and Papert, 1969), three layer networks can discriminate arbitrarily complex classes (Beale and Jackson, 1990; Lippmann, 1987).

Neural Units
Multi-layer networks of linear units can be reduced to single layer networks which can only distinguish linearly separable classes (Minsky and Papert, 1969). Therefore, the neural units of a back-propagation network are nonlinear and have the following form:

\[ y_j = f[\text{net}_j] \]
\[ \text{net}_j = \sum_{i=0}^{M} w_{ji} x_i \]  

(7.20)

where \( f[\cdot] \) is some nonlinear, non-decreasing function. The back-propagation learning algorithm further requires that \( f[\cdot] \) is differentiable and, typically, a sigmoid (Fig. 7.3d) is employed:

\[ f[\text{net}] = \frac{1}{1 + e^{-\text{net}}} \]  

(7.21)

Learning Algorithm
The back-propagation learning algorithm (Rumelhart et al., 1986) is a supervised learning algorithm that modifies the weights \( w_{ji} \) by an amount \( \Delta w_{ji} \) in order to minimize the mean squared error function \( E \). The mean squared error for a given training pattern \( x_p \) is:

\[ E_p = \frac{1}{2} \sum_j (t_{pj} - y_{pj})^2 \]  

(7.22)

where \( t_{pj} \) is the target output of neural unit \( j \) for pattern \( p \) and \( y_{pj} \) is its actual response. The mean squared error for the entire training set of \( P \) patterns is the summation of the error for each pattern:

\[ E = \sum_{p=1}^{P} E_p \]  

(7.23)
The back-propagation algorithm minimizes the error function $E$ by gradient descent, adjusting the weights by an amount proportional to $-\partial E/\partial w_{ji}$:

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}}$$  \hspace{1cm} (7.24)

where $\eta$ is the learning rate ($\eta > 0$). The partial derivative $\partial E/\partial w_{ji}$ can be evaluated using the chain rule (Kreyszig, 1983) and the equation for a neural unit (7.20) as follows:

$$\frac{\partial E}{\partial w_{ji}} = \sum_{p=1}^{P} \frac{\partial E_p}{\partial w_{ji}} = \sum_{p=1}^{P} \frac{\partial E_p}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial w_{ji}} = \sum_{p=1}^{P} \frac{-(t_{pj} - y_{pj}) f'[net_{pj}] x_{pi}}{P}$$  \hspace{1cm} (7.25)

where $f'[x]$ is the derivative $df/dx$. If $f[\cdot]$ is a sigmoid then its derivative is:

$$f'[net_{pj}] = y_{pj}(1 - y_{pj})$$  \hspace{1cm} (7.26)

The weights are adjusted according to the rule:

$$\Delta w_{ji} = \eta \sum_{p=1}^{P} (t_{pj} - y_{pj}) f'[net_{pj}] x_{pi}$$  \hspace{1cm} (7.27)

Defining a new variable $\delta_{pj} = (t_{pj} - y_{pj}) f'[net_{pj}]$ allows the weight change to be expressed as:

$$\Delta w_{ji} = \eta \sum_{p=1}^{P} \delta_{pj} x_{pi}$$  \hspace{1cm} (7.28)

However, in a multi-layer network the target outputs for neural units in the hidden layers are not known and the partial derivative $\partial E_p/\partial y_{pj}$ must be determined in terms of the units of the output layer. For example, $\partial E_p/\partial y_{pj}$ for units of the last hidden layer can be evaluated as follows:

$$\frac{\partial E_p}{\partial y_{pj}} = \sum_{k} \frac{\partial E_p}{\partial y_{pk}} \frac{\partial y_{pk}}{\partial y_{pj}} = \sum_{k} (t_{pk} - y_{pk}) f'[net_{pk}] w_{kj} = \sum_{k} \delta_{pk} w_{kj}$$  \hspace{1cm} (7.29)

where the summation with respect to $k$ is over all neural units in the output layer. The error between the target pattern and the actual output is effectively propagated backwards through the network to the neural units of the previous layer. In fact, the result of (7.29) holds for neural units of all hidden layers and, thus, $\delta_{pj}$ is given by:

$$\delta_{pj} = \begin{cases} (t_{pj} - y_{pj}) f'[net_{pj}] & \text{for output layer} \\ f'[net_{pj}] \sum_{k} \delta_{pk} w_{kj} & \text{for hidden layers} \end{cases}$$  \hspace{1cm} (7.30)
where the summation with respect to $k$ is taken over all units in the following layer. All neural units adjust their weights by an amount $\Delta w_{ji}$ which is determined by the equation:

$$\Delta w_{ji} = \eta \sum_{p=1}^{P} \delta_{pj} x_{pi} \quad (7.31)$$

Because the weights are only adjusted after an entire pass through the training set, this is known as the batch version of the algorithm. However, the weights can be adjusted after presentation of each pattern as follows:

$$\Delta w_{ji} = \eta \delta_{pj} x_{pi} \quad (7.32)$$

This is known as pattern learning and, provided $\eta$ is sufficiently small, also performs gradient descent. However, when $\eta$ is small, learning is very slow. By introducing a momentum factor, $\eta$ can be increased without causing instability. The learning algorithm becomes:

$$\Delta w_{ji} = \eta \delta_{pj} x_{pi} + \alpha \Delta w_{ji}^{old} \quad (7.33)$$

where $\alpha$ is a positive constant ($0 \leq \alpha \leq 1$).

Eberhart and Dobbins (1990) included a momentum term in the batch mode learning algorithm as follows:

$$\Delta w_{ji} = \eta \sum_{p=1}^{P} \delta_{pj} x_{pi} + \alpha \Delta w_{ji}^{old} \quad (7.34)$$

where again $\alpha$ is a positive constant ($0 \leq \alpha \leq 1$). This helps increase learning speed when the gradient of the weight space is small.

**Collective Phenomena**

The back-propagation learning algorithm provides a method for training multi-layer neural networks, which are considerably more powerful than single layer networks. Multi-layer networks are able to learn arbitrary associations between input and output patterns by forming internal representations of the input patterns. In effect, the hidden layers encode the features of the input patterns that the neural network considers to be important. Unfortunately the training of back-propagation networks is slow and the network weights may not converge to the global minimum of the error function but may become trapped in a local minimum.

**Alternative Implementations**

The collective computational power of back-propagation networks is relatively insensitive to the details of the neural transfer function $f[\cdot]$. Therefore, functions other than the sigmoid can be used. For example, the tanh function (Fig. 7.3g) is commonly used and Tepedelenlioglu and Rezgui (1989) showed that non-differentiable piecewise linear functions could also be used provided that a suitable derivative function $f'[\cdot]$ was defined.

Several modifications to the back-propagation learning algorithm have been proposed in an attempt to decrease training time. For example, the introduction of a momentum term as in (7.33) allows the learning rate $\eta$ to be increased without causing the algorithm to become unstable. Use of $\tanh$ as the transfer function of neural units tends to decrease training time, because it does not approach zero at $-\infty$. However, at the beginning of training, when the weights are small and all net inputs close to zero, learning may be slow. When the output of a neural unit is very close to 0 or 1 very little learning takes place because the value of the derivative is almost zero. Ttvetter (1991) suggested that this problem may be overcome, to some extent, by adding a small positive constant to the derivative. Similarly, Haffner et al. (1989) proposed adding a linear function to the transfer function, which results in a constant being added to the derivative.
In another approach, Haffner et al. (1989) used a modified error function:

\[ E = \sum_p \sum_j \log_e \left( 1 - (t_{pj} - y_{pj})^2 \right) \]  \hspace{1cm} (7.35)

Consider the usual situation where the target value for each neural unit of the output layer is either 0.0 or 1.0. As \( |t_{pj} - y_{pj}| \to 1.0 \), the derivative \( \frac{\partial E_p}{\partial y_{pj}} \to 0.0 \), which compensates for the fact that \( f'[net_{pj}] \to 0.0 \).

The training strategy employed can also improve training speed. For example, it is common for back-propagation networks to be trained on first easy examples before more difficult or borderline examples are presented. This is similar to the strategy proposed by Caudill (1991a) where the error is set to zero if it is less than a given error tolerance. Initially the error tolerance is large but is decreased over the training session.

It is possible to overtrain a back-propagation network so that it begins to memorize the training patterns and, therefore, loses its ability to generalize. Caudill (1991a) suggested adding small amounts of noise to the training patterns so that the network never sees exactly the same pattern twice. In this way, generalization rather than memorization is encouraged.

7.8 HARDWARE IMPLEMENTATIONS

One of the major advantages of neural networks is their speed due to their massively parallel architecture. Because of the difficulty implementing neural networks in hardware, computer simulations are often performed. However, such simulations are very slow because all computations must be carried out sequentially. The architecture of the hardware is not compatible with the problem and the result is a considerable reduction in the speed of computation. A similar problem is humans performing mental arithmetic, which they do very slowly.

Hardware implementations of neural networks is an active area of research. The major difficulty facing researchers is the huge number of connections required between neural units. Researchers are presently exploring many computing hardware ideas including electronic, optical, mechanical, acoustic and chemical implementations (Hecht-Nielsen, 1990).

At present electronic and optical implementations are the most common. In essence, electronic models consist of thresholding amplifiers interconnected by a resistive feedback network. However, such networks are constrained by their two dimensional geometry and by the need to isolate individual connections. Electronic neural networks are often employed as coprocessors for conventional serial computers which essentially operate as input/output servers.

In optical and electro-optical networks light is used as the communication medium between neural units. Light rays provide high speed connections between neural units in a three dimensional network. The neural units of such a network can be implemented either optically or electronically (Clark, 1991).

Research into hardware implementations of neural networks is still in its infancy but will play an important role in the application of neural networks to real-world problems. For the time being, neural networks continue to be implemented on serial computers both with and without neural network coprocessors.
7.9 COMPARISON WITH THE BRAIN

Because the emphasis of the neural networks presented in this Chapter is technological achievement, there are significant differences between these neural networks and the brain. However, this does not mean that these neural networks cannot provide insights into the operation of the brain. In fact, despite considerable deviations from the anatomy and physiology of the brain, neural networks often perform remarkably like the brain.

Most neural networks have a narrow biological base, being almost entirely inspired by the cerebral cortex. This is because the cerebral cortex evolved relatively recently and is anatomically rather homogeneous with only a few neuron types and a few standard connection patterns. Furthermore, the cerebral cortex appears to carry out most of the complex functions of humans (e.g., speech, language, perception).

Although the brain contains many varieties of neurons, neural networks usually comprise a single type of neural unit. Whereas neurons in the brain operate asynchronously and continuously, neural units tend to be updated synchronously and at discrete time intervals. Communication between neurons in the brain is by way of chemical synapses. A single neuron tends to either inhibit or excite all neurons with which it forms synapses. In neural networks, synapses are modelled by multiplicative weights. However, there are usually no restrictions as to which weights can be positive (excitatory) or negative (inhibitory) and, hence, a given neural unit may excite some neural units while inhibiting others. In addition, during learning, a weight may change from being excitatory to being inhibitory and vice versa. Finally, neural networks are extremely small compared with the brain. Whereas neural networks do not usually consist of more than a few hundred neural units, the brain contains around $10^{11}$ neurons.

Despite the structural and operational differences between neural networks and the brain, there are often similarities in their behaviour, which further suggests that collective phenomena are relatively independent of the detailed operation of the individual elements. Linsker (1988) showed that a multi-layer network, learning according to a simple unsupervised coincidence rule, developed feature-sensitive neural units. Furthermore, the response properties of these feature-sensitive units were qualitatively similar to those of the first few processing stages of the mammalian visual system. These properties included sensitivity to light-dark contrast and sensitivity to edge orientation. These properties developed even in the absence of structured input patterns, that is, when only random activity was used as the input to the network. Again this is very similar to the visual system of mammals, where feature-sensitive cells are already present at birth.

The self-organizing map (§7.6) has several features in common with the brain. In many places, the brain is organized so that aspects of the sensory environment are represented in one or two dimensional maps (e.g., tonotopic map of the auditory cortex). The particular location of neural response corresponds to a specific feature of the environment. In addition, a form of competitive learning may be used in the brain. In the cerebral cortex, neurons are essentially arranged in two dimensional layers and are densely interconnected through lateral feedback. There is both anatomical and physiological evidence that this lateral interaction is of the following form: (1) short range lateral excitation reaching a radius of up to 50–100 \( \mu \text{m} \) surrounded by (2) inhibitory action up to 200–500 \( \mu \text{m} \) which is in turn surrounded by (3) weaker excitation up to a radius of several centimetres. Kohonen (1988) demonstrated that this type of lateral interaction between neural units produced spatially bounded clusters of activity with the size of the cluster being dependent upon the ratio of excitatory to inhibitory connection strengths. The physiological learning rule proposed by Hebb (1949) can
be expressed mathematically as:

\[
    w_{ji} = w_{ji}^{\text{old}} + \eta y_j x_i
\]

(7.36)

where \( \eta \) is the learning rate, \( x_i \) and \( y_j \) describe the pre-synaptic and post-synaptic activity respectively. When coupled with lateral feedback, this rule essentially results in competitive learning with the term \( y_j \) restricting learning to the spatially bounded cluster of activity. Furthermore, units within the cluster learn at a rate proportional to their distance from the unit with maximum activity (i.e., centre of cluster).

Although the detailed implementation of back-propagation may not be biologically plausible, the essential principle behind it appears to be important in human learning. This principle can be stated as follows:

"Adjust the parameters of the mind in proportion to the extent to which their adjustment can produce a reduction in the discrepancy between expected and observed events."

McClelland (1989) demonstrated some of the similarities between back-propagation neural networks and the brain by comparing the performance of a back-propagation neural network with that of children on a balance beam task. The task required the subject to predict which side of the balance beam would fall. The neural network was trained in a series of epochs and its performance evaluated after each epoch. McClelland found a high correspondence between the learning of the network and that of children. Initially the network, like children of 5–6 years, responded as though distance from the fulcrum was completely irrelevant, always predicting that the side with the greatest weight would fall. As training continued the network, like older children, began to also consider distance. Development progressed from an initial over-focussing on the most salient dimension of the task through a sequence of stages where reliance on the initially unattended dimension gradually increases. Thus, the neural network captured the stage-like development observed in children while at the same time exhibiting an underlying continuity. Similarities between back-propagation neural networks and the brain have also been uncovered when teaching networks to read aloud (Sejnowski and Rosenberg, 1987) and to form the past tense of words (Rumelhart and McClelland, 1986).

These results have already had an important influence on cognitive science with a whole new class of theories of brain function emerging. In fact, there has been a shift from rule-based and symbol manipulation theories to theories based on or, at least, consistent with neural networks.

7.10 SUMMARY

Neural networks are adaptive parallel processing systems inspired by the anatomy and physiology of the brain. There are many different types of neural networks — for example, the Hopfield network, the self-organizing map and back-propagation networks — each of which can be described by its architecture, its neural units and its learning algorithm. Although these neural networks have been developed for technological applications (rather than as models of biology), they can still provide insights into the operation of the brain.
CHAPTER 8

APPLICATIONS OF NEURAL NETWORKS

"The real danger is not that computers will begin to think like men, but that men will begin to think like computers."

S. J. Harris

8.1 INTRODUCTION

The resurgence of interest in neural networks has been due to their potential for practical applications. Because of their ability to learn from examples, neural networks can be applied to a wide range of tasks and are particularly applicable to those problems for which mathematical algorithms do not exist or are unsatisfactory.

Neural network applications are many and varied; examples include identification of radar patterns, mortgage risk assessment, DNA sequencing, automatic vehicle guidance, content addressable memory, recognition of hand writing, monitoring respiratory function during surgery, image restoration, circuit board layout, adaptive equalization, parts inspection in manufacturing, intelligent spreadsheets for financial analysts, detection of explosive substances at airports, classification and interpretation of mass spectroscope data, fault diagnosis of engines, speech recognition, target identification in sonar images, image compression and recognition of star patterns for on-board satellite navigation.

Despite the diversity of application domains, these tasks can be divided into a number of categories: memory, optimization, categorization, control, data processing, predictive modelling and pattern recognition. This Chapter discusses application examples from each of these areas, in each case detailing network architecture, training and performance as well as problem representation. Particular attention is paid to applications of back-propagation neural networks and to the problem of detecting epileptiform transients in EEGs. A number of suggestions are made regarding modifications to back-propagation learning, which should aid in achieving the desired network performance. The Chapter concludes with a brief discussion of the limitations of neural networks.

8.2 GENERAL CONSIDERATIONS

Neural networks are able to perform some tasks that are extremely difficult to implement successfully on conventional digital serial processors. However, neural networks perform poorly on arithmetic tasks at which serial computers excel. Therefore, neural networks must be considered as complementary to, not replacements for, traditional computers and should only be applied to those problems for which conventional computing paradigms are inadequate. Such problems are typically characterized by fuzzy, imprecise or imperfect knowledge or data.
There are a wide variety of neural networks available, only three of which were considered in Chapter 7. Each type of neural network has its own functional capabilities and, consequently, one type of neural network cannot perform all tasks. Thus, choice of a neural network depends on the application. For example, if a network is required to discover features in a set of input patterns, then a network employing an unsupervised learning algorithm should be chosen, whereas if the network was to classify input patterns into known categories then a supervised learning algorithm should be employed.

Having chosen a neural network, it is necessary to specify its architecture. Although the architecture may, to some extent, have been determined by the choice of a particular network, its details remain to be decided. For example, the Hopfield network has a single layer, fully connected architecture but the number of neural units must be specified. Having adopted a back-propagation network, it is necessary to determine the number of layers, the number of neural units in each layer and the number of feedforward connections. The number of layers required depends on the complexity of the problem. For example, a single layer network can only distinguish between input classes that are linearly separable, while three layer networks with sufficient neural units can separate classes of any arbitrary shape (Beale and Jackson, 1990; Lippmann, 1987). The numbers of input and output elements depends on the representation of the problem, while the number of hidden units tends to be a somewhat arbitrary decision. Although in a typical back-propagation network neural units are connected to all units in the following layer, this is not a mandatory requirement. Additional feedforward connections may be included between input and output layers, and connections between consecutive layers may be restricted. McClelland (1989) trained a back-propagation neural network to predict which side of a balance beam would fall. The architecture of the network was such that the weight and distance information were analysed separately, before being combined (Fig. 8.1). In fact, when neural units were fully connected to those of the following layer, the network did not perform the task as well. Thus, the architecture of the network imposes constraints on the training process, which may facilitate learning and generalization if appropriate to the task (McClelland, 1989). Selecting an appropriate network architecture has been described as “inspired guesswork” (McShane, 1992).

Figure 8.1. The balance beam task where the network is required to predict which side of the beam will fall. (a) Balance beam and (b) Network architecture.
All practical applications of neural networks critically depend on the way the problem is represented (Anderson and Rosenfeld, 1989). Inputs to a neural network can take a variety of forms, such as raw data, features extracted from raw data or some combination. For example, a network to detect epileptiform transients may be presented with raw data or with features such as the amplitude, sharpness and duration of individual waves and of the background activity. Most neural networks perform better if they are provided with some ideas about the problem (Caudill, 1991b). Although the way to solve a problem may not be known, some features that are likely to be helpful may be known, and this information should be supplied to the network. Pre-processing can also be used to reduce the complexity of the task to be solved by the neural network. Consider the problem of recognizing handwritten characters drawn on a digitizer tablet. A high resolution binary image (2048x2048 pixels, say) is obtained, where all pixels are zero except those through which the character passes (Fig. 8.2). The vertical and horizontal extents of the character can easily be determined and the character placed in a box. This box can then be subdivided into an array of 15x10 smaller boxes. A binary vector \( x \) consisting of 150 elements can then be constructed, with \( x_i = 1 \) if the character passes through the \( i \)th box and \( x_i = 0 \) otherwise. This vector can be used as the input to a neural network. Thus, the neural network does not have to be insensitive to scaling or to translation.

![Figure 8.2. Pre-processing of characters so that the input vector is independent of scaling and translation.](image)

Most neural networks require normalized inputs. For back-propagation networks, inputs typically lie in the range 0.0-1.0. This normalization can be performed on individual input channels or across several input channels. Eberhart and Dobbins (1990) have found that normalizing related inputs (e.g., all amplitude measures) as a group is usually the most effective approach.

Neural networks learn from examples and, therefore, sufficient data must be available both to train the network and to evaluate its performance. The number of examples required to train a network successfully is very dependent on the network architecture, learning algorithm and specific application. The training set implicitly contains information on how to generalize (Hinton, 1989). Therefore, to ensure that the network does not learn any quirks of the training examples, large training sets should be used. In particular, there should be more training examples than hidden units to encourage generalization rather than memorization (Caudill, 1991a). Eberhart and Dobbins (1990) found that, for back-propagation networks, at least 10 examples of each output class are required to train the network adequately. The performance of the network must be evaluated on data not used for training. The amount of test data necessary depends on the user requirements. For example, in order to specify the percentage of correct classifications to a precision of 5%, 20 examples of each output class must be included in the test set.
8.3 APPLICATION EXAMPLES

Neural network applications can be divided into a number of categories: memory, optimization, categorization, control, data processing, predictive models and pattern recognition. Although the examples presented in this section illustrate typical problems from each of these categories, the distinction between classes is not clear cut, and some would argue that most applications are, in fact, pattern recognition tasks.

8.3.1 Memory

Neural networks can operate as associative memories, learning to associate input patterns with given output patterns. There are two types of associative memory, the first of which is known as heteroassociative. An m-dimensional input pattern $\xi_p$ is associated with an n-dimensional output pattern $\zeta_p$. Thus, when an input pattern $x$ is presented to the network, the output pattern $\zeta_p$, whose key $\xi_p$ most closely resembles the input pattern $x$, is produced.

The second type of associative memory is termed autoassociative. In this type of network, a number of patterns $\xi_p$ (p=1...P) are stored and the pattern recalled is the one that most closely resembles the input pattern $x$. Such a network is, therefore, able to restore an image contaminated by noise to its original condition. Furthermore, an autoassociative network is able to retrieve an entire memory from any subpart of sufficient size. Thus, these networks perform pattern completion and are often termed content-addressable memories. For example, the Hopfield network (§7.5) operates as a content-addressable memory (Hopfield, 1982). However, the performance of the Hopfield network deteriorates once the number of patterns stored exceeds 0.15N, where N is the number of neural units (McShane, 1992). Consequently, the Hopfield network is not suitable for large scale storage, but attempts continue to increase its storage capacity.

