THE UNIVERSITY OF HULL

The Application of Time Encoded Signals to Automated Machine Condition Classification using Neural Networks

being a Thesis submitted for the Degree of

Doctor of Philosophy

in the University of Hull

by

Walter Lucking
BEng (Hons) DipEE Hull

October 1997

"The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them"

Sir William Bragg

DECLARATION

I Walter Lucking, do declare that this work has not been submitted for any other degree at this or any other University.

ACKNOWLEDGEMENTS

I owe a debt of gratitude to all those people who helped me during the course of this research and also those who have helped since during the compilation and completion of this Thesis. I would also like to take this opportunity to thank Doctor David Chesmore who has provided both guidance and support during the course of the work.

Thanks must also go to the Science and Engineering Research Council without whose funding none of this work could have been undertaken.

Finally to my wife and my family who have been so patient and supportive during the course of writing this Thesis I offer my utmost thanks. Without their ever present support and encouragement the writing would have been so much more difficult.

SYNOPSIS

This thesis considers the classification of physical states in a simplified gearbox using acoustical data and simple time domain signal shape characterisation techniques allied to a basic feedforward multi-layer perceptron neural network. A novel extension to the signal coding scheme (TES), involving the application of energy based shape descriptors, was developed. This sought specifically to improve the techniques suitability to the identification of mechanical states and was evaluated against the more traditional minima based TES descriptors. The application of learning based identification techniques offers potential advantages over more traditional programmed techniques both in terms of greater noise immunity and in the reduced requirement for highly skilled operators. The practical advantages accrued by using these networks are studied together with some of the problems associated in their use within safety critical monitoring systems.

Practical trials were used as a means of developing the TES conversion mechanism and were used to evaluate the requirements of the neural networks being used to classify the data. These assessed the effects upon performance of the acquisition and digital signal processing phases as well as the subsequent training requirements of networks used for accurate condition classification. Both random data selection and more operator intensive performance based selection processes were evaluated for training. Some rudimentary studies were performed on the internal architectural configuration of the neural networks in order to quantify its influence on the classification process, specifically its effect upon fault resolution enhancement.

The techniques have proved to be successful in separating several unique physical states without the necessity for complex state definitions to be identified in advance. Both the computational demands and the practical constraints arising from the use of these techniques fall within the bounds of a realisable system.

CONTENTS

Page

Chapter 1

Page

1. Introduction

This thesis details research undertaken during a three year Science and Engineering Research Council funded contract into the application of digital signal processing and neural networks to the problems associated with monitoring the condition of machinery acoustically. It describes the development and evaluation of a number of novel techniques which enable intelligent automation of the monitoring process. The thesis is arranged into five main subsections which describe the key areas of research undertaken during the study. The first three Chapters present the theory and technical considerations involved in the development of the novel analysis techniques whilst Chapters 5 and 6 focus on a series of practical trials directed at evaluating the application of these techniques to a simplified gearbox testbed system.

The work begins with a short introduction detailing the fundamental difficulties associated with the monitoring of machinery together with an outline of proposed methods of approaching some these problems. Chapter 2 gives a detailed account of a range of currently available techniques which can provide feedback on the physical condition of a range of machine types.

Chapter 3 introduces the signal conversion techniques which will be employed as a means of presenting signal characteristics to a classification system and discusses the implementation of these techniques using digital signal processing hardware. This type of implementation could eventually lead to the development of more operationally flexible low cost on-line real time monitoring systems.

Chapter 4 discusses some of the aspects associated with the neural techniques available to perform the final classification of the condition status. Considerations involved in the successful application of this classification mechanism are also discussed.

Chapters 5 and 6 deal with the practical implementation issues involved in the evaluation of different application mechanisms based around the processing techniques discussed in the earlier Chapters. In particular some of the potential pitfalls of specific implementation procedures are examined as are some of the benefits accrued by these techniques. In each case a series of results are presented which seek to highlight the findings of the research.

Finally in Chapter 7 the conclusions of the work contained within this thesis are discussed together with an outline of areas in which future research is required to further evaluate the potential of automated monitoring and management of machinery in situ. Appendices are included at the end of the thesis which detail work performed and techniques used during the course of the work undertaken which it is felt would detract from the general flow of the main body of the text.

1.1 Condition Monitoring

Condition monitoring is a relatively new field of science created primarily out of the necessity for improved efficiency in modern machinery which is becoming increasingly more complex. This advancement in sophistication and consequently cost brings with it a necessity not only to operate more efficiently but also to extend the life cycle of sub-components and enhance safety. Making decisions about the condition of machinery whilst it is in operation based upon the characteristics of data acquired from it is essential to this process. Traditionally this aspect of condition identification was highly dependant upon the knowledge and experience of skilled personnel. More recent advances in computer technology have enabled increasingly more sensitive and responsive applications to be developed. Such systems can not only operate continuously without impairing performance but also reduce reliance upon costly experienced personnel.

All monitoring techniques depend upon one essential premise, that there is a measurable symptom produced as a direct result of each fault condition of the system being monitored. The capability of the monitoring system is then dependant upon the manner in which the characteristics of the measurable symptoms are examined and upon the frequency with which they are examined. Figure 1.1 illustrates the association between the various elements which characterise an operational machine and the monitoring system tasked with identifying its instantaneous status. A particular monitoring system implementation may utilise one or more of these components each of which may be measured in a number of ways.

Condition monitoring has expanded into ever more demanding and cost sensitive fields together with the rapid development of affordable computing power and now extends to a wide range of machine types. For this reason the work contained within this thesis has

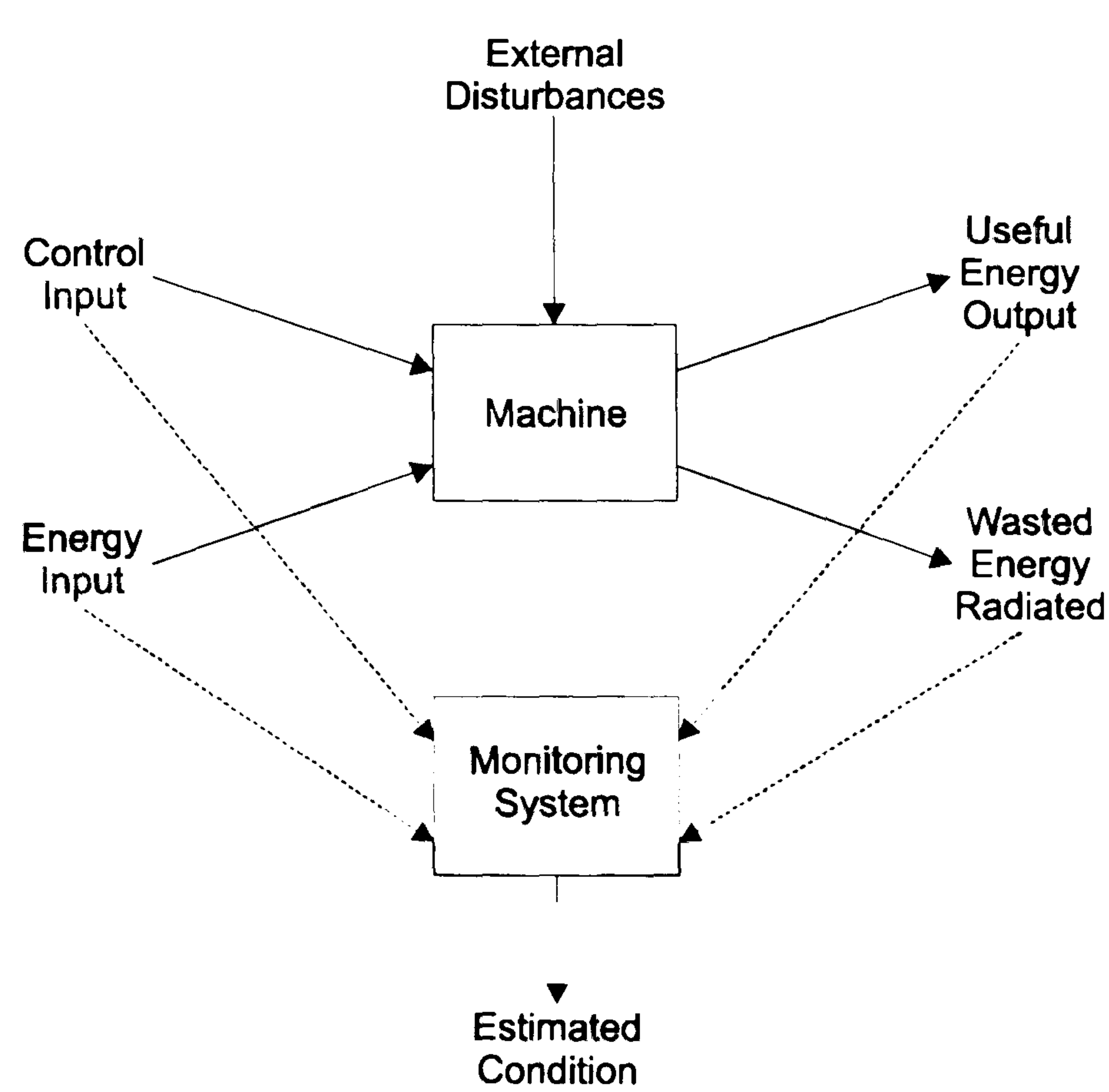

Figure 1-1 A generalised condition monitoring system

focused upon the understanding of a unique subsection within this emerging sector. Although consideration will be given to some of the techniques used in other areas, such as the internal combustion engine, the main focus of the work concentrates on continuous rotating machinery such as turbines, drills, and more specifically gearboxes. Gearboxes represent one of the most common types of machine subsystem and are, like all mechanical systems, susceptible to a range of common faults which may reduce their efficiency, or in some circumstances cause physical failure. Because they are employed in situations where large rotational forces must be transferred they are exposed to high stress loads which increase the likelihood of wear and failure. The application of monitoring techniques to these mechanical devices provides potential for a reduction in the necessity for periodic maintenance as well as the capability for optimising the mechanical performance of systems whilst they are in operation.

The intention of this work is not only to evaluate some of the techniques currently employed for monitoring but also to identify new techniques which may be applicable to the field. The ultimate goal being the development of techniques which make possible on-line automated monitoring of machines either in periodic or continuous modes without the necessity for complex sensor arrangements and highly trained personnel.

1.2 Neural Classifiers

Traditionally condition classification has been performed in one of two ways. The first is manually derived with human evaluation of the available data and the second is an automated algorithmic technique developed as a result of knowledge and/or trials. Whilst the first method is generally more robust as a result of the high level of human intervention, the second is both cheaper and less prone to many types of error associated with human involvement. As technology progresses and the drive for industrial efficiency accelerates, the number of instances where human intensive systems are either feasible or cost effective becomes more limited. However the more traditional automated methods which include techniques such as template correlation and spectral analysis can suffer as a result of inflexibility and their inherent lack of "understanding". They are by definition dependant upon algorithm(s) which are generally tailored by an expert with detailed knowledge of the system under observation. Once tailored to a specific application they may be difficult to modify and can be sensitive to sources of both internal and external interference which have not been specifically catered for.

More recent development of neural networks presents the possibility of implementing monitoring systems which can offer significant reductions in operator overhead, whilst at the same time reintroducing some of the human-like strengths vital to flexible solutions. Neural networks are simple computational models inspired by the human brain which attempt to mimic some of the behavioural aspects of these biological systems. In common with the brain these models comprise many simple processing elements, or nodes, which when combined are able to perform highly complex functions. In contrast to more traditional monitoring techniques, which are composed of pre-programmed sequential computing modules, these networks can be trained and are inherently parallel in nature. They represent not only an alternative method but also a potential means of improving response.

For the purposes of this thesis the specific types of network which will be discussed are supervised multilayer perceptron networks. Prior to operational use they require a period of supervised training in which a problem is presented to the network as a series of exemplars which enables the network to "learn" the problem. The perceptron nodes within these networks are grouped into distinct layers, each of which is connected to nodes within neighbouring layers. Each individual interconnection has associated with it a weight which is used to modify the stimuli passed between itself and other unique processing nodes. This combination of multiple processing nodes in multiple layers interconnected by configurable weight factors modifying stimuli which are transmitted between nodes in the network imparts the network with an innate learning capability. For anything but the simplest of problems two or more layers are required to provide sufficient degrees of freedom in the classifier to adequately learn the problem. Once a network size and architecture has been selected, the training period is used to modify the individual interconnecting weights using an error back-propagation algorithm in conjunction with a set of data exemplars which describe the problem. As each of the exemplars within the training set is presented to the network in turn the network "learns" by evolving its weights to suit the exemplar. Providing the network is able to reach a single weight position which satisfies the demands of all exemplars in the training set the network is able to "understand" the problem and can subsequently be used to classify previously unseen data.

If the problem has been sufficiently described by the training exemplars the network should also then be capable of making decisions upon similar data as well as incomplete or noisy data. Thus neural techniques should provide a mechanism with which to implement a more robust and flexible solution without the necessity for human supervision. They combine the repeatability and consistency of a programmed implementation together with some of the human-like characteristics of knowledge based estimation.

1.3 Time Encoded Signal (TES) Condition Characterisation

Independent of the method used to perform condition classification there is a fundamental requirement to present detailed characterisation data to enable an accurate physical description to be collated. The key facets of this data are that it encompasses sufficient indicators of the physical condition and can be transformed into a format suitable for presentation to the neural classifier. For the sake of simplicity a neural classifier containing neither nodal memory nor feedback was selected for this latter classification stage. This simple architectural constraint confines the decision making process to time independent state classification. Essentially this means that condition data must be separated into discrete packets each of which are processed by the classifier completely independently to determine a corresponding state.

Time encoded signals, or TES, refers to a time domain technique originally developed by King *et al* [1] to convert a speech signal into a series of discrete shape descriptors which are subsequently used to identify a simple word vocabulary. This algorithm provides an ideal mechanism with which to achieve both of the previously mentioned fundamental requirements of a neural network based condition monitor, that of data characterisation and presentation. TES consists of subdividing a discretised analogue signal into a number of unique elements, or epochs. These epochs are then analysed and converted into TES symbols according to their physical shape characteristics. Two unique conversion algorithms are discussed in this thesis, each of which is developed into two separate characterisation formats. Both algorithms focus on different aspects of the signal epoch shape characteristics. So called minima coding is associated with the harmonic content and amplitude coding with the energy content. Unlike King's work the target signal is not human speech but mechanical emissions. Once the conversion is completed the symbol stream so produced is post-processed to generate data in one of the previously mentioned presentation formats.

Two presentation formats were evaluated, the first containing simple symbol histogram information and the second, the A-matrix, more complex histogram and shape information. Both generate data matrices which can be considered as data signatures each of which corresponds to a unique physical condition or state. These signature matrices provide an ideal means of applying condition data to a neural network, each element of the matrix being represented by a single node in the input layer. Prior to application however each matrix element requires normalisation to fulfil the physical requirements of the individual network processing nodes transfer functions which for the purposes of this work demand inputs to be within the range 0-1.

One advantage of employing this simple time domain discrete signal conversion mechanism is the ease with which it may be implemented in digital signal processing hardware. With currently available hardware it is feasible for an entire classification system to be implemented in real-time on a single digital signal processing (DSP) board. This would include the discrete acquisition and pre-filtering of a condition signal as well as the conversion into a TES representation, the transformation into a series of presentation matrices and the final classification by a neural network.

1.4 Proposed Method of Evaluating Automated Condition Classification

Because the field of condition monitoring is wide and varied the intention of this thesis is to identify a subsection from within this field with which to evaluate a series of novel techniques used to classify physical system state. A simplified gearbox was constructed to act as the testbed from which these studies could be performed. This particular type of mechanical subsystem was selected for its relative dynamic simplicity and widespread use within the industrial environment. The dynamic simplicity is imparted by the continuous nature of the rotational movement which reduces the complexity of the subsequent data processing required to generate the relevant condition signatures. Unlike cyclic devices such as combustion engines or pumps the continuous nature of gear rotation introduces fewer constraints upon the selectivity with which data must be acquired to provide acceptable condition information.

The proposed method of acquiring the condition information is via a single acoustic microphone the output of which will, for evaluation purposes, be recorded onto high quality audio tape prior to processing. The recordings are subsequently used to generate acoustic signatures using the TES conversion algorithms prior to final application to a neural network post-classifier for state identification. The core of the system is developed from within a PC type environment which provides the additional benefits of both low cost and industrial suitability. The combination of acoustically derived TES data and neural classifiers offers the potential for the generation of a flexible, trainable, low cost system which would require significantly less overhead in terms of skilled manpower for operational use than any currently available.

A block diagram of the various stages contained within the processing and classification mechanism, all of which are discussed in detail in the following Chapters, is illustrated in Figure 1.2. The initial stage, data capture, depicted in the top left of the diagram will acquire the classification associated acoustic emissions which will ordinarily contain additional acoustic disturbances not associated with problem under observation. Following data capture the analogue signal is converted into a discrete representation before TES conversion is carried out. The filtering of the raw emissions required to condition the signal may be performed before or after ADC conversion depending upon whether it is implemented in hardware or software. An algorithm selector is required to control the conversion from a series of discrete samples into specific types of TES shape descriptor symbols prior to the neural network conditioning. This conditioning module will also require a control input to identify the type of conditioning required for network application. The training input depicted in the final processing module, at the network application stage identifies the control required during the pre-operational system

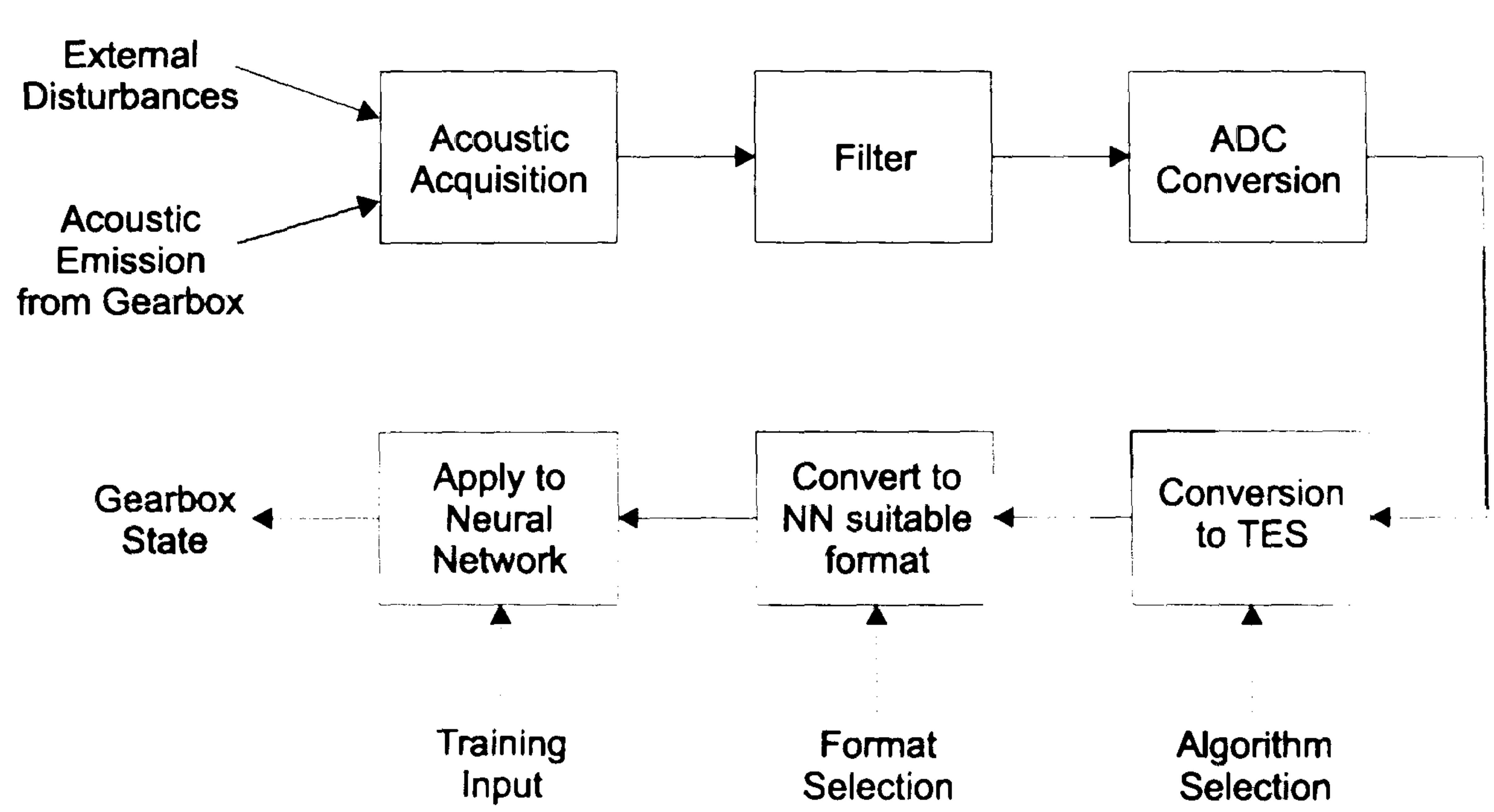

Figure 1-2 Illustration of the gearbox TES classification system

training phase. This is the element of the application which requires input from an operator to direct the neural networks learning and thus its subsequent state classification capability. However once the initial training phase has been suitably completed this control input is unnecessary.

There are several aspects of the work contained in this thesis which are both novel and original. Whilst TES has been applied to numerous speech applications it has so far been limited to a single application within the field of condition monitoring using a so called minima TES conversion mechanism to define the waveform shape descriptor symbols. In this thesis a new technique has been developed which is based upon the energy characteristics of the waveform, termed amplitude TES. Whilst it retains the essential simplicity of conversion characteristic of minima TES it is more applicable to the field of condition monitoring where signal energy fluctuations are the common result of variations in physical condition. This new conversion scheme is evaluated during practical performance trials, detailed in Chapters 5 and 6, against the minima technique.

In addition to this new conversion mechanism a more simplistic neural data application format was evaluated. Rather than the more commonly employed A-matrix presentation format a more basic, and somewhat more compact presentation format, the histogram matrix was evaluated. Whist this method retains only the most basic physioacoustical cue information its represents a significant simplification of the neural network complexity required to perform the subsequent classification. In cases where the physical states are clearly separated this technique may prove to be adequate.

The third novel aspect of the work is the implementation of the techniques using dedicated digital signal processing (DSP) hardware which is becoming increasingly more powerful and widely available. It offers the opportunity to perform the acquisition, conversion and classification of acoustic data from machinery to identify the system state both on-line and in real time. When compared to some of the tools currently available to the condition monitoring fraternity this type of system would offer significant enhancements in operational cost, portability, flexibility and most importantly performance.

A list of the material published during the course of these studies and containing the results of a variety of investigations is included at the end of the thesis in Appendix A.

Chapter 2

Page

2. Condition Monitoring In Industry

Condition monitoring is not a recent addition to the industrial workplace. Monitoring and overhaul of machinery has however historically been carried out manually by groups of highly skilled personnel. These personnel have accumulated, over many years of servicing, experience which has proved invaluable in the continued operation of industrial plant. However with the ever increasing volume, variety and complexity of machinery in use today this role becomes ever more specialised and challenging. Combined with these increasing demands comes the constant drive for improved productivity and enhanced efficiency.

Prior to the recent acceleration in modern computer technology monitoring was restricted to the use of this skilled manual identification which included procedures for regular overhaul, visual inspection of parts and lubrication fluid contamination as well as audible fault identification. The labour intensive nature of these techniques is not only wasteful of resource and parts but also is heavily dependant on the level of skill of operators gained through their experience. More recently the manufacturing process has become increasingly complex with reductions being made in operating tolerances both of plant and of products to increase efficiency. In response to these pressures there is now a growing need in manufacturing industry for advanced monitoring tools with which to carry out these manual tasks; the aim being to reduce both production costs and the dependency upon the highly skilled personnel to carry out these tasks.

Developments have now reached the point at which the manual procedures and skilled personnel who were relied on previously can be outperformed through the application of modern computer technology. This technology is focused around the rapid development of PC based hardware with ever increasing levels of computational as well as storage capacity. Also included amongst these developments is another significant advancement which will no doubt play a key role in the development of reliable and cost effective tools, that of the digital signal processor (DSP). These high performance processors have been developed specifically for the purposes of implementing "real time" signal processing algorithms.

Many earlier developments relied on expensive, bulky, emerging computer technology so the applications to which they were put was in heavy industry and air/sea transport where mission and human safety were critical. However, the advancements in computer technology have produced ever more powerful devices at much reduced prices and sizes. This in turn has led to advancements in monitoring in areas once considered either impractical or uneconomically viable. Consequently the whole field of condition monitoring is expanding into newer and more cost sensitive product areas.

Condition monitoring has now developed into a specialist branch of science solely devoted to the enhanced understanding and recognition of performance degradation and failure modes in machinery. The purpose of this research branch has been to further evaluate the widely understood patterns of wear and damage with the intention of being able to determine in a more methodical and consistent manner the identifiable points on the wear cycle. The goal is to develop simple, objective, cost effective methods of fault identification and early warning that provide the gains in productivity and efficiency that are achievable. In association with these requirements they must be designed in such a way as to be easily integrated into harsh industrial environments.

The research is based not only on the application of modern computational power to known distinguishing measures of machine status but also to the development of additional novel techniques. The aim is to find ever more accurate measures of system state able to recognise the significant, not necessarily quantitatively large, variations in condition so that identification of irregularities can be made earlier, and if possible automatically. If failure modes can be determined in advance of catastrophic failure then remedial action can be planned and prepared for together with any necessary parts. This provides certain key medium and long term benefits to users:

(i) A reduction in economic losses incurred through unplanned stoppage could be expected, particularly in complex plant facilities where failure of a single component can halt production. In turn the cost of periodic overhaul, parts wastage and downtime can be minimised since expectation is replaced by identification. There is no longer a need to be driven by the estimated mean time before failure (MTBF) of parts, which themselves are only an indication of expected lifetime, to plan servicing.

(ii) The stock of spare parts which are usually kept to provide cover for emergency backup of key plant could be reduced if monitoring can provide enough lead time to order and receive spare parts and plan for a controlled shutdown.

(iii) A reduction in the level of costly industrial accidents caused by excessive wear, faulty parts or human error during servicing could be expected.

(iv) A reduced requirement for the highly skilled personnel who would otherwise be needed to provide systems diagnosis and servicing backup.

The greatest benefits, of course, will be achieved where the monitoring and overhaul of large numbers of machine stock distributed over a plant may be rationalised through the application of automated techniques. For example in the oil and gas industry many valves, turbines, pumps and other system sub-components may be distributed over a large site. The environmental and safety requirements mean that without automated monitoring much time and capital expenditure can be consumed on the types of manual monitoring and regular overhaul which have been described. There is therefore a clear requirement for the widespread availability of automated techniques for monitoring. How these requirements are satisfied is of course open to further discussion. Both the direct physical measurement of the machines mechanical parameters (e.g. power output and maximum revolutions) and physical measurement of the changes in mechanical condition would require costly shut-downs. In contrast, the measurement of residual changes as a result of degradation in machine condition do not require a shutdown of the plant. As a consequence this is by far the most promising technique and is where most research is now concentrated. It covers both intrusive (sensors placed on or inside the device) and non intrusive (sensors located in the vicinity of the device), periodic and continuous monitoring configurations.

If the decision is taken to monitor the subject periodically then the periods between monitoring samples being acquired must be selected carefully so as to provide sufficient warning of performance degradation whilst at the same time producing sufficient return on the investment in applying monitoring. If performance degradation is more difficult to predict or particularly mission critical then continuous monitoring, though more expensive, is probably required. The work of Rose [2] on failure modes of helicopter components completed in conjunction with Boeing helicopters highlights this criticality aspect. They report key componentry progressing to failure in time frames shorter than the average flight time rendering ground based tools inadequate. As a result their recommendations were for the development of real-time airborne diagnostics.

The intention of this Chapter is to introduce to the reader some of the wide range of techniques which have been developed through this research to apply tests of a non-destructive nature to determine system state. It is not the intention to discuss the relative merits of each technique with regard to the novel work covered later in this thesis but merely to present the work as a background to the field of condition monitoring at this time.

The first and by far the most commonly studied field among researchers is the development of tools to study the *residual processes* resulting from changes in the system state. This refers to the study of the vibroacoustical emissions from a target system to estimate the level of wear or damage. The technique is probably better described as *measurement analysis* and encompasses the extraction of condition information from data acquired using sensors either in the time domain or frequency domain.

The second and generally less common method is *prediction analysis*. This encompass any technique which seeks to develop a mathematical model of a particular target subject. These models are generally used to simulate certain failure modes in order to compare them with results obtained during testing. Due to its potential complexity this technique is generally limited to the simpler devices in which all the system interactions can be adequately modelled and combined.

2.1 Measurement Analysis of Machine Condition

Measurement analysis describes the family of techniques which attempt to separate and decode the cause-and-effect chain of events underlying changes in vibroacoustic emissions. The diagnosis begins with the selection and positioning of the necessary sensor(s). As would be expected both the choice and positioning of these devices can affect the overall performance of the diagnosis. The selection may be influenced by the requirements of the diagnosis system whilst the physical properties of the device, which affect the transmission paths of emissions, may limit the effective sensor positions.

Once the raw signals have been captured they can be utilised either in their raw basic format to produce a prognosis or, as is more usually the case, post-processed to extract the useful condition information and reduce noise effects. It is these processing techniques which provide the key to a successful system implementation combining accuracy, speed and simplicity. Each type of processing algorithm makes its own particular demands upon the computational equipment available and is dependent upon the level of complexity. The following two sub-sections introduce some of the techniques which have previously been developed by other researchers and are intended to provide the reader with some background knowledge regarding the work which has been completed previously within the field.