8.3.2 Optimization Problems

Neural networks can be used to solve complex optimization tasks in which the aim is to minimize a cost function of many independent variables (Shackleford, 1989; Tank and Hopfield, 1987). The classic optimization task is the travelling salesman problem, where it is necessary to find the shortest route to visit N cities, finally returning to the starting city.

In such problems, the data determine the network architecture required. For example, the travelling salesman problem can be represented by an $N \times N$ matrix of neural units (Fig. 8.3), where the rows represent the N cities and the columns the order in which they are visited. Thus, Fig. 8.3a shows a path C, A, E, B, D. Using this representation, only one neural unit in each row and column can be active at any one time and, therefore, inhibitory connections are placed between units of the same row and of the same column. The distances between cities (normalized to lie in the range 0.0–1.0) are used to form an additional set of inhibitory connections between neural units of adjacent columns (Fig. 8.3b). Because the salesman must return home, the columns at each edge of the matrix are also considered to be adjacent. A solution to the problem is obtained by initializing the network to an arbitrary initial state and allowing it evolve to a stable state. Although the network may not find the best solution, it rapidly finds a near-optimal route. Furthermore, because no search is required, the time taken to find a solution does not increase exponentially with N, the number of cities to visit.
8.3.3 Categorization

Neural networks, such as the self-organizing map and adaptive resonance theory or ART (Grossberg, 1988), that employ unsupervised learning algorithms can be used to categorize or cluster input data. For example, Kohonen (1990) presented a self-organizing map with short-time spectra of continuous Finnish speech. The spectra were calculated every 9.83 ms using a 256 point fast Fourier transform, from which a 15 component spectral vector was formed. All such spectra were presented to the network in the natural order of utterance. During training the neural units became tuned to the acoustic units of speech, known as phonemes. Although the self-organizing map is only a single layer network, it is able to discover features in the input data and display the relationships between them.

8.3.4 Control Problems

Neural networks can be used to solve nonlinear control problems, such as balancing a broomstick on a cart (Fig. 8.4) and reversing a trailer (Fig. 8.5). In a typical control problem, the state of the system $z_{k+1}$ is a function $F[\cdot]$ of the previous state $z_k$ and the input $u_k$:

$$z_{k+1} = F[z_k, u_k]$$

The aim is to provide the system with input vectors $u_k$ that produce a desired state $z^*$. In many control problems, the input required to produce the desired state is not known. However, it is usually possible to provide the network with information regarding its performance. Consider the one-dimensional problem of balancing a broomstick on a movable cart (Fig. 8.4). Given information regarding the current state of the network (e.g., $\phi$, $x$, $\dot{x}$), the force $F$ to be applied to the cart in order to balance the broomstick must be determined. In general, the force required is not known but a measure of performance, such as the angle $\phi$ that the broomstick makes with the vertical, can be calculated. A neural network that employs a graded training algorithm should be able to solve such a problem. However, the...
few networks of this type that do exist are still under development and have not reached the stage of application.

Consequently, neural networks employing supervised learning algorithms are presently being used to solve control problems. For example, Tolat and Widrow (1988) trained a back-propagation network to balance a broomstick on a cart using training examples from human operators. However, because humans are unable to perform this task in real-time, a slowed down computer simulation of the system was used to obtain training examples. Given a sufficient period of training, the network learnt to emulate the human operator.

In other control applications the aim is to reach a final desired state $z^*$ in some arbitrary time. For example, consider the problem of reversing a truck trailer up to a loading dock (Fig. 8.5). The state of the system can be described by the position of the trailer $(x_T, y_T)$, the angle $\phi_T$ of the trailer relative to the dock and the angle $\phi_C$ of the cab relative to the dock. From an arbitrary initial state, it is necessary to reach the final state $\phi_T=0^\circ$, $x_T=x_D$, $y_T=y_D$ where $(x_D, y_D)$ describes the position of the loading dock. In order to train a back-
propagation network an error signal is required at each time step, but in this situation only
the error in the final state is known. Nguyen and Widrow (1990) overcame this problem
by first training a two layer back-propagation network to emulate the system; given the
current state and the input or steering signal the network learnt to predict the next state
of the system. Having trained the emulator, another back-propagation neural network with
4 inputs \((x_T, y_T, \phi_T, \phi_c)\), a single hidden layer of 25 units and one output unit was trained
to produce an appropriate steering signal. The error in the final state was back-propagated
through time using the emulator to determine the effect of a particular steering signal on the
state of the system. Thus, an error signal and the appropriate weight changes at each time
step were evaluated. Training took the form of 16 lessons, beginning with simple examples and
progressing to arbitrary initial positions of the trailer. After training, the network was able
to perform the task from an arbitrary initial condition (provided the trailer was a sufficient
distance from the dock) with a root mean squared (r.m.s.) error in \(\phi_T\) of 7° and in \(y_T\) of 3%
of the truck length.

8.3.5 Data Processing

Traditional signal or data processing techniques tend to be linear (e.g., filtering and coding).
Neural networks can be used to perform nonlinear signal processing on structured signals,
that is, signals that are always qualitatively similar, such as human ECG signals. However,
because of the nonlinearity of neural networks, the principle of superposition does not hold
and, therefore, it is necessary to control the average power level of the input signal.

Noise removal techniques, for example, tend to be linear and rely on differences between
the frequency content of the noise and that of the signal. However, when the frequency
spectrum of the noise significantly overlaps that of the signal, the filtering process not only
removes some of the noise but also modifies the signal itself. Neural networks can be used
as nonlinear filters and, if all signals to be considered are qualitatively similar, are able
to remove noise that lies in the same frequency range as the signal. Hecht-Nielsen (1990)
described a back-propagation neural network with a single hidden layer that was trained to
filter ECG signals contaminated by white noise. Preliminary results produced signal-to-noise
ratios which were 2–10 times better than those achieved using a traditional linear filter.

Neural networks can also be used to perform nonlinear data compression on structured
signals. Typically, a back-propagation networks with \(m\) inputs, \(n\) hidden units and \(m\) outputs
\((n < m)\) is trained using the input pattern as the target output pattern. Once trained, the
output of the hidden layer is a compressed form of the data. Iwata et al. (1990) used such
a network to compress ECG data before storage on tape. Prior to analysis the data were
reconstructed using the weights between the hidden and output layers. Data compression
rates of 1/15 – 1/100 were achieved with r.m.s. errors in the reconstructed data of 0.1–0.5%.

8.3.6 Predictive Models

Neural networks can be used as predictive models. For example, in the trailer reversing
problem (§8.3.4) a back-propagation network was used to model the behaviour of the system
and predict its next state. Similarly, neural networks can be used to predict subsequent
values of a time series. Elsner (1992) trained a back-propagation network to predict the next
value of a chaotic time series generated from the Lorenz equations (§9.4.1). The network
was provided with the eight previous values of the time series. With three hidden units, the
network predicted the next value of test data with an r.m.s. error of 0.072, indicating that
the network was capable of capturing the underlying chaotic dynamics.
In other applications, the network may be required to extract meaningful predictive information from a large amount of less valuable raw data. Consider the problem of scoring loan applications. Humans are particularly poor at this task because their decisions are too often influenced by factors such as appearance. Hecht-Nielsen (1990) described a back-propagation neural network to perform the task. The network was provided with relevant information from the loan application form, such as annual income, type of residence, length of residence, and produced a credit rating (e.g., dollar profit per dollar loaned per year). The training data were previous loans that had either been repaid or written off. Typically, 10 000 such examples are used to train the networks. Currently, many such networks are in regular use and perform significantly better than humans and statistical analysis techniques.

8.3.7 Pattern Recognition

Since neural networks are attempts to replicate some of the information processing capabilities of the brain, it is not surprising that many of their applications are in the field of pattern recognition (i.e., vision and speech processing). In these areas humans can easily outperform the fastest computers employing the most advanced algorithms. In fact, pattern recognition appears to be one of the most promising fields for application of neural networks.

Burr (1988) trained back-propagation neural networks with a single hidden layer to recognize handwritten digits and letters. The digits were presented to the network in the form of a seven element shadow code. The network was trained on five examples of each digit and tested on a different five examples of each digit. Between 95-98% of the test set were recognized correctly depending on the number of hidden units. A 13 element shadow code was used to represent handwritten letters. The network was trained and tested on four examples of each letter and correctly classified up to 94% of the test examples.

Gorman and Sejnowski (1988) used a neural network to classify sonar returns from two undersea targets. A back-propagation network with one hidden layer was trained on sonar returns obtained from a variety of angles. The input to the network was a 60 element vector describing the power spectral density of the returns. The performance of the network tended to improve with increasing number of hidden units and 90% of test examples were classified correctly when 12 hidden units were used.

8.4 MODIFYING BACK-PROPAGATION

At present, back-propagation neural networks are the most powerful and most widely applicable because they can learn arbitrary associations between input and output patterns. However, they do not always produce the required performance. A number of modifications to back-propagation are suggested which should aid in achieving the desired performance for given tasks.

8.4.1 Learning Steepness Parameters

Consider the problem of detecting whether or not there is a peak at the centre of an input pattern. Although this is not an appropriate application for a neural network because it can be implemented by a simple algorithm, it is related to the problem of detecting epileptiform transients in EEGs and provides insights into back-propagation neural networks. To solve this problem, neural units must be able to determine whether one input, say $x_1$, is larger than another input $x_2$. 
The output $y$ of a typical neural unit is given by:

$$y = f \left[ \sum_i w_i x_i \right]$$  \hspace{1cm} (8.2)

which for two inputs becomes:

$$y = f \left[ w_0 + w_1 x_1 + w_2 x_2 \right]$$  \hspace{1cm} (8.3)

If $f[\cdot]$ is the step function $\mu[\cdot]$ and $w_0 = 0, w_1 = -w_2 = w$ then the neural unit:

$$y = \mu[w(x_1 - x_2)]$$  \hspace{1cm} (8.4)

produces a one when $x_1 > x_2$ and a zero otherwise. Typically, however, $f[\cdot]$ is a sigmoid and so the neural unit produces an output in the range 0.0–1.0, depending on the value of $w(x_1-x_2)$. If the neural unit is required to produce an output exceeding 0.8 when $x_1 - x_2 > 0.01$ and of less than 0.2 when $x_1 - x_2 < -0.01$, then the minimum magnitude of weights needed is $w = 139$. However, the maximum size of weights produced by back-propagation tends to be about 20.

Because the magnitude of weights is not important when a step function is employed, it may be useful to introduce a steepness parameter $\beta$ into the sigmoid function so that as $\beta \to \infty$ the transfer function approximates a step function (i.e., $f[\cdot] \to \mu[\cdot]$). Thus, the sigmoid becomes:

$$f[\text{net}] = \frac{1}{1 + e^{-\beta \text{net}}}$$  \hspace{1cm} (8.5)

This corresponds to scaling the net input by $\beta$ before passing it through the sigmoid function. Therefore, the neural unit can be expressed as:

$$\text{net} = \sum_i w_i x_i$$
$$y = f[\beta \text{net}]$$
$$f[x] = \frac{1}{1 + e^{-x}}$$  \hspace{1cm} (8.6)

In this case, the back-propagation learning algorithm (§7.7), in batch mode, becomes:

$$\Delta w_{ji} = \eta \sum_p \delta_{pj} x_{pi} \beta_j$$  \hspace{1cm} (8.7)

with $\delta_{pj}$ given by:

$$\delta_{pj} = \left\{ \begin{array}{ll} (t_{pj} - y_{pj}) f'(\beta_j \text{net}_{pj}) & \text{for output layer} \\ f'(\beta_j \text{net}_{pj}) \sum_k \delta_{pk} w_{kj} \beta_k & \text{for hidden layers} \end{array} \right. \hspace{1cm} (8.8)$$

where the summation with respect to $k$ is taken over all units in the following layer.

Furthermore, it is possible for each neural unit to learn an appropriate value for $\beta_j$. The change in steepness for each neural unit can be derived by considering the partial derivative $\partial E/\partial \beta_j$:

$$\Delta \beta_j = -\eta \sum_p \frac{\partial E_p}{\partial \beta_j}$$
$$= -\eta \sum_p \frac{\partial E_p}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial \beta_j}$$
$$= -\eta \sum_p \frac{\partial E_p}{\partial y_{pj}} f'(\beta_j \text{net}_{pj}) \text{net}_{pj}$$  \hspace{1cm} (8.9)
Thus, the learning algorithm for $\beta$ becomes:

$$\Delta \beta_j = \eta_1 \sum_p \delta_{pj} \text{net}_{pj}$$  \hspace{1cm} (8.10)

Alternatively, the steepness parameters can be adjusted after each pattern and a momentum term can be introduced as follows:

$$\Delta \beta_j = \eta_1 \delta_{pj} \text{net}_{pj} + \alpha_1 \Delta \beta_j^{old}$$  \hspace{1cm} (8.11)

Using this algorithm, each neural unit can learn an appropriate value for its steepness parameter. The value acquired depends on the task of the neural unit. For example, if $\beta_j$ is large then the neural unit is effectively making a 'yes/no' decision and is able to detect very small differences in its inputs, while if $\beta_j$ is small then the neural unit is essentially making a fuzzy decision.

It is important that the steepness parameters are learnt at a much slower rate than are the weights. This is because, as $\beta_j$ increases, $f'(\beta_j \text{net}_{pj})$ approaches zero for a wide range of net inputs and, thus, learning of weights is slowed considerably.

Despite this, the ability of neural units to learn a steepness parameter may prove extremely useful, particularly when inputs need to be treated differently. Furthermore, it enables neural networks to implement and integrate conventional and fuzzy logic. Used in conjunction with modifications designed to increase training speed by preventing $\beta_j$ from reaching zero or by modifying the error function (§7.7), the problems encountered with learning may be overcome.

### 8.4.2 Tuning Network Performance

The performance of a back-propagation neural network can be improved as required by modifying the error function. Typically, back-propagation minimizes the mean squared error:

$$E = \frac{1}{2} \sum_p \sum_j (t_{pj} - y_{pj})^2$$  \hspace{1cm} (8.12)

However, in some instances, it may be advantageous to weight the error of each input pattern by a factor $W_p$ as follows:

$$E = \frac{1}{2} \sum_p W_p \sum_j (t_{pj} - y_{pj})^2$$  \hspace{1cm} (8.13)

The weights $W_p$ may be used (1) to indicate the certainty with which the target output pattern $t_p$ is known, (2) to indicate the relative importance of correctly classifying each type of input pattern or (3) to ensure each type of input pattern contributes equally to the error measure even when there are different numbers of training examples for each output class. In this case the learning algorithm becomes:

$$\Delta w_{ji} = \eta \sum_p \delta_{pj} x_{pi}$$

$$\delta_{pj} = \begin{cases} W_p(t_{pj} - y_{pj}) f'(\text{net}_{pj}) & \text{for output layer} \\ f'(\text{net}_{pj}) \sum_k \delta_{pk} w_{kj} & \text{for hidden layers} \end{cases}$$  \hspace{1cm} (8.14)

Alternatively, the situation may arise where elements of the output pattern do not hold equal importance and the following error function can be used:

$$E = \frac{1}{2} \sum_p \sum_j W_j(t_{pj} - y_{pj})^2$$  \hspace{1cm} (8.15)
Finally, the case may arise where output elements hold different importance depending on the type of input pattern. For example, for a given input pattern some elements of the output pattern may be irrelevant (i.e., "don't cares"). In this situation, the error function becomes:

\[ E = \frac{1}{2} \sum_p \sum_j W_{pj} (t_{pj} - y_{pj})^2 \]  

Again the only change in the learning algorithm is in the calculation of \( \delta_{pj} \) for neural units of the output layer:

\[ \delta_{pj} = \begin{cases} W_{pj} (t_{pj} - y_{pj}) f'([net]_{pj}) & \text{for output layer} \\ f'([net]_{pj}) \sum_k \delta_{pk} w_{kj} & \text{for hidden layers} \end{cases} \]  

Thus, the error function can easily be altered to suit the application and to achieve the desired network performance.

8.5 DETECTION OF EPILEPTIFORM TRANSIENTS

Although the expert system approach to the detection of epileptiform activity in EEGs has proven very successful, development time has been excessive. This is a common problem with expert systems. For example, THEO is an expert system that forecasts solar flares. Although heralded as one of the most rapid developments of a usable expert system, THEO took over one person-year to develop. A back-propagation neural network with a single hidden layer was trained to perform the same task. THEONET was completed in less than a week and outperforms the expert system. Furthermore, THEONET requires only a few milliseconds to make a prediction whereas THEO requires approximately 5 minutes (Eberhart and Dobbins, 1990). For this reason, a growing number of neural networks are replacing more costly, slower expert systems.

Expert systems provide a logical, symbol processing approach to solve problems. Thus, expert systems attempt to reason their way to a solution and excel at tasks that are inherently cognitive. On the other hand, neural networks are able to generalize from specific examples, extracting general principles from training data rather than memorizing individual examples. One of the most promising application areas for neural networks is the field of pattern recognition.

Because of the different abilities of expert systems and neural networks, a hybrid system may provide the optimal solution to many problems. Whereas some aspects of a problem may lend themselves to a rule-based or expert system approach, others may be better handled by a neural network.

The detection of epileptiform activity in EEGs is essentially a pattern recognition task and, therefore, is an appropriate problem for a neural network. However, the classification of epileptiform events as focal or non-focal can readily be implemented by a few rules based on the polarity and spatial relationships between constituent epileptiform transients. Consequently, a hybrid system consisting of a neural network followed by a rule-based system may provide the best solution to the task.

8.5.1 Neural Network Approaches

A number of researchers have examined the possibility of using neural networks to detect specific EEG waveforms. Jansen (1990) attempted to train a back-propagation neural network...
to detect K-complexes, which are relatively large waves of 500–1500 ms in duration that arise during stage II sleep and are usually accompanied by sleep spindles. EEGs were low-pass filtered to 30 Hz, sampled at 100 Hz and down sampled by a factor of two. A network with three hidden layers was presented with 10 seconds of raw EEG data from a single channel and was required to decide whether or not there was a K-complex within this window. After 3600 cycles through the training set, output units still produced values of close to 0.5 and training was discontinued. In another attempt, EEG data were band-pass filtered between 0.5–2.3 Hz to capture K-complexes and between 10–17 Hz to capture sleep spindles. A 40 element input vector was derived from these data and presented to a network with either two or three layers. After training, the network correctly classified 95% of the training data, but on test data only classified 42–67% of the patterns correctly. Not surprisingly, Jansen concluded that the neural network approaches explored were not adequate to detect K-complexes. He suggested that the poor performance of the network may be due to the fact that the network was required to be insensitive to the location of K-complexes within a 10 second window, and proposed that improved performance may be achieved if the network was only required to detect K-complexes when they were more or less centred in the data window.

Eberhart et al. (1989a) have demonstrated the potential of neural networks for detecting epileptiform transients in EEGs. Back-propagation neural networks with a single hidden layer were trained to detect epileptiform transients in individual EEG channels. Although the network was initially required to detect whether there was an epileptiform transient within a 240 ms window, this approach was discontinued because it required the network to be insensitive to the location of epileptiform transients with durations of less than 240 ms. Instead they decided to centre candidate epileptiform transients within a 240 ms window. A network with 8 hidden units and 2 output units was trained on data from three patients and tested on novel data from the same three patients. The network detected 94–97% of epileptiform transients with a selectivity of 64–76%. In another approach, a network with 4 hidden units and 2 outputs was trained to detect epileptiform transients on the basis of nine parameters of individual waves. On the same data, this network detected 88–100% of epileptiform transients with a selectivity of 63–96%.

The promising nature of these results led to the development of a multi-channel version. Raw data from four EEG channels were presented to a network with 16 hidden units and 2 output units. The network was trained on 80 patterns from a single EEG and, after training, the network detected 95% of epileptiform transients with 100% selectivity. The network also performed impressively on test data from the same subject with 96% of epileptiform transients being detected with a selectivity of 96%. Although the networks were only tested on data from the same patients used for training, their performance is very encouraging and illustrates the potential of neural networks in this field.

8.5.2 Preliminary Investigation

The system presented in Chapter 6 for the detection of epileptiform transients consists of three stages: data collection, feature extraction and event detection. The feature extraction stage essentially performs pattern recognition, detecting candidate epileptiform transients on individual channels. Although the feature extractor detects over 80% of epileptiform transients, it also reports many non-epileptiform waves. While the selectivity of the feature extractor can reach 80%, in some EEGs it is as low as 2%. Therefore, it was decided to investigate the possibility of replacing the feature extractor with a neural network. The neural network would then be required to detect at least 80% of epileptiform transients, but hopefully with greater selectivity.
Because of the difficulties encountered when neural networks are required to be insensitive to location, the network was only required to make a detection when an epileptiform transient was in the centre of the data window. Epileptiform transients have durations of up to 250 ms which, therefore, is the minimum amount of data required by the network. A 250 ms window of raw EEG data (50 data points at 200 Hz) was presented to a two layer back-propagation neural network. The network was trained on epileptiform transients and similar appearing background waves from 8 patients. Training was in a number of stages, beginning with easier examples and progressing to the more difficult. After training on the entire training set, the network was tested on data from other patients and detected 70% of epileptiform transients.

This preliminary study revealed that the ability of the network to generalize to novel patterns improved as more examples were included in training and, therefore, indicated that large training sets should be used. However, the overall detection rate of the network was only 70%, which suggests that it may be necessary to alter the error function (§8.4.2) to indicate that the detection of epileptiform transients is more important than the rejection of non-epileptiform waves, particularly when there are more examples available of non-epileptiform waves.

Much of the artifact rejection and event detection performed by the expert system relies on knowledge of the polarity and size of waves and on the activity on other channels. Therefore, if a neural network is to replace the feature extractor, it may need to provide contextual information to the expert system. Efforts are underway to develop a neural network that is able to provide the necessary information.

8.6 LIMITATIONS OF NEURAL NETWORKS

Neural networks effectively provide a ‘black box’ solution to a problem. Because the knowledge of the network is embodied in a large number of weights, it is difficult to determine exactly what the network has learnt.