As discussed previously, measurement analysis techniques can be separated into two sub-categories, those involving processing in the time domain and those employing frequency domain manipulation. The uses to which each have been put will be discussed in isolation before summarising the relative merits of each at the end of the Chapter.

2.1.1 Time Domain Signal Processing

Many of the earliest techniques developed to recover health status information from the machinery were based upon the study of spatio-temporal or time domain variations in the sensor data. This concentration of effort was in most part due to the tools available to early workers in the field. The lack of widespread commercially viable tools to extract the spectral aspects of this data limited the early development of this type of study. Even today with the increasingly widespread availability of computational tools able to extract spectral information the time domain remains a key aspect of much new work. In fact the implementation of accurate temporal synchronisation with the monitored system provides an opportunity to apply time selection to the filtration of sensor data. It also enables the accurate reconstruction of a time history of events taking place within the monitored system enables us to highlight any functional variations in structure. This last point is of special note in the study of variations within reciprocating machinery where each cycle is composed of several distinctly separate events which each contribute to the "group" emissions.

This section covers many of the significant processing sub-sets encompassing the temporal study of machinery together with examples of their use either within the research community or in "real world" development applications. It is not the intention here to detail the acquisition of data *per se* for this itself depends upon the sensor selected for the application. Instead most of the commonly employed methods (vibration, acceleration and acoustic acquisition) will be covered during the discussion of the processing strategies. Two simple estimators, namely root mean square (rms) and peak signal measures are presented initially. From these simple techniques more complex estimators are developed which are directed towards providing less load sensitive discriminant measures.

2.1.1.1 Root Mean Square Signal Assessment

Stronach *et al* in their work on rolling element bearings [3] report this, one of the simplest and most common approaches to time domain analysis, the estimation of overall intensity of a wideband vibration signal. The calculation of the root mean square (rms) of the input condition signal, $s(t)$ over an observation period T, defined in (1) provides useful information relating to the general health state of a bearing.

$$\tilde{s} = \sqrt{\frac{1}{T}\int_{0}^{T} s^2(t)dt} \qquad (1)$$

Stronach *et al* found the measure to be particularly responsive to shaft alignment errors but less so to light wear damage to the bearing itself. This simple measure has been so widely employed that several standards (API611, 612, 613, 616; ISO3945, 2372; VDI2059) have been developed to provide recommended boundary conditions for various groups of machine type. These standards however give only generalised guidelines for monitoring as they have been derived as averages over a large number of machine types. When attempting to use such measures for the identification of condition states in a specific machine two important factors must be considered which can affect the accuracy of the measurements.

(i) The local environment surrounding the target machine which may have several other sources of potentially destructive additive noise.

(ii) Any time varying loading constraints imposed upon the machine.

As a result rms measurement provides an uncomplicated measure requiring simple hardware and a minimum of software processing. This, however, can be offset against some of the problems which could be encountered if a tool was developed to monitor a machine type to be placed in a range of diverse environments. Before the emergence of fast signal processing hardware the use of such an uncomplicated measurement technique would ensure a rapid response even from equipment with relatively modest processing capabilities. This speed of response was considered an important criteria in the work done by Dong *et al* [4] on adaptive drill system monitoring. With only 286 based PC hardware Dong was able to develop a system capable of monitoring chip congestion at the drill bit to workpiece interface. This proved that even relatively simple hardware can achieve a level of fault identification and that given a controlled environment rms measures should not necessarily be restricted to generalised monitoring of machine health status.

2.1.1.2 Peak Signal Analysis

Another method which bears consideration in terms of its widespread use and relative simplicity is signal peak assessment. This refers to the measurement of the maximum signal amplitude achieved in a given calculation time frame. It is generally more sensitive to the changes in condition, being less susceptible to external sources of disturbance. Consequently it should provide a more reliable diagnostic measure than rms. Changes in peak levels between the "new" state and just prior to overhaul of a bearing can be of the order of 10dB. An example of using this measure to determine the health or otherwise of two bearing types using just such variations in vibration velocity levels is shown in Figure 2.1 below. This is taken from [5] and relates to measurements taken from the bearings in a mechanical pump. The vertical line to the right of the diagram indicates the point at which a system overhaul was carried out resulting in an immediate reduction in vibration of about 10dB in both bearings.

However Martins in [6] argues that even 10dB amplitude variations can be experienced in bearings containing no faults due to the harmonic nature of the process. This is further backed up by reference to some international guidelines which were developed in an attempt to rationalise the ranges of variation expected in differing groups of machinery (e.g. ISO3945/2372, API611). These guidelines vary so widely that the use of such simple measures in tools may require the implementation of per machine system tolerance adjustments. Without such adjustments the measures would need to be tuned to account for all the varying mechanical tolerances expected during the production process as well as the environmental perturbations expected in-situ. This could impact severely on the potential sensitivity as a consequence.

Koizumi *et al* [7] carried out some work on journal bearing signature diagnostic measurement which provides some advancement on the basic use of peak analysis. He developed two alternative variations upon the measures we have discussed up to this point. The first was the use of rms and peak measurement comparisons to classify

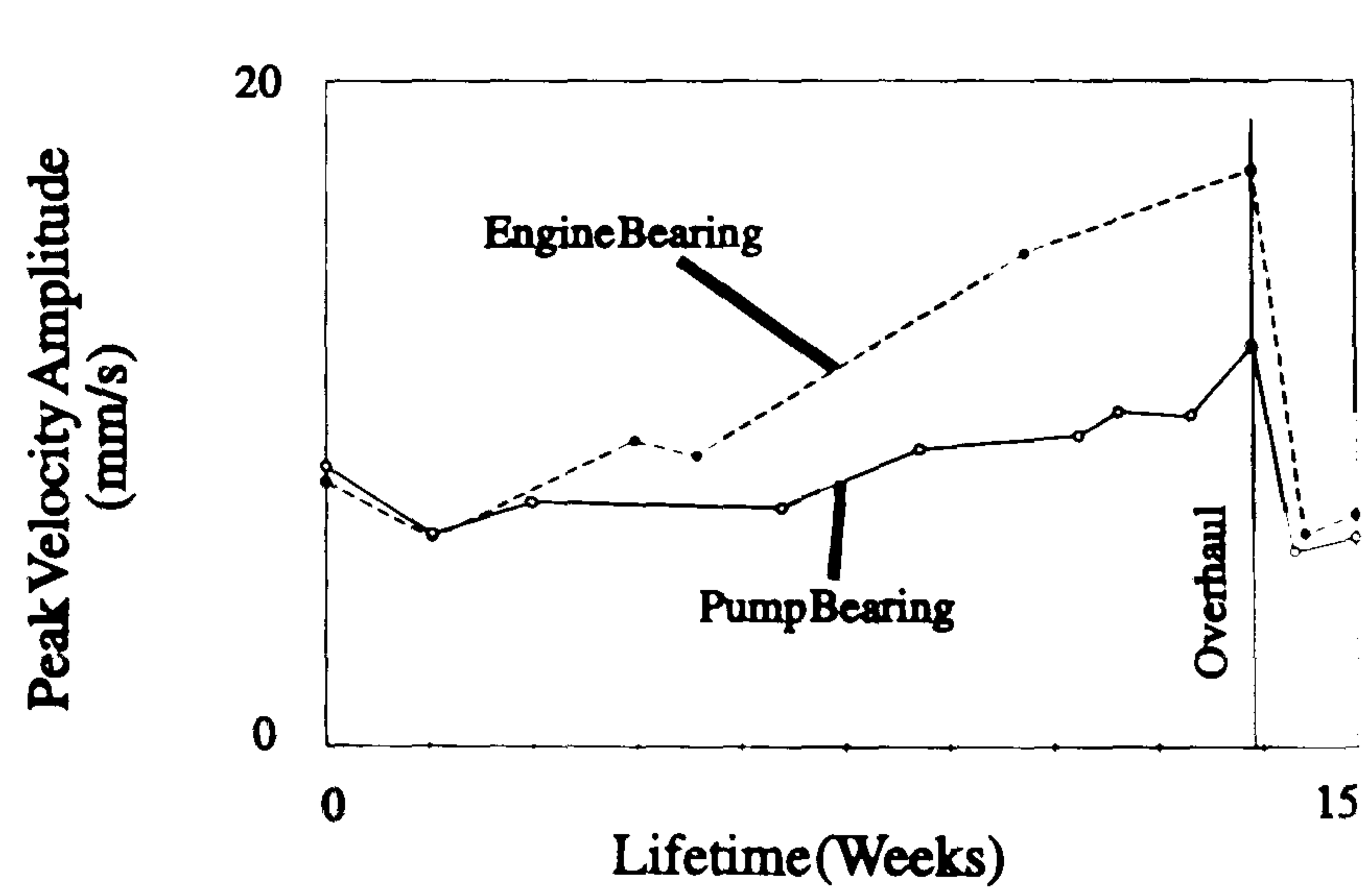

Figure 2-1 Change in peak vibration velocity of a pump bearing during utilisation [5]

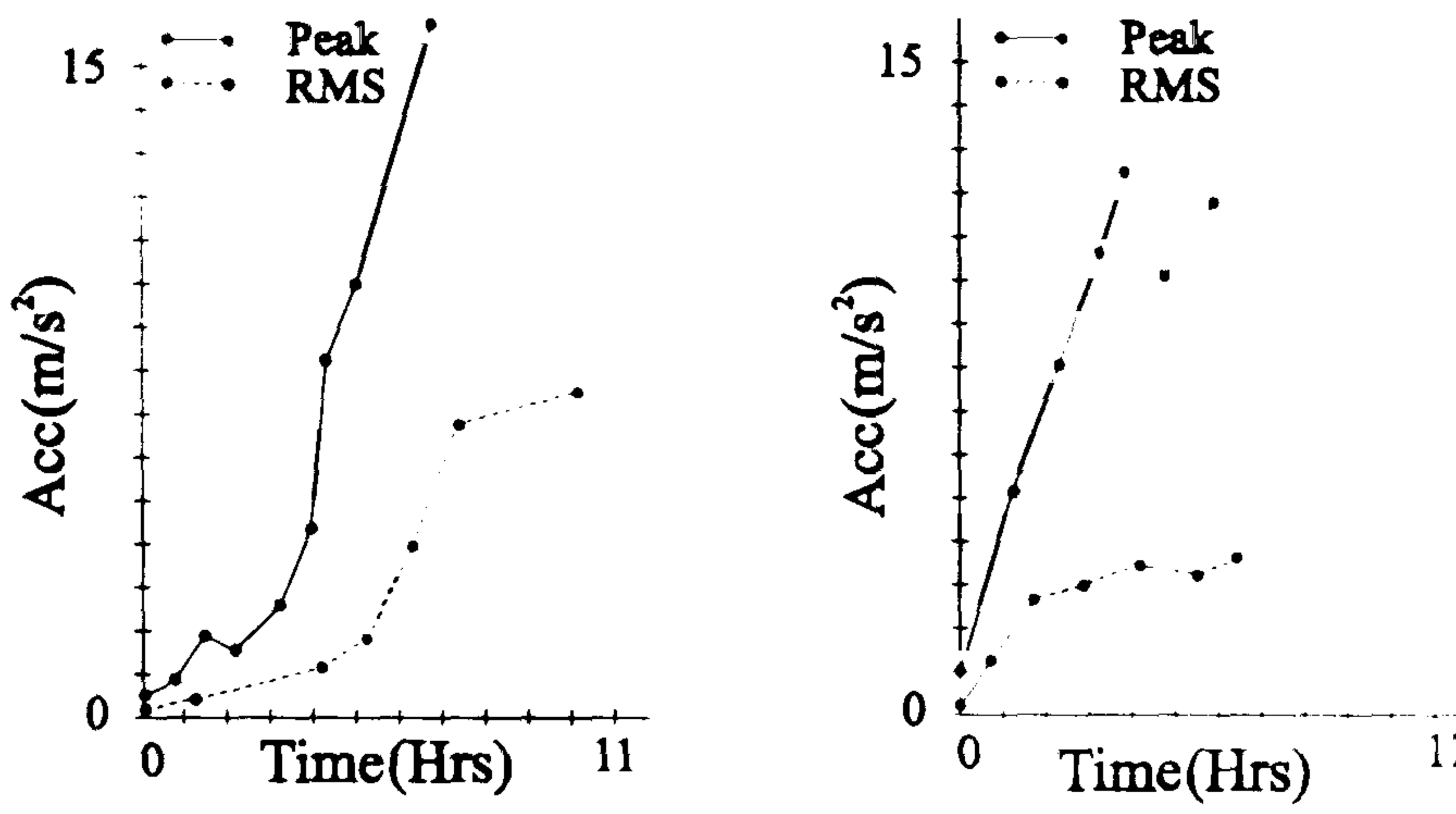

Figure 2-2 Improper Lubrication (left), and Foreign Material Inclusion in Journal Bearings (right). [7]

differing types of failure which provide some interesting observations. Figure 2.2 taken from their work illustrates graphically the difference between bearings containing foreign material and those with improper lubrication. Whilst improper lubrication will cause a gradual increase initially in vibration, the inclusion of foreign material produces some relatively rapid variations in vibration level. In the case of the plot on the right the initial increase is due to the inclusion of dust whilst the later reduction, at approximately four hours, is an indication of self exhaustion of the material by the bearing.

The second technique which they experimented with was the use of spectral selection prior to processing. Two filters were applied to the raw signal, one passing 0.5-20kHz and the other 10-20kHz, after which each was applied to a standard peak vibration measure. Figure 2.3 again taken from their work shows that the failure mode can be estimated by comparison. When components of both the wide band and high frequency agree the most likely failure modes are excess thrust loading or foreign material

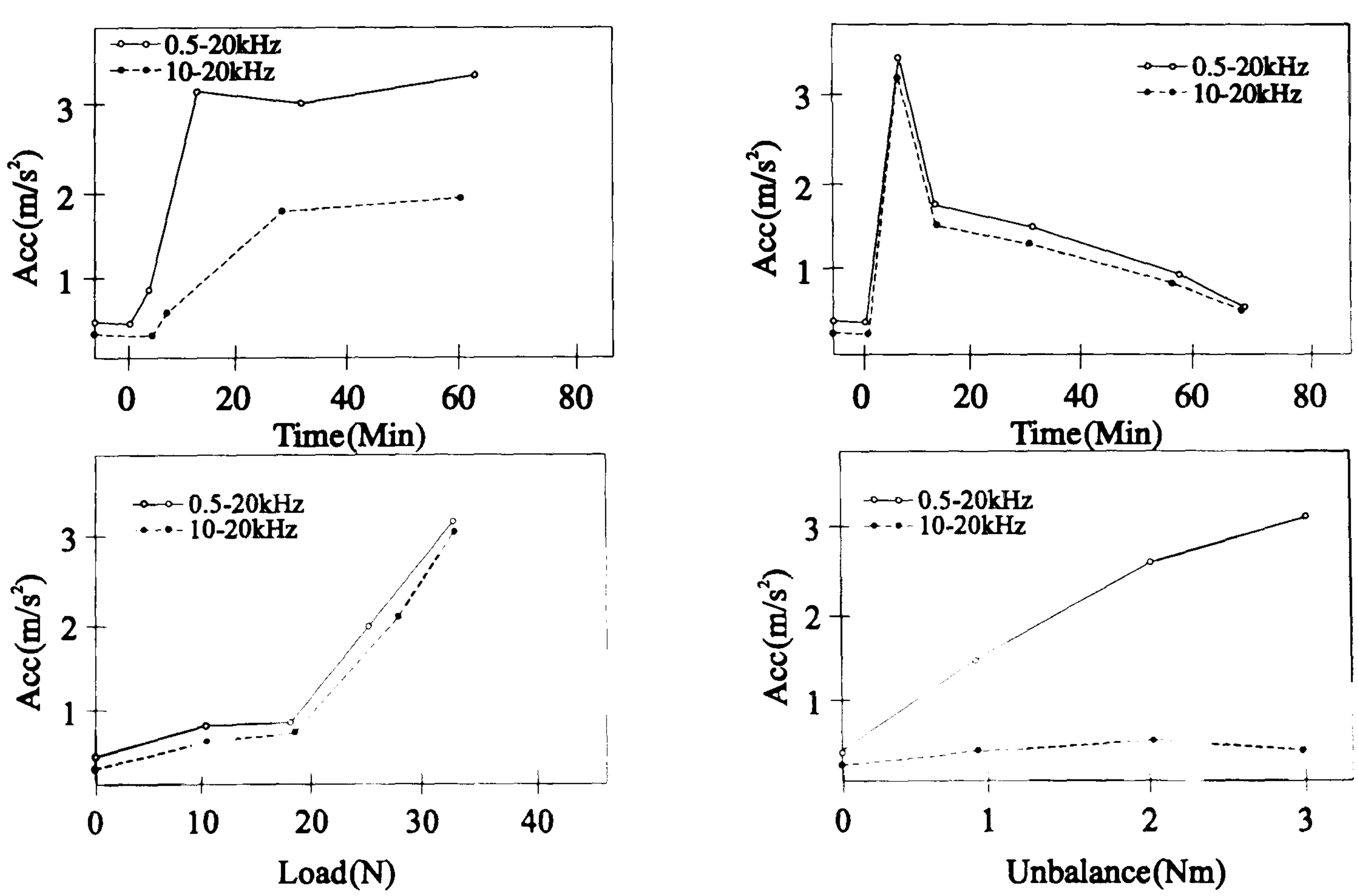

Figure 2-3 Structural response of bearings to Improper Lubrication (t-l), Foreign Material Inclusion (t-r), Excess Thrust Loading (b-l), and Unbalance (b-r), [7]

inclusion. By contrast if the two components do not agree then the most likely failure modes are improper lubrication or system unbalance. To separate each of these pairs of modes, further study would be required on the original waveform in the time domain.

2.1.1.3 Signal Kurtosis Analysis

What should be noted from the two measures described previously is the potential level of diagnostic supervision which would be required to implement such absolute measures in a diverse range of machine types. The development of dimensionless measures, particularly those that are insensitive to speed variations are much more readily applicable to automated diagnostics. A more recent addition to the diagnostics toolset is just such a measure. The statistical *kurtosis* measure is based upon the fundamental premise that a perfect component, for example a bearing, generates vibroacoustical signals in a random manner. Any change in this essential "randomness" caused by wear or failure brings with it corresponding changes in the statistical distributions of the amplitudes of the sensory signal (Figure 2.4).

Mathematically we can express these characteristic amplitude distributions in terms of a

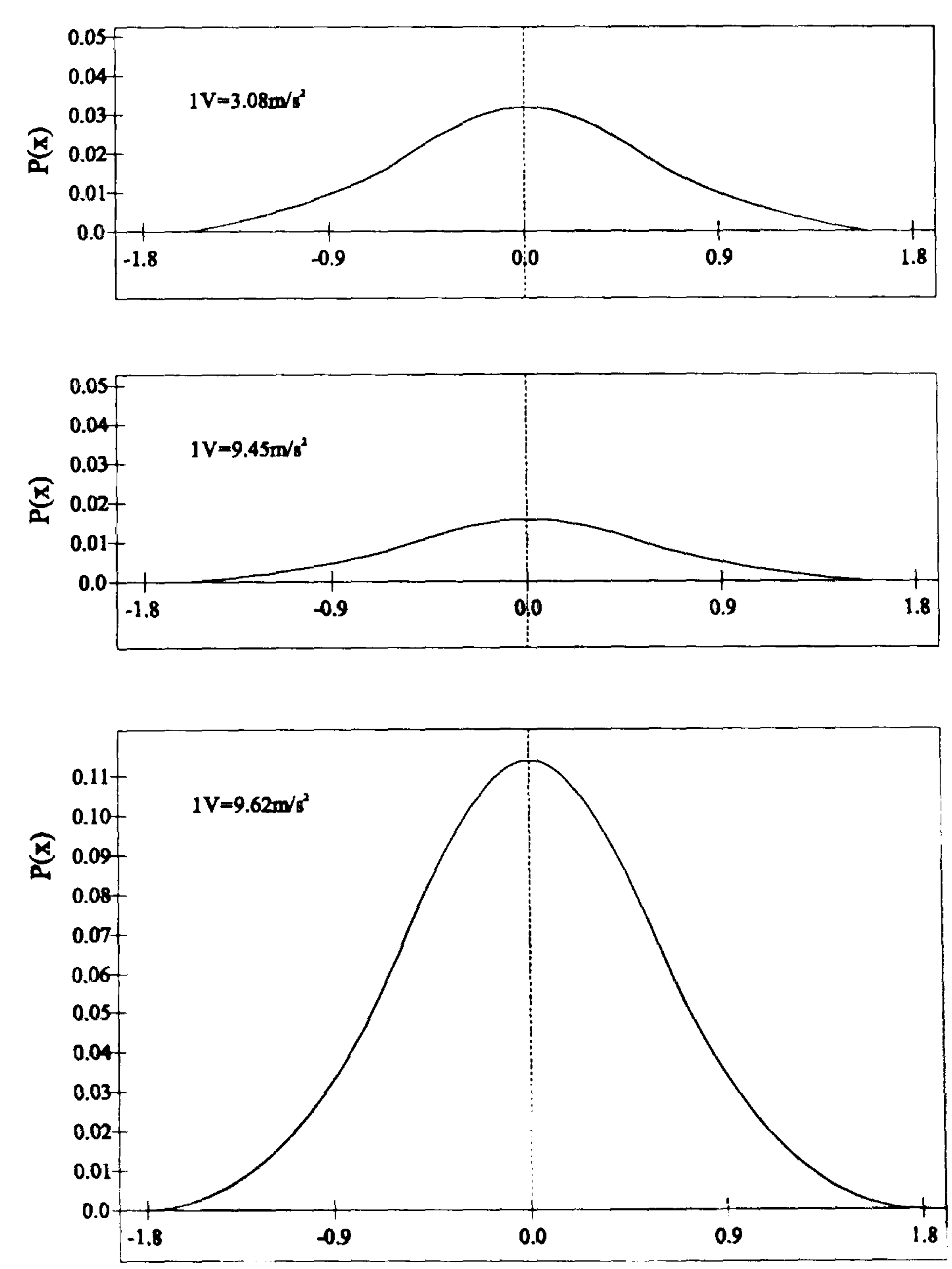

Figure 2-4 Statistical variations in signal amplitude for a perfect bearing (top), a bearing with incipient wear characteristics (middle), and displaying significant damage (bottom). [6]

probability density function, the statistical moments of which are defined by the general integral (2).

$$M_n = \int_{-\infty}^{\infty} (x - \bar{x})^n P(x) dx \tag{2}$$

Where,

$x, \bar{x}$ are the vibrational amplitudes and the mean vibrational amplitude respectively.

$P(x)$ is the instantaneous probability of the event x.

The kurtosis value, β_2, of the signal is defined as the fourth moment (M_4) normalised with respect to the standard deviation ($\sqrt{M_2}$) as expressed in (3).

$$\beta_2 = \frac{\int_{-\infty}^{\infty} (x - \bar{x})^4 P(x) dx}{\left\{ \int_{-\infty}^{\infty} (x - \bar{x})^2 P(x) dx \right\}^2} \tag{3}$$

Stronach *et al* [3] and Dyer [8] both reported that using this equation a nominally "perfect" bearing would exhibit a kurtotsis value of 3-4 whereas a worn or damaged one would be in the region 10-20. This would provide a sufficient level of stable separation to be incorporated within an automated diagnostic system. They also report that the discriminant measures are relatively independent of prevailing load or speed conditions and that they remain relatively unaffected by the transmission path effects introduced by the sensors. The distribution of amplitudes remains relatively constant despite variations in vibrational sensor signal path.

Once again, when this measure, like others described previously, is implemented in a "real" system some preliminary filtration is applied prior to measurement to reduce disturbances and enhance separation. It was noted by Stronach *et al* in [3] that trials of this measure by British Steel produced good condition separation when the analysis was performed on four separate frequency bands. They found that damage severity could be determined by examination of the distribution of kurtosis between the four bands providing the bearings were under load.

The experimental simulations, carried out by British Steel, of fault conditions in several bearings took the form of "baseline" kurtosis measurements for undamaged parts, followed by measurements of the same bearings after subjecting them to artificially induced faults. This type of measure, an example of which is depicted in Figure 2.5, proved accurate in correctly identifying roller and inner/outer race defects. However a note of caution is necessary here. Under "no load" conditions they found that impacting

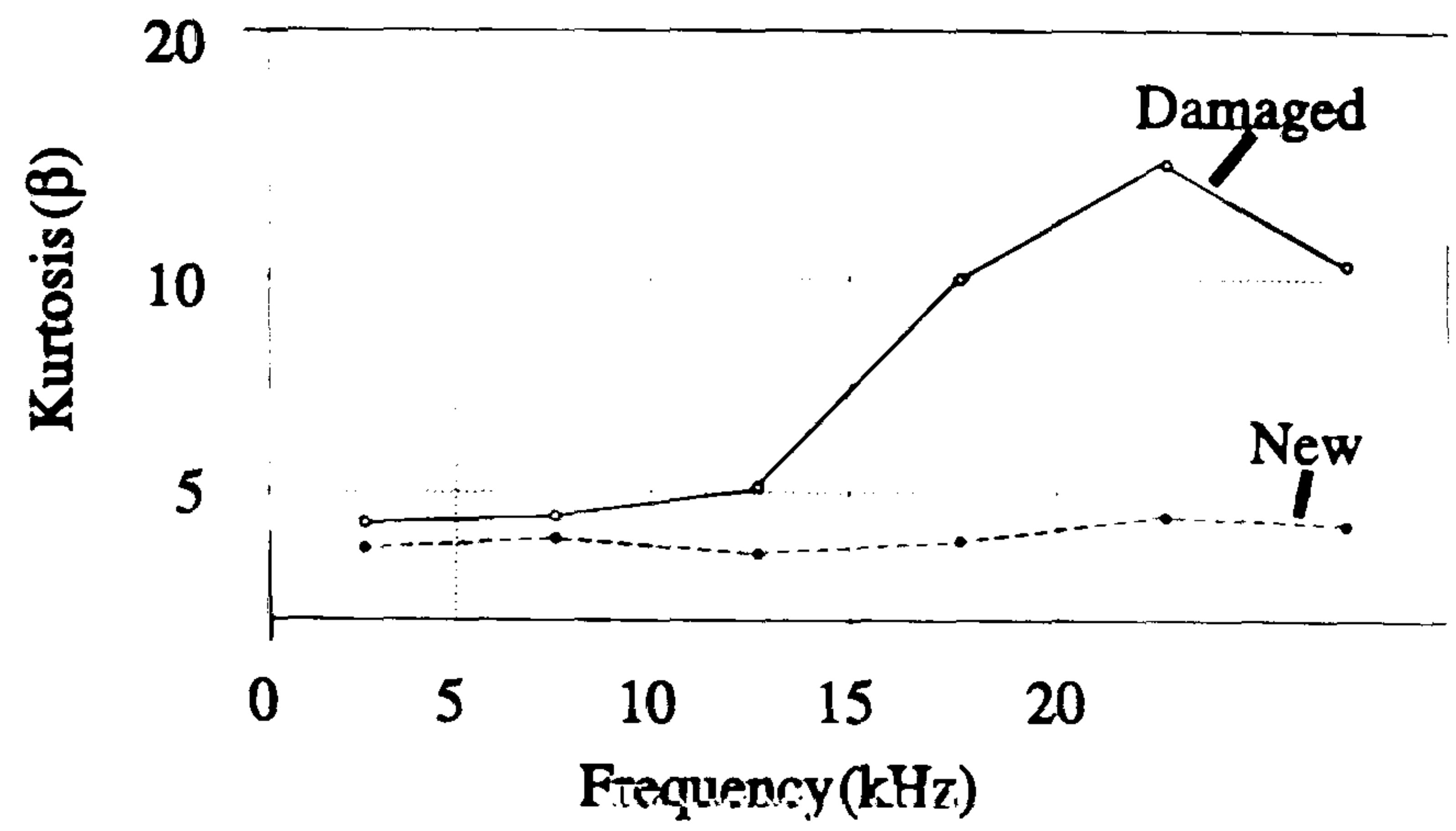

Figure 2-5 Comparison of the value of kurtosis, in six bands measured from vibrations in a bearing housing. [9]

forces caused by wear or damage are sufficiently reduced to limit the magnitude of variations in kurtosis and thereby the separation of states.

2.1.1.4 Combining Signal Measures to Improve Separation

Having now discussed three of the most common types of measure employed to separate the "condition states" of machinery or machine sub-components attention turns to the integration of these measures. In the first two cases, peak and rms, the greatest asset is simplicity rather than guaranteed reliability. The measures are both susceptible to outside sources of interference as well as the prevailing load and speed conditions of the source. The third, kurtosis, was somewhat more reliable and less easily influenced by changes in the prevailing load, speed, and sensor signal path. The discussion however of signal preconditioning and combined measures that have been outlined in both sections 2.2.1.2 (peak) and 2.2.1.3 (kurtosis) deserve further thought. This section will cover the use of such composite methods in more detail. McFadden in his work on fatigue crack analysis of helicopter gearboxes [10] makes a particularly relevant point regarding some of the simpler parameters we have been discussing. He suggests that advanced wear or damage can be identified using many of these simpler techniques but that very early detection, necessary in some critical applications, requires the use of more sophisticated signal conditioning strategies to enhance the information content prior to the application of a suitable separation measure or measures.

The Curtis-Wright sonic analyser, Figure 2.6, is an analysis system combining spectral filtering of the acoustic signal and the use of rms level measurements. It was originally developed for use in monitoring gas turbine systems and comprises a series of narrow, approximately 15Hz, bandpass filters centred about key component characteristic frequencies (e.g. ball-pass frequency in a bearing). Each of the pre-filtered signals is then applied to a standard rms detector of the type previously described. In this way a measure of condition for one of several constituent parts whose characteristics are contained within a specific band can be derived for the machine.

In order to deploy this type of system in an industrial environment there are two problems to overcome. The first is the selection of the relevant acoustic bands. This

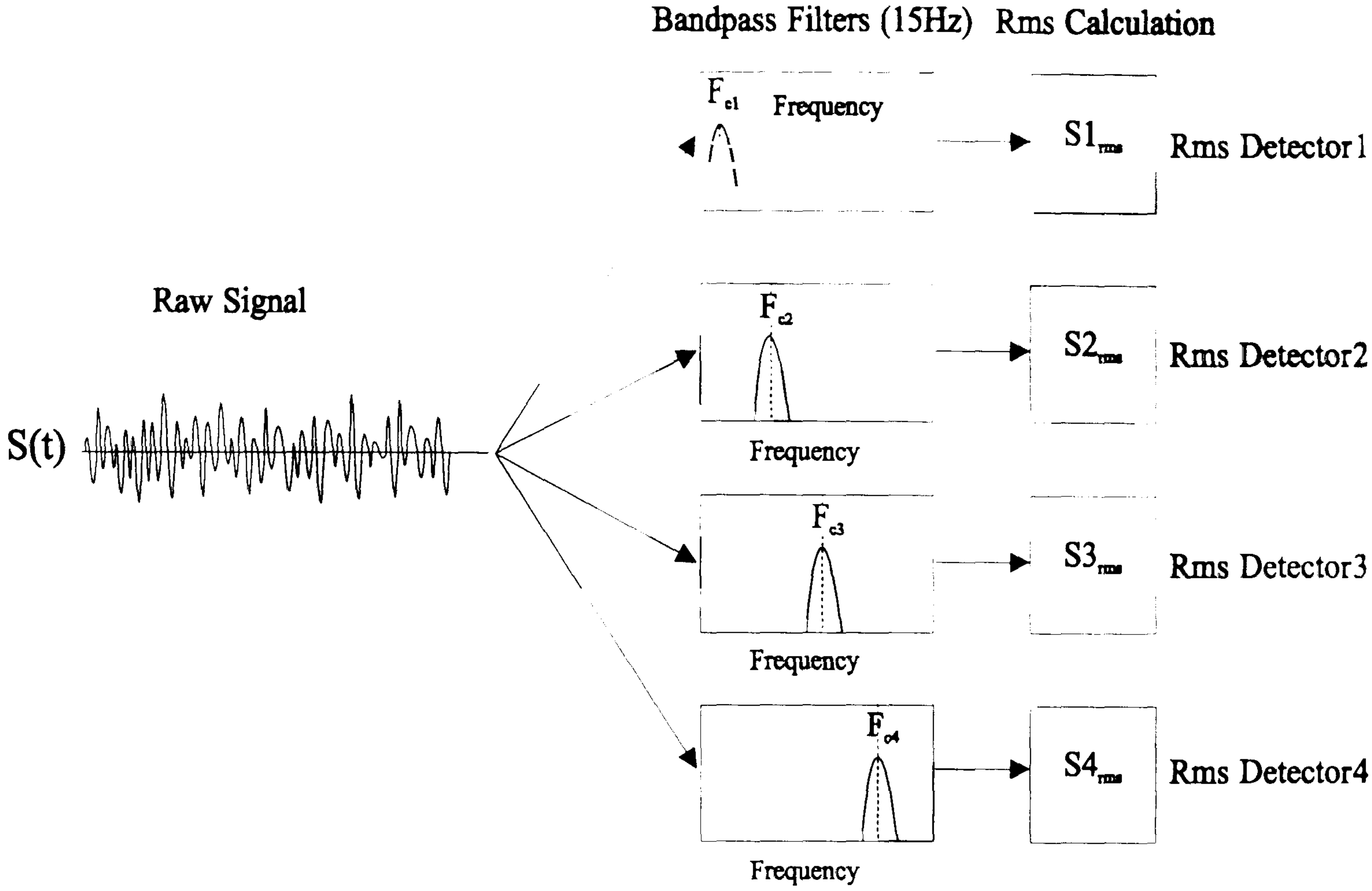

Figure 2-6 Flow Diagram of the Curtis Wright Sonic Analyser

should be done so as to ensure that all failures result in signal perturbations which lie within of one of the selected discreet bands. The second and most difficult aspect of the technique also relates to the selection of specific "failure bands". It is the requirement either for accurate speed stability or dynamic filter tuning in the system under observation since all the measures are highly speed dependant. When this work was first carried out the scarcity of digital signal hardware would have meant performing the dynamic tuning using discreet components. With the advent of cheap DSP devices the dynamic filtering front-end could be more easily implemented in software in real-time so reducing the cost and complexity of this measurement technique.

The peak-to-rms ratio comparison of condition signals, termed figure of merit0 (FOM0) by Gadd and Mitchell [11] in their work with helicopter gearboxes, is a simple development of the peak measure. The premise being that local defects will give rise to a proportionally greater rise in the peak value than they will over the rms . The FOM0 should therefore vary according to the state of the machine under observation. Taking the example of their work with helicopter gearboxes if a tooth breaks on the gear drive system then a significant local increase in the peak level will be recorded due to increased localised impact loading whilst the mean signal level over a longer period will be less significantly affected by such local effects. A consequent rise in FOM0 will be detected highlighting the failure. Wojciechowski [12] reports using this measure to compare the effectiveness and diagnostic sensitivity of the three commonly used sensor types, acceleration, velocity and displacement. Figure 2.7 taken from this work shows the variation in recorded measures taken from the bearing housing of a electric motor for each of the sensor types. C_a refers to acceleration, C_v to velocity and C_d to displacement. Both C_v and C_d show little variation during system degradation whilst

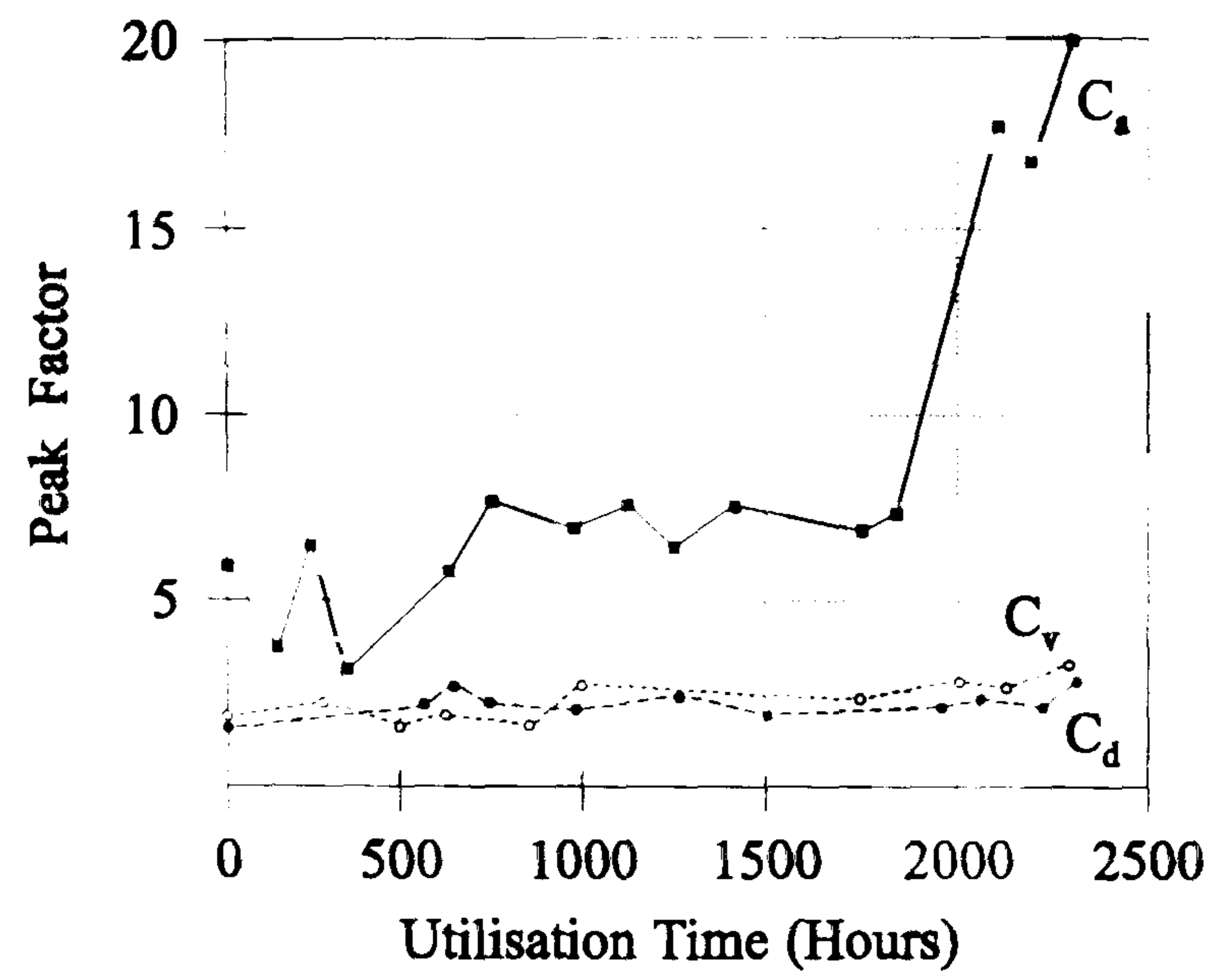

Figure 2-7 Trends in the peak factors of acceleration C_a, velocity C_v, and vibration displacement C_d in the bearing housing of an electric low power motor. [12]

C_a rises from a "normal" value of 5-10 to a worn value of more than 20 giving a clear indication of the change in system condition .

However difficulties can be encountered when implementing this type of measure in bearings with multiple or spreading defects where the rms level may dominate the equation. Gadd and Mitchell also developed further measures; FOM1 to detect the modulation of meshing frequencies which can identify misalignment faults and FOM2 to calculate some zero crossing statistics which are affected by general wear.

2.1.1.5 Improved Separation Using Synchronised Noise Reduction Techniques

In almost all the practical applications of condition monitoring we have described, or would expect to encounter, the effects of external disturbances can impair the reliability of any separation measures. In severe cases the noise effects can reduce sensitivity to such an extent that wear progresses to catastrophic failure prior to detection. Astridge [13] reports that to circumvent these problems in a health monitoring system for helicopter mechanical systems the raw vibration data was enhanced by averaging the signal over several recorded "frames". Each recorded frame refers to the time history of a single machine cycle. Successive frames are summed to produce an enhanced time history of a single cycle. The summation process, illustrated diagramatically in Figure 2.8, reduces the non-synchronous random noise elements within the frame which are responsible for reduced sensitivity. Once these disturbances have been reduced the remaining signal residue will exhibit more clearly defined localised variations where any wear or damage has occurred to the system under observation. The application of measures to this residue to subjectively test the condition status shows improved separation in most cases.

In trials using an accelerometer attached to a helicopter main rotor gearbox casing Mcfadden [10] employed just such a system of synchronised signal averaging to reduce the effects of noise on the digitally sampled signal This signal was then digitally

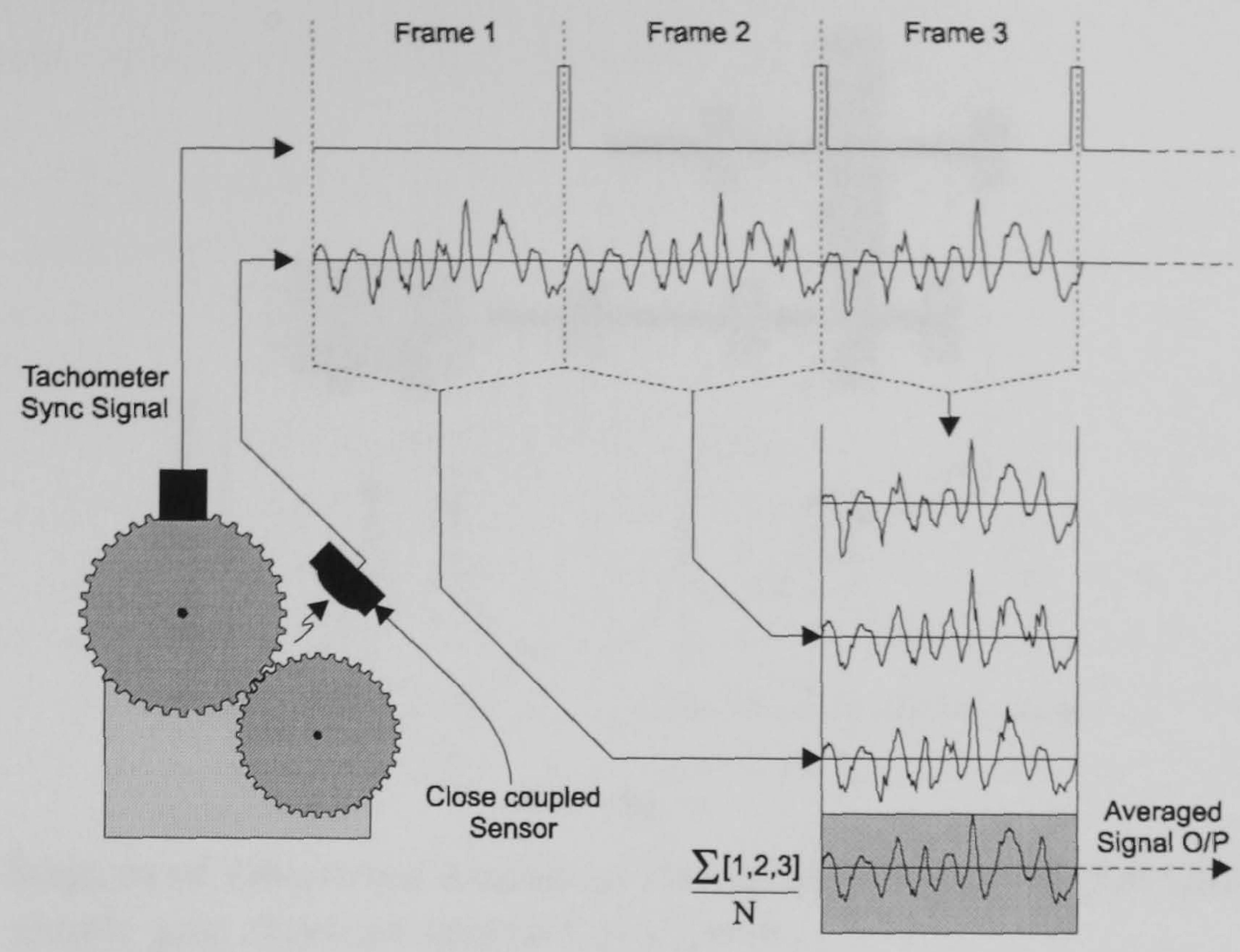

Figure 2-8 The application of synchronised signal averaging as a means of enhancing a condition signal and thus improving separation

bandpass filtered about the second harmonic and a Hilbert transform applied to provide an amplitude and phase modulated analytic signal. Kurtosis analysis, based upon comparisons with a "normal" baseline, was then performed on each of the traces separately, the phase trace providing an earlier warning of damage than did the amplitude trace. McFadden notes also that the more accurately machined the gear arrangement is the earlier a deviation from a "healthy" state can be made. These results provide proof that improvements upon the use of kurtosis on the raw, unpreprocessed signal can be made with the simple addition of synchronised averaging. A further improvement of note is the advantage accrued by the use of the phase information. In the trials it displayed the potential to distinguish between crack growth which introduced a phase-lag and foreign material build-up which introduced a phase-lead.

2.1.2 Frequency Domain Signal Processing

The second condition analysis method available is the extraction of relevant information by spectral decomposition [14, 15, 16, 17]. In contrast to the temporal selection of data discussed previously, spectral decomposition can be applied to identify specific failure modes with less prior knowledge of a system or its fault parameters. Specific physical parameters of the system under observation can be identified on a spectral plot with reference to the running conditions and system structure. Figure 2.9 shows this graphically, identifying the shaft, bearing and gear mesh frequencies of a simplified system.

The ability to differentiate in this manner between specific physical parameters may be particularly relevant in complex situations where the sequence of vibroacoustical events is not clearly identified but the form of the events is known. Biswas *et al* [18] used these principles to study the failure of a high speed industrial turbine gear coupling in a

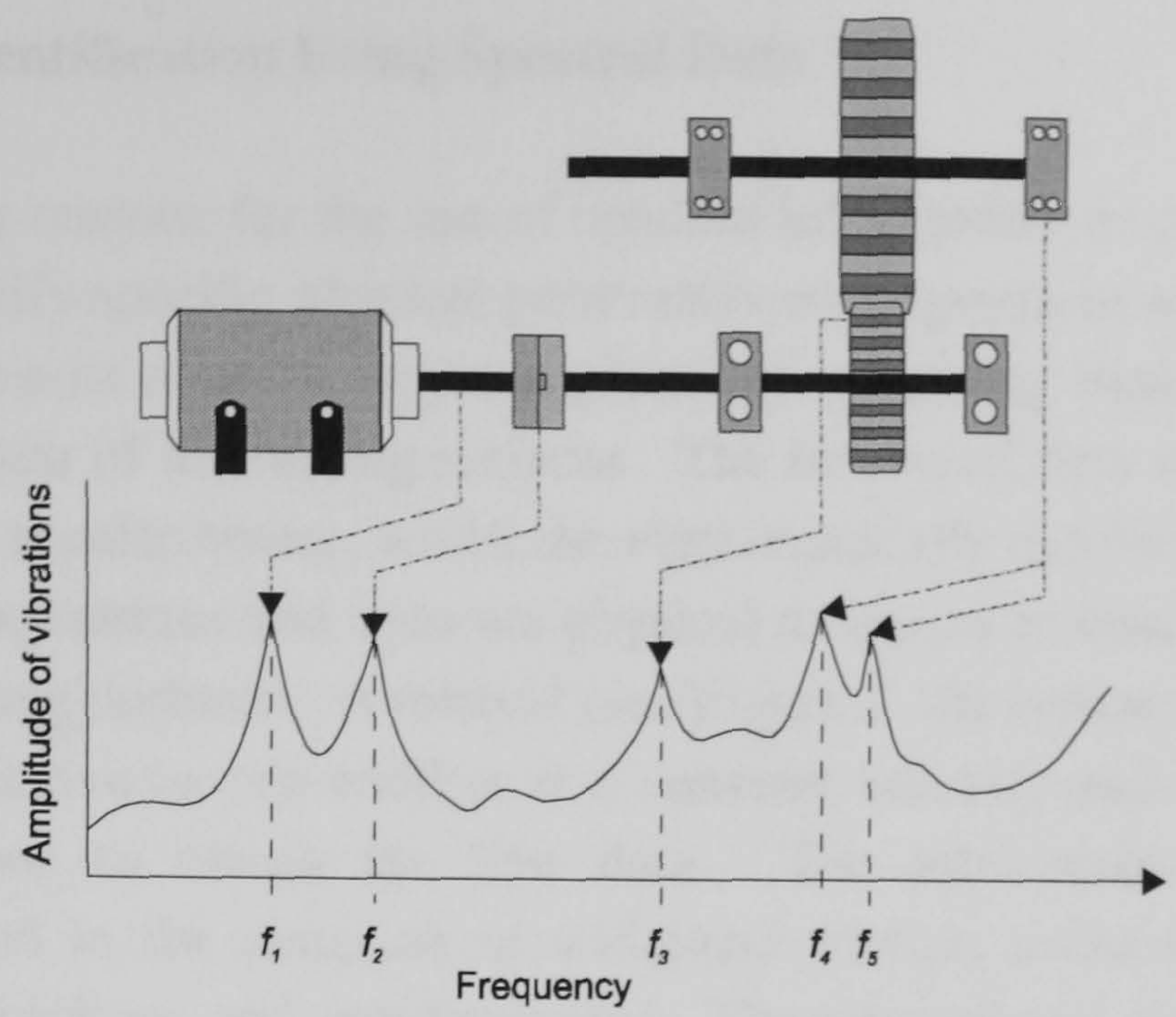

Figure 2-9 Sources of vibrational excitation identified in a spectral plot taken from a simple gear drive set attached to a motor

sugar plant which was resulting in heavy losses in production. As would be expected the spectral analysis provided data which could be traced back to certain characteristics of the machine. Components such as rotation frequency and thrust pad passing frequency could have been predicted, but they also noted the presence of some non-synchronous components which indicated that rubbing was taking place in the coupling. This evidence resulted in the overhaul of the coupling and replacement of damaged parts restoring the machine to full health with a resultant increase in plant productivity.

Simplifying this type of spectral component identification still further requires an understanding of the behaviour of sub components. The wear and damage introduced into these components has two primary effects upon the spectrum of emissions from a machine. The lower frequency components, caused by bearing or shaft inconsistencies for example, can modulate the higher frequency, carrier-like, meshing components of a system. Envelope detection or amplitude demodulation may be used to study these lower frequency perturbations. The second effect is caused as a result of inconsistencies in componentry, particularly gear teeth which are subject to substantial cyclic load variations. Physical irregularities in stiffness for example induce flexing effects in the teeth which in turn lead to phase variations in the emissions. These can be detected and analysed using frequency demodulation techniques.

The increasing availability of cheap, powerful processors together with the development of the Fast Fourier Transform (FFT) has meant that frequency domain analysis can now be performed on-line in real time with dedicated digital signal processors where previously it would have been processed off-line. Increasing availability makes possible the development of low cost spectral analysis identification systems. Sections 2.1.2.1 and 2.1.2.2 investigate some of the wide variety of work completed on the application of spectral measures to condition monitoring under steady state, runup and rundown conditions.

2.1.2.1 System Identification Using Spectral Data

One of the defining reasons for the use of spectral information in condition monitoring is its ability to identify specific physical parameters on a system in motion. Mohamed *et al* [19] report the results of work they completed in comparing measured and simulated spectral data for a pair of interacting surfaces. The measured data was recorded using a microphone and an accelerometer, whilst the mathematically derived data was simulated using simple approximations and accurate physical measures to describe the interactions between two opposing surfaces. A testbed (see Figure 2.10) comprising of two abrasive surfaces moving relative to one another at a constant velocity and under constant load conditions was used to obtain the live data. The intra-surface interactions (see Figure 2.11) resulted in the emission of wideband friction noise which was recorded using both the microphone and accelerometer. The microphone was mounted in close proximity to the two rotating surfaces, whilst an accelerometer was mounted on the testbed support frame to record the localised acoustic and vibration emissions. Spectral analysis consisted of subdividing the signals from each of the two sources into 0.04 second frame segments and performing a fast fourier transform on each. The frame spectra, each representing a 0.04 second signal segment, were then summed over several such segments, reducing the noise effects, to produce an averaged spectral signature frame for each sensor signal.

The spectra taken from each sensor highlight different interactions between the surfaces. Whilst both the microphone and accelerometer derived spectra have peaks at the support beam fundamental frequency, the vibration spectrum also contains the higher harmonics of this. In contrast the noise spectrum contains a broad band emission which is centred about the surface asperity impact centre frequency, f_C given in (4).

$$f_c = \frac{V}{\mu_{sp}} \tag{4}$$

They report that the centre frequency f_C and noise spectra correlate well with the mathematically simulated emissions based on measurements of the mean surface

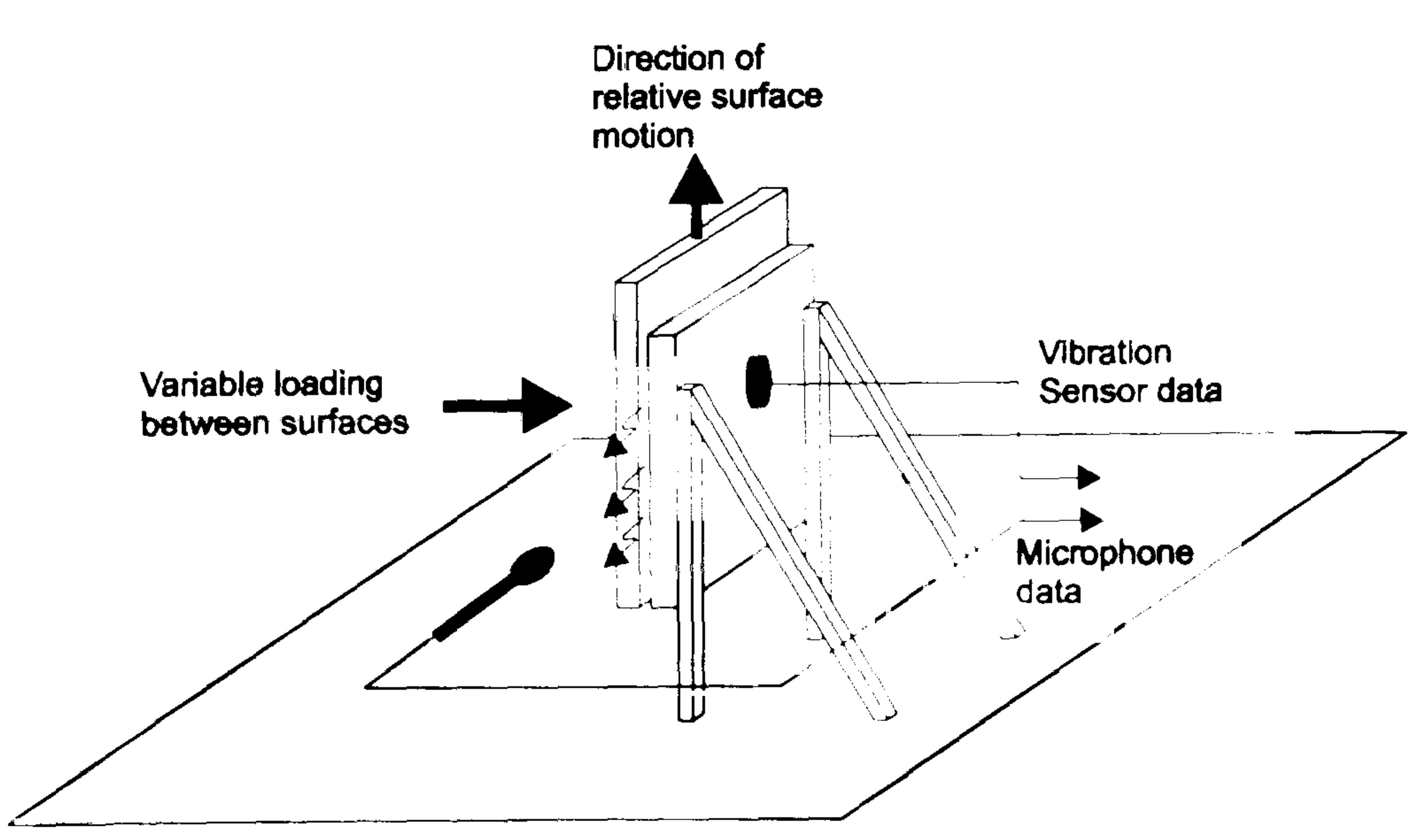

Figure 2-10 Schematic of the testbed system used by Mohamed et al in their work on friction noise. [19]

asperity spacing, μ_{sp}, and relative velocity, V, of the opposing surfaces whilst the loading affects the signal energy content. The ability to relate machine loading, velocity and surface interface condition to the group emissions is an integral part of the condition monitoring process and as such this work provides proof that both physical interactions and system conditions can be estimated from vibrational and acoustic emissions.

The necessity for greater efficiency in system componentry has led towards a reduction in operating tolerances which bring with them an increased risk of wear through rubbing. Increased risk brings with it a need to more accurately estimate the condition of sub components as a result of continuous usage. When working with the reduced operating tolerances required to improve efficiency even small inaccuracies introduced either through wear or inconsistent manufacture in areas such as thermal mismatch, unbalance, misalignment can cause potentially catastrophic wear patterns. Beattty [20] completed some work on rotor response due to radial rubbing and as with Mohamed used spectral analysis to determine the state, in this case, of high speed, high performance turbomachinery rotors. As with Mohamed the data taken from system trials was then compared with simulated results.

Beatty employed a real time data analyser to acquire spectral plots both in steady-state and transient operation which were subsequently used to analyse the harmonic components of the emissions. The harmonic amplitude content was used to estimate the arc length of radial rubbing occurring between the rotor and housing of a high pressure turbopump. When compared with results gained by mathematical modelling they showed good correlation. This work backs up the findings of Mohamed in proving that accurate system state information can be determined through the use of simple spectral techniques. Beatty went one stage further and was also able to identify certain boundary measures relating to rubbing force and arc length after which severe or catastrophic damage would be caused to the parts under observation. These measures would provide an adequate basis with which to develop automated monitoring techniques although this was not specifically discussed in this work.