Neural networks are not able to perform all tasks. For example, they are particularly poor at arithmetic and formal logic, and they are not good at storing and retrieving very large amounts of information with high accuracy. At present, the learning algorithms available do not lend themselves to teaching neural networks to store or recognize temporal sequences. In fact, such tasks are usually performed by mapping temporal patterns into spatial ones.

The theoretical and practical limitations of neural networks are not well understood. In particular, it is unclear what can and cannot be learnt by a given network, and what architectures are required for the acquisition of given sorts of skill. Although theorems show that three layer networks can separate arbitrary classes of input patterns, it is not known how to determine the architecture or set of weights required to solve a given problem.
Although neural networks designed for specific applications have produced promising results, even hardware implementations of these networks do not appear to have realized the full potential of adaptive parallel processing. One reason for this may be the simplistic models of neurons employed.

There is no doubt that an individual neuron has dynamics of great complexity which cannot be reproduced by conventional neural units. While the simplicity of conventional neural units is a computational advantage, it can also be considered a drawback preventing them from exhibiting the rich variety of behaviour observed in neurons.

Recently, it was discovered that relatively simple systems, such as nonlinear recursive equations, can produce a wide range of behaviour including constant, periodic and apparently random activity known as deterministic chaos. Consequently, such a system may provide the basis of a useful model of a neuron.

This Part introduces a neural unit based on the simple nonlinear recursive equation, known as the logistic map. While this neuron model, termed the versatile neural unit, retains characteristics of conventional neural units, it is also able to produce a wide range of behaviour, reminiscent of that observed in neurons. The versatile neural unit provides a means for introducing chaos into neural networks, which may enhance their performance and, at the same time, provide insights into the roles of chaos in the brain.

Again this Part begins with the necessary background material, introducing deterministic chaos and its properties in Chapter 9, before examining in Chapter 10 the evidence for and possible roles of chaos in the brain. The versatile neural unit is introduced in Chapter 11 and incorporated into neural networks in Chapter 12.
9.1 INTRODUCTION

In the early 1960s, Edward Lorenz developed a model of the weather which used 12 equations to express the relationships between temperature, pressure and wind speed. The model behaved in a recognizable manner, displaying familiar patterns over time with pressure rising and falling and air streams swinging north and south. The weather followed a general pattern but the repetitions were never quite exact. Given a particular starting point the weather would unfold in exactly the same way each time, because of the deterministic set of equations governing its behaviour. However, in 1961, Lorenz started midway through a run, giving the system initial conditions from an earlier printout. The new run should have exactly duplicated the old, but in fact the weather diverged rapidly. Lorenz realized that the computer memory stored 6 decimal places but only 3 appeared on the printout. The slightly different initial conditions resulted in the weather unfolding in a completely different way. Lorenz had accidentally discovered the phenomenon now known as deterministic chaos.

Deterministic chaos, or simply chaos, is the phenomenon whereby completely deterministic systems can generate complicated, apparently random behaviour, that is sensitively dependent on initial conditions and, therefore, unpredictable.

This Chapter provides an introduction to deterministic chaos, with a number of chaotic systems being used to illustrate its properties. Particular attention is paid to the logistic map which forms the basis of the versatile neural unit introduced in Chapter 11. A number of quantitative measures which can be used to characterize chaotic systems are then discussed and algorithms that enable these quantities to be estimated from a single scalar time series introduced. The chief merit of these algorithms is their ability to distinguish between chaotic and random signals. Consequently, they have been used to establish the presence of chaos in the brain (Chapter 10).

9.2 THEORETICAL BACKGROUND

Before continuing it is necessary to discuss some theoretical concepts that arise in the study of chaos. The larger framework from which chaos emerges is the theory of dynamical systems.
9.2.1 Dynamical Systems

A system is composed of a number of parts which, in some way, can be perceived as a single entity. For example, the solar system consists of the sun and the planets. A dynamical system is one that changes or evolves with time. The state of a dynamical system is described by its $M$ independent variables $x_1(t), x_2(t), ..., x_M(t)$, which can be represented as a state vector $x(t)$. For example, the state of a chemical system is described by the concentrations of the various reactants. A deterministic system is one where the entire past and future are uniquely determined by the present state.

Dynamical systems are often studied by developing mathematical models which describe some interesting facet of their behaviour. A discrete dynamical system generates a sequence of states $x(1), x(2), ..., x(N)$, and can be modelled by difference equations or discrete maps of the following form:

$$x(n+1) = F[x(n)] \quad (9.1)$$

where the next state of the system $x(n+1)$ depends only on the present state $x(n)$. For example, biological populations with non-overlapping generations can be modelled by the following difference equation:

$$y(n+1) = y(n) \exp \left[ r \left( 1 - \frac{y(n)}{C} \right) \right] \quad (9.2)$$

where $y(n)$ is the population of the $n$th generation, $r$ is the usual growth rate and $C$ is the maximum sustainable population or carrying capacity. The continuous dynamical systems considered here can be described by first order ordinary differential equations of the following form:

$$\dot{x}(t) = F[x(t)] \quad (9.3)$$

where $\dot{x}(t)$ is the time derivative of the state vector $x(t)$ (i.e., $\dot{x}(t) = dx(t)/dt$). For example,

![Figure 9.1. Pendulum](image)

the motion of a pendulum (Fig. 9.1) is described by:

$$\ddot{\theta}(t) + \frac{g}{l} \sin \theta(t) + f \dot{\theta}(t) = 0 \quad (9.4)$$

where $\theta(t)$ is the angle the mass $m$ makes with the vertical, $g$ is the gravitational acceleration, $f$ is the coefficient of friction and $l$ is the length of the pendulum. This equation can be rewritten as a set of first order differential equations by making the substitution $x_1(t) = \theta(t)$.
and \( x_2(t) = \dot{\theta}(t) \):

\[
\begin{align*}
\dot{x}_1(t) &= x_2(t) \\
\dot{x}_2(t) &= -\frac{g}{l} \sin x_1(t) - f x_2(t)
\end{align*}
\] (9.5)

Dynamical systems may be either conservative or dissipative. *Conservative systems* conserve energy; for example, a frictionless pendulum \((f=0)\) once set in motion swings forever, repeating the same pattern (Fig. 9.2a). If the pendulum is perturbed (bumped) it adopts a new pattern of motion which it retains ever afterward, provided it suffers no further disturbance. On the other hand, *dissipative systems* dissipate energy. A pendulum with friction \((f > 0)\) dissipates energy in the form of heat. The friction damps the motion of the pendulum which eventually comes to rest (Fig. 9.2b). If the pendulum is supplied with an external driving force then it adopts the period of motion imposed by the force (Fig. 9.2c). In this situation, any disturbance to the motion of the pendulum is only transient, with the driving force quickly reasserting its characteristic pattern of motion.

\[\text{Figure 9.2. Motion of (a) a frictionless pendulum, (b) a pendulum with friction and (c) a pendulum with friction driven by an external force.}\]

*Linear systems* constitute a very small subset of dynamical systems. A linear system is one which can be expressed in the form:

\[
\dot{x}(t) = A x(t)
\] (9.6)
where $A$ is an $M \times M$ matrix with constant coefficients. The importance of linear systems is that their behaviour is completely understood. The solution of the general linear system (9.6) is given by:

$$x(t) = e^{At}x(0) \quad (9.7)$$

where $x(0)$ is the initial state of the system and $e^{At}$ is defined by the expansion:

$$e^{At} = \sum_{k=0}^{\infty} A^k t^k / k! \quad (9.8)$$

where $A^0$ is the unit matrix. The properties which make linear systems relatively simple to understand are (a) that the sum of any two solutions is also a solution (i.e., the principle of superposition) and (b) that any constant multiple of a solution is also a solution. Therefore, knowing the behaviour of solutions for a small set of initial conditions near the origin implies knowledge of the behaviour of the system for all initial states.

All systems which cannot be expressed in the form of (9.6) are nonlinear. Most physical systems are nonlinear but, for the sake of simplicity, are often modelled by linear equations. Because for small angles $\sin \phi \approx \phi$, low amplitude oscillations of a pendulum can be approximated by the linear system:

$$\begin{align*}
\dot{x}_1(t) &= x_2(t) \\
\dot{x}_2(t) &= -\frac{g}{l} x_1(t) - f x_2(t)
\end{align*} \quad (9.9)$$

For a frictionless pendulum ($f=0$), the result of this linear approximation is that the period of motion appears to be independent of its amplitude. Galileo discovered experimentally that the period of a pendulum was independent of the amplitude of motion. However, this regularity is only an approximation, with the resultant error being small for low amplitude oscillations but increasing with amplitude.

It is not uncommon for experimenters, in their search for regularity, to disregard nonlinearities, ignoring any slight deviations from linearity as experimental error. However, there are many important phenomena which cannot be explained by analysis of linear approximations.

### 9.2.2 Phase Space

The instantaneous state of a system can be represented in an $M$-dimensional Euclidean space by a single point with co-ordinates $x_1(t), x_2(t), \ldots, x_M(t)$ which correspond to the components of the state vector $x(t)$. This $M$-dimensional space is known as the state space or phase space of the system and represents the set of all possible states. The state of a dynamical system changes with time and so the evolution of the system maps out a trajectory in phase space. The dynamics of a system can, therefore, be represented by the collection of all trajectories, which is known as the flow. Geometrically, the flow can be visualized from a phase space portrait consisting of only a few pertinent trajectories.

The pendulum of (9.5) has a two-dimensional phase space with axes $x_1=\theta$ and $x_2=\dot{\theta}$. The phase space portrait of a frictionless pendulum is shown in Fig. 9.3a; the closed curves correspond to the pendulum swinging backwards and forwards, while the on-going curves represent complete circles. As $t \to \infty$, the trajectories of a pendulum with friction approach the equilibrium points at $x_1 = 2\pi n$, $x_2 = 0$ ($n=\ldots - 1, 0, 1\ldots$); Fig. 9.3b shows trajectories approaching such an equilibrium point. For a driven pendulum with friction all trajectories approach the closed curve corresponding to the periodic motion imposed by the driving force (Fig. 9.3c).
9.2.3 Attractors

The previous examples of the motion of pendulums illustrate some of the differences between conservative and dissipative systems. The frictionless pendulum is a conservative system and its phase space portrait shows that the pattern of motion depends on the initial state of the system. On the other hand, trajectories of dissipative dynamical systems tend to converge (as $t \to \infty$) to a subset of phase space. The dissipation of energy causes trajectories to be attracted to a finite region of phase space known as an attractor. Thus, an attractor describes the non-transient, asymptotic behaviour of a system. A system may have several attractors, in which case different initial states may converge to different attractors. The set of points that evolve to a certain attractor constitute its basin of attraction. Several definitions have been proposed for attractors (Milnor, 1985) but, although no single definition has been universally adopted, most include the following features (Ruelle, 1981):

- Invariance: points on the attractor remain on the attractor for all time,

- Compactness: the attractor is of finite size and has a boundary,
- Attractivity: a neighbourhood or basin of attraction exists where trajectories converge to the attractor as $t \to \infty$,
- Irreducibility: the attractor cannot be separated into disjoint sets.

There are three conventional types of attractor:

1. A fixed point attractor represents a constant state; for example, a pendulum with friction has fixed point attractors at $x_1 = 2\pi n, x_2 = 0$ where $n = \ldots -2, -1, 0, 1, 2, \ldots$

2. A periodic attractor, as its name suggests, represents periodic motion and is a closed curve in phase space. For example, the attractor of a driven pendulum with friction is a periodic attractor.

3. A quasi-periodic attractor or $q$-torus represents quasi-periodic motion which arises when a system contains $q$ independent frequencies. Although motion on the attractor is aperiodic, it is predictable because trajectories that start close together remain close together.

### 9.3 THE DISCOVERY OF CHAOS

In contrast to linear systems, the behaviour of nonlinear systems is not completely understood. Nonlinear systems can exhibit complicated, seemingly random behaviour that is sensitively dependent on initial conditions and, therefore, unpredictable. This is the phenomenon known as deterministic chaos.

Jules Henri Poincaré was the first to realize the existence of chaos, writing in 1921:

"... it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible..."

However, after his death this finding was virtually forgotten, and was not rediscovered until the 1960s when Lorenz realized that his model of the weather displayed a sensitivity to initial conditions. Lorenz's work was published in meteorology journals and, consequently, went largely unnoticed by those physicists and mathematicians who could have appreciated its significance. As a result deterministic chaos did not become an area for research until the 1970s when scientists took a renewed interest in irregularity and disorder.

The delay in realizing the significance of chaos was probably due to its counter-intuitive nature. As Robert May states:

"Most of our intuitions formed in mathematics and physics courses are that if you have a fairly simple description of something then it's going to describe fairly simple behaviour." (Gleick, 1984)

But he goes on to explain that, in fact, exactly the opposite is true of chaos — very simple equations can generate complicated and apparently random behaviour. However, underlying the randomness there is order; patterns repeat themselves but are never quite the same twice.

With the discovery that simple systems can produce complicated, unpredictable behaviour came the realization that chaos is everywhere we look — in the turbulent flow of water, in
the changing weather patterns, in the motion of planets, in biological populations, in the spread of epidemics, in chemical reactions ... — and the hope that many complex natural phenomena could be simplified and understood.

The discovery of chaos changed the thinking that precise predictability could be achieved, that it was only necessary to gather and process enough information. In 1776 Pierre-Simon Laplace expressed his view that the universe was strictly deterministic and completely predictable:

"The present state of the system of nature is evidently a consequence of what it was in the preceding moment, and if we conceive of an intelligence which at a given instant comprehends all the relations of the entities of this universe, it could state the respective positions, motions and general affects of all these entities at any time in the past or future." (Crutchfield et al., 1986)

However, 20th century science has seen both quantum mechanics and chaos question this philosophy. A central dogma of quantum mechanics is the Heisenberg uncertainty principle, which states that there is a fundamental limit to the accuracy with which the position and velocity of a particle can be measured. The more accurately the position of a particle is known the more uncertain is its velocity and vice versa. The Heisenberg uncertainty principle means that initial measurements are always inaccurate. Chaos ensures that any error or imprecision in the measurement of the state of a system increases exponentially with time — long-term prediction is impossible.

9.4 Chaotic Systems

Chaos manifests itself whenever certain types of nonlinearities govern the behaviour of a system. The following examples of relatively simple nonlinear systems that produce chaotic solutions serve to introduce some of the properties of chaos.

9.4.1 Lorenz Model

Having discovered that his model of the weather displayed a sensitive dependence on initial conditions, Lorenz searched for simpler systems with similar properties. In 1963 he published a simple model of thermal convection:

\[
\begin{align*}
\dot{x}(t) &= \sigma [y(t) - x(t)] \\
\dot{y}(t) &= -x(t)z(t) + r x(t) - y(t) \\
\dot{z}(t) &= x(t)y(t) - b z(t)
\end{align*}
\] (9.10)

Lorenz solved this system of first order differential equations by numerical integration and showed that, for certain ranges of parameters \((\sigma, b, r)\), the solutions were aperiodic (Fig. 9.4a). Plotting the solutions in phase space revealed the intricate structure of the system (Fig. 9.4b); two mutually connected spirals were traced out such that the outer portions of one spiral subsequently mapped to more inner parts of the other spiral. The state of the system converged to a finite subset of phase space, the attractor, within which neighbouring trajectories diverged. This type of attractor has since been termed strange (§9.5).
Figure 9.4. Lorenz model for $\sigma = 10$, $b = 8/3$ and $r = 28$, (a) aperiodic time series and (b) projection of Lorenz attractor onto $xz$-plane.
Figure 9.5. Rössler system for $a = 0.15$, $b = 0.2$ and $c = 10$, (a) aperiodic time series and (b) projection of Rössler attractor onto $(x+z)y$-plane.
Figure 9.6. Hénon map for $a = 1.4$ and $b = 0.3$ (a) aperiodic time series (b) Hénon strange attractor reveals its self-similar structure, with (iii), which is the inset of (ii), having the same structure as (ii).
9.4.2 Rössler System

Otto Rössler (1976), having noted that the Lorenz model contained 2 nonlinear terms, developed a simpler set of equations with a single nonlinear term:

\[
\begin{align*}
\dot{x}(t) &= -[y(t) + z(t)] \\
\dot{y}(t) &= -x(t) + a y(t) \\
\dot{z}(t) &= b + z(t)[x(t) - c]
\end{align*}
\]  

(9.11)

This system still produces chaotic solutions (Fig. 9.5a), but the resultant attractor consists of only a single spiral, where the outer portions of the spiral undergo a twist and return to more inner parts (Fig. 9.5b).

9.4.3 Hénon Map

Michel Hénon also searched for ways to simplify the Lorenz system but maintain its chaotic properties. In 1976 he published a two-dimensional discrete map:

\[
\begin{align*}
x(n+1) &= y(n) + 1 - a x(n)^2 \\
y(n+1) &= b z(n)
\end{align*}
\]  

(9.12)

which produced chaotic solutions (Fig. 9.6a) for certain values of parameters \(a\) and \(b\). Furthermore, Hénon was able to show that the resultant attractor had infinite levels of detail and possessed a self-similar structure (Fig. 9.6b).

9.4.4 Logistic Map

Probably the simplest chaotic systems are nonlinear recursive equations. The nonlinear recursive loop of Fig. 9.7 generates a sequence of numbers \(y(n)\) \((n = 0, 1, \ldots, N)\). When \(y(n)\)

\[
\begin{align*}
\text{Figure 9.7. A recursive loop with nonlinearity } f[\cdot], \text{ gain } g \text{ and unit time delay } \tau \text{ produces a sequence of numbers } y(n) \text{ } (n = 0, 1, 2, \ldots N).
\end{align*}
\]

appears on the left hand side of the nonlinearity \(f[\cdot]\), \(y(n+1)\) immediately appears on the right. After a unit time delay \(\tau\), \(y(n+1)\) is transferred to the left of the nonlinearity and \(y(n+2)\) appears on the right. The sequence is started by inserting an arbitrary \(y(0)\) on the left. Successive values in the sequence are related by a nonlinear function \(f[\cdot]\) as follows:

\[
y(n+1) = f[g, y(n)]
\]  

(9.13)
where \( g \) is a gain parameter.

For certain nonlinear functions \( f[\cdot] \), recursive loops are capable of displaying deterministic chaos (May, 1976), the most well-known of which is the logistic map:

\[
y(n+1) = f[g, y(n)] = 4 \cdot g \cdot y(n) \cdot [1 - y(n)]
\]  

(9.14)

This first order difference equation has been extensively studied as a model of biological populations with non-overlapping generations, where \( y(n) \) is the normalized population of the \( n \)th generation and \( g \) is a parameter which determines the dynamics of the population.

This very simple system is capable of displaying rich dynamics, producing chaotic activity for \( 0.89 < g \leq 1.0 \). The sensitive dependence on initial conditions of chaos is illustrated in Fig. 9.8. After 10 iterations the solutions are seen to be following very different paths even though the difference in initial values was only 0.001.

The behaviour of the logistic map changes dramatically as the parameter \( g \) is altered. When \( g < 0.25 \) the output or population dies away to zero (Fig. 9.9a). For \( 0.25 < g < 0.75 \) the output converges to a single non-zero value (Fig. 9.9b). As \( g \) is increased beyond 0.75 the output begins to oscillate between first 2 values (Fig. 9.9c), then 4 values, then 8 values ... until for \( g > 0.89 \) the output becomes chaotic (Fig. 9.9d).

The way in which the behaviour of the system depends on the gain \( g \) can be displayed as a bifurcation diagram (Fig. 9.10), which clearly shows the period-doubling route to chaos. As \( g \) is increased, the output, initially a constant value, becomes a period-2 cycle, then period-4, then period-8 ... until finally it is highly irregular and aperiodic. The system attractor undergoes a series of pitchfork bifurcations (Fig. 9.11), each of which results in a doubling of the orbit period. The bifurcations occur at well defined parameter values \( g_1, g_2, \ldots, g_\infty \) and accumulate at a particular parameter value \( g_\infty = 0.89248 \ldots \), after which chaotic behaviour is observed.
Figure 9.9. Behaviour of logistic map for (a) $g = 0.2$, (b) $g = 0.7$, (c) $g = 0.85$ and (d) $g = 0.95$. 
Mitchell Feigenbaum (1978) studied this period doubling behaviour and found that, for sufficiently high periods, the structure of one cycle (period-2048, say) is indistinguishable from that of the previous cycle (in this case 1024), provided that the appropriate scaling is performed. The period doublings display a self-similar structure with a scaling in $x$ of $\alpha = 2.502908...$ and in $g$ of $\delta = 4.66920166...$. As a result, it is possible to predict the value of $g$ at which the next period doubling occurs and the new values of $x$ that the sequence adopts. Furthermore, Feigenbaum discovered that the scaling parameters $\delta$ and $\alpha$ are universal for all systems that can be approximately modelled by a nonlinear map with a quadratic extremum.

Figure 9.10. Bifurcation diagram for the logistic map showing the dependence of the activity produced on the value of gain $g$.

Figure 9.11. Pitchfork bifurcation resulting in a doubling of the orbit period.
9.5 STRANGE ATTRACTORS

The evolution of a dissipative system is governed by a phase space attractor. The examples of the previous section illustrate that not all attractors are of a conventional form (i.e., fixed point, periodic or quasi-periodic). In the literature, attractors which are not fixed points, periodic or quasi-periodic are often termed strange. However, this includes both attractors that exhibit a sensitive dependence on initial conditions and those that have a non-integer dimension. The distinction between these two categories is not always clear, probably because most known attractors that are sensitively dependent on initial conditions also have a non-integer dimension. A counter example is the hyperbolic toral automorphism (Devaney, 1987) which displays a sensitive dependence on initial conditions but has an integer dimension. Similarly, not all attractors with a fractional dimension are sensitively dependent on initial conditions (Grebogi et al., 1984). Throughout this thesis the term strange attractor refers to attractors exhibiting sensitivity to initial conditions. The ‘strangeness’ of an attractor is, therefore, due to its dynamics or flow rather than to its geometry in phase space.

Therefore, chaotic motion evolves on a strange attractor. As $t \to \infty$ trajectories converge to the attractor but nearby trajectories diverge at an exponential rate (Fig. 9.12); to remain finite these diverging trajectories must eventually fold back on themselves. A strange attractor results from a balance between an unstable growth process (stretching) and a stable limiting process (folding). The combination of these processes produces a complex geometrical structure with infinite levels of detail (as was illustrated by Fig. 9.6b).