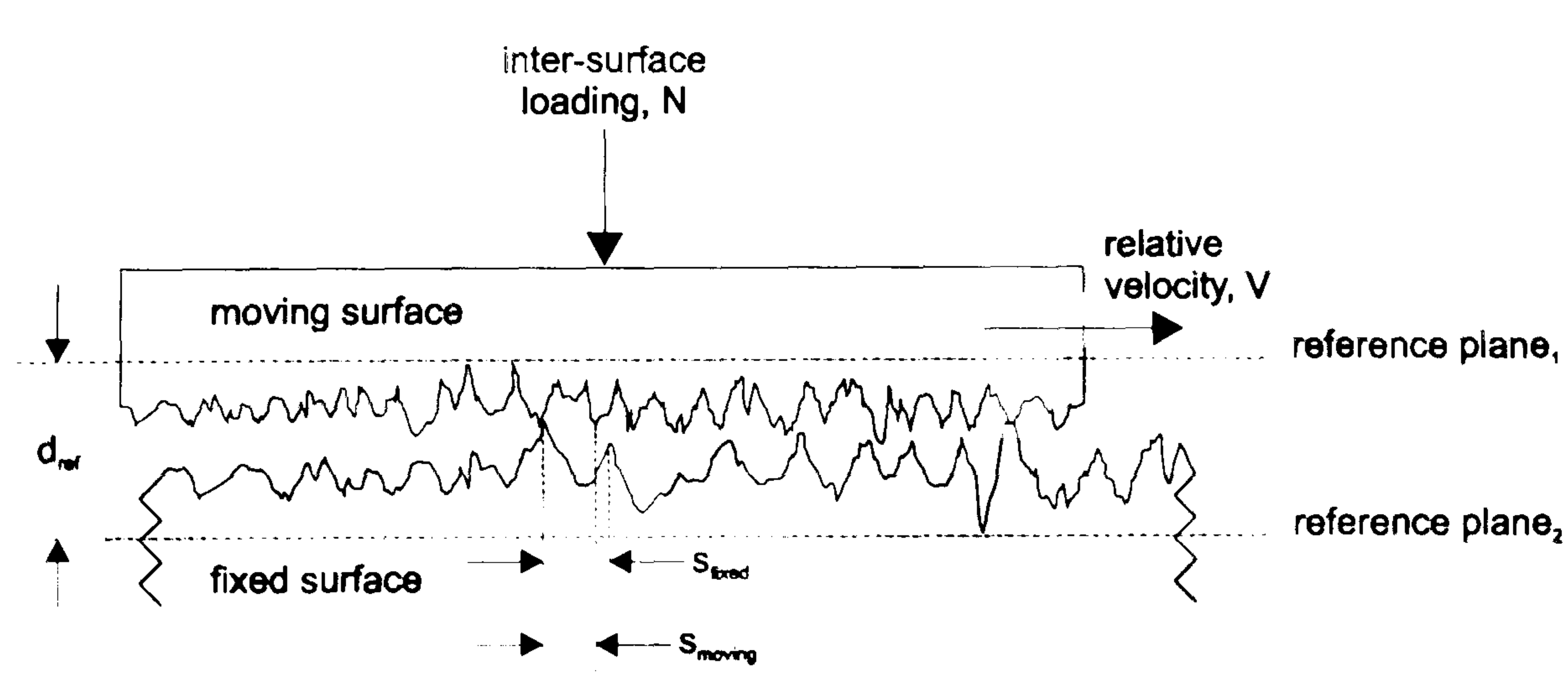

Figure 2-11 Simplified view of the intra-surface interactions on the testbed system used by Mohamed et al. [19]

2.1.2.2 The Application of Spectral Data to Condition Monitoring

The work done by both Mohamed and Beatty provided a good introduction to the principle of system diagnosis using spectral data. Zhaoqian *et al* [21] used these same principles when developing a diagnosis system for the toolpiece of a CNC lathe. The raw signal data source in this application was an accelerometer installed on the lathe in the feed direction. Again the same principles of signal averaging to those we detailed earlier in the discussion of time domain applications was used to reduce the effects of noise interference. During initial trials they were able to identify tool piece characteristics, such as natural frequencies and higher harmonics as peaks on the recorded spectra. They report that wear during usage causes the relatively sharp characteristic spectral peaks produced by a new toolpiece to spread, causing a broader spectral peaks.

They went on to investigate the concept of spectral component division. The spectral components of a worn toolpiece are divided by the fingerprint spectral components characterised by a new toolpiece. The results from this study highlighted the importance, in this application, of the spectral ranges 1-3 kHz and 16-20 kHz which proved most sensitive to changes in wear. The higher band itself was more appealing in terms of an in-situ diagnosis system since a considerable amount of the external noise interference could be eliminated by using a high pass filter. Any other local high frequency components not attenuated by the high pass filter and not relevant to the system condition are readily attenuated by the structure of the machine itself. This premise relies on the accelerometer being placed sufficiently close to the toolpiece that the path attenuation effects do not impact on those signal components containing the condition information. The concept of sub-division of spectral bands is not a novel one and is in fact used to a greater or lesser extent in a substantial number of applications to rationalise the quantity of data required for diagnosis. There are two reasons why this technique may be useful.

(i) Variations in group emissions caused by wear or damage are focused on specific spectral bands.

(ii) Certain spectral bands provide improved conditions for signal separation, perhaps through the elimination of external noise effects.

In the case of [21] the elimination of the majority of external disturbances made the 16-20 kHz band preferable. Any such compression in the spectral data requirements for adequate fault separation also has a useful secondary effect, that of a reduction in diagnosis system complexity and response time. For example, if the group emissions from a complex gear cluster arrangement were to be analysed for the purposes of monitoring a particular gear within the cluster there is a risk that any diagnosis system will be overloaded with sensor data. However if the assumptions made above are applied then damage will affect certain elements of the spectrum of the group emissions more than others so subdivision and extraction of specific spectral elements should provide a better analysis solution.

However by sub-dividing the spectrum for the purpose of monitoring we can incur the additional problem of non-static measurement instability described by Randall [22]. These effects are characterised by a spectral emphasis shift as a result of the machine under examination being subjected to variations in either load or speed which have an impact upon the spectral signature. Hence the choice of the spectral components to use in determining system state can be further complicated but still remains the key to a successful spectral monitoring system strategy. Most researchers agree though that the spectral components which are of most interest in a system composed of a rotating shaft(s) are generally the first few harmonics of the shaft rotational speed namely half, first, second and third.

Imam *et al* [23] extracted precisely these components from a rotating shaft using encoders attached to the machine to synchronise the frame sampling for their sub-band and spectral signature comparison measurements. They employed time domain signal averaging methods to reduce the effects of background noise prior to spectral decomposition. Prior to monitoring on-line a series of recordings are performed to obtain an initial baseline data set for later comparison. Each sample frame, relating to a single shaft cycle, recorded whilst the shaft was in a steady-state mode, is then synchronously summed over a number of cycles to provide a single histogram signature. For the on-line monitoring a similar process of synchronous sampling and summing is used to provide an in-use signature. The baseline signature is subsequently subtracted from the in-use signature to produce an "error signature".

The analysis is then completed by performing a fourier transform on this error signature to provide a set of differential histogram harmonics. The variations in the extracted harmonic components can provide sufficient detail to identify faults such as crack growth and shaft imbalance. In field trials Imam *et al* noted that early detection of cracks of the order of 1-2% of the 7" shaft diameter could be made by studying the variation in the first and second harmonic components. The initial crack development causes asymmetric shaft imbalance which gives rise to a rapid rise in the second harmonic amplitude component. As the depth of the crack increases the shaft flexibility begins to dominate so reducing the second harmonic component and giving rise to proportionally greater changes in the first harmonic. They also noted during the course of their trials that these variations are position sensitive. For cracks near couplings or bearings the dominant initial changes are noted in the second harmonic whereas the closer the crack is to the mid-span of a shaft the more dominant the effects of flexibility and hence the first harmonic. The monitoring tools which they developed using these techniques and which were later used in trials had three main modes of operation.

(i) Manual mode, where the measures discussed previously are extracted for manual recognition by a trained expert.

(ii) Automatic mode, where an internal rule set is used to identify likely fault patterns.

(iii) Analytical mode, which allows a trained user to estimate parameters regarding the current system state, simulate the expected effects of these system conditions and then compare them with the actual system signatures.

These monitoring tools have been successfully implemented in the field on turbine systems in the power generation industry and provide a significant improvement over those which were otherwise available.

Another method whereby spectral data can be used to monitor condition is unitary spectral coefficient tracking as Herbert [24] reports. As with Imam *et al.* Herbert used a tachometer attached to the shaft under observation to co-ordinate the sampling of sensor data. In this case though, the timing is not used to reduce noise effects but to continually monitor the shaft speed during a rundown or runup cycle. At selected rotational velocities corresponding to shaft resonance frequencies a data acquisition process is initiated. This consists of sampling the signal data and extracting the amplitude and phase components. Initially measurements are then used to develop a system state historical archive baseline. This archive is then used to determine, by comparison with the baseline, the current state of the system either in an on-line, or off-line mode.

The techniques discussed up to now have shown the potential of spectral analysis in determining the health or otherwise of unitary cycle rotating machinery. The issues relating to the use of spectral data in the case of reciprocating machinery are somewhat more involved. Reciprocating machinery itself provides a greater challenge to those who wish to study the spectra for monitoring purposes and in this respect the work of Chaudhuri and Serridge [25,26] should be considered. The added complexity to which Stronach, *et al* [27] also refers in applying spectral techniques to reciprocating machines is focused around the discontinuous series of events taking place during a single machine cycle. Each cycle consists of at least one, and usually many more, instantaneous detonation events produced by a machine which is also inherently mechanically unstable. As with the continuous case each machine is made up of many sub-components, all of which have specific harmonic properties. However in this case the excitations are further complicated by the series of instantaneous events each of which excites characteristic machine resonance's producing a group emission which exhibits a wide dynamic range. The spectral information content is consequently of a more complex and non-steady nature. Again as with many of the previous discussions the principles of time synchronous averaging of input data to reduce the effects of noise provides a good basis from which to start the analysis. Both Serridge and Stronach then apply full and partial cycle spectral analysis to the pre-processed data. The terms full and partial represent the length of the time segment, full being one rotation cycle, over which the FFT analysis is performed. Full cycle analysis can, as a result of the non-steady nature of each cycle, quickly become highly complex and sometimes misleading. On the other hand partial cycle analysis sub divides the rotation into its constituent events, making the study of dynamic wear and damage effects simpler. This requires the synchronous subdivision of the sampling cycle into shorter, typically 10-20ms.

frames which can be referred to particular mechanical "events" such as inlet valve opening or exhaust valve closing.

Chaudhuri applied various sensor types depending upon the type of faults being identified. He found acceleration mode sensors to provide greater resolution for faults involving higher frequency effects, whilst displacement and velocity sensors are more suited to lower frequency faults such as structural support problems. Despite the complexity Chaudhuri was able to determine several failure modes. One particular example was cylinder liner looseness, which was identified by studying the spectrum measured using both acceleration and velocity vibration sensors attached to the cylinder head. In this case the fault manifests itself in the higher frequency end of the spectrum making the acceleration mode sensor more sensitive than the velocity measurement.

In contrast to Chaudhuri's manual comparisons of spectral content to determine system state, Stronach refers to the obvious advantages which could be afforded by the implementation of a form of pattern matching to the spectral data to provide automated diagnosis. His preferred method is to further simplify the process by producing a difference spectra. The difference spectra would be produced by subtracting the baseline spectral components from the on-line spectral components taken during machine usage. With this system a series of fault patterns, each defined by its difference spectra, could be compared to the current pattern and a decision on its distance measure be obtained to classify the condition. Such a system would of course be of considerable attraction in terms of both simplicity and a reduced requirement for skilled personnel to operate the machinery. However the major disadvantage afforded by such a system is the necessity to provide a complete set of failure patterns for each machine prior to usage.

2.2 Simulation Analysis of Machine Condition using Mathematical Modelling

This is arguably the most deterministic technique but relies on the development of sometimes complex mathematical descriptors of the machine to be monitored. Obviously the more complex the machine, the more convoluted the mathematical representation. There are two variations upon this theme, the first is transfer function estimation and the second is component part simulation. The first uses the emissions from the device to derive an estimation of the function describing the variations from source to sink whilst the second relies on knowledge of the device sub-componentry and an understanding of the emission characteristics of each of these within the device in its running state.

2.2.1 Transfer Function Modelling

The dynamic response of a complete system can be modelled in terms of a transfer function in which the behaviour of sub-components are characterised in mathematical terms. These include the characteristic frequencies of sub-components, steady state gains, time constants and damping effects caused by both material content and system design. Once a model has been produced these sub-components can be individually identified from each other by the application of the transfer function to the output signal from the machine being monitored. This function estimation may be done in the time or frequency domain and relies on the accurate sensing of both the signal source and its sink to the outside world.

In [28] this technique is applied to an electro-hydraulic servo system. The key to the success in this instance is to minimise complexity whilst at the same time ensuring that all relevant dynamic parameters are included in the function. Firoozian uses (5) with which to model a simple system comprising two critical components with different time constants (T_1 and T_2). Monitoring is then a simple matter of calculating the baseline values on the system when new and periodically comparing the in-use values with the baseline whilst in service. This comparison would highlight changes in dynamic behaviour of the system. Limits could be associated with changes in these parameters in order to determine the state the component parts of the machine. The major advantage this has over many techniques is that each element within the model can be individually monitored for variations separately from one another. It has one major disadvantage in that the more complex the system the more convoluted and complex the ensuing model is to estimate. It does however in this form provide a somewhat modular approach to the system which could be beneficial if a sub-component was to be replaced within the system. This technique does not necessarily require the generation of a wholly complete model of the system under observation but must provide a representative model from which the relevant machine parameters may be identified.

Bartelmus's development of coherence function measures used to generate condition measures from multiple sensor systems attached to gearbox drive systems is based upon the modelling of physical interactions of meshing teeth. In [29] he describes mathematically the development of a condition measure based upon the acquisition of a pair of inputs from a gear system. The measure is based upon the separation of the correlated and uncorrelated parts of the two signals. He found the measure to provide good feedback for assessing the state of the gear mesh providing that the measure is

$$\frac{\theta_{out}}{\theta_{in}} = \frac{A}{\left(T_1 s + 1\right)\left(T_2 s + 1\right)\left\{\frac{1}{\omega_n^2} s^2 + \frac{2\xi}{\omega_n^2} s + 1\right\}} \tag{5}$$

Where,

- A is the steady state gain.
- T is the time constant.
- ω_n is the natural frequency of one component.
- ξ is the damping ratio of the same component.
- θ_{in} is the demand signal, centrifugal force, and θ_{out} is the output variable, measured vibration.

made under normal working conditions.

2.2.2 Component Modelling

This technique is based upon the premise that the group emissions from a mechanical device are made up from the emissions from each of a number of constituent mechanical sub-components. In other words each of the separate sub-component excitations provides some part of the total emission. The proportion that each part provides is not necessarily solely related to its size. If each of the constituent sub-components can be modelled mathematically then it follows that a model may be derived that provides a complete system description built up from each of these individually modelled constituent parts. This is what is meant by the term component modelling.

In [30] McFadden and Smith model the vibratory emissions from a rolling bearing and attempt to relate modelled characteristic emissions to the measured characteristics in the case of a single defect on the inner race of a roller bearing. They model the effect by assuming the excitation caused by the rolling ball impacting upon a point defect is an impulse function, δ_t. The magnitude of the impulse is proportional to the severity of the defect and the load exerted upon it. Several other factors must be taken into account when modelling this simple system. The first is the number and velocity of the rolling elements in the bearing and the constantly varying angular load exerted upon the defect by the rolling elements as the bearing rotates (assumed to follow the Stribeck equation). The second factor which must be included in this model is the effect introduced by cyclic variations in the transmission path between the defect impact emissions and the transducer used to sense them. The last point noted by the authors is that "real" damage cannot be considered a point source but can be modelled by assuming it is a series of adjacent point defects

Each of these effects is applied in turn to the initial impulse function to produce a comprehensive model. In order to determine the accuracy of this model they compare the modelled emissions with experimental measurements taken from a real system. The agreement between the two results is reasonable. However considering such a simple model we should be able to estimate the resultant emissions with relative accuracy. In many of the types of machinery in which we would wish to monitor condition we may have several tens of bearings together with many hundreds of other separate components as well as casings, mountings and the effects from surrounding plant machinery may be involved. In this situation the building of a reasonable model is no longer a trivial exercise. The complexity involved would most likely ensure that component modelling would be uneconomic to apply. If this is the preferred option a conscious effort should be made to apply it to the correct type of application whilst at the same time ensuring that only relevant constituent parts of the model be estimated for the purposes of identification.

2.3 Chapter Summary

This Chapter has provided an overview of some of the techniques which have been applied to the extraction of condition data for use as a means of estimating system condition. The requirement for such systems is becoming more widespread in today's safety and cost conscious environment. A good system provides savings in the cost of direct ownership through reduced maintenance and labour as well as enabling operators to run the plant closer to the system operating tolerances resulting in increased efficiency. Whether we apply the techniques on-line or off-line, continuously or periodically depends upon key factors such as likely cost benefits, the expected mean time before failure (MTBF), and safety criticality.

All the techniques rely primarily on the premise that any component, or group thereof, being monitored produces a characteristic signature which is modified as a result of any change to the system. Which techniques are most applicable to particular situations depends in most part on the nature and type of variations which we expect to encounter and the level of external interference which may be introduced to the raw signals by the local environment. Some of the simple time domain processing techniques such as peak and rms are easily implemented but are affected by local noise and can be somewhat machine specific. The use of combined measures can improve the condition separation as can the use of dimensionless measures such as kurtosis. This technique in particular is reported to be relatively insensitive to changes in both load and speed as well as the transmission path effects due to sensor positioning. Other techniques combining the use of spectral selection with time domain measures, as in the case of the Curtis-Wright analyser, rely heavily on the use of accurate speed control or dynamic filter tuning for their effectiveness. In most cases the measures can be enhanced by the application of synchronised filtering which reduces the random noise elements within the raw condition signals. The synchronisation information can also be used to specify the initiation or position of a specific fault by reverse referencing.

Spectral processing in contrast can be used to provide system information based on a physical understanding of the processes under observation. Primary rotation frequencies and harmonics of each of the components can be estimated and compared to the extracted data to provide status information. In complex systems the condition of individual parts may be difficult to extract from the acquired spectra. As in the case of time domain techniques though some selection may be applied to reduce the complexity of the data. The model based techniques are complicated still further by the requirement to simulate each of the sub-components accurately enough to make relevant comparisons with recorded data. For this reason their application has yet to become widespread. With the more recent introduction of modern digital signal processors cheaper, simpler and more compact devices are now feasible, bringing with them an opportunity to develop monitoring systems for an increasingly wider range of machinery.

Chapter 3

Page

3. Time Domain Signal Coding

The concept of coding a signal in terms of its shape and duration between amplitude zero crossings has its roots in some early work by Licklider [31] on the intelligibility of infinitely clipped human speech. Studies of clipped speech showed that removal of signal shape information had little effect on the intelligibility. Whilst connected sentences were identified in 90% of test cases this level was reduced to 75% in tests on isolated words. However this relatively small effect on intelligibility is outweighed by a marked reduction in speaker identification by listeners. It was apparent from this initial work that the removal of shape information removes data vital to voice classification. In 1978 King and Gosling went on to develop this signal coding technique for use in speech encoding and identification [1]. It was for this reason that the technique became known as time encoded speech, or TES, as it more usually termed. Their research centred around maintaining more of the shape information found to be vital to the speech quality and subsequent speaker identification.

Initially the work concentrated on applications in automated tactical military communication systems where the emphasis is not focused on telephone quality speech but on intelligible speech which may be transmitted over low bit rate channels (e.g. HF radio). TES coding not only produces the necessary reduction in bandwidth requirements but also permits the subsequent application of complex digital encoding algorithms to provide secure communications which are essential in the military arena. They went on to study the technique's applicability to direct voice input (DVI) systems for use in 'hands off' control of tactical military equipment. Such systems are ideally suited to reducing loading on personnel in high stress hostile environments. This application requires a robust classification engine which must be able to respond quickly to user commands. Additional constraints are placed on the system by the necessity to perform the classification in different environmental conditions introduced by the variety of circumstances in which it would be used. For example during military engagements a subject's speech can become highly distorted due to stress, the effects of protective clothing (e.g. respirators) or the necessity for whispered speech. In all cases the classification must take account of these effects and still be able to respond accurately. This work was then extended to voice recognition of severely handicapped people suffering from disorders which have reduced their ability to interact using speech. In some cases this leaves severely impaired speech as the only form of communication for a patient. The work carried out by Warner *et al* [32] centred on the development of a system which enabled skilled staff to train a speech recogniser for a particular patient depending upon their specific needs. The recogniser is trained using a series of utterances from the patient with the help of a staff member. This routine identifies a small, sometimes less than ten, set of unique patient utterances which can be associated with a particular sentence or action. The results of this study identified the effectiveness of such a system to a group of people who would otherwise be unable to communicate. All of the work discussed up to this point regarding the development of utterance archetypes and archetype comparisons employ simple difference sum calculations to determine the likeness of a particular utterance to each of a number of predetermined archetypes. As will be discussed later in this Chapter a more recent addition to the field, namely that of neural classifiers, offers further room for future development of such systems in terms of flexibility and accuracy.

This work has shown that the coding technique itself is compatible in a number of different applications and has resulted in the development of several robust and accurate systems implementations encompassing it. All these initial developments have been based upon applications to human speech processing but there is obviously no underlying reason why such techniques could not be applied to signals in other fields. The term TES, or Time Encoded Speech, was originally chosen simply because it described the application of the technique to speech in particular. However the technique may equally well be described as Time Encoded Signal where the signal can be taken from any number of a wide range of sources and using a variety of sensor types. Automating the application of condition monitoring to industrial machinery as previously discussed in Chapter 2 is a highly desirable goal. It requires the development of a system which is able to acquire the primary measurement data then process it and extract the information pertinent to classifying the state of a monitored target as efficiently as possible. In each case the location and dynamic range of the transducer chosen to capture the raw signal will vary with respect to the type of machinery and the choice of monitoring strategy. TES provides an ideal means of both compressing and conditioning a wide range of raw signal sources into a format which can be applied to simple automated recognition techniques. Given the level of processing power now available at relatively low cost TES based condition monitoring has the potential to be performed continuously in real-time whilst a system is in operation. In this Chapter both the theoretical and practical aspects of the application of an acoustical TES based system to the automated management of condition classification in a simulated gearbox will be discussed.

3.1 The Principles of a TES Coding Scheme

The definition of the principle of TES coding can be described as *"the conversion of a digitally sampled signal into a series of shape codes representative of the original signal in terms of one or more of its physical attributes"*. The points in the sample stream which define the limits between which the attributes for each TES code are measured are the real zero crossings which are identified by sample polarity inversions. In TES the periods between these polarity inversions are termed *epochs* which means literally *"an extended period of time characterised by a memorable series of events"*. The TES conversion consists of allocating a unique symbol to each individual epoch in the stream dependant upon the chosen set of measured attributes specific to each conversion type. Figure 3.1 describes graphically this signal segmentation and conversion process based either on the amplitude or minima shape characteristics of each respective epoch. The details of these two particular techniques will be covered more fully later in the Chapter. However once these attributes have been identified the conversion is a simple matter of using a pre-configured look-up table to select each symbol depending upon its shape characteristics. The resulting symbol stream is denoted in Figure 3.1 by the numbers seen below the raw digitised source signal.

In order to encode a signal source into a representative series of TES symbols for the purposes of condition monitoring the shape attributes and their relevance to the changing machine condition of the source signal must be defined. The selection of a particular set of attributes is the first important aspect of the monitoring process. Without sufficient primary information about the source signal the identification of specific faults becomes impossible. The three attributes which are considered to be most relevant to condition monitoring are duration, harmonic composition and energy content.

The duration attribute contains fundamental frequency and noise content information and requires only a straightforward measurement of the number of discrete samples in an epoch. That duration measures can be used to estimate signal to noise content as well as explicit frequency data may seem abstract initially. However given that signal generation is the result of physical interactions certain assumptions regarding the type of epochs which can physically be generated during the normal course of events can be made. Certain epochs falling outside these predefined physical boundaries may quite legitimately be attributed to the introduction of additional unwanted noise. The noise content of a specific signal may be estimated by measuring the frequency with which low energy, short duration (1-2 samples) epochs occur. Such epochs are found to be indicative of additive noise effects having caused "spurious" crossings in the signals being analysed. Phipps [33] notes in some of his work on human speech conversion that such noise impairs the general speech quality. For his applications in low rate speech

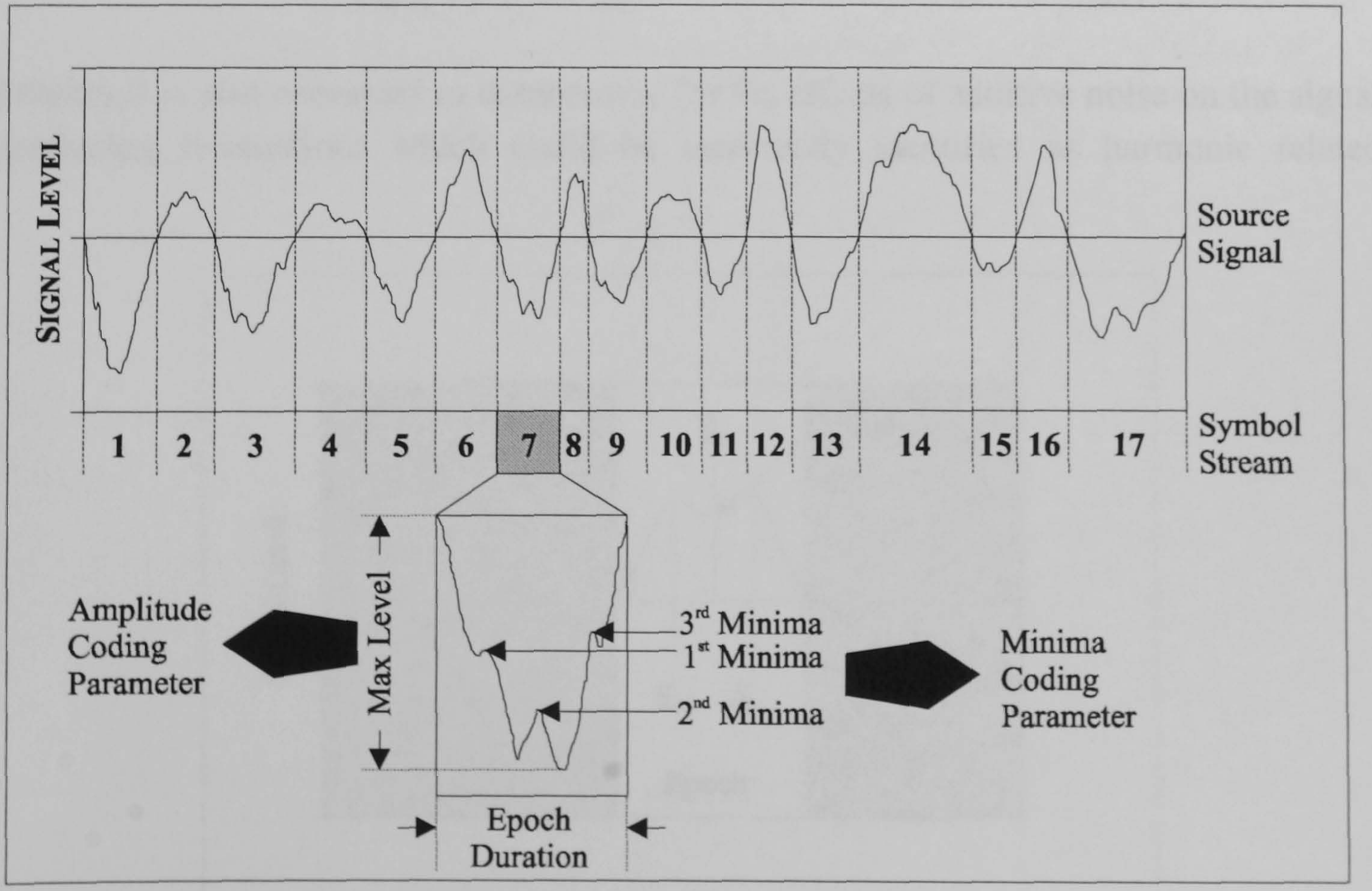

Figure 3-1 A graphical representation of the signal attributes used to convert a raw source signal using the TES technique

transmission he employed a more complex duration measurement technique aimed at reducing these external noise effects. By measuring the duration and energy of successive epochs against selected thresholds he was able to identify suspected noise induced epochs. By then interpolating between successive adjacent epoch maxima the original crossover points position can be estimated. However this type of added complexity is considered to be unnecessary for an initial monitoring investigation. It would entail additional computational overhead to what is intended to be a simple application. Since human intelligibility is unnecessary it is thought better to provide conversion and manipulation tools which take account of this additive noise component in the condition signal.

The harmonic attribute which as the name implies conveys additional information regarding the spectral content is portrayed through the measurement of epoch minima characteristics. The minima techniques used by King and others described earlier which were applied to their speech work were based upon some earlier work performed by Bond and Cahn [34] and later developed further by Voelcker [35]. It was Bond and Cahn who first introduced the concept of complex signal zeros, or minima, and their relationship with the harmonic content of the signal. In general they found that the number of minima measured for a given epoch duration was indicative of the presence of certain specific higher harmonics. Consequently these minima characteristics can be used as an indirect measure of a signals harmonic content. In the discrete domain the minima themselves are identified by comparing successive samples within an epoch for gradient reversals. If the continuous stream of samples, S_n, in Figure 3.2 is considered then the test for a minima condition consists of the simple comparison detailed in Eq.1

$$|S_{n-1}| \geq |S_n| < |S_{n+1}| \qquad (1)$$

In reality it is also necessary to compensate for the effects of additive noise on the signal introducing fluctuations which could be incorrectly identified as harmonic related

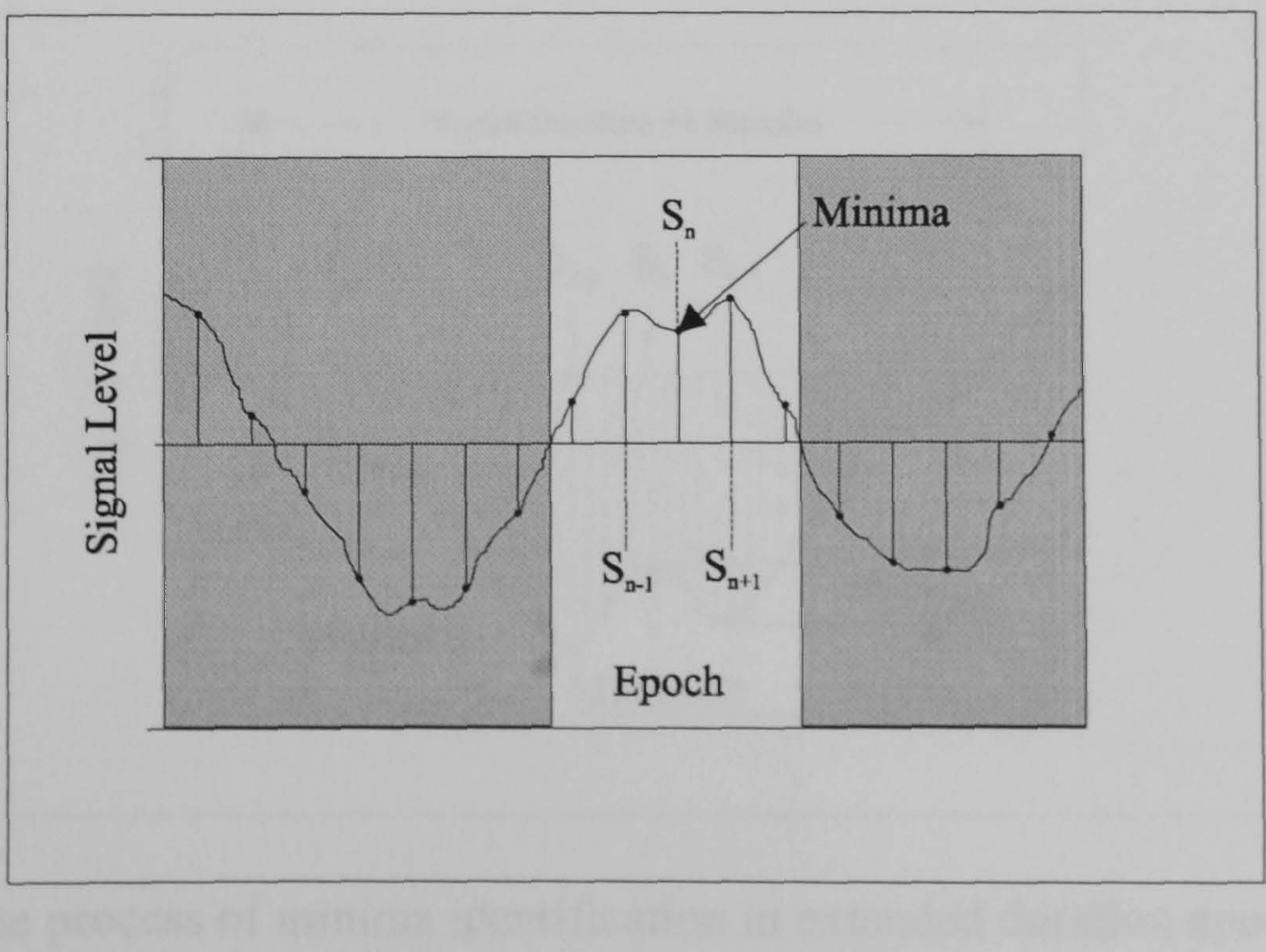

Figure 3-2 Simplified view of the process of minima identification in a discrete signal

minima. This correction is implemented by applying the depth threshold test described in Eq.2 to each potential minima identified.

$$\left(|S_{n-x}| - |S_n|\right) \geq Minima_{thresh} \leq \left(|S_{n+x}| - |S_n|\right) \qquad (2)$$

In the simplest case this measure involves comparing the adjacent maxima samples (S_{n-1}, S_{n+1}) with the central minima sample S_n to calculate the depth. In cases where a minima extends over several samples this comparison must search outwards from the minima position to the surrounding maxima samples at S_{n-x} and S_{n+x} to measure the depth. Figure 3.3 illustrates this search graphically, identifying two potential minima only one of which is classified as a minima after the depth test has been performed.

The third and final attribute to be considered is the energy attribute. As the name implies this attribute is intended to convey information about the energy content of each individual epoch and is estimated by identifying the maximum epoch amplitude (A_{max}). Taken in isolation this measure can convey information about the peak-to-mean ratio of the signal but when combined with the duration component it provides an estimate of signal energy. Measuring the amplitude attribute itself requires a simple sample by sample linear comparison search to be performed on the magnitude of individual samples from one zero crossing point to the next within each epoch. In fact this description of the estimation of the energy content attribute is a somewhat simplified view of the application required to provide useful classification data. This is because, unless carefully controlled, the measurement is prone to variations caused by sensor position and equipment characteristics. These variations are caused as a result of the potential fluctuations in measurement conditions between successive signal acquisitions. To circumvent some of these problems the magnitude measure is made with respect to a selected signal reference. Providing this reference measure is acquired satisfactorily the conversions can be made sufficiently repeatable as to be acceptable for signal

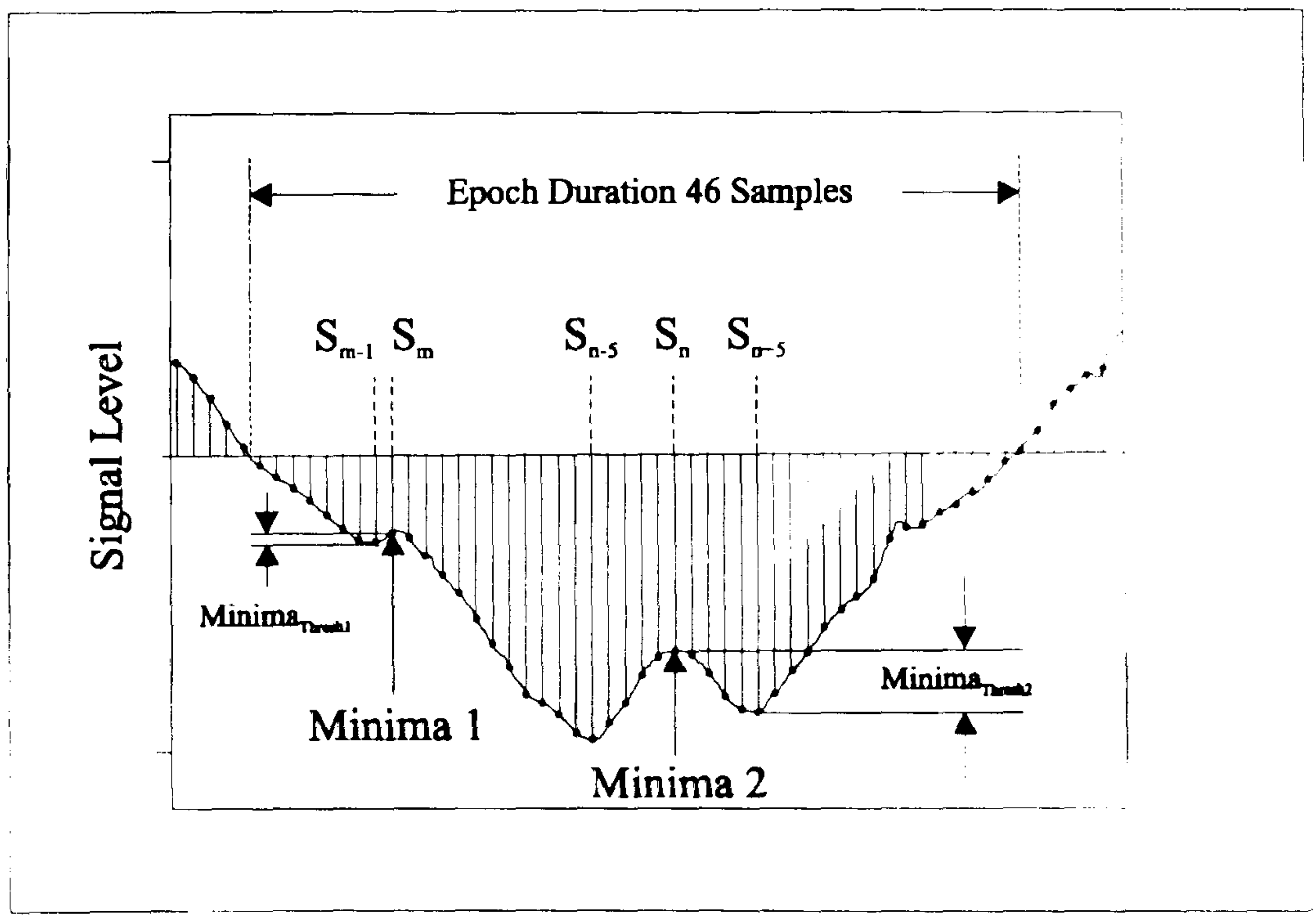

Figure 3-3 The process of minima identification in extended duration epochs under varying noise conditions

classification. The acquisition of the reference level will be considered in depth in the implementation section later in this Chapter.

As a result of selecting these three epoch attributes the description of shape can be made in several fundamentally distinct ways. Excluding for the moment the complexity of defining a coding scheme which uses all three attributes simultaneously two separate methods were developed both of which employed the duration component as a fundamental feature. The first method, which uses the extrema information in conjunction with the duration measure is known as *minima coding*. The second technique, which replaces the extrema information with the energy composition characteristics has been called *amplitude coding*. Both of these coding schemes result in a stream of uniquely defined epoch symbols each conveying information about a fundamentally different aspect of the signal source. Figure 3.4 graphically describes these two schemes both of which are intended be used independently to determine the health status of machines in use. For the purposes of this diagram the amplitude coding, using a 300 element allocation table, has been applied to positive epochs and minima coding, using a 150 element table has been applied to the negative epochs.

As has been described in this section TES coding is able to provide a simple means by which an analogue signal may be converted into a format suitable for its subsequent classification. In Chapter 2 some of the methods which have been employed previously to provide mechanical status information on machinery were discussed. Some of the ways in which these wear and damage patterns in machine parts introduce variations in their vibroacoustical emissions were also described. It should now be apparent that TES implicitly provides such frequency, harmonic and energy content information without the necessity for extended or complex processing. In addition the techniques described

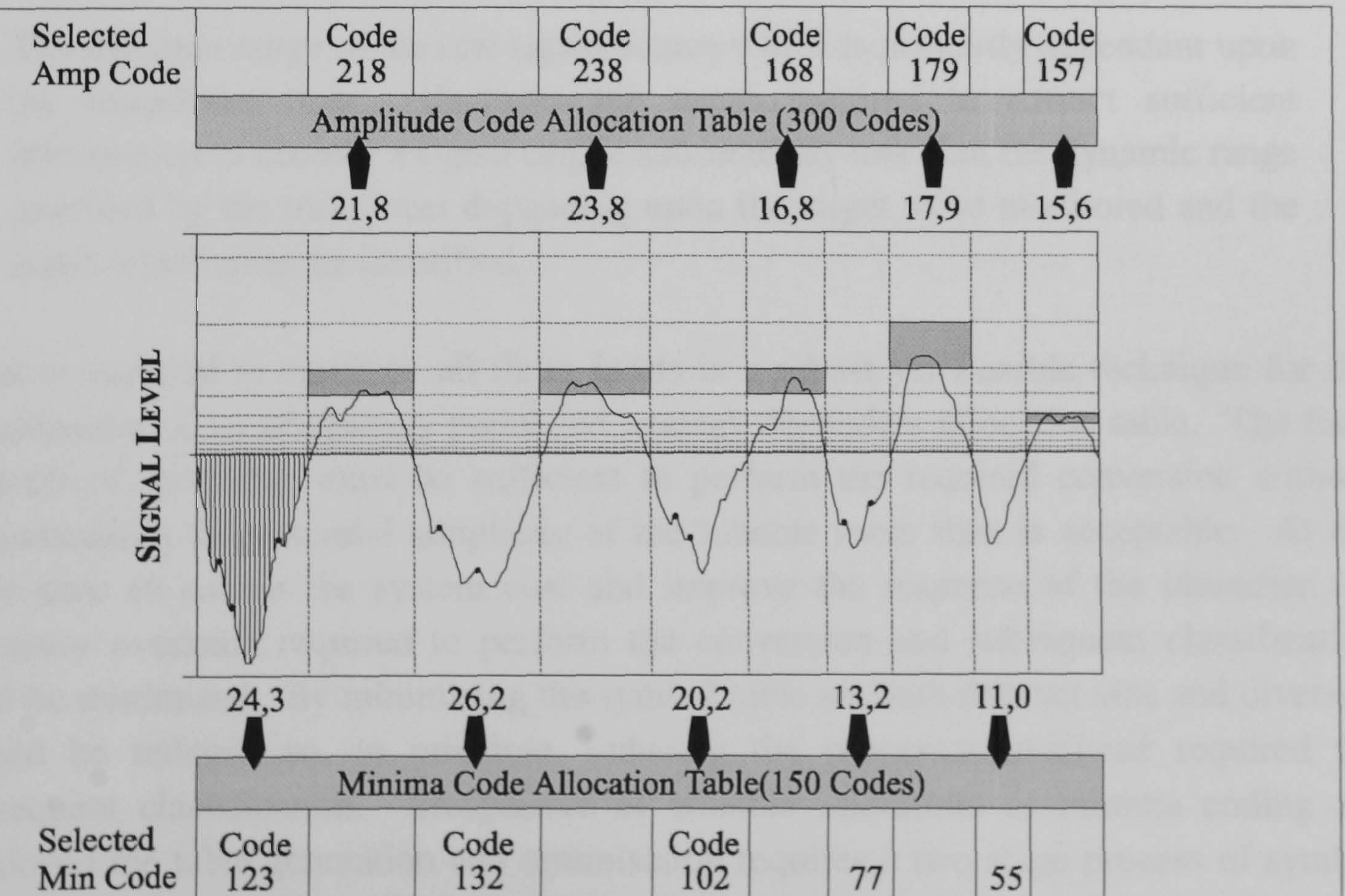

Figure 3-4 Graphical representation of both of the coding schemes under consideration

so far are ideally suited to the wide range of digital signal processing hardware which is becoming available at ever more competitive prices.

3.2 Allocation of TES Symbols to Signal Epochs to Generate a Symbol Stream

The process of coding each epoch requires the development of a suitable allocation table which defines for each particular coding mechanism the association between unique symbol descriptors and the epoch parameters used to select them. Coding then consists of measuring the relevant parameters from the discretised signal and selecting the symbol referred to by this set of parameters from the allocation table. The number of descriptor types or epoch symbols resident in each allocation table is dependant upon the signal source and additionally the states which must later be identified during classification. The basic rule of thumb is that there must be sufficient to adequately describe a particular source and this is itself dependant upon three key factors.

1) The definition of the term "adequately described source" which is application specific and difficult to define without practical trials. The basic concept itself is that the more coarsely the signal is encoded the smaller the set of symbols required to describe the signal becomes. However any reduction in the symbol set is made at the expense of increased signal distortion.

2) The coding strategy to be employed. In terms of the two methods being proposed in this work, amplitude coding requires a larger table than minima coding simply because the minima diversity of the gearbox acoustic emissions is less than the amplitude quanta diversity for the same emissions.

3) The dynamic range of the raw signal source which is primarily dependant upon the transducer type. However the range required to extract sufficient information to classify a signal can be substantially less than the dynamic range provided by the transducer depending upon the target to be monitored and the states which must be identified.

What is required to combine all these facets is a robust yet flexible technique for the development of an adequately populated strategy dependant allocation table. The final contents of this table must be sufficient to perform the required conversion without compromising the essential simplicity of the scheme more than is acceptable. At the same time to reduce the system cost and improve the response of the identifier the processor overhead required to perform the conversion and subsequent classification must be minimised. By minimising the symbol table set both data set size and diversity should be reduced so, in principle, reducing the processor overhead required for subsequent classification. Irrespective of whether amplitude or minima coding are employed the table generation and optimisation requires a two stage process of symbol rationalisation. The first stage applies a set of absolute boundary conditions for each table parameter based on knowledge of the signal source and of the information requirements of the classification mechanism. These initial boundary conditions are defined for each table parameter independently. They effectively outline the perimeters of a fully populated symbol table and should exclude only descriptors which occur spuriously or convey no information pertinent to the classification process. The second stage of rationalisation is concerned with optimising this fully populated table to produce an allocation table containing a minimal symbol set. This second phase results in the generation of a partially populated table containing only those symbols which are essential to the subsequent classification process.

In performing this table boundary definition and the subsequent optimisation it is necessary to be aware of the two most basic limitations involved and develop optimisation techniques which take account of them. The first is the specific physical limitations associated with the emissions and the second is the application limitations related to a particular device or machine implementation. The epoch duration boundary parameter, for instance, may be estimated using information known about the band-pass filters used to condition the signal prior to discrete sampling. These filters will impose specific high and low frequency limitations on the source which will result in specific epoch duration limitations. As such this restriction is described as a physical boundary. In the case of the other two coding parameters, extrema and energy, limits are more usually constrained by the requirements of the classification mechanism since it is difficult to specify physical boundaries on these attributes. The potential number of minima occurring in an epoch or the number of energy quantisation levels which are used to code epochs is application specific and difficult to quantify without prior evaluation. As such the type of symbol optimisation most applicable to these particular parameters is statistical analysis. In practice this evaluation is performed on a series of test segments acquired from the particular source involved to identify these bounds. Because of its basis on statistical analysis these limits are termed application limitations and are dependant upon the characteristics of a particular signal source as well as the requirements of the monitoring application.

3.3 Statistical Analysis of the Signal Source to Define the Symbol Table

As would be expected the potential size of an allocation table describing all possible signal epoch parameter combinations is vast. One of the primary objectives of employing TES is to reduce the volume of data required to represent a signal in an effective manner and since this goal is at odds with the principle of a fully populated allocation table optimisation is essential. In previous speech work carried out by a number of researchers this tuning has been performed by statistical means. Although this does not necessarily identify the suitability or otherwise of each of the symbols within the table in the post-coding classification process it does limit the symbol set based on the structure of the source itself. It also has the advantage of providing a simple yet flexible means of automating the process of table construction which for the purposes of an automated system requiring minimal user intervention would be of additional benefit.

A key concern when applying statistical analysis to assist in the table optimisation process is to ensure that the acoustic sample(s) used to carry out the analysis are deemed to be a sufficiently significant subset of the acoustic source to be subsequently classified. This should be accomplished by including at least one extended sample set from each of the classification states which must be identified. Once this stipulation has been satisfied the analysis itself simply consists of performing an evaluation of the relevant symbol table parameters from the sample segments acquired from the source. Initially the sample set is analysed to extract all the relevant shape information in order to define parameter quantisation boundaries for each parameter in isolation. The second stage focuses on analysing the symbols generated by a sample set using a base table generated with the initial quantisation boundary selection. This stage enables the symbol set to be reduced still further by considering the parameters together rather than in isolation.

3.3.1 Initial Parameter Quantisation to Minimise Signal Distortion and Table Size

This initial phase of the table rationalisation is essentially part of what was previously termed an application limitation. It employs information extracted from the source samples to identify optimised table boundaries for each parameter. These optimised boundaries are defined as being positioned so as to minimise signal distortion for each given set of table rules. The distortion takes two forms, shape distortion and noise distortion. The TES coding process as with analogue to digital conversion is susceptible to noise as a result of quantisation effects. The amount of quantisation noise introduced during this coding is non-uniform in distribution. It is dependant upon the positions of the parameter boundaries as well as the dynamic properties of the signal. The most effective method of reducing the effects of quantisation errors is to provide a constant signal to quantisation ratio (SQNR) over the entire dynamic range. This is achieved by allocating the signal parameter boundaries in a non-linear fashion across the range depending upon the cumulative distribution of the parameter when measured from the source. This results in allocation tables which have non-linear parameter boundaries.

In the case of amplitude coding where the emphasis is on generating codes which provide an indication of signal energy content this non-linear step-size allocation entails coding the lower energy epochs more accurately than the higher energy ones to maintain the necessary flat SQNR in the converted signal. The nature of the source is important in the allocation of boundaries since the distribution of amplitude levels across the dynamic range should accurately reflect the distribution of epoch energies in the source. Each individual source may have a different set of physical attributes which will cause these boundaries to shift.

Thus to minimise noise distortion in an n-level energy based TES coding scheme it is necessary to identify (n-1) boundary positions based upon the source signal characteristics. Figure 3.5 provides a good example of the boundary positions selected

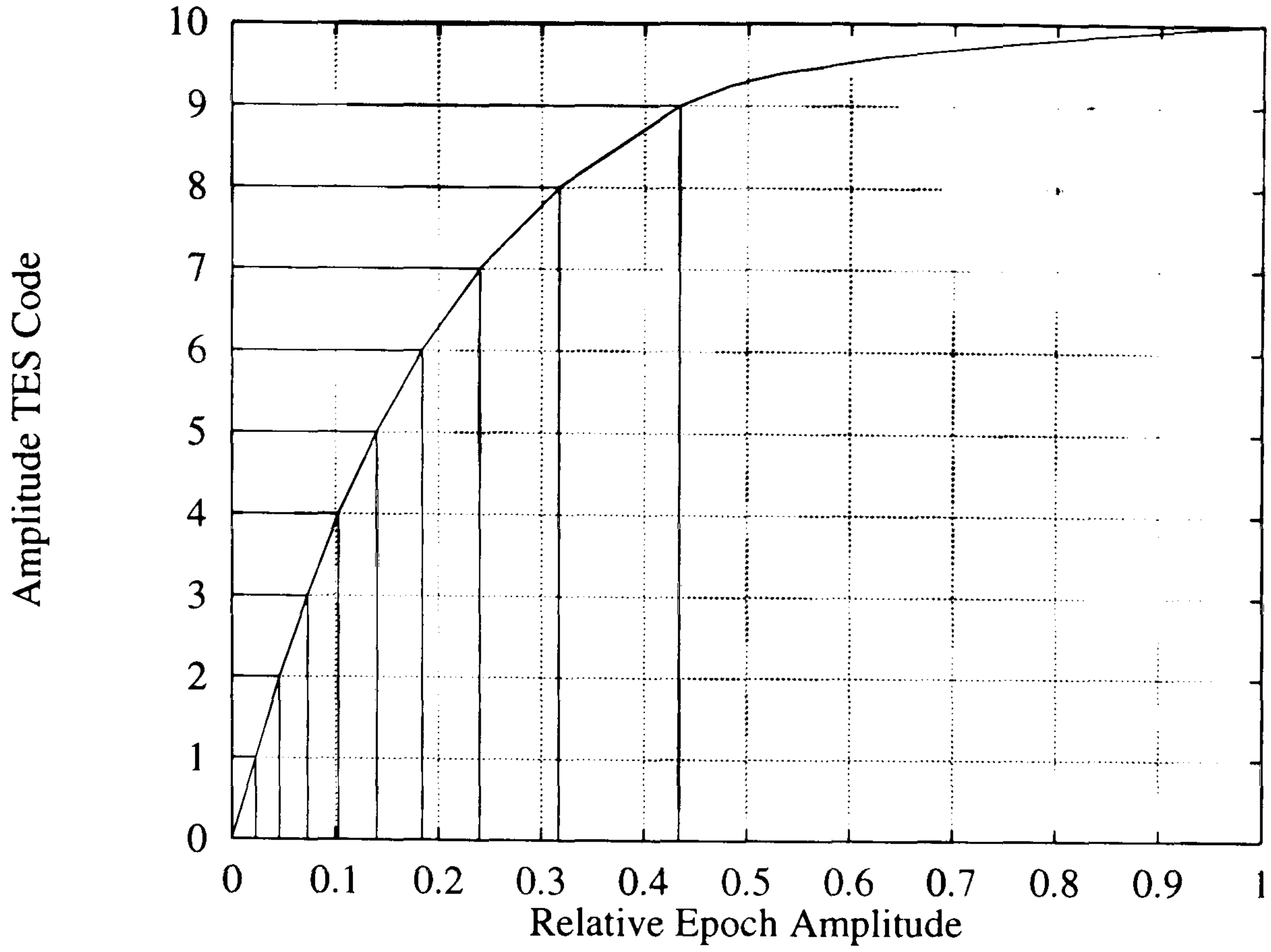

Figure 3-5 Statistical identification of the energy boundaries for application in a ten level quantisation TES conversion scheme

after the statistical analysis of a taped record from the gearbox testbed platform converted using this type of energy based amplitude coding scheme. This particular example is of a ten level (n=10) energy parameter definition plot. It should be noted that all the boundaries lie in the 0-0.5 region with non-linear separations between the successive points. If these boundaries were employed in a practical system then epochs having normalised amplitude peaks in the region 0.5-1.0 may be allocated symbols from the symbol table which significantly distort their energy content with respect to actual levels. Each coding scheme parameter, as with the case of the energy term just described, requires its own particular limitations to be placed on the dimension, n. However the same boundary optimisation principles are used for both of the remaining parameters, minima and duration, which make similar demands on the use of statistical analysis.

The development of a practical recognition system must encompass such tools as are required to perform the analysis and boundary selection discussed in this section. This introduces two considerations which are essential to the viability of a "system". The first is that the statistical pre-processing of sample source signals from a particular device in order to perform the boundary selection satisfactorily should be as automated and as simple to operate as possible. Secondly, since any subsequent classifications are based on the data stream produced by signal encoding using an optimised table any variations in the symbol tables could be expected to impact upon the consistency of the classification itself. Consequently this situation should be controlled by providing as large an initial data set as possible during the calibration phase. The principle being to encompass as many different system conditions so enhancing the range of signals which can be adequately catered for.

3.3.2 Secondary TES Symbol Table Optimisation

Once the primary phase of parameter truncation and quantisation has been performed an initial, or first-stage, symbol table is produced which contains entries for all axis combinations. This table however is generated as a result of statistical analysis of each of the parameters when taken in isolation. As a result there is room for additional rationalisation of the symbol set populating each allocation table. This second stage of optimisation considers symbol rather than parameter distributions in the source. The analysis is performed by passing a sample signal through a TES coder which uses the newly created first-stage allocation table in the coding algorithm to produce a raw symbol stream. Measurements can then be made on the frequency of occurrence of the individual symbols within the allocation table contained in the raw symbol stream subsequently generated. This statistical symbol evaluation provides valuable information on the occurrence of combined signal parameters which in turn can be used to prune the fully populated first-stage table to produce a better optimised second-stage table which is only partially populated.

This secondary pruning process requires the provision of criteria for symbol incorporation into the table. Since the complexity of the signal-to-state relationship is generally non-trivial using a purely physical means to distinguish between a symbol which is useful to the classification procedure and one which may be discarded is not practicable. Instead a more basic frequency based criterion is applied to each symbol from the primary table. In other words each unique symbol is required to attain or surpass a pre-selected cut off frequency prior to inclusion in the final optimised allocation table. This procedure reduces the symbol set to a minimum by excluding all those symbols which do not attain the required selection cut of frequency.

In defining this cut off point a balance must be sought between the primary need for an allocation table containing few symbols which reduces subsequent matrix diversity and the need to minimise the introduction of signal distortion due to symbol warping. Whilst reduced symbol diversity should lessen the demand placed upon the subsequent classifier increased distortion will introduce additional overhead. Symbol warping is the inevitable result of a fragmented table and refers to the distortion caused when an epoch whose associated symbol is missing from the allocation table is reassigned to a neighbouring symbol which has a different shape definition. The process of minimising the effects of the symbol warping phenomenon are discussed in more detail in the next section.