![Figure 9.12. Divergence of nearby trajectories within the Rössler attractor.](image)

Strange attractors which exist in phase spaces of more than three dimensions are particularly difficult to visualize. It is, therefore, necessary to derive quantitative measures which enable recognition, characterization and classification of strange attractors and, hence, chaos — dimension, entropy and Lyapunov exponents are commonly used for this purpose.

9.5.1 Attractor Dimension

The phase space dimension $M$ of a system is the number of independent variables needed to specify an arbitrary initial condition. The dimension of an attractor is often considerably
less than that of the phase space in which it resides (Farmer, 1982) and can be thought of as a measure of the amount of space it occupies (Falconer, 1990). Strange attractors have a complex geometrical structure and, therefore, calculation of their dimension is not straightforward.

Benoit Mandelbrot (1977) studied objects that reveal more and more detail as they are magnified. He termed such objects fractals and found that they tend to display similar patterns over different scales. Because measures of classical geometry (length, depth, thickness) are unable to describe the complexity of these objects, Mandelbrot turned to calculating their dimension, introducing the notion of non-integer or fractional dimension to describe the degree of irregularity. Because the structure of conventional attractors is very regular, they have integer dimensions; for example, the dimension of a fixed point is 0, of a periodic attractor is 1 and of a q-torus is q. However, strange attractors have a complex geometrical structure and their dimension may be fractional.

There are an infinite number of ways to generalize dimension to the fractional case, including the similarity dimension, fractal dimension, information dimension and correlation dimension.

The similarity dimension can be relatively simply explained. If one takes a geometrical shape, say a cube, and doubles its linear size in each spatial direction, then one obtains an object whose volume is greater than that of the original, in this case $8 = 2^3$ times greater. In general, taking an object of dimension $D$ and increasing its linear size in each spatial direction by a factor of $I$ results in a volume increase of $k = I^D$ times. Thus, the similarity dimension can be defined as:

$$D = \frac{\log k}{\log I} \quad (9.15)$$

which need not necessarily be an integer. The classical example of an object with a non-integer dimension is the Cantor set (Fig. 9.13), which has a similarity dimension $D = 0.6309...$. However, the similarity dimension is only meaningful for the small set of objects that display strict self-similarity (Falconer, 1990).

The fractal dimension $D_F$ of an attractor is universally applicable and is a geometric measure of dimension. It can be defined by partitioning phase space into $M$-dimensional hypercubes of side $\epsilon$ and counting the number $N(\epsilon)$ of hypercubes through which the attractor passes. The fractal dimension is then given by:

$$D_F = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log 1/\epsilon} \quad (9.16)$$

Figure 9.13. A Cantor set can be generated by considering the unit interval $[0..1]$, removing the central third $[1/3..2/3]$ and repeating this process on the remaining line segments again and again. The dimension of the limiting set can be determined by considering the interval $[0..1/3]$ and tripling its linear size. The result is two copies of the original and, therefore, the dimension is $D = \log 2/\log 3 = 0.6309...$. The fractal dimension $D_F$ of an attractor is universally applicable and is a geometric measure of dimension. It can be defined by partitioning phase space into $M$-dimensional hypercubes of side $\epsilon$ and counting the number $N(\epsilon)$ of hypercubes through which the attractor passes. The fractal dimension is then given by:

$$D_F = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log 1/\epsilon} \quad (9.16)$$

Figure 9.13. A Cantor set can be generated by considering the unit interval $[0..1]$, removing the central third $[1/3..2/3]$ and repeating this process on the remaining line segments again and again. The dimension of the limiting set can be determined by considering the interval $[0..1/3]$ and tripling its linear size. The result is two copies of the original and, therefore, the dimension is $D = \log 2/\log 3 = 0.6309...$.
Many, but not all, strange attractors have a non-integer fractal dimension. For example, the fractal dimension of the Hénon attractor is $D_F=1.264$, for the Lorenz attractor $D_F=2.07$ and for the Rössler attractor $D_F=2.01$.

The fractal dimension does not take into account the density of points on the attractor (i.e., the number of points in each of the hypercubes). This inexactitude is corrected by the information dimension $\sigma$, which is defined to be:

$$\sigma = \lim_{\epsilon \to 0} \frac{I(\epsilon)}{\log 1/\epsilon}$$

$$I(\epsilon) = - \sum_{i=1}^{N(\epsilon)} P_i \log P_i$$  \hspace{1cm} (9.17)

where $N(\epsilon)$ is again the number of hypercubes of side $\epsilon$ required to cover the attractor and $P_i$ is the probability of a trajectory passing through hypercube $i$. When the density of points on the attractor is uniform each hypercube has equal probability and $\sigma = D_F$, otherwise $\sigma < D_F$.

The correlation dimension $\nu$, introduced by Grassberger and Procaccia (1983a), is closely related to both the fractal dimension $D_F$ and the information dimension $\sigma$, belonging to the same family of generalized dimensions $D_q$ (Grassberger and Procaccia, 1984):

$$D_q = \lim_{\epsilon \to 0} \frac{I_q(\epsilon)}{\log 1/\epsilon}$$

$$I_q(\epsilon) = \frac{1}{1-q} \log \sum_{i=1}^{N(\epsilon)} P_i^q$$  \hspace{1cm} (9.18)

where $\lim_{q \to 0} D_q = D_F$, $\lim_{q \to 1} D_q = \sigma$ and $D_2 = \nu$. In general, $D_{q+1} \leq D_q$ with equality when the attractor is uniform. Therefore, the correlation dimension is a lower bound on both the fractal and information dimensions.

### 9.5.2 Entropy

Entropy is a measure of the amount of information that a system produces. The entropy of a discrete system is usually expressed in bits per symbol. The information contained in a symbol $s_j$ which occurs with probability $P_j$ is defined to be:

$$\text{I}(s_j) = - \log_2 P_j$$  \hspace{1cm} (9.19)

Thus, the rarer a symbol the more information is conveyed by its occurrence. The metric entropy of a system is the average information produced per symbol and is given by:

$$K = - \sum_{j=1}^{N} P_j \log_2 P_j$$  \hspace{1cm} (9.20)

The entropy of a system can be expressed in bits per second by multiplying by the data rate in symbols per second.

The entropy vanishes only if there is complete predictability and is limited by the number of possible symbols $N$ ($K \leq \log_2 N$). Therefore, entropy can be considered as a measure of unpredictability.

The behaviour of systems governed by fixed point, periodic or quasi-periodic attractors is completely predictable and, hence, $K = 0$. Despite being completely deterministic, chaotic systems are only predictable on short time scales and, therefore, have a positive entropy.
The metric entropy \( K \) can be calculated by considering the trajectory \( x(t) \) of a dynamical system on an attractor, which has been sampled at equally spaced time instants \( t = nT \) to produce a sequence \( x(n) \). By partitioning phase space into \( M \)-dimensional hypercubes of side \( E \) the quantity \( P(i_1, i_2, \ldots, i_d) \) can be defined to be the joint probability that \( x(1) \) is in hypercube \( i_1 \), \( x(2) \) is in hypercube \( i_2 \), ..., and \( x(d) \) is in hypercube \( i_d \). The entropy is then given by (Grassberger and Procaccia, 1983b):

\[
K = - \lim_{T \to 0} \lim_{\tau \to 0} \lim_{d \to \infty} \frac{1}{d} \sum_{i_1, \ldots, i_d} P(i_1, i_2, \ldots, i_d) \log_2 P(i_1, i_2, \ldots, i_d)
\]

The entropy of a system is a measure of the rate at which information is created or destroyed and, therefore, determines the time scale on which prediction is possible. For example, the Lorenz system has an entropy of 0.91 bits/sec for parameter values \( \sigma = 10, b = 8/3, r = 28 \) (Terenina, 1984). Therefore, if the initial state was specified to 1 part per million (i.e., 20 bits), the future behaviour could not be predicted after about 22 (=20/0.91) seconds. After this time the small initial uncertainty essentially covers the entire attractor and all that can be said about the state of the system is that it lies somewhere on the attractor.

The metric entropy \( K \) is just one member of a family of generalized entropies \( K_q \) (Grassberger and Procaccia, 1983b) which can be defined as:

\[
K_q = - \lim_{T \to 0} \lim_{\tau \to 0} \lim_{d \to \infty} \frac{1}{d} \log_2 \sum_{i_1, \ldots, i_d} P^q(i_1, \ldots, i_d)
\]

The metric entropy is given by \( K = \lim_{q \to 1} K_q \) and the relationship \( K_{q+1} \leq K_q \) exists between consecutive members.

9.5.3 Lyapunov Exponents

The Lyapunov exponents of a system measure the average rate of expansion or contraction of phase space in a particular direction. A system with an \( M \)-dimensional phase space has a spectrum of \( M \) Lyapunov exponents, which describe the local stability properties of the phase space. These stability properties can be determined by examining the behaviour of the system under small perturbations. A system may be stable to perturbations in a certain direction but unstable to those in other directions. The effect of all possible perturbations can be examined simultaneously by following the evolution of points in a small \( M \)-dimensional hypersphere of radius \( \epsilon(0) \). This hypersphere evolves into an ellipsoid with principal axes \( \epsilon_i(t) \). The Lyapunov exponents \( \lambda_i \), usually expressed in bits/sec, are then defined to be (Farmer, 1982; Wolf et al., 1985):

\[
\lambda_i = \lim_{t \to \infty} \lim_{\epsilon(0) \to 0} \frac{1}{t} \log_2 \frac{\epsilon_i(t)}{\epsilon(0)}
\]

where \( \lambda_i \) can be ordered such that \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M \). Since the orientation of the ellipsoid changes continuously as it evolves, the directions associated with each Lyapunov exponent vary in a complicated way. Positive Lyapunov exponents describe the divergence of points in phase space, while negative ones correspond to convergence. The sum of all Lyapunov exponents is the average divergence of phase space, which is zero for conservative systems and negative for dissipative ones.

For a fixed point attractor all Lyapunov exponents are negative (\( \lambda_i < 0, i = 1, \ldots, M \)) indicating that phase space contracts in all directions to become a single point. All attractors which are not fixed points have at least one zero Lyapunov exponent because motion on
the attractor has neutral stability (i.e., remains on the attractor). For example, a periodic attractor has one zero exponent while all other exponents are negative ($\lambda_1=0$, $\lambda_i<0$, $i=2...M$) and, similarly, a 2-torus has two zero exponents ($\lambda_1=\lambda_2=0$, $\lambda_i<0$, $i=3...M$). A strange attractor is characterized by at least one positive Lyapunov exponent indicating that nearby trajectories diverge in at least one direction. Therefore, in three dimensions a chaotic system would have one positive, one zero and one negative exponent ($\lambda_1>0, \lambda_2=0, \lambda_3<0$). The positive Lyapunov exponents measure the rate at which neighbouring trajectories diverge and, thus, reflect the rate at which predictability is lost. The magnitude of the negative Lyapunov exponents determine the rate at which trajectories converge to the attractor and correspond to the rate at which transient information is lost.

The Lyapunov exponents are believed to be related to other global properties of attractors. Kaplan and Yorke (1979) argued that a meaningful concept of dimension can be defined in terms of the spectrum of Lyapunov exponents of a system. This measure is known as the Lyapunov dimension $D_L$ and is defined to be:

$$D_L = \frac{1}{j} \sum_{i=1}^{j} \lambda_i$$

(9.24)

where $j$ is the largest integer such that $\lambda_1 + \lambda_2 + \ldots + \lambda_j > 0$. Initially, Kaplan and Yorke conjectured that the Lyapunov dimension was equal to the fractal dimension ($D_L = D_F$), but it has proven to be more closely related to the information dimension $\sigma$. More recently the relationship $D_L \leq \sigma$ has been established (Grassberger and Procaccia, 1984) but numerical results suggest that for generic examples $D_L = \sigma$ (Frederickson et al., 1983).

The entropy of a system is also related to the spectrum of Lyapunov exponents. The sum of the positive Lyapunov exponents $\lambda_i^+$ provides an upper bound on the entropy $K$ (Grassberger and Procaccia, 1984):

$$K \leq \sum_i \lambda_i^+$$

(9.25)

Again, numerical results support the conjecture that, for generic examples, the metric entropy is equal to the sum of the positive Lyapunov exponents (Termonia, 1984).

9.6 DISTINGUISHING CHAOTIC FROM RANDOM SIGNALS

In a typical physical experiment, a single scalar variable $u(t)$ is monitored and its value recorded at equally spaced time instants $t=n\tau$ ($n=0...(N-1)$). From the behaviour of the resultant time series, scientists attempt to understand the underlying system. Traditionally, frequency analysis has been used to determine whether time series are periodic, quasi-periodic or random. All time series with broadband frequency spectra were considered to be produced by a random process. However, chaotic systems also generate time series with broadband frequency spectra. Therefore, methods are needed to distinguish chaotic from random activity and, ideally, these techniques would also be able to characterize chaotic systems.

Chaotic motion evolves on a strange attractor of finite dimensionality, while random behaviour occupies an arbitrarily large number of dimensions and does not evolve on an attractor. Therefore, an irregular time series with a broadband frequency spectrum may be considered deterministic if an attractor of relatively low dimension can be reconstructed from the data, and as chaotic if this attractor displays a sensitive dependence on initial conditions. Mathematicians have developed methods for reconstructing attractors from a single scalar time series and estimating their essential properties (i.e., dimension, entropy and
Lyapunov exponents). This Section introduces a number of commonly employed algorithms for distinguishing between chaotic and random activity. These algorithms have been used to establish the presence of chaotic activity in the brain (Chapter 10).

9.6.1 Reconstruction of Attractors

A multi-dimensional phase space can be reconstructed from measurements of the state of a system. Packard et al. (1980) proposed that the state of an $M$-dimensional system at any instant of time can be specified by the measurement of any $M$ independent quantities. Wherever possible, such quantities should be measured but they can be derived from a single time series if necessary (Packard et al., 1980; Eckmann and Ruelle, 1985).

In the experimental situation the number of quantities needed to specify an arbitrary state is not usually known and indeed may be infinite. The motion of the system is assumed to evolve on a finite dimensional attractor and the aim, therefore, is to reconstruct the system attractor rather than the entire phase space.

Whitney (1937) showed that, under certain general conditions, it is possible to embed a $d$-dimensional manifold in an $m$-dimensional embedding space provided $m \geq 2d + 1$. The result is an image of the manifold which possesses the same topology as the original manifold.

When only a single time series $u(t=nt)$ is available it is necessary to derive $m$ quantities $x_k(t)$ ($k = 1, \ldots, m$). Suitable quantities can be obtained in a variety of ways. Packard et al. (1980) proposed calculation of time derivatives (i.e., $x_k(t) = d^{k-1}u(t)/dt^{k-1}$) or use of time delayed versions of the original signal (i.e., $x_k(t) = u(t - (k-1)T)$). Takens (1981) mathematically justified derivation of quantities by both these methods, showing that the resulting reconstructions were topologically equivalent to the original attractor of the system. Thus, although an attractor reconstructed by these methods may not have visual similarity with the original (due to a nonlinear transform between the original system variables and those used for the reconstruction), it does have the same essential properties (e.g., dimension, entropy and spectrum of Lyapunov exponents).

The numerical estimation of time derivatives introduces errors into the reconstruction (Eckmann and Ruelle, 1985; Froehling et al., 1981). Therefore, an image of the system attractor is usually obtained by using time delayed versions of the original signal (e.g., Fig. 9.14). The state of the system is then represented by:

$$x(t) = [u(t), u(t-T), \ldots, u(t-(m-1)T)]$$

(9.26)

where $T$ is a fixed time delay and $m$ is the dimension of the embedding space, known as the embedding dimension.

The idea behind this method is that the evolution of any single variable of a system is determined by the other variables with which it interacts. Knowledge of these variables is implicitly contained in the history of any single variable. For an infinite amount of noise-free data, the time delay $T$ can, in principle, be chosen arbitrarily (Takens, 1981). In the experimental situation these conditions are not satisfied and, consequently, the choice of $T$ becomes important. If $T$ is chosen to be very small the presence of experimental noise means that $u(t)$ and $u(t-T)$ are indistinguishable and all trajectories appear to lie along the main diagonal of the embedding space (i.e., $x_1=x_2=\ldots=x_m$). Therefore, the time delay should be such that the derived quantities $x_k(t)$ are independent and a number of methods for choosing a suitable delay have been proposed (§9.6.5).

Before reconstruction the dimension of the attractor is unknown and, therefore, choice of a suitable embedding dimension $m$ is difficult. Roux et al. (1983) suggested increasing
Figure 9.14. A two-dimensional projection of the reconstructed Lorenz attractor with $T = 0.1s$ (a) reconstruction from $x$ and (b) reconstruction from $z$.

$m$ until no further structure appears in the reconstruction. However, this is very subjective and in higher dimensions or in the presence of noise becomes impractical (Broomhead and King, 1986). Typically, $m$ is increased until measures of the reconstructed attractor, such as dimension, saturate. If the activity is in fact random, no saturation will result; estimates of dimension, entropy and Lyapunov exponents will continue to increase as the embedding dimension is increased. Although Whitney’s embedding theorem states that $m \geq 2d + 1$, in practice, an unambiguous phase space reconstruction can be obtained for smaller $m$ ($m \geq d$) (Roux et al., 1983).

Once an attractor has been constructed which is topologically equivalent to the system attractor, its essential properties (dimension, entropy and spectrum of Lyapunov exponents) can be calculated.

### 9.6.2 Calculation of Lyapunov Exponents

Several algorithms have been proposed for the calculation of Lyapunov exponents from a reconstructed attractor. One of the most commonly used is that introduced by Wolf et al. (1985) to calculate the non-negative Lyapunov exponents of a system. Consider the evolution of points in a small $m$-dimensional ellipsoid. From the definition of Lyapunov exponents (9.23), it is clear that the linear extent of the ellipsoid grows as $2^{\lambda_1 t}$ and that the area defined by the first two principal axes of the ellipsoid grows as $2^{(\lambda_1 + \lambda_2)t}$. Thus, the sum of the first $j$ exponents is defined by the long-term evolution of a $j$-volume element. The largest Lyapunov exponent $\lambda_1$ can be estimated by effectively monitoring the long-term evolution of a single pair of nearby trajectories (Fig. 9.15). An initial point on the attractor is chosen and its nearest neighbour located. The distance between these two points is denoted by $L(t_0)$. The evolution of these points is followed for a time short enough so that only the small scale structure of the attractor is examined. At this time $t_1$ the initial length has evolved to $L'(t_1)$. A replacement point is chosen so that the replacement distance $L(t_1)$ is small and the angular separation between the original and replacement points is minimized. The largest Lyapunov exponent (bits/sec) is then calculated as:

$$\lambda_1 = \frac{1}{t_R - t_0} \sum_{j=1}^{R} \log_2 \frac{L'(t_j)}{L(t_{j-1})}$$ (9.27)
where $R$ is the total number of replacement steps. As the embedding dimension $m$ increases so do the data requirements of this algorithm, because the density of points on the attractor must be sufficient for suitable replacement points to be found.

Figure 9.15. Calculation of largest Lyapunov exponent by tracing the evolution of neighbouring trajectories on short time scales.

Other methods tend to be more computationally expensive but do enable the calculation of several Lyapunov exponents including some negative ones (Eckmann and Ruelle, 1985; Sano and Sawada, 1985). If a sufficient number of the Lyapunov exponents of the attractor can be calculated, then the Lyapunov dimension and an upper bound on the entropy can be determined from (9.24) and (9.25) respectively.

9.6.3 Calculation of Attractor Dimension

The fractal dimension $D_F$ of an attractor can be calculated numerically by a box counting algorithm. The embedding space is partitioned into $m$-dimensional hypercubes of side $\epsilon$ and the number of non-empty hypercubes $N(\epsilon)$ determined. Repeating this process for several $\epsilon$ allows the fractal dimension $D_F = \lim_{\epsilon \to 0} \log N(\epsilon)/\log(1/\epsilon)$ to be calculated from the slope of $\log N(\epsilon)$ vs $\log(1/\epsilon)$. This algorithm is computationally very expensive and requires vast amounts of data, particularly for large embedding dimensions $m$. Greenside et al. (1982) showed that it was impractical to calculate fractal dimensions by this method. Although alternative algorithms have been proposed (Liebovitch and Toth, 1989; Hunt and Sullivan, 1986), they have not been generally adopted. Computation of the information dimension is equally difficult with the number of data points in each hypercube also having to be taken into account.

Grassberger and Procaccia (1983a) demonstrated that of the infinite number of dimensions $D_\varphi$ that can be used to characterize a strange attractor, only the correlation dimension ($D_2 = \nu$) is relatively easy to calculate. Due to the exponential divergence of trajectories most pairs of points are dynamically uncorrelated, but because they lie on an attractor are spatially correlated. The correlation sum measures this spatial correlation and is defined to be:

$$C(r) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{n=1}^{N} \mu \left[ r - ||x(k) - x(n)|| \right]$$  \hspace{1cm} (9.28)

where $x(i)$ is the state of the system at the $i$th time instant, $N$ is the number of points of the attractor and $\mu[]$ is the unit step function. The correlation sum counts the number of points that lie within a radius $r$ of each other and, thus, measures the extent to which the presence of a given point affects the position of its neighbours. For small $r$, the correlation
sum scales as:

\[ C(r) \sim r^\nu \]  

(9.29)

where \( \nu \) is the correlation dimension. The correlation dimension can, therefore, be determined from the slope of the linear region of \( \log C(r) \) versus \( \log r \) (Fig. 9.16). Because \( D_{q+1} \leq D_2 \), the correlation dimension provides a lower bound on both the information dimension \( \sigma \) and the fractal dimension \( D_F \) but, in practice, the inequality has proven rather tight (Grassberger and Procaccia, 1983c).

![Figure 9.16. The correlation dimension can be determined from the slope of the linear region of \( \log C(r) \) versus \( \log r \).](image)

Noise is not spatially correlated and the correlation integral of noise scales as \( C(r) \sim r^m \). Therefore, the correlation dimension of a random signal continues to increase with increasing embedding dimension. Furthermore, the correlation integral of an attractor reconstructed from experimental data has two scaling regions. On very small length scales, where noise dominates, it scales as \( C(r) \sim r^m \), but on longer length scales \( C(r) \sim r^\nu \) (Franaszek, 1984; Ben-Mizrachi et al., 1984).