Once again the success of the secondary pruning exercise is mainly dependant upon the selection of an adequate set of source samples being passed through the coding process to provide the symbol distribution data which is used to identify those symbols warranting inclusion. Each particular recording, or state, will have differing acoustical properties and consequently the symbol characteristics will vary accordingly. To provide a symbol set capable of performing adequately over a range of signal types the source sample set should include sufficient data for each of the required types. What

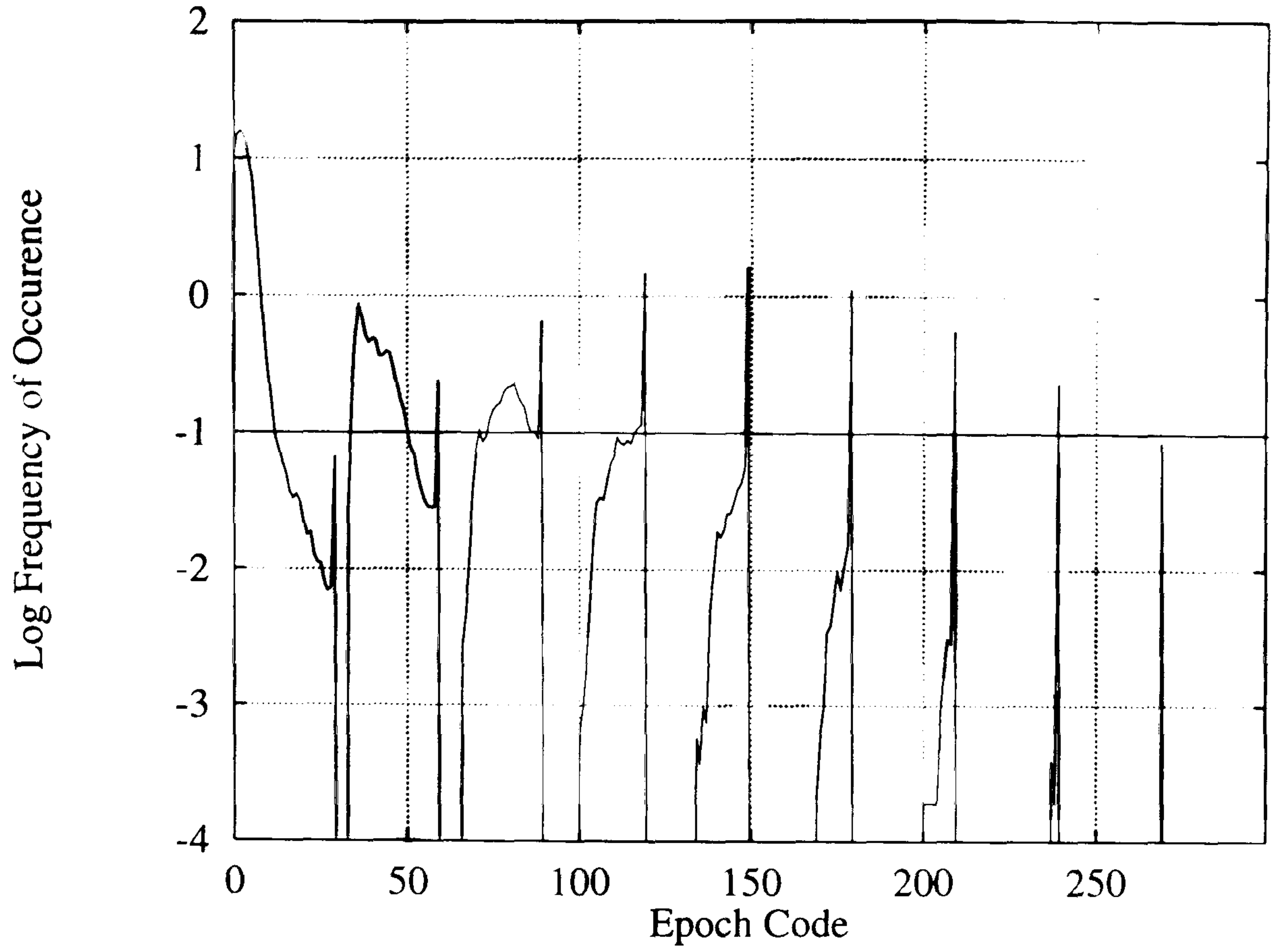

Figure 3-6 Statistical analysis of symbols produced from a coder applying minima based symbol selection to emissions corresponding to an out-of-mesh state

should result if the set selection is performed effectively is an allocation table which is able to code a limited range of source signals whilst at the same time maintaining the signal distortion introduced to an acceptable level. It should also preserve this signal information in as compact a form as possible to enable the development of classification strategies based upon this data which respond in reasonable time and with sufficient accuracy for minimal computational effort.

Graphical examples of the frequency of symbol occurrence produced by segments of

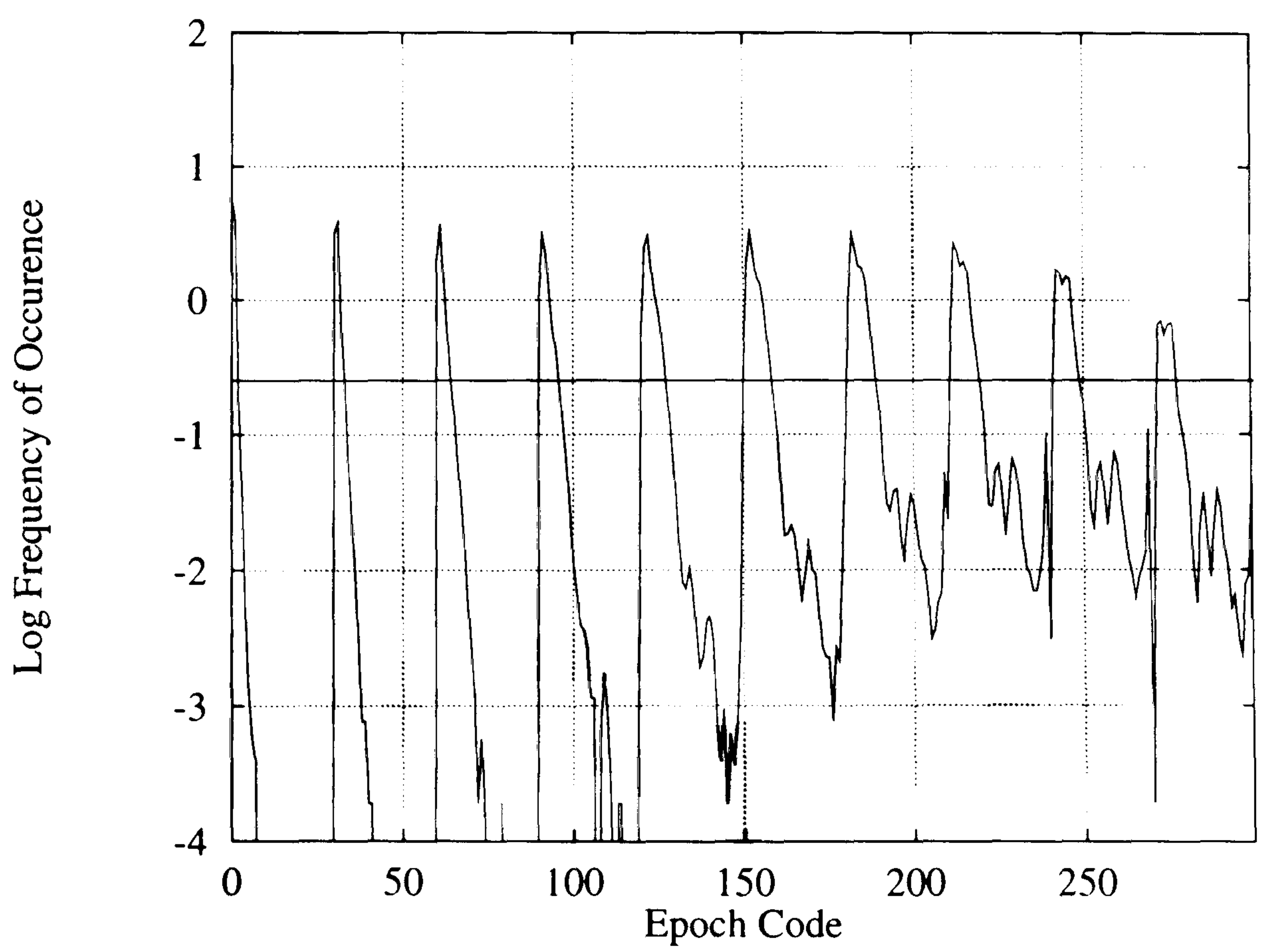

Figure 3-7 Statistical analysis of symbols produced from a coder applying amplitude based symbol selection to emissions corresponding to a partial-mesh state

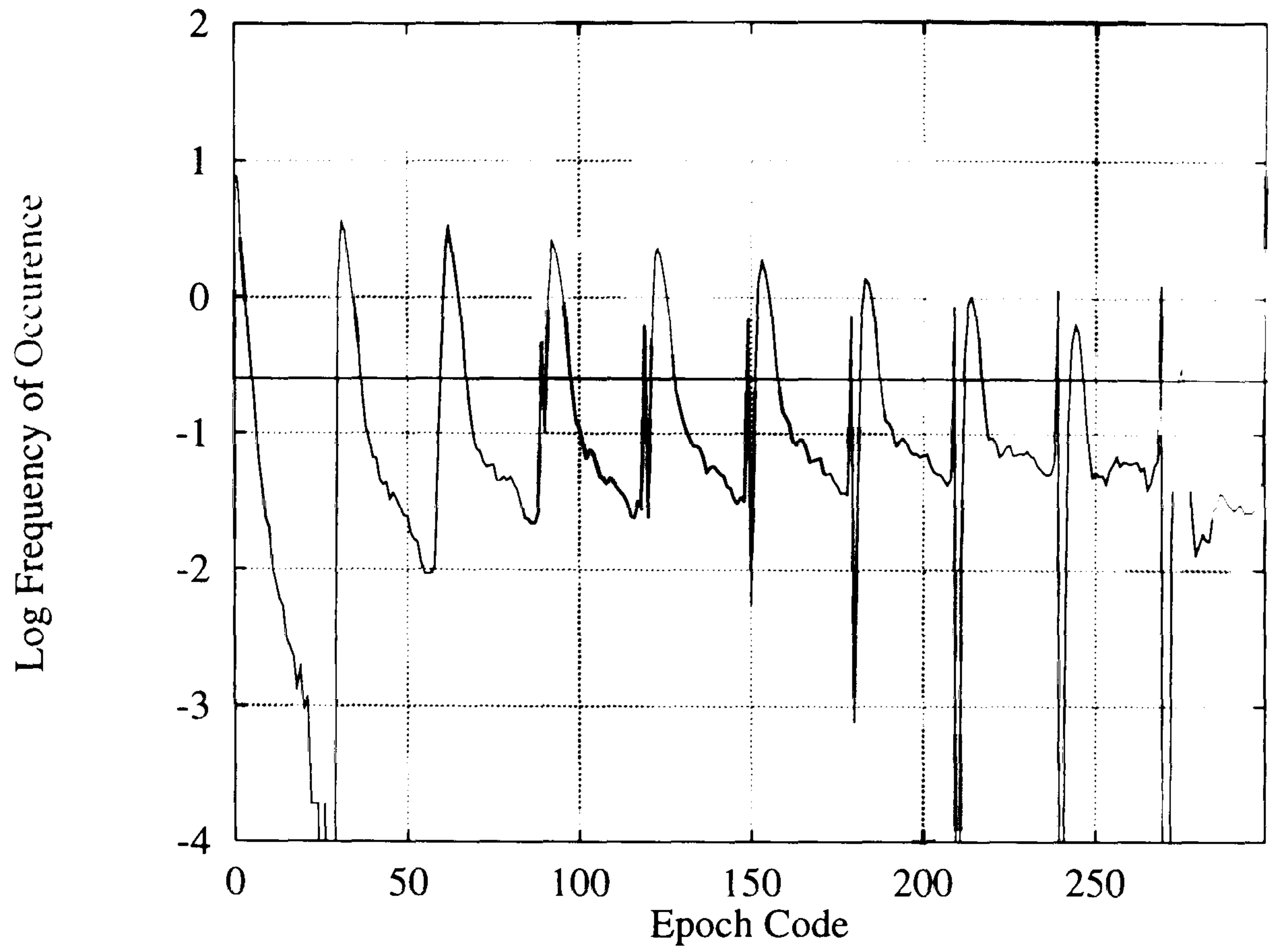

Figure 3-8 Statistical analysis of symbols produced from a coder applying amplitude based symbol selection to emissions corresponding to an out-of-mesh state

acoustical source from the testbed system acquired to perform the table minimisation are given in Figure 3.6, Figure 3.7 and Figure 3.8. As can be seen the distributions vary according not only to the signal source used but also to the coding strategy and symbol tables employed. For example, comparing the minima symbol distributions in Figure 3.6 with those for amplitude coding in Figure 3.8 both acquired from an identical signal source the differences in the distributions are clearly visible. The minima symbols are more evenly distributed across the table when compared to the peaky characteristics of amplitude coding. As a result the cut off for minima coding is placed at logf=-1 which corresponds to a frequency of occurrence of 0.1%. This produces a second-stage table containing just 30 symbols. In contrast the more peaky nature of amplitude coding requires a cut-off to be placed at logf=-0.6 which corresponds to a frequency of symbol occurrence of 0.25% giving a second-stage table containing 40 symbols. However from this it should be noted that the pruning has resulted in the omission of between 85 and 90 percent of symbols from an initial fully populated table of 300.

If the symbol distributions are examined in detail certain key characteristics associated with the source, which will later be used to classify the signals, can be identified. For instance the pronounced peaks which accompany some of the symbol blocks in the distributions are the result of epoch clipping on the tables duration axis. These peaks give an indication of the low frequency content of the signal. If the distributions in Figure 3.7 and Figure 3.8 which were both acquired using the same amplitude coding strategy are compared certain other characteristics can be seen. For example in Figure 3.8 the signal energy is more evenly distributed throughout the frequency range whilst in Figure 3.7 it is biased more towards codes in the range 120-300. This type of data is precisely the type of signal information which can be used in accurately determining the state of the signal source.

3.4 Symbol Allocation within a Fragmented Table

The selection of symbols based on the frequency of occurrence which is so critical to table minimisation poses an additional problem which was alluded to in the previous section, i.e. that the minimised table becomes fragmented as a result of the symbol selection. Thus the encoder used must be developed to account for the possibility that a given set of epoch parameters will not have a unique symbol allocated to them. Under these circumstances the encoder must select a symbol from those remaining in the fragmented table. This process is termed symbol warping. Warping will obviously cause some distortion in the coded signal no matter how it is performed. However the effects of the distortion can be minimised if the manner in which the reallocation is performed is chosen carefully. The solution is to define a set of rules outlining the definition of nearest neighbour symbols to those entries in the table which do not have unique symbols. These rules are then applied when an unallocated table entry is referenced and a replacement symbol is selected instead. Essentially the most important aspect of this rule set identification is to monitor, at each stage, the side effects caused by the inevitable epoch, and in turn signal, distortion. These distortion effects are most easily identified by a reduction in sensitivity, response and accuracy of the subsequent classification process used to separate the TES data sets.

Figure 3.9 graphically displays the scenario of a pair of epoch parameters which relate to a position in the allocation table which is unoccupied. The position is surrounded by eight other symbols which in this instance have all been included in the final table. The rules defining the best choice of substitute symbol, or freedom bound, is dependant upon the emphasis placed upon the information defined by the two axes. In the

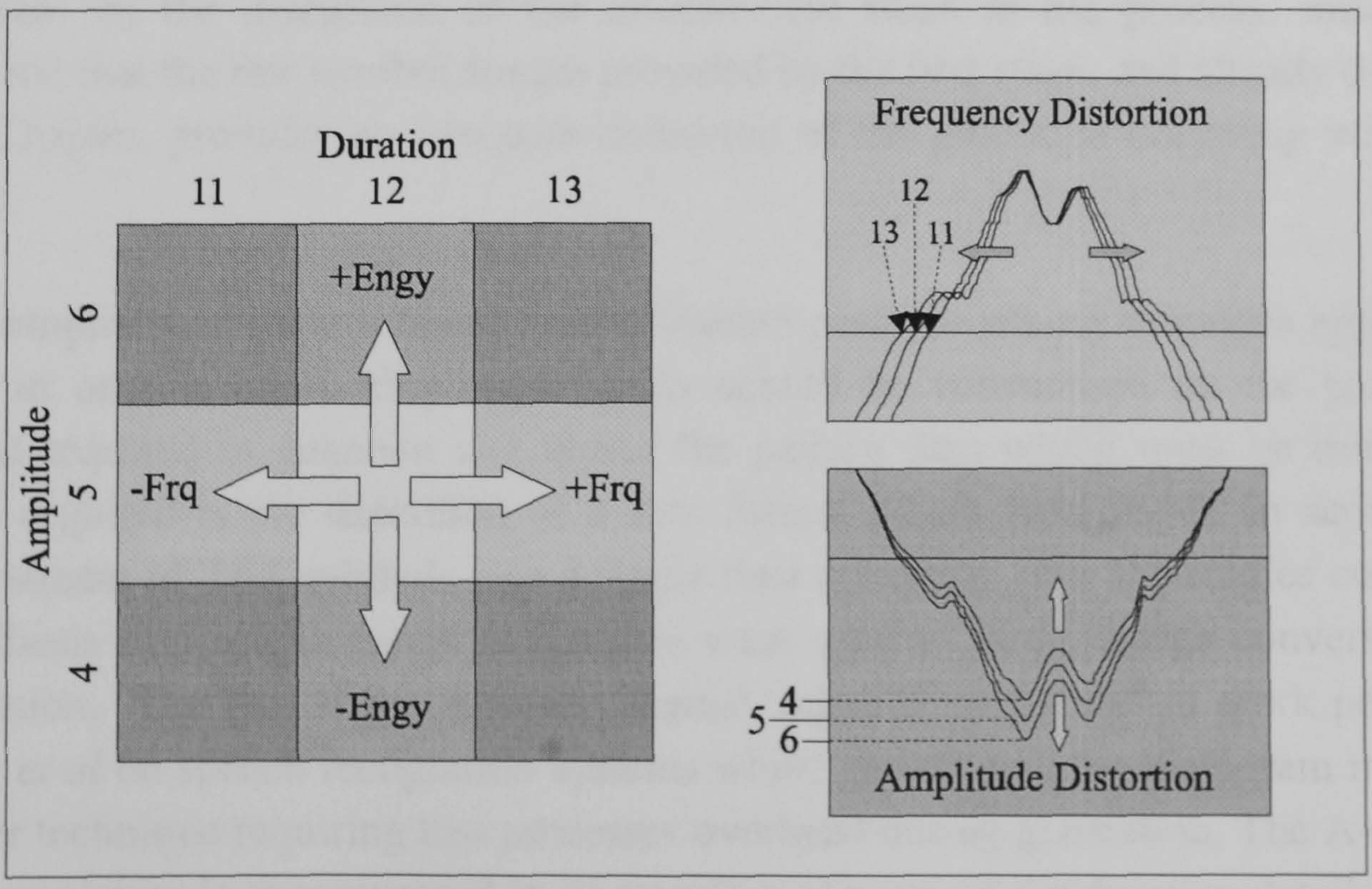

Figure 3-9 Amplitude and frequency code warping as a result of allocation table fragmentation with the amplitude coding strategy

amplitude coded example above one axis describes the duration of the epoch and consequently contains information about the raw frequency content of the signal. The second axis in this scenario defines the amplitude quantisation groups and as such describes the energy state to which the epoch belongs. Since the best secondary choice of symbol is determined to be the nearest neighbour the diagonal symbols may be eliminated as these would constitute a distortion in both information axes. This simply leaves a choice between the vertically or horizontally aligned symbols. The preservation of the epochs frequency information, or duration, is considered to be of greater importance than the energy or minima information conveyed by the remaining epoch parameter. This reduces the choice to the two remaining codes one of which increases the perceived energy of the signal epoch and one which reduces it. Allocating the epoch to the lower band would statistically minimise the error in energy allocation since the allocation is non-linearly quantised.

3.5 The Application of TES Information to Condition Monitoring

The extraction of accurate machine condition information based upon the application of either minima or amplitude derived TES coding algorithms consists of three interrelated sub-areas. The primary stage which has already been discussed is the adequate conversion of the raw signal data derived from the acoustic sensor into a format suitable for post-processing classification. This first stage results in the generation of a continuous stream of TES symbols which are focused on preserving the key aspects of the raw signal. The second stage absorbs the symbol stream taken from the first stage and converts it into a format which may readily be applied to a condition classification based pattern matching algorithm. The final stage in the process centres on the development of pattern matching techniques which are able to provide fast and reliable identification of signal characteristics from the TES data generated. This section will concentrate on the discussion of the intermediate stage in the process, making the assumption that the raw symbol stream provided by the first stage, and already described in this Chapter, provides an adequate definition of the processes occurring within the machine.

The development of pattern based symbol stream post-processing strategies appropriate for use in on-line monitoring applications should be constrained by the processing overhead required to generate and utilise the pattern data which must be minimised. What is required is the definition of a data format which compresses an acoustically derived stream of TES symbols into a single data reference. For the studies completed in this thesis two matrix based techniques were used to perform this conversion and compression. The first is the A-matrix format, taken from the earlier work performed by King *et al* on speech recognition systems whilst the second, the histogram matrix, is a simpler technique requiring less processor overhead during generation. The A-matrix's added complexity is encompassed in its retention of some of the duration detail which is discarded during histogram generation. Both techniques generate matrix format data sets which for ease of visualisation are easily translated into three dimensional contour

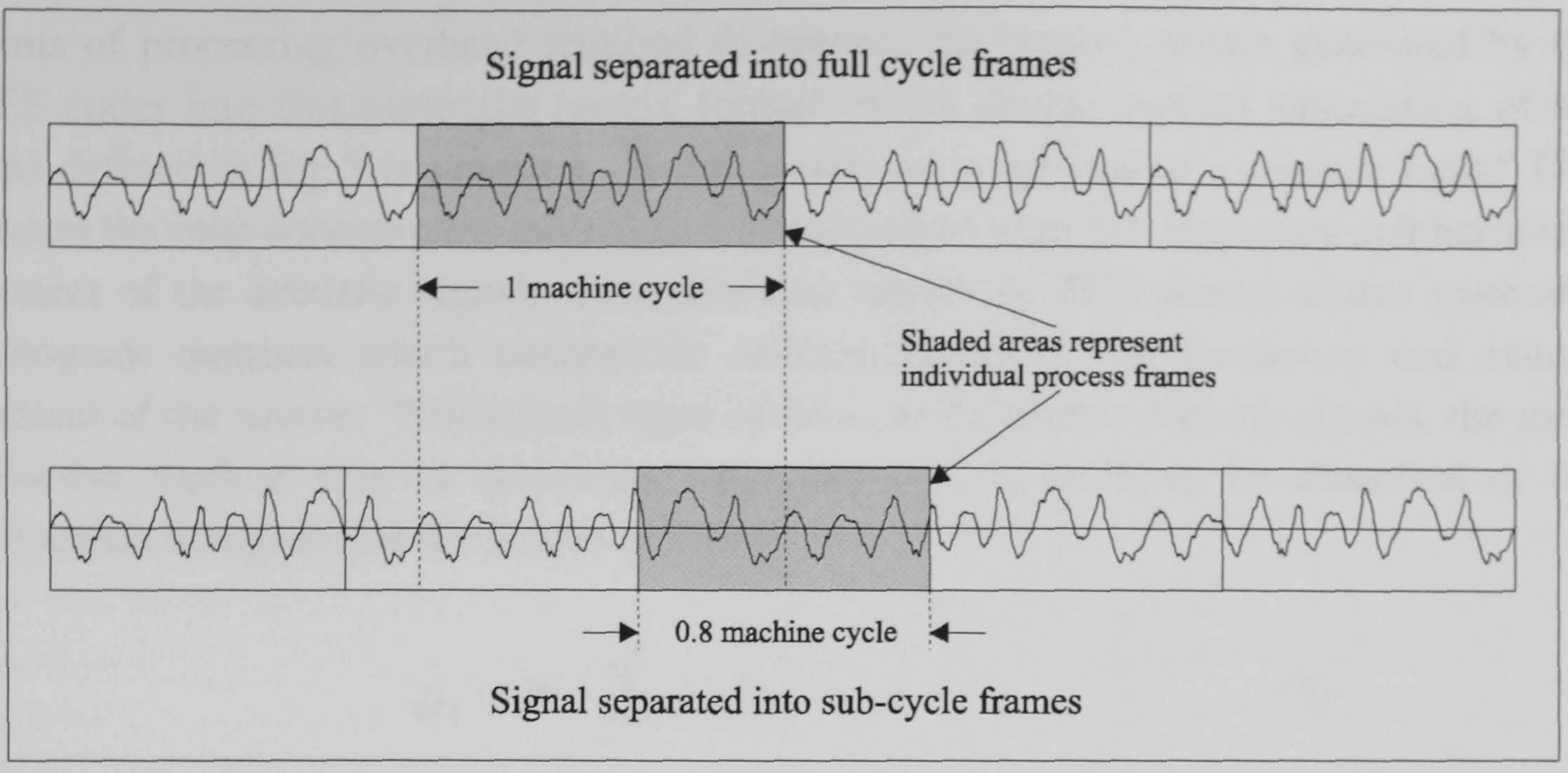

Figure 3-10 Description of the relationship between the cyclic nature of a signal and the way in which it can be separated into token frames to generate condition matrices

maps which will fluctuate with respect to time. These fluctuations are related to the calculation period or framesize used for the matrix generation (see Figure 3.10) as well as on the acoustical stability of the source under observation. In addition to these fluctuations secondary effects which are not the result of predictable instabilities but caused as a result of genuine changes in the system condition due to the introduction of faults are expected. If this is indeed the case then there is no reason why the contour movements within the acoustically derived matrices cannot be used to provide feedback regarding the condition of the target system.

One of the important areas of consideration made during the authors studies was the statistical comparison of condition feedback provided by each of the four different matrix formats applied during this study. These consist of matrices derived using combinations of the two different TES conversion techniques and the two distinct matrix generation strategies. Amplitude and minima are the TES conversion strategies concerned whilst each of these are combined with either histogram matrix or A-matrix template generation schemes to produce classification data. Each of the combinations of coding strategy and matrix generation algorithm were evaluated on acoustical token sets recorded from the testbed gearbox system. These practical comparisons and the discussion of the relative merits of each are discussed in depth in Chapters five and six. The remainder of this section is concerned with the mathematical derivation of the matrix generation strategies which have already been described and are used in the practical trials in Chapters five and six.

3.5.1 The Generation of Simple Symbol Histogram Contour Maps

The generation of histogram contour maps for direct application to pattern matching algorithms provides a simple and yet effective manner of information concentration. In terms of processing overhead required to convert the symbol stream generated by the TES coder into this particular matrix format only a simple symbol summation of the type defined in Eq. 3 is required. If this technique is applied to a minima based TES stream the map concentrates information pertaining to both the frequency and harmonic content of the acoustic signal. In contrast an amplitude TES stream source generates histogram matrices which concentrate information about the frequency and energy content of the source. Whilst both types of data can be used to classify signals, the most effective method may be dependant upon the type of faults to be classified or the properties of signal that the source generates.

$$a_{ij} = N^{-1} \sum_{n=0}^{n=N} X_{ij}(n) \tag{3}$$

where N is the number of TES symbols in a single frame, d_x is the duration parameter of the n^{th} symbol and p_x is the minima or the amplitude parameter associated with the symbol used in matrix generation. X is a function defined by:

$$\begin{aligned} X_{ij}(n) &= 1 \qquad \text{if } d_x\{t(n)\} = i \text{ and } p_x\{t(n)\} = j \\ X_{ij}(n) &= 0 \qquad \text{otherwise.} \end{aligned}$$

To provide the means for histogram map generation from the raw symbol stream using Eq. 3, independent of the coding scheme, the symbol stream must first be subdivided into shorter discrete token frames. Each individual frame is then subjected to the histogram algorithm to generate a simple matrix, or token, which represents the properties of the signal over the frame period. As discussed earlier the selection of this frame length should be made with consideration for the cyclic nature of the source to be monitored. It should be short enough to reduce the likelihood of signal averaging over

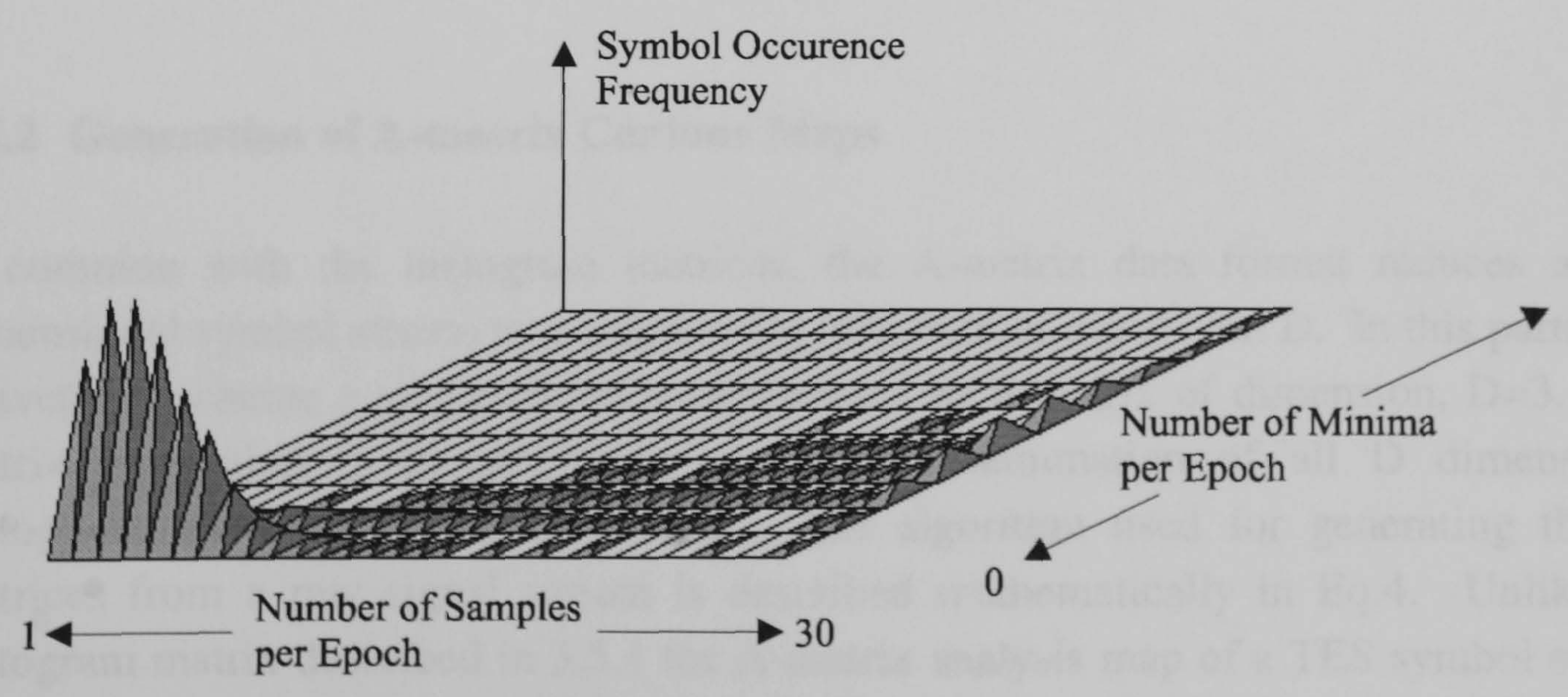

Figure 3-11 A histogram matrix generated from a gearbox acoustic recording using a Minina coding scheme

multiple cycles which may swamp intermittent faults yet long enough to eliminate the predictable variations due to expected periodic events in the system under observation.