### 9.6.4 Calculation of Entropy

As with Lyapunov exponents and dimension, several algorithms have been proposed to compute entropy. Termonia (1984) proposed a variation on box counting, while Fraser (1986) estimated the entropy from the mutual information of the signal. Grassberger and Procaccia (1983b) found that of all the generalized entropies \( K_q \) only \( K_2 \) is relatively easy to calculate. The algorithm proposed for calculating \( K_2 \) is very similar to that employed to compute the correlation dimension, but instead of measuring the correlation between points,
the correlation between pieces of trajectories of length \( d \) is calculated as follows:

\[
C_d(r) = \lim_{d \to \infty} \frac{1}{N^2} \sum_{n=1}^{N} \sum_{k=1}^{N} \mu \left[ r - \left( \sum_{i=1}^{d} [x(n+i) - x(k+i)]^2 \right)^{1/2} \right]^{1/2}
\]

(9.30)

where \( k \) ranges over one trajectory and \( n \) over the other, \( x(k) \) is the state of the system at the \( k \)th sampling instant. The correlation between trajectories is equal to \( r \nu \exp(-d \tau K_2) \). Therefore, plotting \( \log_{10} C_d(r) \) vs \( \log_{10} r \) for several \( d \) produces lines with slope \( \nu \) displaced from each other by \( \exp(-d \tau K_2) \) and both \( K_2 \) and the correlation dimension \( \nu \) can be estimated.

### 9.6.5 Accuracy of Algorithms

In the event of an infinite amount of noise-free data, the algorithms presented accurately compute the correlation dimension, entropy and Lyapunov exponents of a strange attractor. However, in the experimental situation only a finite amount of data is available and it is unavoidably contaminated with noise. In fact, when considering EEG signals only very short segments of data can be used because the signal is stationary for short periods of time. Small, noisy data sets impose severe restrictions on the accuracy of quantities calculated (Abraham et al., 1986).

The accuracy to which the essential properties of an attractor can be estimated depends, first of all, on the reconstruction of the attractor. The accuracy with which the reconstructed attractor preserves the topology of the actual system attractor depends on both the embedding dimension \( m \) and the time delay \( T \) used to derive additional quantities from a time series. Although Whitney’s embedding theorem states that \( m \geq 2d + 1 \), where \( d \) is the dimension of the attractor, in practice an unambiguous reconstruction can be obtained for smaller \( m \), provided that \( m \geq d \) (Roux et al., 1983). However, \( d \) is unknown and so choice of a suitable embedding dimension is difficult. Typically, \( m \) is increased until measures of the reconstructed attractor saturate.

Choosing a suitable time delay \( T \) is equally difficult, but several methods have been proposed. Fraser and Swinney (1986) thought that \( T \) should be chosen so that \( u(t) \) and \( u(t-T) \) are independent. While the first zero crossing of the autocorrelation function corresponds to linear independence and may provide a suitable time delay, Fraser and Swinney suggested that the first local minimum of the mutual information would provide a more suitable value, because the mutual information measures the general dependence of two variables. Fraser (1989) extended this method to higher dimensions, employing redundancy analysis to obtain a measure of the general dependence of \( m \) variables. However, Destexhe et al. (1988) found that these methods do not always provide a satisfactory time delay and suggested using a fraction (e.g., 25%) of the pseudo-period of the signal (a similar approach was adopted by Froehling et al. (1981)). In another approach, Liebert and Schuster (1989) defined a generalized correlation integral and found that the first local minimum provides a useful value for the time delay.

Several groups have found that the choice of \( T \) and \( m \) are not independent and that an important parameter for obtaining reliable reconstructions is the \textit{window time} \( T_W = (m-1)T \) (Broomhead and King, 1986). The window time corresponds to the amount of data used to specify the state of the system.

Estimation of properties for a given reconstruction depends on the density of points on the attractor which in turn depends on the embedding dimension \( m \) and the length of the time series. The accuracy with which the Grassberger-Proccacia algorithm (§9.6.3) is able to estimate the correlation dimension has been extensively studied. Smith (1988) stated that to
determine the correlation dimension to an accuracy of 5%, a time series of at least $42^L$ points would be required, where $L$ is the greatest integer less than the dimension of the attractor. However, in practice, reasonable results can be obtained with considerably less data. An additional problem is that, for finite data sets, the Grassberger-Proccaccia algorithm produces a biased estimate of the correlation dimension. The correlation dimension of noise tends to be underestimated, while that of low-dimensional attractors is overestimated (Ramsey and Yuan, 1989). The bias increases with embedding dimension but decreases with increasing amounts of data. Several modifications have been proposed to minimize these bias effects (Dvořák and Klaschka, 1990; Theiler, 1986), but have not been generally adopted. Rapp et al. (1989) imposed several practical conditions on the convergence of the algorithm to minimize the possibility of obtaining spurious results:

1. the scaling region must be linear (e.g., less than 10% variation in slope),
2. the scaling region must be of a certain minimum length (e.g., a factor of 2 in $\log_e r$),
3. the calculated correlation dimension must be stable for increasing embedding dimension $m$ (e.g., less than 10% variation over 4 consecutive embeddings).

As embedding dimension $m$ increases the density of points on the attractor decreases and, consequently, the scaling region decreases. Recently, a technique known as singular value decomposition has been proposed that legitimately reduces the size of the embedding dimension required. After reconstructing the system attractor in an $n$-dimensional embedding space by the method of time delays, singular value decomposition is applied, which effectively rotates the attractor in this space. In so doing, orthogonal directions in the embedding space are identified which can be ordered according to their singular values (Albano et al., 1988). The singular value associated with a particular direction is a measure of the variance of the attractor in that direction. The spectrum of singular values gives a crude indication of the distribution of the attractor in the embedding space (Broomhead et al., 1987). The dimension of the smallest space that contains the attractor can be estimated from the number of ‘large’ singular values. In general, most of the information is concentrated in $m$ dimensions and the remaining $n - m$ components can be discarded with minimal information loss. Because noise is evenly distributed throughout the embedding space, the discarded dimensions correspond to those with the worst signal-to-noise ratios. While the attractor is concentrated in $m$ dimensions, the remaining $n - m$ dimensions are dominated by noise. Thus, singular value decomposition also improves signal-to-noise ratios (Landa and Rosenblum, 1991).

9.7 SUMMARY

Deterministic chaos is the term used to describe activity produced by a completely deterministic system, that is sensitively dependent on initial conditions. Such activity is seemingly random, but underlying the disorder there is order — patterns recur but never exactly repeat themselves. Although chaotic activity is very complicated, it can be produced by extremely simple nonlinear equations, such as the logistic map.

Despite its deterministic origins, chaos is only predictable on short time scales. Therefore, chaotic activity has a positive entropy and is constantly producing information.

Chaotic motion evolves on a strange attractor which can be characterized by its dimension, entropy and Lyapunov exponents. A number of algorithms have been proposed to detect strange attractors and estimate their essential properties. Application of these algorithms to EEG signals has established the presence of chaotic activity in the brain (Chapter 10).
CHAPTER 10

CHAOS AND THE BRAIN

"May your chaos be always of high dimension."

Farmer (1986)

10.1 INTRODUCTION

Complicated and unpredictable behaviour with deterministic origins abounds in nature; examples include the changing weather patterns, the motion of the planets, the spread of epidemics, the turbulent flow of water and the variations in biological populations. However, chaos is often seen in terms of the limitations it implies, such as lack of predictability — but it is possible that nature uses chaos constructively. Although chaos places severe limits on our predictive abilities, it may be the major factor contributing to the constant variety and richness of our experience.

There is no doubt that chaos ensures the wide variety of our weather (Palmer, 1989); although seasonal patterns recur, accurate forecasting even a few days in advance is impossible. However, chaos may also fulfil other roles in nature. Farmer (1986) suggested that because chaos produces bounded randomness, it may provide the genetic variability required for biological evolution. Conrad (1986) thought that, through the amplification of small fluctuations, chaos may provide biological systems with access to novelty, which may enhance exploratory behaviour, while Crutchfield et al. (1986) proposed that chaos may help prey evade capture through unpredictable changes of direction. There have also been many speculations regarding the presence and possible roles of chaos in the brain.

There is a growing body of evidence that chaotic activity is present in the brain. Studies of the behaviour of individual neurons have revealed that neurons are complex nonlinear devices capable of generating chaotic activity, while analyses of EEGs have suggested that much EEG activity is governed by chaotic dynamics. This experimental evidence of the existence of chaos in the brain is supported by theoretical models that display chaotic activity over certain parameter ranges.

10.2 CHAOTIC ACTIVITY OF NEURONS

There is no doubt that the dynamics of an individual neuron are very complex. The details of neuron operation began to be uncovered in the 1940s with the rediscovery of the giant axon (~1 mm in diameter) in squid and the subsequent development of the microelectrode. Based on results obtained from microelectrode studies of squid giant axons, Hodgkin and Huxley (1952) developed a mathematical model of the electrical behaviour of the neuron membrane. The model consists of a set of nonlinear differential equations, which describe the
flow of ions across the membrane and the resultant change in membrane potential. Hodgkin and Huxley verified their equations by comparing simulated action potentials with those recorded experimentally. Because of the close correspondence between simulations and experimental observations, the Hodgkin-Huxley equations (as they are now known) have been used extensively to study neuron operation and to direct experimental research.

Aihara et al. (1984) examined the numerical solutions of the Hodgkin-Huxley equations with a sinusoidal stimulating current. Depending on the stimulus parameters (i.e., amplitude and frequency) either synchronized, quasi-periodic or chaotic oscillations of the membrane potential were observed. As predicted by these simulations, a wide range of behaviour can be observed by subjecting a neuron to periodic stimulation and altering the amplitude and frequency of the stimulus. Aihara and Matsumoto (1986) subjected a squid giant axon to a sinusoidal current input and observed either synchronized, quasi-periodic or irregular oscillations in membrane potential. Matsumoto et al. (1987) performed similar experiments stimulating the squid giant axon periodically with current pulses. For certain ranges of stimulus amplitudes and periods, $p : q$ phase locking was observed with $p$ action potentials being produced in response to every $q$ input pulses ($p \leq q$). However, for other parameter values irregular bursts of action potentials were generated. The similarity between these observations and numerical simulations led to the conclusion that all observed behaviour was generated deterministically and, furthermore, the irregular activity was interpreted as deterministic chaos (Aihara and Matsumoto, 1986).

Chaotic activity has also been observed in several types of molluscan neurons (Holden et al., 1982; Hayashi et al., 1982a; Hayashi et al., 1983) and in neurons of the motor cortex of the squirrel monkey (Rapp et al., 1985), as well as in other biological membranes including cardiac cells (Guevara et al., 1981) and Nitella internodal cells (Hayashi et al., 1982b).

Because the chaotic activity of neurons has usually been observed under artificial conditions, there is still much debate as to whether neurons produce chaotic activity under normal physiological conditions in the brain. However, the experimental results do reveal that neurons are complex nonlinear devices capable of generating chaotic activity.

### 10.3 CHAOS IN THE EEG

The EEG provides a measure of the overall level of activity of a large number of neurons and, therefore, enables brain activity to be studied on a larger scale. Until very recently the EEG was considered to be an essentially random signal (Jansen, 1991) and, as such, has been subject to considerable statistical analysis. However, progress in nonlinear dynamics has provided new methods for studying complex systems and has enabled the EEG to be viewed as the output of a deterministic nonlinear system. The chief merit of these new techniques, which were introduced in Chapter 9, is their ability to discriminate deterministic from random activity.

In an effort to establish the chaotic nature of EEG activity, several groups have applied algorithms for calculating correlation dimension and Lyapunov exponents to portions of EEG recordings. Although results vary between research groups, they all support the hypothesis that the EEG is generated by a deterministic system.

Several groups have employed the Grassberger-Procaccia algorithm (§9.6.3) to estimate the correlation dimension of attractors reconstructed from EEG data, claiming that a fractional dimension implies sensitive dependence on initial conditions (i.e., chaos). Although Rapp et al. (1989) have found that experimental evidence supports such a relationship between fractional dimension and chaos, Grebogi et al. (1984) have demonstrated that not all
attractors with a fractional dimension are chaotic. A low correlation dimension, considerably less than the embedding dimension (i.e., \( \nu \ll m \)), does, however, provide evidence that the underlying system is deterministic (and it may well be chaotic). Babloyantz and Destexhe (1987) found the dimension of alpha activity to lie between 6.1–7.0, Mayer-Kress and Holzfuss (1987) obtained a similar result of 6.4 ± 1.2. However, Rapp et al. (1986) calculated the dimension of alpha activity to be the much lower value of 2.4 ± 0.2. Soong and Stuart (1989) demonstrated the effect of filtering on correlation dimension, obtaining dimensions of 7.1–7.8 for unfiltered alpha activity and 2.6–2.7 for the same activity after band-pass filtering (5–15 Hz). They compared these results with those obtained from synthetically generated signals to show that the calculated dimension of alpha activity was not attributable to noise nor induced by the filtering process.

Similar calculations have been performed for the various stages of sleep (Babloyantz and Destexhe, 1987; Röschke and Aldenhoff, 1991) and reveal that dimension decreases (from ~7 to ~4) as sleep sets in. Babloyantz and Destexhe (1986) found the dimension of the EEG to drop dramatically during petit mal seizures to a value of approximately 2, while Rapp et al. (1990) showed that the dimension of the EEG increases during periods of cognitive activity. It appears that much EEG activity is governed by deterministic dynamics, the dimension of which changes with levels of consciousness and attention.

The only definitive evidence of chaos is the existence of at least one positive Lyapunov exponent. It is intriguing that in most of the forementioned experiments the largest Lyapunov exponent was only calculated when an integer correlation dimension was encountered. In these cases, the Lyapunov exponent was computed to distinguish quasi-periodic from chaotic activity. However, the difference between these two possibilities should be evident from their power spectra (which are considerably easier to calculate). Nevertheless, positive Lyapunov exponents have been obtained for the various stages of sleep (Babloyantz et al., 1985), during periods of petit mal epilepsy (Babloyantz and Destexhe, 1986) and for alpha activity (Soong and Stuart, 1989) and establish the chaotic nature of EEG activity.

These experimental results are supported by theoretical models of regions of the brain. Having studied the olfactory system of rats, Freeman (1987) assembled a set of ordinary differential equations based on its anatomical and physiological properties and showed that they accurately represented its operation. Furthermore, he discovered that, for certain parameter domains, the model generated chaotic activity which was indistinguishable from activity observed in the resting EEG, while for other parameter ranges activity reminiscent of epileptic seizures was produced. Based on extensive experimental research into neurons of the hippocampus, Traub et al. (1988) developed a realistic model of a neuron, considering it to be made up of a cell body and a collection of branching cylinders, each of which was characterized by several parameters (e.g., length, capacitance, resistance). Almost 10,000 such neurons were interconnected by way of excitatory and inhibitory synapses. The behaviour of this network was shown to be consistent with numerous experiments on the response of hippocampal slices to brief electrical shocks (Traub et al., 1989). Furthermore, the model displays oscillatory behaviour consistent with that observed in the EEG. By virtue of the many thousands of coupled nonlinear differential equations, the model produces chaotic activity for certain parameter domains.

10.4 PURPOSE OF CHAOS IN THE BRAIN

Although the existence of chaos in the brain is becoming more widely accepted, many questions concerning its purpose remain unanswered. There is much debate as to whether the presence of chaos in the brain is merely a by-product of the interaction of many nonlinear
elements or whether it has evolved for a specific purpose.

Freeman (1987) notes that chaos is unavoidable in dynamics of large scale interactive systems, but believes it to be more than an accidental by-product of the brain's complexity. He proposes that chaos maintains a steady level of unpatterned basal activity, which is essential for exercising neurons (Freeman, 1989) and provides a variability that allows the system to respond quickly to a variety of stimuli (Yao and Freeman, 1990).

The ability of chaos to produce information (§9.5.2) led to the idea that it underlies creativity (Farmer, 1986). Kurten and Clarke (1986) thought that "chaos could, in moderation, be beneficial particularly in the context of higher mental processing where it is advantageous to have the potential of breaking out of rigid, stereotyped behaviour". Therefore, chaos could provide a mechanism for free will in a world governed by deterministic laws (Crutchfield et al., 1986; Farmer, 1986). Freeman (1989) proposed that it is the ability of chaos to create information that endows biological intelligence and suggested that chaos may be the chief property that makes the brain different from artificial intelligence machines (Freeman, 1991).

Gallez and Babloyantz (1991) described chaos in the brain as a compromise between "an absolute determinism, with a poor information processing capability, and a total randomness, which would be useless for an intelligent behaviour". Rapp et al. (1986) speculated that chaos may be useful both for searching memories and in the early stages of decision making, while Kepler et al. (1990) proposed that chaos may be the property which allows the brain to discriminate between two very similar inputs (e.g., face recognition).

Although speculations regarding the functions of chaos in the brain are rife, they tend to be rather vague and unsubstantiated. A better understanding of the roles of chaos in the brain is needed and may be provided by mathematical models.
CHAPTER 11

CHAOTIC NEURAL UNITS

"The latest authors, like the most ancient, strove to subordinate the phenomena of nature to the laws of mathematics."

Isaac Newton (1642–1727)

11.1 INTRODUCTION

An individual neuron has dynamics of great complexity which are unable to be reproduced by conventional neural units. While the simplicity of conventional neural units is a computational advantage, it can also be considered a drawback, preventing them from exhibiting the wide variety of behaviour observed in neurons. Because neurons are capable of generating chaotic activity, chaos may be an important feature to include in models of neurons.

There are two approaches to modelling neurons, both of which can produce neural units capable of displaying chaos, that is, chaotic neural units. In the first approach, certain operational features of a neuron are modelled and as more features are incorporated the behaviour of the resultant neural unit becomes more realistic. However, considerable complexity is required before chaotic activity is produced. The second approach concentrates on replicating the overall behaviour of a neuron, with minimal regard for its internal operation. The advantage of the latter approach is that the resultant neural units can be considerably simpler and easier to understand.

This Chapter proposes that a simple chaotic system may provide the basis of a useful model of neuron behaviour and, specifically, introduces a neural unit based on the logistic map — the versatile neural unit. While this neural unit retains characteristics of conventional neural units, it also exhibits a wide variety of behaviour reminiscent of that observed in neurons. Furthermore, this chaotic neural unit is considerably simpler and its behaviour more easily understood than those developed by modelling neuron operation.

11.2 MODELS OF NEURON OPERATION

Neural units based on modelling operational features of neurons produce only simplistic behaviour unless considerable complexity is introduced. This Section reviews a number of such neural units and their behaviour, beginning with conventional neural units and progressing to those capable of generating chaos.
11.2.1 Conventional Neural Units

McCulloch and Pitts (1943) proposed the first model of a neuron, which received both excitatory and inhibitory inputs. Provided no inhibitory inputs were active, the neural unit produced an output pulse when a certain minimum number of its excitatory inputs were active. An active inhibitory input prevented the neural unit from generating an output pulse. This neural unit has since been modified with the inputs $x_i$ being multiplied by positive or negative weights $w_i$ corresponding to excitatory and inhibitory inputs respectively. The neural unit produces an output pulse when the linear summation of its weighted inputs exceeds a threshold $\theta$ as follows:

$$
\text{net}(n) = \sum_{i=1}^{M} w_i x_i(n)
$$

$$
y(n) = \mu(\text{net}(n) - \theta)
$$

where $M$ is the number of inputs and $\mu[\cdot]$ is the unit step function.

A more general form of the conventional neural unit (Fig. 11.1) allows the output $y$ to be continuous valued rather than binary. In this case, the unit step function $\mu[\cdot]$ is replaced by a nonlinear function $f[\cdot]$, which may take a variety of forms, but is essentially monotonic and non-decreasing. For example, Hopfield (1984) used the sigmoid function:

$$
f[z] = \frac{1}{1 + e^{-z}}
$$

Because of their simplicity, the behaviour of conventional neural units is extremely limited; for a constant net input, a constant output is produced.

![Figure 11.1. Conventional neural unit. Inputs $x_i$ are received via weighted connections $w_i$ and summed to produce a net input net, from which the threshold $\theta$ is subtracted. The result is passed through a nonlinear function $f[\cdot]$ to produce an output $y$.]

11.2.2 Temporal Summation

Conventional neural units model only the spatial summation of inputs, but neurons also integrate inputs temporally. Inputs received by a neuron produce a change in its membrane potential. When the membrane potential exceeds a threshold an action potential is generated and the membrane potential returns to its resting value (§1.2.1). The temporal summation of inputs can be modelled, as shown in Fig. 11.2, by maintaining an internal state $p$, which changes as a function of the inputs and is reset by generation of an output pulse. This neural
unit can be represented as:

\[ \text{net}(n) = \sum_{i=1}^{M} w_i x_i(n) \]

\[ p(n) = \sum_{k=1}^{n-T-1} R^{k-1} \text{net}(n-k) \]  \hspace{1cm} (11.3)

\[ y(n) = \mu[p(n) - \theta] \]

where \( T \) is the instant of time at which the last output pulse was generated and \( R \) is a constant \((0 \leq R \leq 1)\) that determines the relative effect of previous inputs. If \( R=0 \) then this neural unit reverts to being a conventional neural unit.

This neural unit \((R > 0)\) displays more realistic behaviour than the conventional neural unit, producing either constant or periodic activity in response to a constant net input \( \text{net} \). For \( \text{net} \geq \theta \), a constant output is produced. However, periodic activity is produced if \( \text{net} < \theta \), with the period being dependent on the values of \( R \) and \( \text{net} \).