If the statistical symbol data contained within the histogram map is then plotted in three dimensions a visual description of the acoustical processes taking place over the calculation frame is obtained. The map is subdivided into individual elements, a_{ij}, each of which corresponds to a unique symbol resident in the allocation table used for TES conversion. The magnitude of each matrix element ($|a_{ij}|$), denoted by the height above the matrix plane in visual terms, corresponds to the frequency with which its associated symbol has occurred during the frame used to generate the matrix. Thus peaks are expected in the matrix map corresponding to the most common source symbols and smaller hills for those less commonly used symbols. Each matrix map has a fixed boundary on the horizontal plane due to the constraints imposed by the fixed symbol set but the contours contained within each of the maps is dependant on the nature of the signal from which it was derived.

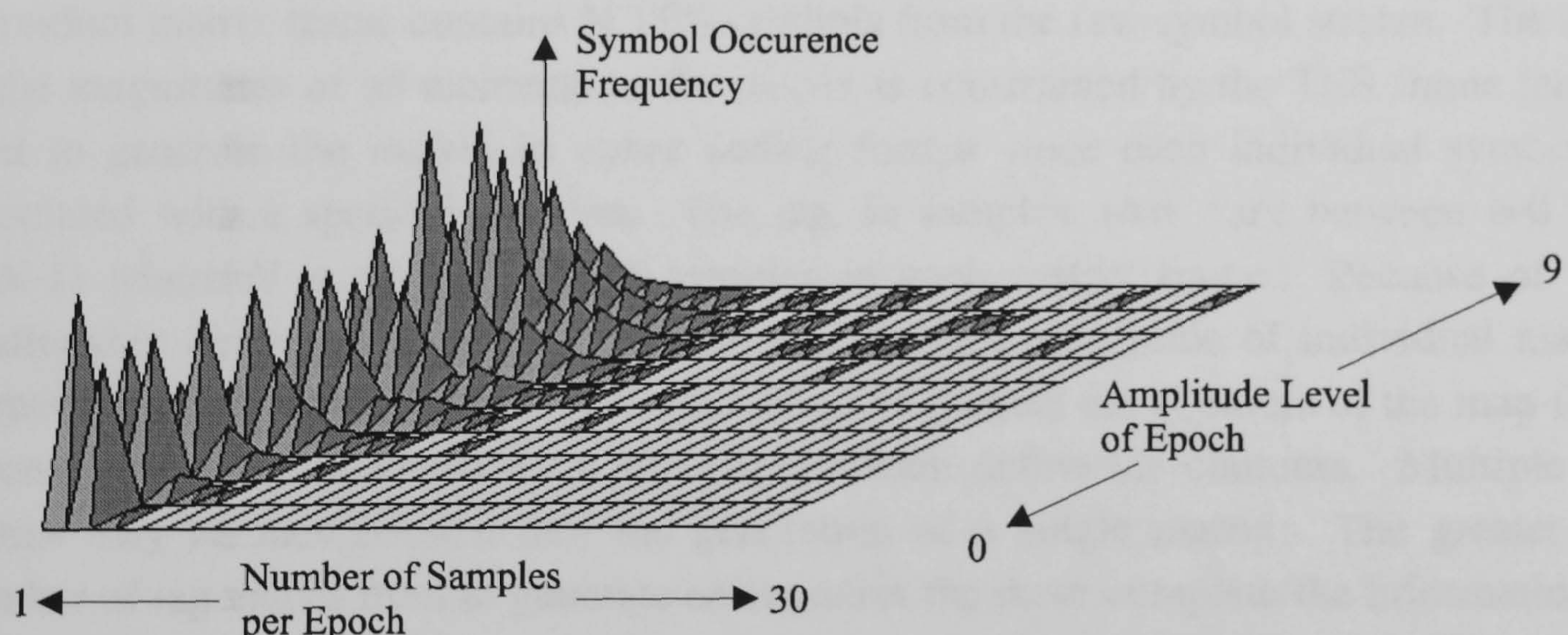

Figure 3-12 A histogram matrix generated from a gearbox acoustic recording using an Amplitude coding scheme

3.5.2 Generation of A-matrix Contour Maps

In common with the histogram matrices, the A-matrix data format reduces an N-dimensional symbol stream to a single point reference of dimension D. In this particular conversion scheme a single frame is represented by a matrix of dimension, D=3. The matrix generation is performed by an ordered summation of all D dimensional histograms associated with the stream. The algorithm used for generating the A-matrices from a raw signal stream is described mathematically in Eq.4. Unlike the histogram matrix described in 3.5.1 the A-matrix analysis map of a TES symbol stream retains some of the general signal shape and timing information from the original signal. This information would have been lost in the less complex symbol histogram contour map described previously.

$$a_{ij} = (N - l)^{-1} \sum_{n=l+1}^{n=N} X_{ij}(n) \qquad (4)$$

Where N is the number of TES symbols in a frame and l is the lag parameter defining the distance, in symbols, between symbols pairs in the stream which are compared. X is a function defined by:

$$X_{ij}(n) = 1 \qquad \text{if } S(n) = i \text{ and } S(n-l) = j$$
$$X_{ij}(n) = 0 \qquad \text{otherwise.}$$

And $S(n)$ refers to the n^{th} symbol in the TES symbol stream.

As can be seen it defines a matrix whose *i* and *j* boundaries are symmetrical and equal to the number of TES symbols resident in the requisite allocation table. The magnitude of each element, X_{ij}, of the matrix is dependant upon the frequency of occurrence of pairs of TES symbols, S(n) and S(n-*l*), separated by a pre-defined symbol lag, *l*. Each individual matrix frame contains N TES symbols from the raw symbol stream. The sum of the magnitudes of all elements in the matrix is constrained by the TES frame length used to generate the matrix in either coding format since each individual symbol is associated with a specific duration. The lag, in samples, may vary between l=0 and l=(N-1) where N is the number of samples in each matrix frame. Because of this relationship between the symbol comparisons and the magnitude of individual matrix elements the lag term used in matrix generation will affect the contours of the map for a given frame since it affects the comparisons which define the contours. Multiple lag values may be incorporated into the generation of a single matrix. The greater the number of lag values used to generate each matrix the more complete the information in the matrix corresponding to the source becomes. However, work both in the speech recognition field and in early trials on machinery suggest that a single value of lag, usually l=1, is sufficient to derive an A-matrix adequate for signal classification whilst

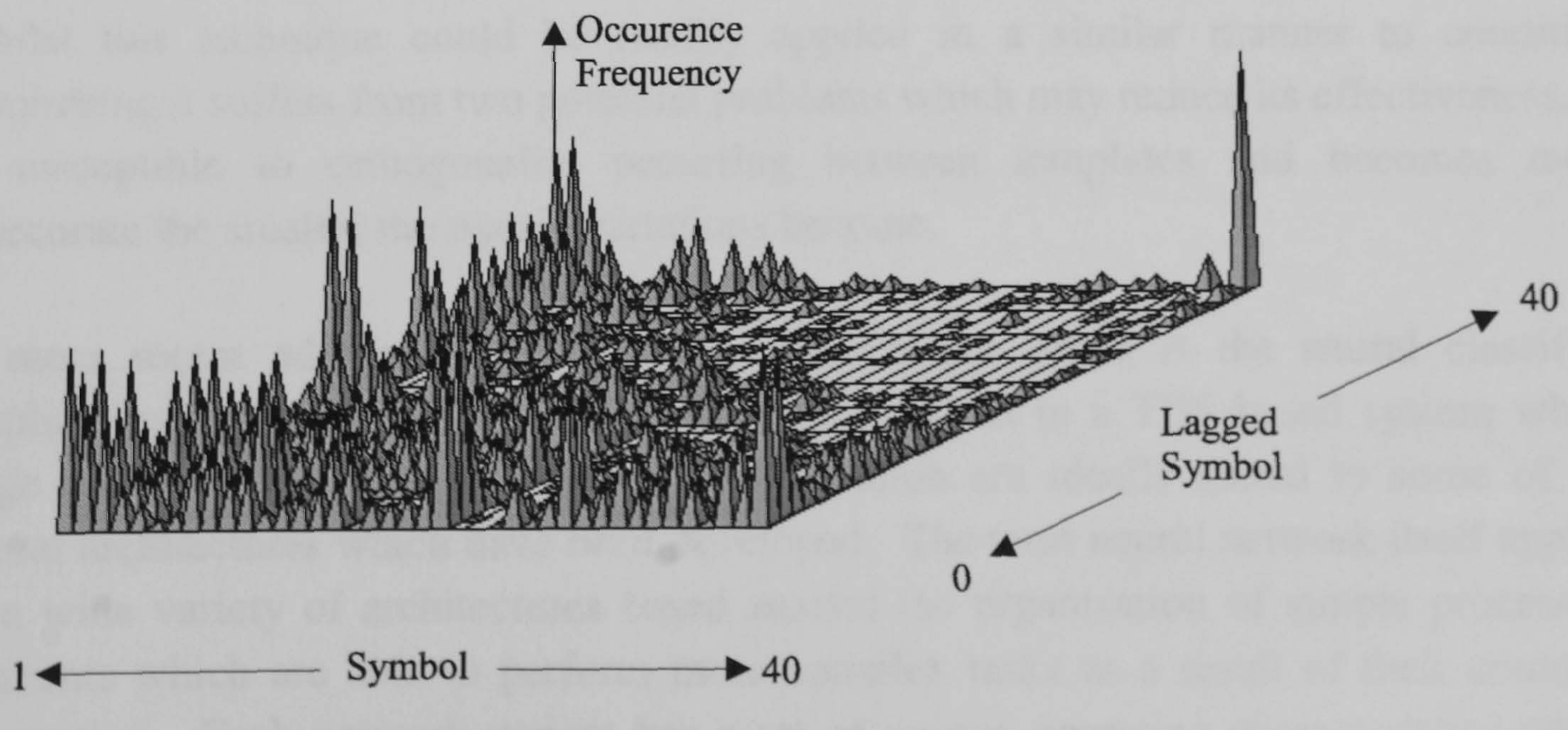

Figure 3-13 An A-matrix generated from a gearbox acoustic recording using an Amplitude coding scheme

minimising the generation overhead. This selection is ideal since it minimises processor overhead in matrix generation whilst at the same time retaining sufficient signal information to classify.

3.5.3 The Identification of a Suitable Pattern Matching Tool

The final stage in the process of signal classification is the application of a suitable pattern matching strategy to the contour map style matrices that have been constructed using either the simple histogram or more complex A-matrix algorithms. From the point of view of industrial applications this final classification mechanism should combine accuracy with ease of use. As such the development of an automated technique for template and matrix correlation to identify signal class would improve still further the suitability of a system, particularly to unskilled users in the industrial workplace. In the early work carried out in [32,36,37] on TES speech techniques these matrix template comparisons were performed using a simple distance based algorithm of the type described in Eq.5.

$$c(A,B)=\frac{\left(\sum_{ij} a_{ij} b_{ij}\right)^2}{\left(\sum_{ij} a_{ij}\right)^2\left(\sum_{ij} b_{ij}\right)^2} \tag{5}$$

This algorithm was successfully used by the authors to classify a range of spoken words in both continuous and broken speech. It is invariant to scale changes and generates a distance measure, c, which is used to perform the classification. The measure compares the generated contour map, A (with elements a_{ij}), with each one of a predefined range of templates, B_i (elements b_{ij}). Each of these templates, which must be generated prior to the comparisons, contains details of a specific signal type. In the case of the work performed by King and others these signals were segments of human speech utterances. Whilst this technique could be readily applied in a similar manner to condition monitoring it suffers from two potential problems which may reduce its effectiveness. It is susceptible to orthogonality occurring between templates and becomes more inaccurate the smaller the matrix variations become.

A more recent addition to the pattern classification arena is the neural classifier. Applications such as the matrix comparisons required in a TES based system where large and repetitive data sets require classification are ideally suited to some of the neural architectures which have been developed. The term neural network itself applies to a wide variety of architectures based around the organisation of simple processing elements which are able to perform more complex tasks as a result of their common interaction. Each network variant has a set of unique operating characteristics which predefine the uses to which it may be put. Supervised network architectures are ideally suited to the application of the TES data which has been described previously in this

Chapter. These configurations which are able to separate a complex decision space into relevant classification groups are termed supervised networks because of their requirement for an initial period of training. They are representative of a type of classifier known as an adaptive pattern recogniser which is specifically aimed at pattern separation. The application of such networks to pre-processed matrix templates of human speech by others have shown classification accuracy's of the order of 90%. Following on from this, some very early work was completed by Vu *et al* [38] using similar neural techniques to identify simple acoustic characteristics of a diesel engine with excellent results. These were achieved with a code table containing only 29 codes to compress the 4kHz band limited data stream and to generate the A-matrix templates. From these positive early results it would seem that the technique may indeed prove to be a viable alternative to the more traditional techniques which have hitherto been available. The adaptive nature of the classification process also lends itself well to industrial applications where ambient conditions are rarely ideal and where there is a desire to reduce the necessity for skilled labour in the workplace. For these reasons the author has considered the application of neural techniques for this last phase of the classification process with a view to understanding the underlying principles and any limitations in their use. The theoretical and practical concerns associated with the application of these techniques to monitoring applications will be developed in full in Chapter 4. For the moment the neural classifier is assumed to be a black box implementation which takes pre-processed matrix tables as an input and produces a system state decision output based upon the data contained in the matrix.

Due to the nature of this particular neural architecture there are two phases to the application of a classification tool based upon this technique in a monitoring system, the first being supervised learning by the neural processor of the specific patterns, or contours, in the matrices produced by each state, or class, of signal. This requires the application of a predefined training data set which adequately defines each particular signal class. Once the neural processor has completed this training phase successfully it should in theory at least be capable of identifying and categorising a specific matrix presented to it. In the remaining part of this Chapter the practicalities of combining all of the stages so far outlined to provide a system which could be developed as a monitoring package will be discussed. These practicalities cover such problems as speed of conversion and compression, storage limitations, accuracy, usability and response. Further discussions on the specific problems associated with the application of neural classifiers is included in Chapters four, five, and six.

3.6 Practical Considerations in Implementing a Real Time TES Processing Engine

In order to perform a series of studies into the practical application of TES coding to machine condition classification and make objective statistical comparisons of the different processing techniques over a range of simulated machine conditions an efficient experimental toolset is required. For the purposes of this thesis a simple practical testbed system was constructed. The constraints imposed upon this testbed implementation were more relaxed than those which would be required for a more dedicated commercial system. Rather than simplicity of use or cost, the main concern at this stage is the ease with which the system can be reconfigured to study the effects of variations in classification strategies. User friendliness, responsiveness and the cost implications associated with changes made to the system are considered to be of secondary importance. These secondary performance criteria are however taken into consideration when characterising and evaluating any processing modifications. The modifications themselves range from changes in recording conditions, coding schemes, matrix template generation and compression schemes to the type and arrangement of the pattern matching neural network used to finally classify the data.

The TES coding system which was produced to satisfy all these diverse requirements employed a modular approach whereby each stage of the process could be considered in relative isolation. This maximised the ease with which the system could be modified

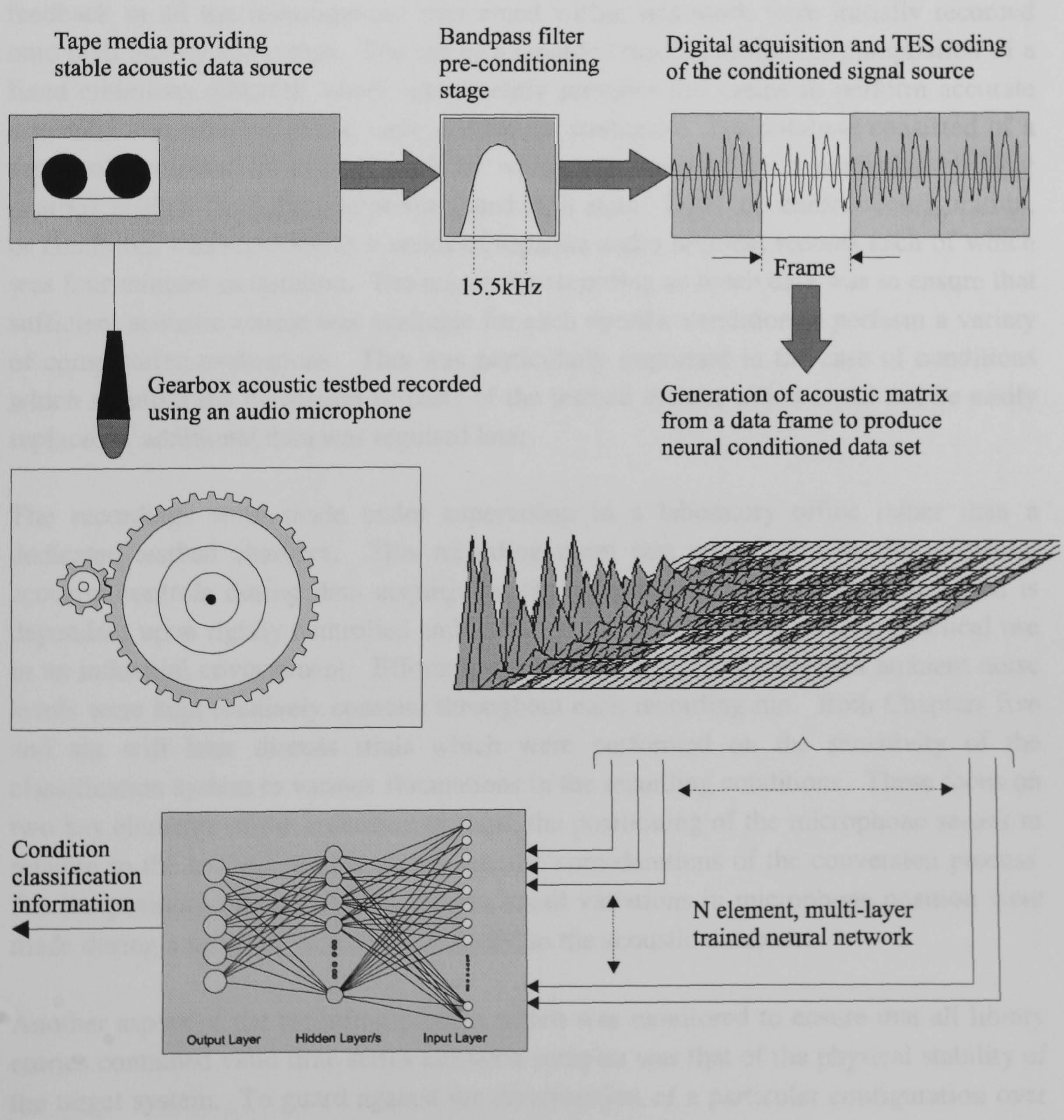

Figure 3-14 Modular description of the signal acquisition and condition identification system used for the trials

but at the expense ultimately of simplicity. However for study purposes this modular system, seen graphically in Figure 3.14, provides a reasonable compromise between flexibility and acceptable response. For a more detailed description of the individual modules in the test and evaluation system the reader is directed to refer to Appendix A. The remainder of this Chapter discusses one of the most important aspects of practical trials of TES for classification purposes, that of acoustic data stability.

3.6.1 The Acoustic Library used for Practical Trials

In order to enable repeatable comparative tests of the various TES techniques against several different machine conditions and in various configurations it is essential to provide a stable acoustic source against which comparisons may be made. To achieve this fundamental stability the testbed acoustic emissions used to provide the condition feedback in all the investigations performed within this work were initially recorded onto high quality audio tape. The use of a recorded media enabled the compilation of a fixed emissions database which subsequently provides the means to perform accurate statistical comparisons of the various different strategies. The database consisted of a series of extended recordings, each of which corresponded to a particular physical gearbox testbed configuration or simulated fault state. Each mechanical configuration, or condition, was recorded as a series of separate audio segment records each of which was four minutes in duration. The reason for recording so much data was to ensure that sufficient acoustic source was available for each specific condition to perform a variety of comparative evaluations. This was particularly important in the case of conditions which required the mutilation of parts of the testbed system which could not be easily replaced if additional data was required later.

The recordings were made under supervision in a laboratory office rather than a dedicated testbed chamber. This recording room was not subject to any particular acoustic controls during data acquisition, the principle being that a system which is dependant upon rigidly controlled ambient acoustic noise levels is of little practical use in an industrial environment. Efforts however were made to ensure that ambient noise levels were kept relatively constant throughout each recording run. Both Chapters five and six will later discuss trials which were performed on the sensitivity of the classification system to various fluctuations in the recording conditions. These focus on two key elements of the recording process, the positioning of the microphone sensor in relation to the testbed and the signal specific considerations of the conversion process. For the positional sensitivity evaluations small variations in microphone position were made during a series of recordings included in the acoustic database.

Another aspect of the recording process which was monitored to ensure that all library entries contained valid time series emission samples was that of the physical stability of the target system. To guard against the deterioration of a particular configuration over each of the recordings made for a unique gearbox state checks were made after each and every run when the testbed was shutdown. After each recording and shutdown cycle measurements were made to ensure that the configuration had remained sufficiently stable throughout the complete recording. This ensures that each four minute acoustic state segment acquired from the testbed contains only data relating to a single system condition state.

3.7 Chapter Summary

This Chapter has dealt with the historical and theoretical description of TES coding as well as the practicalities of implementing a system with this type of signal coding scheme for the purposes of monitoring the condition of a simple gearbox testbed system. Licklider, King and Gosling all contributed to the early development of this simple signal coding scheme. Their early work focused on the development of the technique as a means of developing tactical military communications systems which were able to operate over low bit rate channels. Later work considered the techniques applicability to human speech recognition for which it proved well suited. Early indications of the extension of the technique into condition monitoring are provided by the work of Vu *et al* [38]. However the technique is still in its infancy in this respect.

All the work carried out by previous researchers up to this point had centred around the use of the minima characteristics of the discrete signals to specify TES shape descriptor symbols. For the purposes of condition monitoring at least there would appear to be certain advantages in developing the technique and employing the energy rather than the minima characteristics of signals to provide shape information in the coded signal. Both this and the minima based techniques have been discussed during the course of the Chapter together with illustrations of each of the coding processes involved. The application of both techniques to practical signal conversion is discussed in association with some of the problems faced in selecting and optimising the various allocation tables required for each method. This discussion covered both the initial imposition of physical constraints on the boundaries of the tables as well as the subsequent statistical optimisation for particular signal types. The optimisation is particularly pertinent to the development of the A-matrix conversion which requires a much smaller allocation table if the subsequent matrix size is to be limited. A mechanism for the allocation of symbols within a fragmented symbol table is also provided.

The main objective of employing TES is to provide a means of signal application to a classifier for the purposes of condition monitoring. The evaluation of two matrix compression techniques for this purpose, histogram and A-matrix, is covered together with examples acquired from the gearbox testbed system used for practical evaluation of the monitoring techniques. Neural networks are presented as a means of condition classification using the matrix data produced as a result of TES coding. They are compared to the more classical means of pattern identification such as simple distance measures. The main advantage of utilising neural network classifiers is twofold. Firstly they can offer substantial improvements in terms of processing overhead due to their inherently parallel nature and secondly they do not require an explicit definition of the classification problem itself. Both of these are of interest to the automation of complex condition monitoring problems.

Finally the implementation of an evaluation system is outlined. The necessity for a modular architecture which provides an easy means of reconfiguration is recognised as is the necessity for a stable library of acoustic recordings with which to evaluate each stage of the monitoring process. The separate modules within this evaluation system are briefly described in Appendix A and provide the reader with a basic insight into the means by which the TES based condition monitoring evaluation trials discussed in later Chapters were performed.

Chapter 4

4. The Application of Neural Networks to TES Condition Monitoring

After initial conversion of the acoustic source which is performed using a TES encoder the next important aspect in the implementation of an automated monitoring system is the signal classification. Considering the complexity of some of the physical abnormalities and faults which can occur and the acoustical effects which arise as a result, it is quite apparent that state identification is unlikely to be trivial. What the TES conversion and subsequent matrix encoding generates is a series of technique dependant matrices which contain basic information pertaining to the source. The quantity of data generated for each unique source token varies between 300 to 1600 points depending upon the matrix encoding algorithm employed. Each of the data matrices, corresponding to particular machine states must then be analysed and identified if the associated condition is to be correctly ascertained.

Traditional pattern classification methodologies are available and could indeed have been used in this identification phase. These techniques could be used to calculate distance metrics between an acquired data matrix and a series of predefined templates corresponding to each of the known system states. However the more valid states there are in the system and the shorter the distance metric is between these states the more fine grained the classifier resolution must become in order to correctly select the associated state. These pressures result in a classification space which becomes ever more complex as a consequence of the volume of data necessary to adequately represent the source initially. This in turn is likely to lead to an extension of the time taken to perform the template mapping from the representation to classification space due to the inherently non-parallel nature of these traditional methodologies. However the more recent developments in neural based techniques offer some potentially useful enhancements to this latter mapping stage.

Neural networks, more usually termed artificial neural networks, are computational engines loosely modelled on the human brain. Their development stemmed originally from a desire to both understand more fully the human brain and to emulate some of its key strengths. Initially the innovative developments by Rosenblatt [39, 40, 41] on perceptrons in the later part of the 1950's were hindered by the fundamental problem surrounding the procedures for training of these networks. Such difficulties were overcome with the development of the now commonly used back-propagation learning algorithms developed by Rumelhart *et al* [42]. This comparatively recent breakthrough triggered an explosion in research into the practical application of such networks which has since lead to their widespread acceptance in a variety of fields. These include applications as diverse as process control, financial analysis and signal processing where they are able to provide significant financial benefits over more conventional means. One of their key strengths, and the reason for their widespread application, is their ability to model the arbitrarily complex multidimensional non-linear functions found in some real world problems, without the necessity for an explicit definition of the relationship between the variables involved by the implementer. More conventional methods rely upon a more in-depth understanding of the underlying principles and relationships involved in a particular process to provide a central processor with the necessary artificial intelligence to make decisions based on input stimuli. In contrast a neural based system relies on a much larger network of less complex processing elements and can be applied in situations where the knowledge base is incomplete. These characteristics of neural networks can provide distinct advantages in both the application and subsequent classification response even to relatively complex data sets.

The biological similarity already alluded to originates from certain distinguishing features which are present in the brain itself. The first is that signals are passed between individual processing elements, or nodes, by interconnections. Each of these connections has an associated weight factor which conditions the signal passed between individual nodes. The processing nodes themselves then apply a simple activation function to their net input to generate an output signal which is passed on to further nodes in the network. Whilst individual nodes are themselves only able to apply relatively simple activation functions to their net input signal the strength of the network as a whole is imparted by the manner in which these individual elements are interconnected. The resulting interconnected networks are able to demonstrate not only a rich variety of flexibility by identifying patterns and forming associations between individual data occurrences through experience but are also capable of doing so in a robust manner. The inherent robustness is imparted by the massively distributed parallel nature of the architecture which ensures that, given a reasonable network size, individual nodes only contribute to the output rather than dictate it. This not only enables decisions to be made on incomplete or noisy input data, which is common in real world applications, but may also provide a degree of graceful performance degradation in situations where individual nodes fail.

The rapid expansion in the field of artificial neural network research and development has produced now a wide range of network types each with different functional properties. Characterisation of each of these network types is generally performed by defining their nodal connectivity, or architecture, the specific nodal activation function used in each of the individual network elements or the training mechanism applied to the network to prepare it. The architectures vary from simple single layer networks to more complex multi-layer networks which can be fully or partially interconnected and which can include feedback paths with or without nodal memories. The training algorithms, used to pre-condition the networks prior to application, generally fall into one of three clearly defined categories. Unsupervised learning networks, as the name implies, require no output control and rely instead upon inherent properties of the input stimuli to identify common properties or features amongst them. A common use of such architectures, typified by Kohonen self-organising map networks, is the clustering, or partitioning of an input space to produce a series of data exemplar vectors corresponding to similar input vectors within the input space. In contrast supervised learning networks such as the multi-layer perceptron require each of the known states or output vectors which the network is expected to be capable of identifying to be backed up by a suitable number of input excitation vectors which the network is trained to associate with the relevant output states. This is a much more rigid and convoluted technique requiring a suitably controlled regime of data application to obtain acceptable results. The third and final type is the reinforcement learning network in which a penalty function is applied to weight updates according to the performance of the network to input stimuli during the training phase.

Having now identified some of the key incentives behind the selection of neural classifiers as well as introducing some of the historical and biological origins of these techniques the remainder of this Chapter will cover some of these aspects in more detail. The first section, 4.1, deals with the basic theoretical principles of the technique focused specifically on the multi-layer perceptron (MLP) implementation. Section 4.2 forms a general review of the practical considerations associated with applying TES data to these MLPs for the purposes of monitoring a simple gearbox. Section 4.3 considers the impact of variations in both the internal and external architectures of the networks on the systems performance. Section 4.4 examines the necessity for adequate training procedures to be defined and the contribution this has on the subsequent performance during comparative trials.

4.1 A Theoretical Introduction to the Multilayer Perceptron (MLP)

One of the most common types of neuro-computational network used in the field of pattern recognition, the feed-forward multilayer perceptron or MLP, network has undergone rapid development since Rosenblatt's early perceptron work. This is a network containing many simple individual processing nodes called perceptrons, groups of which are ordered into distinct layers. The layers are then organised into specific functional types. The primary layer acts a passive data presentation type layer which performs no other function than to pass the input vector, or data pattern, onwards to the first dedicated processing layer. This first processing stage, and any other subsequent internal stage which is not directly exposed as an input nor an output layer is termed a hidden layer. In multi-layer architectures there may be any number of these so called hidden layers depending upon the level of data abstraction required to map input to output. The final stage in the network is called appropriately enough the output layer and its function is to present the result of the networks input-to-output vector mapping. The networks feed-forward definition refers to the fact that data is passed sequentially through the layers from input to output in a forward direction during operation. In this most basic format there are no feedback links between layers nor memories within individual nodal elements.