### 11.2.3 Refractory Period

Nagumo and Sato (1972) proposed a neural unit which models the relative refractory period of a neuron, that is, the period following generation of an action potential during which the neuron operates with elevated threshold (§1.2.1). In their neural unit (Fig. 11.3), the threshold \( \theta \) changes as a function of previous outputs, increasing whenever an output pulse is generated and then decaying towards its usual value \( \Theta \). The neural unit can be represented mathematically by:

\[ \text{net}(n) = \sum_{i=1}^{M} w_i x_i(n) \]

\[ \theta(n) = \Theta + \alpha \sum_{k=1}^{n} R^{k-1} y(n-k) \]  \hspace{1cm} (11.4)

\[ y(n) = \mu[\text{net}(n) - \theta(n)] \]
where $\alpha$ is a positive constant determining the effect of previous outputs on the threshold, $R$ is a constant ($0 \leq R \leq 1$) that describes the rate at which the effect of previous outputs decays and $\mu[\cdot]$ is the unit step function. For a constant net input $net$, almost all responses are periodic or constant, but chaotic solutions do arise for a very limited set of parameters (Aihara et al., 1990).

\[
\begin{align*}
\text{Figure 11.3. Neural unit developed by Nagumo and Sato to model the relative refractory period of a neuron.}
\end{align*}
\]

Inputs $x_i$ are received via weighted connections $w_i$ and summed to produce a net input $net$, from which the threshold $\theta$ is subtracted. The result is passed through a step function $\mu[\cdot]$ to produce a binary output $y$. The threshold $\theta$ depends on previous outputs, increasing when an output pulse is produced by an amount determined by the constant $\alpha$ and then decaying towards its asymptotic value $e$. The effect of previous outputs decays over time with only a fraction $R$ remaining after a fixed time delay $\tau$.

Aihara et al. (1990) modified this neural unit, proposing that the threshold $\theta$ changes in relation to a graded potential $p$ rather than to the output pulses themselves. The resulting neural unit (Fig. 11.4) can be expressed as:

\[
\begin{align*}
net(n) &= \sum_{i=1}^{M} w_i x_i(n) \\
\theta(n) &= \Theta + \alpha \sum_{k=1}^{n} R^{k-1} p(n-k) \\
p(n) &= f[net(n) - \theta(n)] \\
\end{align*}
\]

where $\Theta$ is the asymptotic value of the threshold, $\alpha$ is a positive constant (usually unity) determining the effect of previous graded potentials $p$, $R$ is a constant describing the rate at which the effect of previous graded potentials $p$ decays and $f[\cdot]$ is a nonlinear function. Although this neural unit generates a graded potential $p$, it transmits binary pulses to other neural units. The output is given by:

\[
y(n) = \mu[p(n) - \vartheta] \\
\]

where $\vartheta$ is the threshold for transmission of an output pulse and $\mu[\cdot]$ is the unit step function.

For certain nonlinear functions $f[\cdot]$ (e.g., a sigmoid), this neural unit produces either constant, periodic or chaotic activity in response to a constant input $net$. For example, Fig. 11.5
Figure 11.4. Neural unit developed by Aihara et al. to model the relative refractory period of a neuron. Inputs $x_i$ are received via weighted connections $w_i$ and summed to produce a net input $net$, from which the threshold $\theta$ is subtracted. The result is passed through a nonlinear function $f[\cdot]$ to produce a graded potential $p$, which alters the threshold $\theta$. An output pulse is produced when $p$ exceeds a value $\theta$. Again $\tau$ is a unit delay, $\alpha$ is a constant that determines the effect of $p$ on the threshold and $R$ is a parameter describing the rate at which the effect of previous potentials $p$ decays.

shows a chaotic sequence of output pulses produced by this neural unit. Furthermore, by systematically altering the value of the net input $net$, Aihara et al. demonstrated that alternating periodic and chaotic sequences of output pulses can be produced (Fig. 11.6), similar to those observed in squid giant axons.
Figure 11.5. Chaotic activity of the neural unit developed by Aihara et al.: (a) output $y$, (b) graded potential $p$ and (c) threshold $\theta$, for parameters $R=0.65$, $\sigma=1.0$, $\Theta=0.0$, $\psi=0.5$, net=1.0 and a sigmoidal nonlinear function $f(x)=1/(1+e^{-x/0.82})$. 
Figure 11.6. Bifurcation diagram for the neural unit developed by Aihara et al. showing alternate periodic and chaotic activity: (a) threshold $\theta$ and (b) graded potential $p$, for parameters $R=0.65$, $\alpha=1.0$, $\Theta=0.0$, $\vartheta=0.5$ and a sigmoidal nonlinear function $f(z) = 1/(1 + e^{-z/0.02})$. 
11.3 MODELS OF NEURON BEHAVIOUR

Aihara et al. (1990) demonstrated that, with the introduction of additional structural complexity, neural units can display a wide range of neuronal activity. However, because the emphasis of these neural units is on modelling operational features of a neuron, complex configurations are required to reproduce complicated observed behaviour.

By concentrating on the observed behavioural characteristics of neurons rather than on their operation, it is possible to devise relatively simple neural units that display a rich range of behaviour. The inspiration for such neural units is the type of nonlinear recursive equations capable of generating chaos (§9.4.4).

\[ z(n) = 1 - a z(n-1)^2 \]  

Figure 11.7. Bifurcation diagram for \( z(n) = 1 - a z(n-1)^2 \) showing that the parameter \( a \) determines the type of activity produced.

Kaneko (1990) developed a neural unit based on the simple nonlinear recursive equation:

\[ z(n) = 1 - a z(n-1)^2 \]  

which produces chaos when \( 1.4 < a < 2 \) (Fig. 11.7). The resultant neural unit is shown in Fig. 11.8. In addition to receiving inputs \( x_i \) through weighted connections \( w_i \), the neural unit receives internal feedback regarding its previous output. The relative effects of the external inputs and the internal feedback are determined by the parameter \( \alpha \) as follows:

\[
\begin{align*}
\text{net}(n) & = \sum_i w_i x_i(n) \\
y(n) & = f[\alpha \text{net}(n) + (1 - \alpha) y(n-1)] \\
f(z) & = 1 - a z^2
\end{align*}
\]  

(11.8)

where \( \text{net}(n) \) must lie in the range \(-1.0 \ldots 1.0\) and \( 0 \leq \alpha \leq 1 \). The behaviour of this neural unit depends entirely on the parameters \( a \) and \( \alpha \) rather than on its inputs. Again consider a
Figure 11.8. Neural unit developed by Kaneko based on a simple nonlinear recursive equation. Inputs $x_i$ are received via weighted connections $w_i$ and summed to produce a net input $net$. The net input is combined with the previous output according to a constant $\alpha$ and the result passed through a nonlinear function $f[z] = 1 - az^2$ to produce an output $y$. Again $\tau$ is a fixed time delay.

constant net input $net$. For $\alpha = 1$ there is no internal feedback and the output of the neural unit is a constant. At the other extreme $\alpha = 0$ and the external inputs have no effect. In this case the behaviour of the neural unit is either constant, periodic or chaotic depending on the value of $a$ (Fig. 11.7). Thus, the behaviour of the neural unit depends on both $\alpha$ and $a$ and not on the value of $net$.

Figure 11.9. Bifurcation diagram of the logistic equation which becomes the transfer function of the chaotic neural unit.

If the net input is used to set the value of a bifurcation parameter, then a neural unit is obtained whose output depends on its inputs. For example, the net input $net$ can be used to
set the gain $g$ of the logistic map (§9.4.4) as follows:

$$
net(n) = \sum_i w_i x_i(n)
$$

$$
y(n) = f[net(n), y(n-1)]
= 4 net(n) y(n-1) [1 - y(n-1)]
$$

(11.9)

where $net_i$ must lie in the range 0.0...1.0. Thus, the inputs themselves determine the type of behaviour exhibited by the neural unit. In response to a constant net input $net$, either constant, periodic or chaotic activity is produced. For $net \leq 0.75$ the output is constant, while for $0.75 < net < 0.89$ it is periodic and when $0.89 < net \leq 1.0$ chaotic activity is generated (Fig. 11.9).

If the output of the neural unit is not available until a time $T$ after the inputs have changed, then, provided $T$ is sufficiently large to ensure that transients have decayed, the bifurcation diagram of the logistic map effectively becomes its transfer function (Fig. 11.9). It can be seen that for $0.0 \leq net \leq 0.75$, this neural unit is qualitatively similar to conventional neural units, with the output being a monotonic non-decreasing function of the net input, but for $net > 0.75$ it displays entirely different behaviour.

![Figure 11.10. Inverted bifurcation diagram of the logistic equation, which is effectively the transfer function of the versatile neural unit.](image)

This neural unit can be modified to accord with the experimental observations of Freeman (1987), who found that when a rat was presented with a novel odour the activity of its olfactory neurons appeared chaotic but when a familiar odour was presented their activity was more ordered. In a neural network, a pattern that is familiar to a neural unit produces a large net input. Therefore, by inverting the bifurcation diagram of the logistic equation, as in Fig. 11.10, it is possible to produce a neural unit that behaves in a way similar to the
neurons observed by Freeman. The resulting neural unit (Fig. 11.11) can be expressed as:

\[
\begin{align*}
\text{net}(n) &= \sum_{i=1}^{M} w_i x_i(n) \\
y(n) &= f[\text{net}(n), y(n-1)] \\
&= 1 - 4 \left[ 1 - \text{net}(n) \right] y(n-1) \left[ 1 - y(n-1) \right]
\end{align*}
\] (11.10)

In this case, the neural unit is qualitatively similar to conventional neural units for \(0.25 \leq \text{net} \leq 1.0\), with the output being a non-decreasing function of the net input. However, for \(0.11 < \text{net} < 0.25\) the neural unit displays periodic activity and when \(0.0 \leq \text{net} < 0.11\) chaotic activity is generated.

![Figure 11.11. The versatile neural unit based on inverted logistic equation. Inputs \(x_i\) are received via weighted connections \(w_i\) and summed to produce a net input \(\text{net}\), which is passed through a nonlinear function \(f[\cdot]\) to produce an output \(y\). The function \(f[\cdot]\) is the inverted logistic equation \(f[\text{net}, y] = 1 - 4 [1 - \text{net}] y [1 - y]\) and \(\tau\) is a fixed time delay.](image)

I have termed this neural unit the **versatile neural unit** because of its ability to reproduce neuronal activity while retaining the features of conventional neural units that enable it to be incorporated into established neural networks (Chapter 12).

### 11.4 SUMMARY

A new neural unit has been proposed based on a chaotic system. This versatile neural unit is relatively simple, particularly in comparison with the neural unit developed by Aihara et al. (1990) (§11.2.3), yet still displays a rich range of behaviour. Furthermore, the inputs themselves determine the type of activity produced by the neural unit. This neural unit illustrates the difference between the two approaches to modelling neurons. When emphasis is placed in modelling neuron operation, only simplistic behaviour is obtained unless considerable extra complexity is introduced. However, if instead emphasis is on replicating observed neuron behaviour, then a simple chaotic system, such as the logistic map, can be used as the basis of a neural unit.

A single neuron operating in isolation is hardly meaningful yet networks of neurons (e.g., the human brain) are able to perform complicated pattern recognition tasks with apparent ease. Thus, although the behaviour of chaotic neural units is in itself intriguing, of greater interest is the collective behaviour of networks of these units, which is considered in the following chapter.
CHAPTER 12

CHAOTIC NEURAL NETWORKS

"To the theoretical question, Can you design a machine to do whatever a brain can do? the answer is this: If you will specify in a finite and unambiguous way what you think a brain does do with information, then we can design a machine to do it. Pitts and I have proved this constructively. But can you say what you think brains do? .”

W. S. McCulloch (1898–1969)

12.1 INTRODUCTION

The information processing capabilities of biological systems are far superior to those of machines. The promise of neural networks is that they may be able to replicate some of the information processing abilities of the brain.

Although neural networks designed for specific applications have produced promising results (Chapter 8), even hardware implementations of these networks do not appear to realize the full potential of adaptive parallel processing. Zak (1990) believes that the fundamental limitation of neural networks is their rigid behaviour compared with even the simplest biological systems.

The introduction of chaos into neural networks would overcome the problem of their rigid behaviour and, in so doing, may enhance their performance and provide insights into the roles of chaos in the brain.

This Chapter begins by considering ways of introducing chaos into neural networks. The behaviour of networks of chaotic elements is then examined and, in particular, networks of versatile neural units are shown to produce EEG-like activity. Finally, versatile neural units are incorporated into the self-organizing map, which consequently displays an improved ability to cluster input patterns.

12.2 CHAOS IN NEURAL NETWORKS

Chaos can be introduced into neural networks, which are then termed chaotic neural networks, in a number of ways. For example, a fully connected network of conventional neural units can produce chaotic activity if the weight matrix \( W = \{w_{ij}\} \quad i, j = 1..M \) is asymmetric. Sompolinsky et al. (1988) interconnected neural units of the form:

\[
net_j = \sum_{i=1}^{M} w_{ji} x_i
\]
\[ \dot{p}_j = -p_j + \text{net}_j \quad (12.1) \]
\[ y_j = \tanh [g p_j] \]

where \( p_j \) is the rate of change of the internal state \( p_j \), by pseudo-randomly choosing the weights \( w_{ji} \) and found that chaotic activity was generated once the gain parameter \( g \) exceeded a critical value. Stariolo (1991) introduced asymmetry into the weight matrix of a Hopfield network (§7.5) consisting of graded response neural units of the form:

\[ \text{net}_j = \sum_{i=1}^{M} w_{ji} x_i \]
\[ y_j = \tanh [g \text{net}_j] \quad (12.2) \]

A number of patterns \( \xi_p \) were stored by setting the weights as:

\[ w_{ji} = \begin{cases} \frac{1}{M} \sum_{p=1}^{P} \xi_{pi} \xi_{pj} & i \neq j \\ 0 & i = j \end{cases} \quad (12.3) \]

where \( M \) is the number of neural units, \( P \) is the number of patterns \( \xi_p \) to store, and \( \xi_{pi} = \pm 1 \). Asymmetry was introduced by setting a proportion \( D \) of the weights \( w_{ji} \) to zero. The network produced chaotic activity for high values of dilution \( D \).

Not surprisingly, networks of chaotic neural units can also display chaos. Aihara et al. (1990) added temporal summation to their neural units (11.6), before connecting three of them into a loop \( (w_{21}=w_{32}=w_{13}=0.5) \). For certain parameter values, the graded potential \( p_j \) of each neural unit was chaotic. Kaneko (1990) interconnected his neural units (11.9) via uniform weights \( (w_{ji}=1/M, w_{jj}=0) \) which ensured the net input to each neural unit remained in the range \(-1.0 \ldots 1.0\). The parameters \( \alpha \) and \( \beta \) of the neural units were chosen to be in the chaotic region and, consequently, all neural units produced chaos.

Versatile neural units can also be interconnected via weights \( w_{ji} \) to produce a neural network displaying chaotic activity. As mentioned in §11.3, the net input to each neural unit must lie in the range 0.0...1.0. This is achieved by choosing the weights such that:

\[ |w_{ji}| \leq \frac{\alpha}{M} \quad (12.4) \]

where \( M \) is the number of neural units in the network and \( \alpha \) is a constant \( (0 \leq \alpha \leq 1) \) describing the maximum coupling between neural units. The behaviour of this network depends on the parameter \( \alpha \). For example, the average activity of 10 neural units interconnected via uniform weights \( (w_{ji}=\alpha/M, w_{jj}=0) \) is either constant, periodic or chaotic depending on the value of \( \alpha \) (Fig. 12.1). Alternatively, the weights can be pseudo-randomly selected according to a given probability density function \( \rho(x) \); for example:

\[ \rho(x) = \begin{cases} \frac{M}{\alpha} & 0 \leq x \leq \alpha/M \\ 0 & \text{otherwise} \end{cases} \quad (12.5) \]

Again the behaviour of the network depends on the parameter \( \alpha \), with the average activity of these networks being qualitatively similar to EEG signals, as can be seen from Fig. 12.2. Thus, the versatile neural unit is able to model a wide range of the observed behaviour of biological neural systems without the introduction of significant extra complexity. Furthermore, it may be possible to enhance the established computational advantages of neural networks by employing chaotic rather than conventional neural units.
Figure 12.1. Average activity of 10 versatile neural units interconnected via uniform weights: (a) $\alpha=0.2$, (b) $\alpha=0.5$ and (c) $\alpha=0.6$.

Figure 12.2. EEG-like activity produced by a network of 10 versatile neural units interconnected via pseudo-randomly selected weights. (a) $\alpha=0.4$, (b) $\alpha=0.6$, (c) irregular EEG activity and (d) alpha rhythms.
12.3 THE CHAOTIC SELF-ORGANIZING MAP

Although it has been demonstrated that certain neural networks can produce chaos, the function of chaos in such networks has been little investigated. Stariolo (1991) showed that increasing the dilution of a Hopfield network resulted in chaos, but also decreased the storage capacity of the network. Kaneko (1990) studied the collective properties of networks of his chaotic neural units and found that a number of neural units, termed a cluster, produced the same chaotic output sequence. He classified the state of the network by the number of clusters $k$ and the number of elements in each cluster $N_k$. Further simulations revealed emergent properties such as switching among attractors and hierarchical dynamics, both of which may have applications to biological information processing. Although this network displays a variety of interesting collective phenomena, it does not include a learning capability. By incorporating versatile neural units into the self-organizing map, a chaotic neural network — the chaotic self-organizing map — is produced, which is able to learn from experience.

The conventional self-organizing map (§7.6) is made up of linear neural units, which can be expressed as:

$$
y_{pj} = \text{net}_{pj} = \sum_{i=1}^{M} w_{ji} x_{pi} = w_j \cdot x_p = \|w_j\| \|x_p\| \cos \phi_{pj}\tag{12.6}
$$

where $\phi_{pj}$ is the angle between the input pattern $x_p$ and the weight vector $w_j$. Typically the vectors $x_p$ and $w_j$ are of unit length and, therefore, $y_{pj} = \cos \phi_{pj}$. When an input pattern $x_p$ is presented to the network, neural units compete for the privilege to learn — the winning neural unit being the one with the largest output $y_{pj}$. Because $y_{pj} = \cos \phi_{pj}$, the winning neural unit is the one whose weight vector is closest to the input pattern. The weight vectors of the winning neural unit and its neighbours are modified according to the rule:

$$
w_j = \frac{w_{pj}^{old} + \eta x_p}{\|w_j^{old} + \eta x_p\|} \tag{12.7}
$$

where $\eta$ is the learning rate, which typically decreases linearly over the training session from an initial value $\eta(0)$ to a final value $\eta_{\text{min}}$.

Chaos can be introduced into the self-organizing map by replacing the linear neural units of the conventional network with versatile neural units:

$$
y_{pj}(n) = 1 - 4 \left[ 1 - \text{net}_{pj} \right] y_{pj}(n-1) \left[ 1 - y_{pj}(n-1) \right]\tag{12.8}
$$

Thus, when an input pattern $x_p$ is presented to the network, each neural unit produces a sequence of outputs $y_{pj}(n)$. As with the conventional self-organizing map, neural units compete for the privilege to learn but, in this case, the winning neural unit is chosen to be the one with the greatest average output $\bar{y}_{pj}$ calculated over $T$ elements of the output sequence as follows:

$$
\bar{y}_{pj} = \frac{1}{T} \sum_{n=1}^{T} y_{pj}(n) \tag{12.9}
$$

As $T \to \infty$, the average output $\bar{y}_{pj}$ approaches the mean value of the versatile neural unit (Fig. 12.3). It can be seen that the neural unit with the largest net input has the largest
average output $\bar{y}_{pj}$, provided $net_{pj} > 0.38$. However, if all $net_{pj} < 0.38 \ (j=1...M)$ then a neural unit operating in its chaotic region with $net_{pj} < 0.05$ has the largest average output.

In practice, $T$ is relatively small and the average output $\bar{y}_{pj}$ depends not only on $net_{pj}$ but also on the initial value $y_{pj}(0)$ (Fig. 12.4). For small $T$, the initial state has such a large effect that it is difficult to predict which neural unit will have the largest average output on the basis of its net input. Thus, the competition to learn effectively becomes probabilistic.$^2$ Although the neural unit with the largest net input usually has the greatest probability of winning, all neural units have the opportunity to win the competition to learn. As with the conventional self-organizing map the winning neural unit and its neighbours adjust their weight vectors according to (12.7).

Several practical problems arise when implementing the chaotic self-organizing map. Firstly, the net input $net_{pj}$ must lie in the range 0.0...1.0. However, the quantity $x_p \cdot w_j$, which is employed by the conventional self-organizing map, lies in the range $-1.0...1.0$ for vectors of unit length. A value for $net_{pj}$ can be derived from this quantity in several ways. For example, $x_p \cdot w_j$ can be thresholded as follows:

\[
net_{pj} = \begin{cases} 
0 & \text{if } x_p \cdot w_j < 0 \\
 x_p \cdot w_j & \text{otherwise}
\end{cases}
\]  

or it can be mapped to the unit interval:

\[
net_{pj} = \frac{1}{2} (1 + x_p \cdot w_j)
\]  

$^1$The initial value $y_{pj}(0)$ is the final value of the sequence produced for the previous input pattern.

$^2$Although the outcome of the competition to learn is deterministic, it depends on the initial values $y_{pj}(0)$, which are unknown.
Figure 12.4. Distribution of $\overline{y_{pj}}$ as a function of the net input $net_{pj}$ for 20 different initial conditions $y_{pj}(0)$ uniformly distributed over the unit interval: (a) $T = 1$, (b) $T = 5$, (c) $T = 10$ and (d) $T = 50$.

A further problem is that for all $net_{pj} \geq 0.75$, the average output $\overline{y_{pj}}$ approaches unity (Fig. 12.4). As a result, several neural units may have the same average output and a single winner cannot be chosen. This problem can be solved by scaling the net input by 0.75. Finally, if the output $y_{pj}(n)$ of a neural unit ever reaches unity then it remains so forevermore, regardless of the net input. The simplest solution to this problem is to limit the maximum output of neural units to a value just less than unity (e.g., 0.9999). However, this has the disadvantage that it again limits the ability of the network to choose a winner. For example, if the maximum output is limited to 0.9999 then neural units whose weight vectors are within 0.5° of the input pattern may have the same output, while if the maximum output is 0.99 then neural units within 4.7° of the input vector can produce the same output. Therefore, the maximum allowable output should be as close to unity as possible.

Because there are a number of ways of deriving $net_{pj}$ from $x_p \cdot w_j$, there are several variants of the versatile neural unit, two of which are considered here. The first is termed the thresholded input neural unit because $x_p \cdot w_j$ is thresholded so that it lies in the unit interval, before being scaled by 0.75:

$$net_{pj} = \begin{cases} 0 & \text{if } x_p \cdot w_j < 0 \\ \frac{1}{3}(x_p \cdot w_j) & \text{otherwise} \end{cases}$$

(12.12)

This equation can be rewritten in terms of the angle $\phi_{pj}$ between the input pattern $x_p$ and
the weight vector $w_j$ as follows:

$$net_{pj} = \begin{cases} \frac{3}{4} \cos \phi_{pj} & -90^\circ \leq \phi_{pj} \leq 90^\circ \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (12.13)

The thresholded input neural unit produces chaotic activity if $|\phi_{pj}| > 82^\circ$, periodic activity if $71^\circ < |\phi_{pj}| < 82^\circ$ and a constant output if $|\phi_{pj}| < 71^\circ$.

The second variant of the versatile neural unit is referred to as the *mapped input* neural unit because $x_p \cdot w_j$ is mapped to the interval $[0...0.75]$ as follows:

$$net_{pj} = \frac{3}{8} \left[ 1 + x_p \cdot w_j \right]$$

$$= \frac{3}{8} \left[ 1 + \cos \phi_{pj} \right]$$  \hspace{1cm} (12.14)

In this case, the neural unit produces chaotic activity when $|\phi_{pj}| > 135^\circ$, periodic activity when $110^\circ < |\phi_{pj}| < 135^\circ$ and a constant output if $|\phi_{pj}| < 110^\circ$.

The versatile neural unit is qualitatively similar to the conventional linear neural unit for $net_{pj} \geq 0.25$, with the output being a monotonic non-decreasing function of the net input. For the mapped input neural unit this corresponds to $|\phi_{pj}| < 110^\circ$ and for the thresholded input neural unit to $|\phi_{pj}| < 71^\circ$. Therefore, the mapped input neural unit can be considered as an intermediate step between the conventional linear and the thresholded input neural units.

The self-organizing map is renown for its ability to cluster input patterns, model their probability density function and develop a global spatial order. In the following sections, the performance of the chaotic self-organizing map on each of these tasks is investigated through a number of simulations. In all simulations two-dimensional, unit-length vectors are used so that they can be represented on the unit circle (Fig. 12.5). As a result, each self-organizing map has two fan-out elements in its input layer and each neural unit, of which there may be any number, has a two-dimensional weight vector associated with it.

![Figure 12.5. Two-dimensional unit length vectors can be represented on the unit circle.](image)

### 12.3.1 Clustering Input Patterns

Consider a self-organizing map where only the neural unit that wins the competition to learn modifies its weight vector. Although such a network cannot develop a spatial order (because
all neural units learn independently), it is still able to cluster input patterns. During training, the weight vector of each neural unit approaches the average of the cluster of input patterns for which it won the competition to learn. Thus, each neural unit becomes selectively sensitive to a cluster of input patterns.

In order for the self-organizing map to cluster input patterns optimally, all neural units must participate in learning. Because the neural unit whose weight vector is closest to the input pattern wins the competition to learn, the weight vectors must be initialized so that every neural unit, at some stage, wins the competition and, consequently, learns. However, this is not possible without a priori information about the distribution of input patterns.

![Cluster 2](Cluster 2) ![Cluster 1](Cluster 1)

**Figure 12.6.** Two well-separated clusters of input patterns.

Clustering problems arise when not all neural units participate in learning. For example, when input patterns from two well-separated clusters (Fig. 12.6) are presented to a conventional self-organizing map consisting of only two neural units, the weight vectors typically move to the average of the nearest cluster of input patterns (Fig. 12.7a) and, thus, each neural unit becomes sensitive to a single cluster of input patterns. However, if the initial weight vectors are such that one weight vector is closer to all input patterns than the other, then it approaches the average of all input patterns while the other weight vector remains in its initial position (Fig. 12.7b). Regardless of which input pattern is presented to the network, the same neural unit wins the competition to learn and all input patterns are grouped into a single cluster.

In the chaotic self-organizing map, the neural unit whose weight vector is closest to the input pattern does not necessarily win the competition to learn and, thus, the problem of weight initialization should, at least to some extent, be overcome.

In order to test this hypothesis, input patterns from two clusters were presented to a self-organizing map consisting of only two neural units. Each input pattern was presented to the network 100 times, during which time the learning rate was decreased linearly from $\eta(0)=1.0$ to $\eta_{\text{min}}=0.01$. The initial positions of the weight vectors were pseudo-randomly selected from the unit circle and the network trained from 15 different initial states.

In the first instance, the two clusters of input patterns (Fig. 12.6) were well-separated and patterns from each cluster were presented alternately to the network. As expected, the ability of the conventional self-organizing map to cluster input patterns depended on the initial positions of the weight vectors (Fig. 12.7). In fact, the conventional self-organizing map was only able to cluster input patterns correctly for 8 of the 15 initial states. Because the neural unit whose weight vector is closest to the input pattern always wins the competition
Figure 12.7. Clustering of the conventional self-organizing map: (a) weight vectors moved from their initial positions to the average of nearest input cluster, (b) one weight vector moved to average of all input patterns while the other remained in its initial position.

To learn, the network was not able to cluster the input vectors correctly for the 7 initial states where one weight vector was closer to all input patterns than the other. When the same input patterns were presented to a chaotic self-organizing map consisting of either mapped input or thresholded input neural units, input vectors were correctly clustered for a larger number of initial conditions. When mapped input neural units were employed, the ability of the chaotic self-organizing map to cluster input vectors not only depended on the initial positions of weight vectors but also on $T$, the number of iterations over which the average output $\bar{y}_{pj}$ was calculated. For $T=50$ the network clustered input patterns correctly for 10 of the 15 initial conditions. As $T$ was decreased, the ability of the network to cluster input vectors improved with patterns being correctly clustered for 11/15 initial conditions when $T=10$, 14/15 when $T=2$ and for all 15 initial states when $T=1$. When thresholded input neural units were used, the chaotic self-organizing map correctly clustered the input patterns, regardless of the initial positions of weight vectors and of the value of $T$. Because the neural unit whose weight vector was closest to the input pattern did not necessarily win the competition to learn (particularly for small $T$), the chaotic self-organizing map was able to cluster input patterns correctly for
some initial states where the conventional network failed.

The behaviour of the chaotic self-organizing map can be understood by considering the competition to learn. For large $T$, the initial value $y_{p}(0)$ has little effect on the average output $\overline{y}_{p}$ of a neural unit and, therefore, the outcome of the competition to learn can be predicted (as with the conventional network) solely on the basis of the net input to each neural unit. For an input pattern $x_{p}$, the neural unit with the largest net input $net_{p}$ wins, unless all $net_{pj}$ ($j=1...M$) are less than 0.38 and one neural unit is operating in its chaotic region with $net_{pj} < 0.05$ (Fig. 12.3).

Consider the chaotic self-organizing map consisting of two mapped input neural units, where:

$$net_{pj} = \frac{3}{8}[1 + \cos \phi_{pj}]$$

(12.15)

Thus, $net_{pj} < 0.38$ corresponds to $|\phi_{pj}| > 90^\circ$, while $net_{pj} < 0.05$ corresponds to $|\phi_{pj}| > 150^\circ$. Therefore, the neural unit whose weight vector $w_{j}$ is closest to the input pattern $x_{p}$ wins the competition to learn, unless neither weight vector is within $90^\circ$ of the input pattern and one is more than $150^\circ$ from it. For example, consider the situation illustrated in Fig. 12.8, where $w_{1}$ is closer to all input patterns than $w_{2}$. When an input pattern from cluster 1 is presented, the neural unit associated with $w_{1}$ wins the competition to learn because $|\phi_{p1}| < 90^\circ$. However, when a pattern from cluster 2 is presented, the neural unit associated with $w_{2}$ wins. Although $w_{1}$ is still closer to the input pattern than $w_{2}$, it is more than $90^\circ$ away (i.e., $|\phi_{p1}| > 90^\circ$) and because $|\phi_{p2}| > 150^\circ$, the neural unit associated with $w_{2}$ is operating in its chaotic region with $net_{pj} < 0.05$ and, thus, is able to win the competition to learn. Provided that the learning rate is sufficiently large $w_{2}$ moves close enough to the input pattern to enable it to win the competition when the pattern is next presented. Thus, the chaotic self-organizing map with mapped input neural units is able to cluster input patterns correctly for some initial states where the conventional network failed.

![Figure 12.8](image)

Figure 12.8. Clustering of the chaotic self-organizing map with mapped input neural units. The weight vectors correctly cluster input patterns even though $w_{1}$ is closer to all input patterns than $w_{2}$.

The chaotic self-organizing map employing thresholded input neural units is able to cluster input patterns correctly for an even wider range of initial weight positions (Fig. 12.9). In this
case:

\[ \text{net}_{pj} = \begin{cases} 
\frac{3}{4} \cos \phi_{pj} & -90^\circ \leq \phi_{pj} \leq 90^\circ \\
0 & \text{otherwise}
\end{cases} \] (12.16)

and \( \text{net}_{pj} < 0.38 \) corresponds to \( |\phi_{pj}| > 60^\circ \), while \( \text{net}_{pj} < 0.05 \) corresponds to \( |\phi_{pj}| > 86^\circ \). When neither weight vector was within 50° of the input pattern, a neural unit operating in its chaotic region with \( \text{net}_{pj} < 0.05 \) (i.e., \( |\phi_{pj}| > 86^\circ \)) won. In the event that both neural units were operating with \( \text{net}_{pj} < 0.05 \), the winner was effectively chosen at random\(^3\). Thus, in contrast to the conventional self-organizing map, the weight vectors did not always approach the nearest cluster of input patterns (Fig. 12.9b).

**Figure 12.9.** Clustering of the chaotic self-organizing map with thresholded input neural units: (a) weight vectors correctly clustered inputs regardless of initial positions of weight vectors, (b) weight vectors moved to the average of an input cluster but not necessarily the closest.

Although the chaotic self-organizing map is able to cluster input patterns correctly for a wider range of initial weight positions than is the conventional network, clustering problems

\(^3\) Although the competition is deterministic, the outcome depends on the initial values \( y_{pj}(0) (j=1...M) \) and, therefore, cannot be predicted.
can still arise when one weight vector, \( w_1 \) say, is closer to all input patterns than \( w_2 \). In this situation, \( net_{p1} > net_{p2} \) for all input patterns \( x_p \) \((p=1...P)\). If, in addition, \( net_{p1} > 0.38 \) for all input patterns, then the corresponding neural unit wins the competition regardless of which input pattern is presented. Therefore, \( w_1 \) moves to the average of all input patterns, while \( w_2 \) remains in its initial position. A similar clustering problem arises if \( net_{p2} \) exceeds 0.05 when \( net_{p1} \) is less than 0.38. In this case, when \( w_1 \) is sufficiently far from the input pattern to allow a neural unit with a lower net input to win, the neural unit associated with \( w_2 \) is not operating in its chaotic region with \( net_{p2} < 0.05 \). For example, consider the chaotic self-organizing map with mapped input neural units and the situation illustrated in Fig. 12.10, where \( w_1 \) is closer to all input patterns than \( w_2 \). When an input pattern from cluster 1 is presented to the network, the neural unit associated with \( w_1 \) wins. This neural unit also wins when a pattern from cluster 2 is presented, because \( w_2 \) is within 150° of the input \((i.e., net_{p2} > 0.05)\) and effectively ‘too close’ to win the competition to learn. Therefore, \( w_1 \) moves to the average of all input patterns, while \( w_2 \) remains in its initial position.

Clustering problems did not arise when thresholded input neural units were employed. Because input vectors from the two clusters were separated by at least 80°, it was not possible for one weight vector to be within 60° of all inputs nor for one weight vector to be closer to all input patterns than the other weight vector without the second neural unit operating in its chaotic region with \( net_{pj} < 0.05 \) \((i.e., |\phi_{pj}| > 86°)\) for at least some input vectors. Thus, both neural units participated in learning and input patterns were correctly clustered.

As \( T \) was decreased, the ability of the chaotic self-organizing map with mapped input neural units improved. This is because the competition to learn becomes probabilistic and neural units with a low net input have an increased chance of winning the competition and, thus, participating in learning.

When the angular separation between clusters of input patterns is smaller, clustering problems may arise, even when thresholded input neural units are employed. Input patterns from two clusters (Fig. 12.11) separated by 40° were presented to a self-organizing map again consisting of only two neural units. In this case, the conventional self-organizing map clustered the input patterns correctly for only 4 of the 15 initial states. Because the clusters
were closer together, the situation where one weight vector was closer than the other to all input patterns arose more frequently. The chaotic self-organizing map employing mapped input neural units performed little better than the conventional network. Even for $T=1$, correct clustering was only achieved for 6 of the 15 initial states. The performance of the chaotic self-organizing map with thresholded input neural units depended on the value of $T$. The network correctly clustered inputs for $7/15$ initial states when $T=50$, $10/15$ when $T=20$, $14/15$ when $T=10$ and for all 15 initial conditions when $T \leq 5$. The ability of the network to cluster inputs improved as $T$ was decreased, due to the increased likelihood of neural units with a low net input winning the competition and, thus, participating in learning. Clustering problems arose because input clusters were sufficiently close together so that one weight vector could be within $60^\circ$ of all input patterns (Fig. 12.12a) or one weight vector could be closer to all input patterns without the other neural unit operating in its chaotic region (Fig. 12.12b). Although such situations should arise infrequently, the network only correctly clustered input patterns for 7 of the 15 initial states when $T=50$. In all but one case, both neural units won the competition to learn at some stage. However, the learning rate $\eta$ was not large enough to ensure the weight vector moved sufficiently close to the input pattern so that it was able to win the competition when the pattern was next presented. In fact, the network moved from its initial state to one where clustering problems were encountered (Fig. 12.13). When the initial learning rate was increased to $\eta(0)=2.0$, the network clustered patterns correctly for all 15 initial states when $T \leq 20$ and for $14/15$ cases when $T=50$.

These simulations demonstrate that the ability of the chaotic self-organizing map to cluster input patterns is less dependent on the initial positions of the weight vectors than that of the conventional network. This is particularly apparent when thresholded input neural units are employed, $T$ is relatively small and the initial learning rate $\eta(0)$ is large.

12.3.2 Modelling the Probability Density Function of Input Patterns

Having illustrated the ability of the chaotic self-organizing map to cluster input patterns, it is necessary to consider the accuracy to which the probability density function of input patterns can be modelled. In the following simulations, input patterns were pseudo-randomly selected from the unit circle according to a given probability density function $\rho(x)$, and presented to a self-organizing map where again only the winning neural unit was allowed to modify its weight vector. The learning rate was decreased linearly over the training session from $\eta(0)=1.0$ to $\eta_{\text{min}}=0.01$. The accuracy to which the network modelled $\rho(x)$ was measured by
Figure 12.12. Clustering problems arise in the chaotic self-organizing map with thresholded input neural units (a) when one weight vector is within $60^\circ$ of all input patterns or (b) when one weight vector is closer to all input patterns but the other is not far enough away to be operating in its chaotic region.

Figure 12.13. Clustering problems arise in the chaotic self-organizing map with thresholded input neural units if the learning rate is not sufficiently large.

calculating the root mean squared (r.m.s.) error $\sigma$ between the actual positions of the weight vectors and their ideal positions.

In the first instance, 1000 input vectors were selected from a continuous uniform probability density function:

$$\rho(x) = \begin{cases} 
  1/360 & -180^\circ < x \leq 180^\circ \\
  0 & \text{otherwise}
\end{cases} \quad (12.17)$$

and presented to a self-organizing map consisting of eight neural units. Ideally, the eight weight vectors would become evenly distributed around the unit circle. The initial positions of the weight vectors (Fig. 12.14a) were generated pseudo-randomly and, not surprisingly, did not provide a good model of $\rho$ with the r.m.s. error being $\sigma=36.6^\circ$. During training, the weight vectors of the conventional self-organizing map moved from their initial positions
Figure 12.14. Input patterns were selected from a uniform distribution and presented to a conventional self-organizing map consisting of 8 neural units with $\eta(0) = 1.0$ and $\eta(1000) = \eta_{\text{min}} = 0.01$. (a) Initial positions of weight vectors and (b) final positions of weight vectors.

to those shown in Fig. 12.14b, which are almost evenly distributed around the unit circle ($\sigma = 3.9^\circ$). Because the probability density function of the input patterns was continuous, at some stage during training, every weight vector was closest to an input pattern. Thus, all neural units were able to win the competition to learn and participate in learning. As a result, the weight vectors slowly rotated around the unit circle and the conventional self-organizing map was able to model $\rho(x)$ relatively accurately, regardless of the initial positions of the weight vectors.

The ability of the chaotic self-organizing map to model $\rho(x)$ depended on both the type of chaotic neural unit employed and the number of steps $T$ over which the average output of a neural unit was calculated. For the same value of $T$, networks of mapped input neural units typically produced more accurate models of $\rho(x)$ than did those of thresholded input neural units (Fig. 12.15). The accuracy with which the chaotic self-organizing map modelled $\rho(x)$ improved considerably as $T$ was increased. For $T=1$ and thresholded input neural units, the final positions of the weight vectors did not model $\rho(x)$ any better than randomly positioned vectors (Fig. 12.15b), but for $T=50$ the chaotic self-organizing map performed as well as the conventional network (Fig. 12.15e,f).

For small $T$, the competition to learn depends on the initial values $y_{pq}(0)$ and is effectively probabilistic — neural units with a low net input have an increased chance of winning the competition to learn. For $T=1$, the initial values have such an effect that the neural unit with the largest net input does not necessarily win, even if its weight vector is only a few degrees from the input pattern. This is particularly evident when thresholded input neural units are employed. For a given input pattern, most neural units are operating in their chaotic regions and a winner appears to be chosen almost at random. Consequently, weight vectors tend to ‘jump’ around the unit circle and poor models of $\rho(x)$ are produced. However, as $T$ increases, the neural unit whose weight vector is closest to the input pattern tends to win and the chaotic self-organizing map behaves more like the conventional network. Consequently, more accurate models of $\rho(x)$ are obtained. For the same value of $T$, networks of mapped input neural units tend to produce more accurate models of $\rho(x)$ than do those consisting of thresholded input neural units. This is because the neural unit whose weight vector is
Figure 12.15. Input patterns were selected from a uniform distribution and presented to a chaotic self-organizing map consisting of 8 neural units with $\eta(0) = 1.0$ and $\eta(1000) = \eta_{\text{min}} = 0.01$. Final positions of weight vectors for (a) mapped input neural units and $T = 1$, (b) thresholded input neural units and $T = 1$, (c) mapped input neural units and $T = 10$, (d) thresholded input units and $T = 10$, (e) for mapped input units and $T = 50$, and (f) for thresholded input units and $T = 50$. 

(a) $\sigma = 20.1^\circ$

(b) $\sigma = 37.2^\circ$

(c) $\sigma = 5.3^\circ$

(d) $\sigma = 9.6^\circ$

(e) $\sigma = 4.0^\circ$

(f) $\sigma = 3.9^\circ$
closest to the input vector has a greater chance of winning than it does when thresholded input neural units are used.

The ability of the chaotic self-organizing map to cluster input patterns improves as $T$ decreases, while the accuracy to which it is able to model the probability density function of input patterns improves as $T$ increases. Therefore, it is necessary to make a compromise and for the remaining simulations $T$ was chosen to be 10.

Although the conventional self-organizing map is able to model a continuous probability density function relatively accurately, it does not perform as well when $\rho(x)$ is discontinuous. Because some neural units never win the competition to learn and those that do can only move towards the nearest cluster of input patterns, the accuracy to which $\rho(x)$ is modelled depends on the initial positions of weight vectors. In the chaotic self-organizing map, neural units do not necessarily become sensitive to the nearest cluster of input patterns and, therefore, the chaotic self-organizing map should outperform the conventional network when $\rho(x)$ is not continuous.

![Figure 12.16. (a) Four equiprobable clusters of input patterns and (b) the corresponding ideal positions of weight vectors.](image)

In order to test this hypothesis, input patterns were pseudo-randomly selected from four equiprobable clusters (Fig. 12.16a). Again 1000 input patterns were presented to a self-organizing map consisting of eight neural units, where only the winning neural unit was allowed to learn. Ideally, the weight vectors of these neural units would have moved to the positions indicated by Fig. 12.16b, so that each neural unit would be sensitive to the same number of input patterns. However, during training, the weight vectors of the conventional self-organizing map tended to move from their initial positions towards the nearest cluster of input patterns, with some weight vectors remaining in their initial positions because the corresponding neural units were not able to win the competition to learn. For example, in one simulation, the weight vectors moved from their initial positions (Fig. 12.17a) to those indicated in Fig. 12.17b; while one weight vector remained in its initial position, the others moved towards the nearest cluster of input patterns. Thus, the accuracy with which the network is able to model $\rho(x)$ depends on the initial positions of the weight vectors. Similarly, the performance of the chaotic self-organizing map depended on the initial positions of the weight vectors. Networks of mapped input neural units often produced slightly better models of $\rho(x)$, as measured by $\sigma$, than did conventional networks (e.g., Fig. 12.17), because all neural units participated in learning. However, the weight vectors still tended to move towards the
closest cluster of input patterns. When thresholded input neural units were employed, again all neural units were able to win the competition and participate in learning. Furthermore, the weight vectors did not necessarily approach the closest cluster of input patterns. Although all neural units became sensitive to a single cluster of input vectors (Fig. 12.17d), relatively poor models of $p(x)$ were still produced because once weight vectors became 'attached' to a cluster of input patterns they were unable to 'escape'.

![Figure 12.17](image-url)

Finally, 1000 input patterns were drawn from four clusters with different probabilities (Fig. 12.18a) and presented to a self-organizing map consisting of nine neural units. The weight vectors should have modelled the distribution from which the patterns were selected (Fig. 12.18b) but, in practice, they tended to move to positions that modelled $p(x)$ only slightly better than the pseudo-randomly generated initial positions (e.g., Fig. 12.19). Again the performance of the networks depended on the initial positions of the weight vectors, with neural units tending to become sensitive to the nearest cluster of input patterns.
Figure 12.18. (a) Four non-equiprobable clusters of input patterns and (b) the corresponding ideal positions of weight vectors.

Figure 12.19. Input patterns from four non-equiprobable clusters were presented to a network of 9 neural units with $\eta(0)=1.0$ and $\eta(1000)=\eta_{\text{min}}=0.01$. (a) Initial positions of weight vectors, (b) final positions for the conventional network, (c) for the chaotic network with mapped input neural units and $T=10$, and (d) for the chaotic network with thresholded input neural units and $T=10$. 

\[ \sigma = 66.0^\circ \]
\[ \sigma = 60.7^\circ \]
\[ \sigma = 59.8^\circ \]
\[ \sigma = 43.1^\circ \]
12.3.3 Generating a Spatial Order

Kohonen (1982) emphasized the importance of the global ordering of input patterns which can be produced by the self-organizing map. The formation of such a spatial order relies on the interaction of neural units during learning, which is achieved by allowing neighbours of the winning neural unit to modify their weight vectors. Changes in individual neural units are only reinforced if they result in global order. By enabling the neighbours of the winning neural unit to learn, weight vectors are able to 'escape' from a cluster of input patterns and, therefore, improved models of the probability density function of input patterns should be produced.

![Diagram](image)

**Figure 12.20.** A one-dimensional array of neural units where neighbours of the winner are allowed to learn at the reduced rates $\eta_1$ and $\eta_2$.

In the following simulations, the neural units of the self-organizing map were arranged in a one-dimensional array with two neural units either side of the winner being allowed to modify their weight vectors. While the winning neural unit learnt at a rate $\eta$, its immediate neighbours learnt at a reduced rate $\eta_1$ and their neighbours at a rate $\eta_2$ (Fig. 12.20), where typically $\eta_1 = \eta/4$ and $\eta_2 = \eta_1/4 = \eta/16$. Because the neural units at either end of the array have fewer neighbours, they have fewer opportunities to learn, which tends to cause a 'contraction' of the distribution of weight vectors. In order to avoid these edge effects and allow a fair comparison with earlier simulation results, the array of neural units was considered to 'wrap around', effectively producing a circular arrangement of neural units.

Input patterns were randomly selected from the unit circle and presented to a self-organizing map consisting of eight neural units. Because the network spends some time developing a spatial order, it may take longer to train and so 1500 input vectors were presented. The initial learning rate was chosen to be the lower value of $\eta(0) = 0.5$, since neural units learn more frequently — modifying their weight vectors not only when they win the competition but also when they are neighbours of the winning neural unit. Regardless of
the initial positions of weight vectors, both the chaotic and conventional self-organizing maps developed a global spatial order and, at the same time, were able to model the uniform distribution from which the input patterns were drawn relatively accurately (Fig. 12.21). As with the conventional network, the chaotic self-organizing map may generate a local rather than a global spatial order if there is not sufficient interaction between neural units (Fig. 12.22). This may arise if the learning rate is low (e.g., $\eta(0) = 0.2$) or if too few neighbours of the winning neural unit are allowed to learn.
Figure 12.22. A local spatial order developed by the chaotic self-organizing map employing thresholded input neural units, $T=10$ and $\eta(0)=0.2$.

When input patterns were selected from non-continuous probability density functions, the interaction of neural units not only resulted in a global order, but also produced significantly more accurate models of $\rho(x)$ (Figs 12.23 and 12.24). The interaction of neural units allowed the weight vectors to 'escape' from the nearest cluster and, consequently, the performance of the self-organizing map was less dependent on their initial positions. In both cases, the chaotic self-organizing map employing thresholded input neural units produced the most accurate models of $\rho(x)$. In addition to generating a spatial order, interaction between neural units affects the final positions of weight vectors. In the conventional network, this interaction tends to distribute weight vectors relatively evenly around the unit circle. However, in the chaotic self-organizing map with thresholded input neural units, it tends to draw the weight vectors away from the average of the cluster where they congregated when only the winner was allowed to learn (Fig. 12.17d) and closer to their ideal positions (Fig. 12.23d).

The clustering problems encountered with the conventional self-organizing map are usually overcome by ensuring that every neural unit has the opportunity to learn. This is achieved by defining a large neighbourhood around the winning unit and allowing all neural units in this neighbourhood to modify their weights. Thus, although a neural unit may never win the competition, it can still learn by being in the neighbourhood of a winning unit. However, chaotic self-organizing maps, particularly those employing thresholded input neural units, provide an alternative solution to this problem by giving all neural units the opportunity to win the competition to learn. As with the conventional self-organizing map, a global ordering can be produced by allowing neighbours of the winning neural to learn. In addition, the interaction of neural units tends to improve the accuracy with which the network is able to model the probability density function of input patterns.

12.4 SUMMARY

This Chapter emphasizes the usefulness of the versatile neural unit both as a model of neuron behaviour and as a way of introducing chaos into neural networks. Networks of versatile neural units were shown to produce activity qualitatively similar to EEG signals. Therefore, study of such chaotic neural networks and their behaviour may aid understanding of the patterns of activity observed in biological neural systems.
Figure 12.23. Input patterns from 4 equiprobable clusters were presented to a network of 8 neural units with $\eta(0)=0.5$, $\eta(1500)=\eta_{\text{min}}=0.01$ and $\eta_2=\eta_1/4=\eta/16$. (a) Initial positions of weight vectors, (b) final positions for conventional network, (c) final positions for chaotic network using mapped input neural units and $T=10$ and (d) using thresholded input neural units and $T=10$. Brackets indicate r.m.s. error for the same network when neural units learn independently of each other.

Furthermore, the versatile neural unit provides a simple way of introducing chaos into conventional neural networks that are able to learn. In particular, two variants of the versatile neural unit — mapped input and thresholded input — were incorporated into the self-organizing map. The ability of the resulting chaotic self-organizing map to cluster input patterns, model their probability density function and generate a global spatial order were examined and compared with those of the conventional self-organizing map. The chaotic self-organizing map outperformed the conventional network at clustering input patterns, par-
Figure 12.24. Input patterns from 4 non-equiprobable clusters were presented to a network of 8 neural units with $\eta(0)=0.5$, $\eta(1500)=\eta_{\min}=0.01$ and $\eta_2=\eta_1/4=\eta/16$. (a) Initial positions of weight vectors, (b) final positions for conventional network, (c) final positions for chaotic network using mapped input neural units and $T=10$ and (d) final positions for chaotic network using thresholded input neural units and $T=10$. Brackets indicate r.m.s. error for the same network when neural units learn independently of each other.

particularly for thresholded input neural units, large initial learning rates and small $T$. However, the chaotic self-organizing map produced better models of continuous probability density functions for large $T$. Therefore, a compromise had to be made between clustering input patterns and modelling a continuous $\rho(x)$ and, consequently, $T$ was chosen to be 10. When $\rho(x)$ was not continuous very poor models were produced unless neural units were allowed to interact during learning. This interaction not only produced more accurate models of $\rho(x)$ but also resulted in a global spatial order.
In general, the performance of networks employing mapped input neural units lay between that of networks consisting of linear neural units (i.e., the conventional self-organizing map) and that of the chaotic self-organizing map employing thresholded input neural units.

The introduction of chaotic neural units into the self-organizing map resulted in an improved ability to cluster input patterns and model their probability density function. Further research should include evaluating the performance of the chaotic self-organizing map on higher dimensional problems.
Part V

IMPLICATIONS FOR BRAIN RESEARCH
CHAPTER 13

CONCLUSIONS AND FUTURE RESEARCH

"The highest activities of consciousness have their origins in the physical occurrences of the brain just as the loveliest of melodies are not too sublime to be expressed by notes."

W. Somerset Maugham (1874–1965)

13.1 INTRODUCTION

This thesis examines the contributions engineering makes to brain research and, in particular, focuses on processing electroencephalograms (EEGs), neural networks and deterministic chaos. The two major achievements presented are: (1) a system for the automated detection of epileptiform transients in EEGs and (2) a neural unit which enables deterministic chaos to be introduced into neural networks. A list of the papers and presentations that have been prepared during the course of this research can be found in the Preface. In this Chapter conclusions are drawn from the work presented and aspects of this work warranting further investigation are identified.

13.2 PROCESSING ELECTROENCEPHALOGRAMS

Automated analysis of EEGs can be divided into two categories: (1) analysis of background activity (§5.2) and (2) the detection epileptiform activity (§5.3). In this thesis, emphasis is on the detection of epileptiform activity and, in particular, on the detection of epileptiform transients in inter-ictal EEG recordings.

Despite numerous attempts to automate the detection of epileptiform transients in EEGs (§5.3.2), very few systems have been introduced into clinical use and, even then, with only limited success. The central problem is the high number of false detections, which not only reflects the difficulty in distinguishing epileptiform transients of variable morphology from a wide range of background activities and artifacts, but also that insufficient attention is paid to spatial and temporal contextual cues present in the EEG. In contrast to the EEGer, who relies heavily on the presence of epileptiform transients and artifacts on adjacent channels as well as on the presence and spatial distribution of epileptiform events elsewhere in the EEG, automated systems tend to consider only a very limited context.

A system for the automated detection of epileptiform activity in EEGs has been developed (Chapter 6), which makes considerable use of both spatial and temporal contextual information to detect epileptiform events and to reject artifacts and background activity. This system consists of three stages: data collection, feature extraction and event detection. The data collection stage filters, samples and digitizes EEGs. The feature extraction
stage, which is implemented in C, detects candidate epileptiform transients by calculating and thresholding parameters of individual EEG waves. An optimal set of parameters for distinguishing between epileptiform and non-epileptiform waves were determined by statistical analysis (Appendix B). Thresholds for these parameters were selected so that at least 80% of epileptiform transients are detected. The event detection stage is a rule-based expert system. Extensive collaboration with the EEGer was required to determine relevant knowledge and rules, which were subsequently encoded in Prolog. Development of this system has been in a number of stages; at each stage more data have been analysed and greater emphasis placed on spatial and temporal contextual information present in the EEGs. To date, this system has been evaluated on 21 EEGs used during development. The overall detection rate of 64% with an average 3.5 false detections per hour compares very favourably with the performance of other systems (§6.6). However, the outstanding feature of this system is its ability to detect, on average, 53% of epileptiform events with no false detections.

Although evaluation of this system is still at a preliminary stage, the attainment of a more than satisfactory detection rate, coupled with zero false detections, is a considerable — and unmatched — achievement. It is this ability to eliminate false detections that will make the system applicable to routine clinical recordings. However, more data are required to evaluate its performance further and a clinical trial is underway. Further stages of development include processing referential montages, application to recordings with periods of sleep (long-term monitoring) and integration of the data collection, feature extraction and event detection stages to achieve a system capable of real-time on-line detection of epileptiform activity in EEGs. The application of back-propagation neural networks to the detection of epileptiform transients is also being investigated (§8.5) and a number of modifications have been proposed (§8.4) that should enable the desired network performance to be achieved.

A reliable system for automatic detection of epileptiform activity, such as that presented here, coupled with a background analysis technique (§5.2), such as topographic mapping, would comprise an extremely powerful EEG analysis system. Such a system would not only be a useful clinical tool which would save time and increase objectivity and uniformity, but would also be an invaluable research tool, enabling studies of the relationships between interictal epileptiform activity and medication, seizures or sleep stages, and of the significance of various epileptiform transient morphologies.

13.3 CHAOS IN NEURAL NETWORKS

There is a growing body of evidence that chaos is present in the brain both on the neural level (§10.2) and in EEGs (§10.3). In fact, the discovery of chaos has renewed the hope that the origins and significance of the EEG will one day be understood. Although the existence of chaos in the brain is becoming more widely accepted, many questions concerning its purpose remain unanswered. There is much debate as to whether the presence of chaos in the brain is merely a by-product of the interconnection of many nonlinear elements or whether it has evolved for a specific purpose. Although speculations on the functions of chaos in the brain are rife (§10.4), they tend to be rather vague and unsubstantiated. Better understanding of the roles of chaos in the brain are needed and may be provided by mathematical models. The introduction of chaos into neural units and, subsequently, into neural networks may provide some insight into the functions of chaos in the brain and, at the same time, may enhance the performance of conventional neural networks.

In Chapter 11, a new neural unit was proposed and termed the versatile neural unit. Although this neural unit is relatively simple, it still displays a rich range of behaviour reminiscent of that observed in neurons. Furthermore, the inputs themselves determine the type
of activity produced by the neural unit, which accords with the experimental observations of Freeman (1987).

The versatile neural unit not only provides a simple model of neuron behaviour but also provides a means for introducing chaos into neural networks. The resultant chaotic neural networks display activity qualitatively similar to that observed in EEGs (§12.2). Therefore, study of chaotic neural networks and their behaviour may aid in understanding the origins and significance of EEGs.

Because the versatile neural unit retains characteristics of conventional neural units, it can be incorporated into conventional neural networks. In particular, two variants of the versatile neural unit — the mapped input and the thresholded input — were introduced into the self-organizing map. The resultant chaotic self-organizing map displayed an improved ability to cluster input patterns and to model their probability density function, particularly when thresholded input neural units were employed. Thresholded input neural units effectively introduced more chaos into the self-organizing map than did mapped input neural units and, for the simple two-dimensional problems studied, produced the best network performance. By introducing chaos into self-organizing map, all neural units were afforded the opportunity to learn. Thus, each neural unit became associated with a cluster of input patterns. Furthermore, neural units did not necessarily become associated with the nearest cluster of input patterns and so neural units were effectively able to 'forget' their initial conditions. Chaos provided the variability that allowed the network to forget its initial conditions and to model its new environment. Thus, chaos may provide the mechanism that enables animals to adapt to an ever-changing environment.

Further research should include evaluating the performance of the chaotic self-organizing map on higher dimensional problems, investigating the optimal amount of chaos to be included, and the introduction of versatile neural units into other established neural networks.
APPENDIX A

THE 10–20 INTERNATIONAL SYSTEM

The locations of electrodes on the scalp for EEG recording were standardized by Jasper (1958). Nineteen positions on the scalp were identified, which relate in a proportional way to four reference points — the bridge of the nose (nasion), the bump at the back of the head immediately above the neck (inion) and the depressions above the cheek bones just in front of the ears (left and right preauricular points). In addition to these, a further two electrode positions (a1 and a2) located on the ear lobes are often used.

The locations are labelled with alphabetic characters (f frontal, c central, p parietal, t temporal, o occipital) and numbers (odd on the left and even on the right).

![Diagram of 21 electrode placements according to the 10-20 International System.](image)

Figure A.1. Placement of 21 electrodes according to the 10–20 International System.

The standard positions of electrodes (Fig. A.1) are as follows:

1. The vertex cz is halfway between the nasion and inion, and halfway between preauricular points.

2. The two other midline locations, fz and pz, are 20% of the nasion-inion distance in front of and behind the vertex.
3. $c3$ and $c4$ are 20% of the interauricular distance to the left and right of the vertex respectively on the interauricular line.

4. $t3$ and $t4$ are 40% of the interauricular distance to the left and right of the vertex.

5. $f3$ and $f4$ are 20% of the interauricular distance left and right of $fz$.

6. $f7$ and $f8$ are 40% of the interauricular distance left and right of $fz$.

7. $p3$ and $p4$ are 20% of the interauricular distance left and right of $pz$.

8. $t5$ and $t6$ are 40% of the interauricular distance left and right of $pz$.

9. $fp1$ and $fp2$ are 10% of the nasion-inion distance above the nasion and 10% of the interauricular distance left and right of the midline.

10. $o1$ and $o2$ are 10% of the nasion-inion distance above the inion and 10% of the interauricular distance left and right of the midline.
APPENDIX B

OPTIMAL PARAMETER DEFINITIONS

The wide variety of epileptiform transient morphologies and their similarities to artifacts and background waves makes automated detection far from straightforward. Several approaches have been used to detect epileptiform transients. The most common of these is the mimetic approach, where the EEG is divided into waves and parameters of each wave are calculated, thresholded and compared with measures of the background activity.

It is generally accepted that the important parameters for distinguishing between epileptiform transients and non-epileptiform waves are sharpness, amplitude and duration (Gotman, 1980; Ktonas and Smith, 1974; Saltzberg et al., 1967). However, each of these parameters can be calculated in a variety of ways. Furthermore, these parameters need to be compared with the background activity. There are many methods available to quantify background activity and a number of ways to normalize wave parameters with respect to background measures.

A statistical analysis was performed to determine optimal definitions for parameters in terms of their ability to discriminate between epileptiform transients and similar appearing artifacts or background waves.

Epileptiform transients and non-epileptiform waves were selected from six EEGs and a variety of parameters calculated for each wave. These included parameters of each wave and its constituent halfwaves, measures of the background activity and comparisons of wave parameters with background measures.

B.1 PARAMETERS

For the two halfwaves constituting each wave, a number of amplitude, slope and duration parameters were calculated. These included the peak-to-peak duration and amplitude, the average halfwave slope, the maximum halfwave slope and the peak halfwave slope. The peak slope of each halfwave was calculated by fitting a line to several (2, 3, 4, 5, and 6) points near the peak using a least mean squares technique (see Fig. 6.3b).

For each wave, the amplitude from the peak of the wave to a floating mean — the average EEG value over 75, 100, 125 or 150 ms — was determined. Other wave parameters were derived from those of the constituent halfwaves; for example, sum of halfwave durations, average halfwave amplitude, angle between halfwave slopes, product of halfwave slopes and sum of halfwave slopes.

A number of measures of the background activity were also computed. These were calculated over one second, centred on the wave under consideration. The mean and standard deviation of the peak-to-peak halfwave durations, amplitudes and slopes were computed along
with the average distance of the EEG from the floating mean and the average slope of the EEG. The coefficients of variation (i.e., standard deviation divided by mean) for halfwave durations, slopes and amplitudes were determined to provide measures of the rhythmicity of background activity.

Several techniques were employed to normalize the parameters of individual waves with respect to the background measures. These included dividing parameters of individual waves by those of the background activity, computing the difference between wave and background parameters (i.e., difference from mean), and dividing this difference by the corresponding standard deviation (i.e., number of standard deviations from mean).

In total 297 parameters of each wave were calculated and included in the statistical analysis.

**B.2 DISCRIMINANT ANALYSIS**

Discriminant analysis is a statistical technique that enables a researcher to study the differences between two or more groups with respect to several variables. In particular, discriminant analysis can be employed to discover useful discriminating variables from among a large number of variables. Although based on the assumption that variables have a normal distribution, it appears to be relatively robust (Klecka, 1988).

The discriminating ability of each variable can be measured in a number of ways (e.g., F ratio, Mahalanobis distance, Wilk's lambda, Rao's V, residual variance). In the following analysis, the F ratio was used which measures the difference between the means of a variable for each group but also takes into account the standard deviation (Fig. B.1).

![Figure B.1. Distribution of two variables. Although the difference between means is greater for Variable #1, Variable #2 has the greater discriminating ability and larger F ratio.](image)

Initially, each EEG was considered separately. Thus, parameters of epileptiform transients were compared with those of non-epileptiform waves from the same EEG. It is unrealistic to present the F ratios for all 297 parameters and for each EEG. Therefore, results are shown for only a selection of parameters.

The measures of wave amplitude and sharpness are:
Table B.1. F ratios for a number of measures of amplitude, relative amplitude, sharpness and relative sharpness for each EEG considered independently.
Amplitude from peak to floating mean calculated over 75 ms
Average of peak-to-peak halfwave amplitudes
Sum of peak-to-peak or average halfwave slopes
Sum of peak halfwave slopes
Sum of maximum halfwave slopes
Angle between peak-to-peak halfwave slopes
Angle between peak halfwave slopes
Angle between maximum halfwave slopes
Product of peak-to-peak halfwave slopes
Product of peak halfwave slopes
Product of maximum halfwave slopes.

These parameters are normalized in a number of ways with respect to the following background measures which are calculated over one second:

- Average distance of EEG from floating mean ($\bar{A}$)
- Average peak-to-peak halfwave amplitude ($\bar{A}_2$)
- Standard deviation of peak-to-peak halfwave amplitudes ($\sigma_{A_2}$)
- Average halfwave slope ($\bar{S}$)
- Standard deviation of halfwaves slopes ($\sigma_S$)

F ratios for amplitude, relative amplitude, sharpness and relative sharpness parameters are presented in Table B.1. F ratios are not comparable between EEGs because of the different number of waves in each group. The largest F ratio for each EEG is shown in large bold type, while the largest F ratio for each parameter type and each EEG is emboldened.

These results reveal that, in general:

1. The amplitude to mean ($A_1$) is better than the average peak-to-peak amplitude ($A_2$).
2. The relative amplitude ($A_1/\bar{A}$) is better than amplitude itself ($A_1$).
3. The best measures of sharpness are those based on the peak halfwave slopes (i.e., $\sum P_k, L P_k, N P_k$).
4. The relative sharpness ($\sum P_k /\bar{S}$) more important than sharpness alone ($\sum P_k$).
5. The relative amplitude ($A_1/\bar{A}$) is more discriminating than the relative sharpness ($\sum P_k /\bar{S}$).

Another discriminant analysis was then performed to compare epileptiform transients from each EEG with a single set of non-epileptiform waves. The set of non-epileptiform waves was chosen from EEG #5 because of the wide variety of background activities and artifacts in this EEG. On the basis of the initial results where each EEG was considered separately (Table B.1), it was decided to use $\bar{A}$ as the measure of background amplitude and not to include the product of halfwave slopes in the analysis. F ratios for the remaining parameters are shown in Table B.2.

The F ratios confirm that the relative amplitude and relative sharpness provide better discrimination between epileptiform and non-epileptiform waves than do the amplitude and
Table B.2. F ratios for a number of measures of amplitude, relative amplitude, sharpness and relative sharpness for epileptiform transients of each EEG compared with the same set of non-epileptiform waves.
sharpness. From these results it is clear that, in general, the optimal parameters are $\frac{A_1}{\bar{A}_1}$ and $\sum P_k / \bar{S}$.

In further analyses, epileptiform transients from each EEG were compared with waves that often closely resemble epileptiform transients, that is, alpha waves and muscle spikes. The results of these analyses indicated that the average halfwave duration, the coefficient of variation (i.e., mean/standard deviation) of halfwave durations and the coefficient of variation of halfwave amplitudes are useful discriminating parameters.

Even at this stage, it is worth remembering that the final system must run in real-time and, therefore, speed is of the essence. For this reason and because the background amplitude and background slope are highly correlated, it was decided to normalize both amplitude and wave parameters with respect to the background amplitude ($\bar{A}_1$).
REFERENCES


J. D. Bronzino (1986) Biomedical engineering and Instrumentation : Basic concepts and applications. PWS Engineering.


B. Saltzberg, R. G. Heath and R. J. Edwards (1967) EEG spike detection in schizophrenia research. *Digest of 7th International Conference on Medical and Biological Engineering* 206.