Each individual layer within the network interacts with its neighbouring downstream layer through a series of weighted interconnections between the individual perceptrons. Each perceptron, or node, acts as a simple processing element absorbing the series of weighted input stimuli from the previous layer, summing them, adding a bias weight and then passing the sum through a non-linear transfer function to produce a cumulative output stimulus. This output stimulus is then fed through to the perceptron elements in the subsequent layer. Whilst the early work with single layers of trained perceptrons was limited to the mapping of simple decision boundaries, multiple layered or cascaded, networks offered the capability to implement far more complex decision boundaries. A 3 layer architecture, an example of which is illustrated in Figure 4.1, contains an input or presentation layer, a single hidden layer, and an output or class layer and provides sufficient decision boundary modelling capability for most data characterisation problems of reasonable complexity. In addition to the perceptrons transfer function and the effects of adding or removing layers the internal connectivity between layers and the number of nodes within each individual layer will also affect the ensuing performance both during the training and operational phases of classification. This will be discussed more fully later in the Chapter in section 4.3

The application of an MLP network to a real world problem must be performed in two phases. Prior to operational classification a period of supervised training is required during which the interconnecting weights between individual perceptrons in adjacent layers are trained on a series of data exemplars. During this phase the network must be presented with a reasonable cross-section of the data it will later be required to classify. Until the early training problems concerning the adjustments of these interconnecting weights acting upon the output stimuli in successive layers of such cascaded networks had been developed by, among others Rumelhart *et al* [43], the theoretical improvements in performance could not be realistically achieved. The most widely used algorithmic approach to this weight tuning phase is a gradient descent algorithm which is used in conjunction with a soft-limiting non-linear perceptron transfer function such as the sigmoid function. This non-linearity is necessary to ensure that the transformations between layers are not simply linear as would be the case with a hard-limited function. It also has the advantage of being differentiable making the

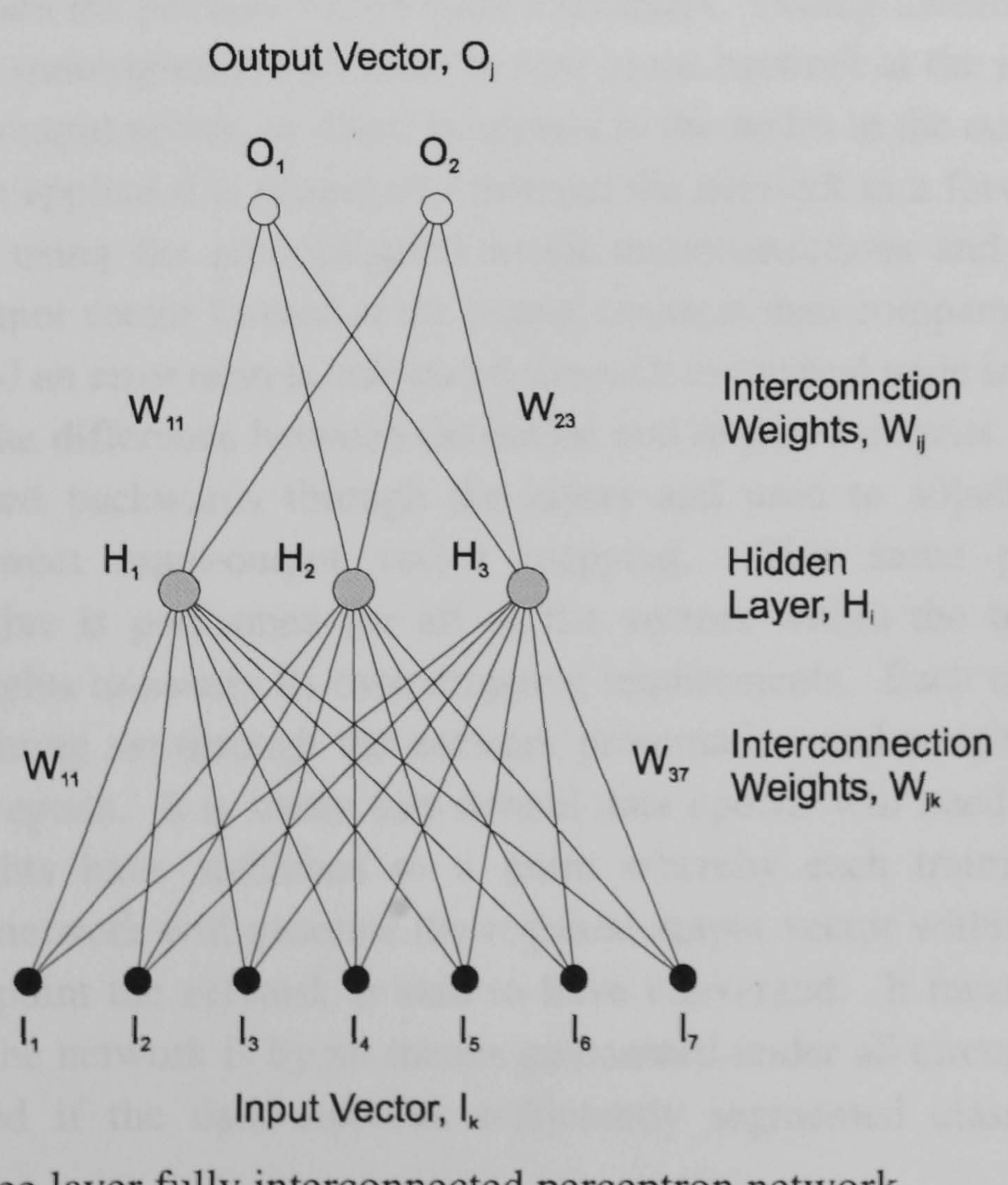

Figure 4-1 A three layer fully interconnected perceptron network

gradient descent approach to learning feasible. It is also useful in a wide variety of situations where a continuous valued output is required rather than a hard-limited binary switching mechanism.

The back-propagation algorithm as this improved training scheme was termed was a central factor in harnessing the potential which the earlier work of Rosenblatt had predicted. It provides a means of apportioning the contribution to an error in an output layer node to the nodes in multiple previous layers by back propagating this error contribution through each of the various layers to the input layer. In this way the algorithm tunes the weights so as to reduce the overall sum-of-squares error in an attempt to identify the global weight error minima within the classification data space for a particular data set. This algorithm is detailed further in Appendix B. In practice the identification of the global minima is rarely achieved since the error surface attributed to a mapping problem is generally non-ideal and contains peaks and troughs which the training algorithm may become trapped in. The effect that such a situation has upon subsequent classification is dependant on the error surface itself and the distance the local minima which the network converges to is from the surfaces global minima.

Prior to the commencement of training the interconnecting weights must be initialised, usually to small random values. The intention of this being to provide a reasonably stable starting point for the network which ensures that input signals to the processing nodes do not fall into the so called saturation region of the transfer function. The training procedure for a specific application requires sufficient data exemplars to be selected for the associated classification space to be adequately described so enabling the network to learn the patterns within these exemplars. During training each exemplar vector within the training set is presented in turn to the network at the input layer nodes and the required output vector, or class, is applied to the nodes in the output layer. Once a vector has been applied it is propagated through the network in a forward direction to the output layer using the preconfigured nodal interconnections and their associated weights. The output vector formed at the output nodes is then compared to the required output pattern and an error term is calculated for each individual node in the output layer by determining the difference between the actual and required outputs. This error term is then propagated backwards through the layers and used to adjust the weights to produce the correct input-output vector mapping. This same presentation and propagation routine is performed for all of the vectors within the training set, each updating the weights to satisfy its own mapping requirements. Each of these passes of the complete training set through the network presentation and weight tuning cycle is called a training epoch. It is likely that several data epochs will need to be performed before the weights have stabilised to a point whereby each training vector when presented to the network will generate the required output vector within a selected error bound. At this point the network is said to have converged. It must be stressed that convergence of the network is by no means guaranteed under all circumstances. It will only be achieved if the data contains sufficiently segmented classes and there is sufficient network flexibility to implement the necessary boundary definitions to separate the classes.

Assuming that the initial problem can be adequately defined so as to be presented to a neural network for classification the network must then be suitably trained to learn the interclass boundaries. There are several pitfalls which can potentially undermine the implementation of a neural classification scheme at this stage when compared with some of the more traditional techniques of class identification. These problems are mainly caused as a result of the semi-autonomous nature of the problem definition by a neural network. Most basic among these is the initial problem associated with convergence. During this stage the weight tuning process itself can be adversely effected by local minima located on the global error surface which can slow or in some cases stop the network from reaching a sufficiently global minimum to perform acceptably. Even if this does not happen and the network is able to reach convergence there is a finite possibility that the network will identify aspects of the data not specifically related to the particular classification problem. This may subsequently cause the system to classify previously unseen data according to a different set of data characteristics so reducing the systems overall capability. This final problem is one of the most difficult to control because there is no physical means of directing the weight tuning during training in this type of network learning algorithm. As a consequence it can produce a network which is well able to identify the elements within the training set but unable to adequately identify similar class exemplars taken from previously unseen data. This situation is described using two terms specifically associated with neural applications, namely generalisation and memorisation.

Memorisation refers to the situation whereby a network is able to classify correctly only those members of the training set. Generalisation on the other hand is a means of defining a networks capability to identify a previously unseen input vector correctly. A network is deemed to generalise acceptably when it can correctly class a series of unseen data vectors which whilst being members of identical class populations may differ slightly in terms of their data definition as a result of noise or operational signal perturbations. The ability to generalise from real world data sets which are almost never ideal is essential to the successful application of the procedure. An inability to generalise may be caused as a result of under or over training or by a fragmented class set whose boundaries cannot adequately be mapped by a specific network architecture. Over training may cause the network to memorise a set of patterns so that when faced with an unseen data pattern it is incapable of identifying the correct class. The ideal solution is one in which a network has learnt sufficiently to class unseen data correctly but has not reached the stage at which memorisation takes place. Identifying this optimum training phase cut-off point is a non-trivial problem which is, amongst other factors, associated with a specific data set, a training regime and the network architecture itself [44, pp. 148-156]. All these factors must be taken into account when considering the development and application of a neural solution to a classification problem.

4.2 Technical Considerations in Applying a Neural Classifier to the TES Data

Having described some of the general theoretical concepts behind the application of neural classifiers in section 4.1 it is necessary now to consider a few of the practical details surrounding the development of a classification system for acoustical condition monitoring. Essentially what a neural approach is able to offer this particular application is a means of extracting information from a complex acoustic signal with a reduced requirement for the operator to predefine the dimensions of freedom of the signal data under observation. They can, with more recent developments in dedicated hardware, provide a means of applying a fast, hard real time classification system to monitoring applications. With such hardware the replacement of microprocessors and digital signal processors which introduce limitations in terms of fixed serial operation becomes feasible. Advancements of this sort bring with them the potential for operational performance several orders of magnitude faster than currently available processor technology can produce for comparable cost. Such development can only accelerate the applications to which neural classifiers become feasible.

The architecture of neural networks, based as they are on multiple discrete processing elements, require the physical application of the signal data to be performed in a similarly discrete manner. For the application to acoustical condition monitoring this could in its most basic format be carried out by simply applying a series of discrete digital samples from an acoustical signal source. However the development of the TES scheme already described in Chapter 3 enables additional acoustical signal information to be encoded into a series of discrete elements represented by unique codes predefined for a signal type or types. The subsequent matrix conversion techniques developed around this initial TES encoding are able not only to present additional signal data to the classifier but also represent an ideal presentation format for direct application to the input layer of a supervised MLP classifier.

Applying this particular type of network leads to the development of an acoustical monitoring system which through necessity must operate in two distinctly different modes. The primary mode is the training mode in which a series of pre-prepared exemplar data sets acquired from the signal source are applied in succession to a network containing as many input nodes as there are data elements in the exemplars and as many output nodes as there are classes to be identified. In this scenario each of the classes which must be identified by the network must be presented to it during the training phase. For practical reasons several exemplars corresponding to each distinct machine state will be required for acceptable training. Once this stage has been satisfactorily completed and the network has learnt the signal types sufficiently the second phase may be entered into. This is the active classification mode whereby live data samples are acquired from the source, converted using the same TES scheme and then applied to the network for classification in the same way that training sets had been previously. The remainder of this section focuses on the practical considerations in applying such TES data acquired from the gearbox testbed system to basic three layer

MLP networks and some of the difficulties which must be overcome to produce an acceptable monitoring solution given these stipulations.

4.2.1 Safety Aspects of Neural Classifiers Applied to Monitoring Applications

Applying what are essentially non-deterministic classifiers to the field of condition monitoring brings with it a heavy responsibility in terms of the associated safety considerations. Condition monitoring itself may be applied to a particular machine for one of a number of reasons. It may be applied as a means of enhancing performance and economy through the identification of a more ideal set of operating conditions. It may be used to improve the economic viability of machine plant by reducing the cost of overhaul and replacement parts which would normally be part of a predetermined service schedule. It may also be used to reduce the overhead of trained operators on a site. Whatever the reason for its application one fact remains, that the implementation must be capable of being adequately controlled. MLP networks employing supervised training are however by nature essentially non-deterministic. Unlike some of the more conventional means of classification they are trained and not programmed. Training does not require the definition, by an expert, of a series of rules which are tested for and analysed by a fixed control algorithm or series of control sub-modules containing problem definitions. Instead they are trained to develop associations between a given set of vectored inputs and specific outputs. Since this mechanism is autonomous it introduces the possibility of a trained classifier making a decision on a given set of inputs which an expert may, given the same information, feel is inappropriate. Such decisions sometimes termed false positives can be catastrophic. In fact an MLP implementation is able to produce such unexpected false positive decisions with a high degree of certainty on previously unseen data sets, particularly if the training is not directed and monitored adequately.

This type of behaviour can be particularly disconcerting when considering some of the applications to which such classification procedures may be applied. For example one field in which such techniques are beginning to be developed is in the monitoring and automated diagnostics of mechanical subsystems onboard aircraft and helicopters, particularly in the military arena. In such applications the capability of a monitoring system is critical to the safety of personnel. Whilst identifying failure modes too early will cause inconvenience and additional service overheads to be incurred, identifying them too late can lead to terminal failure and in catastrophic circumstances the loss of human life. This is obviously an unacceptable situation.

Concern over crashes, ditchings and precautionary landings in helicopters, both military and commercial, which led to tighter regulations has increased funding for research into enhanced monitoring and diagnostics tools [13, 45]. This has produced health and usage monitoring (HUM) systems capable of improving both the financial viability and safety of such equipment. The majority of these tools are based on more conventional techniques many of which are off-line ground based systems dedicated to post-mortem diagnosis. XMAN and CEMS IV [46], developed by the US military for jet engine diagnostics is a combination of control system, historical database, and knowledge base and typifies the use of advanced monitoring and expert systems for the enhancement of system availability and safety. This particular system however is still heavily reliant upon the capability of the technician for its successful application. If neural techniques are to become as accepted as some of these more conventional methods have in recent times it is essential that the safety issues, particularly the possibility of unconstrained false positive decisions, are confronted and overcome. However the overriding potential advantage offered by neural techniques in terms of the reduced necessity for expert knowledge during application or fault identification still makes them attractive to researchers. Two techniques which have been or could be used to constrain the application of neural techniques are self-adaptive monitoring and trend analysis.

Skitt and Witcomb's [47] use of a data compression network combines the advantage of supervised networks with the reduction in the necessity for separate training and operational phases. They employ a five layer MLP which has its input and output vectors constrained to be identical. This removes the necessity for the implementer to predefine the class associated with each input vector. Instead the network, in this case a 64-16-3-16-64 arrangement of nodes, is used to compress the condition data acquired from a jet engine and represented by a 64 element data input vector into a 3 element encoded representation. Each 3 point vector extracted from the hidden layer is then used to define a virtual point in a three dimensional space. During normal operation the engine traces a trajectory in virtual space which can be used to monitor its condition. After an initial period of usage the normal bounds of the engine within the data space can be ascertained and subsequently used to identify rapid deviations from the norm which may correspond to the introduction of faults. This method of application is able to absorb the smaller trajectory deviations resulting from natural wear as the engine ages without necessarily concluding that a fault has occurred. This type of self-adaptive network is ideally suited to the long term monitoring of equipment being able, as it is, to evolve with a particular machine throughout its lifecycle.

In contrast an MLP network trained in the more conventional manner using pre-prepared class exemplars can also be applied in a manner which is commensurate with robust and safe monitoring by applying a degree of post-decision trend analysis. The trend analysis can consist of a few simple rules being applied to the raw output of a neural classifier to filter the decisions made by the network to ensure that they are mechanically consistent. Rather like human experts neural classifiers are prone to making decisions which are difficult to account for sometimes. It is these mechanically inconsistent so called false positive decisions which can reduce the effectiveness of neural implementations and introduce unacceptable classification errors into critical systems monitoring. However providing these perturbations occur relatively infrequently a post analysis filter is capable of eliminating them and identifying the true mechanical trends taking place. In this way the strengths of a neural classifier can be combined with the more conventional wisdom of an expert-like decision classifier. An example of this type of decision filtering may be a situation in which a classifier identifies a single failure pattern in an engine being monitored within a continuous stream of healthy state patterns. Applying the basic mechanical premise that faults are generally not self rectifying it is reasonable to assume that real faults will exhibit a trend towards an unhealthy status rather than occurring aperiodically. The rules defining this mapping function would of course depend upon the type of failure being identified and the pattern with which this fault is expected to exhibit itself.

4.2.2 Implications of using TES to Precondition the Acoustic Signal

Whilst many aspects of applying neural networks to TES data are the same as applying other types of data there are certain aspects of the TES mechanism which will drive the configuration of the network. This study has employed two possible TES coding strategies each of which have been subjected to two different post coding conversion algorithms in order to provide a primary source of conditioning information for a gearbox testbed monitoring system. The first and most rudimentary transformation is the generation of a statistical code likelihood, or histogram, matrix. In both amplitude and minima conversion guises this corresponds to a matrix containing 300 individual elements for each data vector. The second and more computationally intensive is the generation of the A-matrix data presentation format. Significantly this later technique produces a larger data vector format which contains additional signal information intended to further aid the identification of the gearbox condition. Depending upon whether the initial coding scheme employs amplitude or minima signal components this data matrix will vary in size between 900 and 1600 unique code elements.

Having selected the four combinations of TES coding and data compression it is essential then to identify which, if any, of these schemes is best suited to the task of monitoring the condition of the gearbox. The definition of an ideal solution will require a balance to be sought between each of the constituent elements of a successful scheme from initial signal conversion to the final state identification. Network size, network architecture, the data set size, the information content and the training strategy all play an integral part in this balance. A simple example of this balancing problem is the basic data presentation format. Where the basic 300 element histogram matrix can be generated more rapidly and requires a smaller network for data application the A-matrix is more computationally intensive to generate and requires a significantly larger network for application. Given that the networks used to apply these data formats are fully interconnected then this corresponds to networks containing three to six thousand weights each of which must be calculated for each training epoch or operational decision. The application of A-matrix data requires a more than five fold increase in this number of weight calculations per cycle. Since initial development of the classification mechanism will be performed using neural networks simulated on single processor workstations rather than in hardware this represents a significant increase in the system processor overhead. However should the additional signal data provided by the A-matrix conversion provide a better solution in terms of classifier capability then the necessary network expansion and consequent degradation in response as a result of the more intensive data scheme may be acceptable.

The use of the TES scheme also makes certain specific demands regarding the definition of the code look-up tables necessary for the initial conversion. Whilst neural techniques are ideally suited to noisy data they still require sufficient data to be presented to the network if an adequate classification is to be performed. For the gearbox application described in this work a statistical filtering technique was necessary to reduce the symbol set required for A-matrix generation. This filtering must be performed in such a manner as to minimise the number of unique TES codes in the lookup table. However at the same time it must ensure that the distortion and subsequent noise effects introduced to the input vectors applied to the network to not degrade classification performance significantly. It is important also that this phase of the mechanism does not become reliant upon the need for expert knowledge as this can only reduce the effectiveness of what is otherwise intended to be a semi-autonomous monitoring system.

The training of classifier networks in a TES based system does not differ from those of other applications in as much as the scheme can be considered simply as another means of generating an input vector. However for the scheme to be successful there must be sufficient movement within the data sets themselves and the network must have sufficient degrees of freedom to enable the boundaries to be defined. Of interest during this phase of the application is the means of selecting a training set which is able to best fulfil the requirements of the network in terms of optimised generalisation. The best means of defining any such signal parameters affecting the training phase is a process of practical trials. These will be covered in more detail in Chapters five and six which focus on practical trials.

4.2.3 The Development Classification System

For development purposes the neural network architectures employed with the TES coding scheme were simulated using a SUN Sparc 10 workstation. This type of software simulated implementation not only enables the evaluation of a series of architectures within reasonable time scales but also provides the scope for considerable flexibility in terms of the construction of the basic network. Rather than specifically develop a package to perform the neural implementation for the purposes of this thesis one was chosen from the many which are commercially and freely available. The Aspirin/MIGRAINES package [48], available for a multitude of different hardware platforms was selected for its ease of use and architectural flexibility. It uses a scripting language, Aspirin, to define a networks architecture before generating an executable simulation of that network which can be trained and tested. The language provides flexibility in all the key areas of feed-forward network topologies which determine the capability of the subsequent classifier. These are specifically the:-

i) Number of layers.

ii) Number of nodes in each layer.
iii)Transfer function each node applies to its input stimulus.
iv) Inter-nodal connections between layers.

The data patterns required by the networks themselves can be presented to the simulations in a simple space separated ASCII file format. The data pattern or matrix template files themselves are generated separately as defined previously in Chapter 3 by the PC hosted DSP board coder module. Once the data is generated locally by this PC hosted system the files are transported to the Sparc station for presentation to the network simulations. Aspirin also enables the common training parameters associated with the back propagation algorithm to be controlled via a sequence of command line arguments associated with a particular network and data set. This flexible system provides a fast, simple and robust method of developing an understanding of the demands of applying TES data to automated condition monitoring.

MIGRAINES is a utility available with the Aspirin package which itself provides an interface which can be used for interacting and evaluating a particular network configuration both during and after training. It enables a user to study the network internally rather than simply treating it as a black box implementation. This can, for instance, be useful in identifying the evolution of the weights and biases within the network which can provide valuable information as to the means by which a network extracts information from the input layer for classification. Such information may be used for tasks such as pruning whereby individual nodes or connections can be selected for removal if they are deemed to contribute little to the classification process.

The Sparc 10 platform is capable of providing adequate computational power to perform studies into the various architectural configurations of a TES implementation when used in conjunction with the Aspirin/MIGRAINES package. The vast burden of the computational effort is obviously required during the training phase when the demands of the back-propagation of the error contributions from individual nodes have to met. During the trials performed the training phase varied from a few seconds to a few hours depending upon the demands placed upon the network and the configuration used. Once the network had been trained the time taken to classify an unseen data set was of the order of a few milliseconds per input vector. Thus even without specific hardware acceleration this level of classification response, both during training and evaluation, makes the workstation simulation solution acceptable.

4.2.4 Potential for Hardware Optimisation of Networks

Until relatively recently the full benefit offered by the massively parallel nature of neural network implementations has been difficult to achieve. Most developments in the field of neural computation have been performed by simulating the architectures on serial processor based machines as indeed has been done with the TES based system considered in this work. The main problem with the development of dedicated VLSI hardware devices has always been the need for very large numbers of interconnections between individual processing elements. These often amount to some fraction of the square of the number of nodes and has thus limited the potential size of the networks. However the development of a range of optical based [49] as well as VLSI [50] implementations now present the possibility of achieving the types of training and response speeds which until now had been unattainable with serial processor systems.

Many of the major semiconductor companies now have neural hardware implementations available for a range of network types. For example Bell Laboratories developed the Analogue Neural Network Arithmetic chip (ANNA) which supports a range of optical character recognition (OCR) algorithms. Intel have developed the Electrically Trainable Analogue Neural Network (ETANN) which supports a range of network architectures including back-propagation. These are just two examples of a wide range of products which are becoming available from these manufacturers to meet the increasing demand from developers for use in a diverse range of products. No doubt the increasing availability of such hardware will bring with it further applications which until now would not have been feasible due to excessive training requirements. In some cases applications which could require training phases stretching into weeks or years on conventional serial processor systems could conceivably be performed in real-time with hardware assistance.

4.3 Architectural Considerations of Practical MLP Networks

Up to now the architectural considerations of a specific TES focused neural implementation have not been covered in detail. The ability of an MLP network to approximate an arbitrary non-linear mapping given an adequate period of training and reasonable data set is not in question. However there are a number of practical concerns which must be considered if the network is to be optimised for any specific application. The first is the matter of selecting a physical network configuration. As we have already said the number of individual layers within the network, the number of elements within each layer and the interconnections between layers all play a part in defining the characteristics of a network. Selecting a suitable size and construction is essential if key areas of concern are to be addressed.

The first of these is the question of training. The important aspects here are the networks ability to learn the classification space and the time taken to do so. Once trained it is also important that the network is able to generalise from the knowledge it has acquired and classify accurately data from outside of the training set itself. If the network is too small it will be incapable of forming a sufficiently accurate model of the problem whilst if it becomes too large it can become over capable. An over capable network is able to implement numerous solutions, all of which fit the training data but which may ultimately be poor approximations to previously unseen data. Architectural optimisation consists of seeking the solution which best balances these various requirements for each specific application. The remainder of this section covers in more detail the selection of the architectural parameters of networks used in conjunction with TES data during practical trials.

4.3.1 Selection of a Suitable Basic Network Architecture

The first stage in defining the architecture of an MLP network is the selection of the basic organisation of the nodal layers within the structure. This definition refers specifically to the separation of the individual nodes into a series of ordered processing layers. Cybenko [51] states that architectures containing two hidden layers are sufficient to model any function with arbitrary accuracy. In general however, one hidden layer is sufficient for most practical applications [52, 53]. Consequently despite the statement made by Fausett [page 324, 54] that in certain situations a second layer may improve the networks general training capability most applications use only a single hidden layer. In the interests of reduced complexity and computational overhead, the networks to which TES data has been applied in this work have been limited to these more basic single hidden layer implementations.

Once the basic configuration of the network has been defined the number of nodes within each layer must be carefully selected. The best method for estimating this optimum network size and configuration remains practical trial and error. This entails starting with a reasonable definition and evaluating the performance at each stage of an optimisation process. A few simple guidelines exist for the definition of this architectural starting point prior to practical evaluation [54, pp. 298, 55]. Most are related to the number of data patterns in the data set used to train the network. Widrow [56] states that the number of training samples required is approximately ten times the number of weights in the network. Baum and Hausler [57] provide a more theoretical insight into the determination of this pattern and data set size relationship. The subsequent process of network tuning may be achieved either by starting with a minimal sized network and gradually increasing the number of nodes or by starting with a large network and pruning nodes which contribute little to the decision process.

In the case of TES data networks the definition of the basic network is driven by the data itself. The two signal conversion strategies impose certain predefined limitations on the flexibility. Whilst the simple histogram matrices consist of 300 individual data items the A-matrices contain between 900 and 1600 unique elements depending upon the conversion strategy employed. These data matrices therefore fix the size of the input stage of the network. In a similar manner the output stage is defined by the number of different states which must be identified. Since each output node corresponds to a unique state the number of nodes is again fixed for each application. The hidden layer is a little more difficult to define. This layer is not bounded in the same way as the input or output layers but is instead dependant upon the complexity of the data space which the network is expected to classify. It is this layer which provides the flexibility necessary to define the class boundaries within the data space. There will be a lower limit on the size of this layer as a result of the location of each of the class boundaries but no fixed upper bound. However as this layer increases in size so the number of nodal calculations for each data presentation increases and corresponding network response is reduced. In addition an oversized network can become over capable which itself will impair the classification performance.

The practical selection and evaluation of network architectures suitable for the varying TES data matrix types is discussed in sections 5.2.2 and 6.6. Section 5.2.2 covers the evaluation of the hidden layer configuration in an histogram matrix system. Section 6.6 approaches the problem from the opposing direction and considers the configuration in terms of the data space applied to a fixed network architecture.

4.3.2 Implications of Reduced Nodal Connectivity

Many MLP's are used in their most basic configuration in which each and every node in adjacent layers is connected together. However there is no reason to suggest that all these interconnections are indeed necessary for a specific data class boundary definition. The concept of reducing the number of interconnections within the network is an attractive one [50, pp. 30-36]. It not only reduces complexity and the associated computational overhead but can provide a means of subdividing the data classes and improving separation. In some ways the subdivision of the network through the removal of nodal interconnections can be compared to an expert system in which several knowledge sources, or experts, exist to provide an integrated solution. In the case of a neural classifier this subdivision can be achieved by separating one large network with a single classification space up into several smaller knowledge expert like subnets each of which is tasked with a subset of the space. The conflicting demands placed upon the single classifier architecture to separate each of the classes during the training and classification phase can thus be reduced through the removal of certain connections which subdivide the network and thereby the classification space itself. In addition to this the smaller network so created requires fewer weights so reducing the computational demands both during training and operation.

The manner in which the connections are removed can have a profound effect on the way in which the data is to be applied and the way in which the network performs as a result. For example a partially interconnected network similar to the example in Figure 4.2 which subdivides the interconnections at the hidden to output layer interface requires little or no change in the actual manner of application to that of a fully interconnected network. Such a variation is likely to affect the back-propagation procedure modifying the time required to converge as a result of the change in classifier demands placed on the reduced connectivity network.

In contrast a network similar to that in Figure 4.3 will require consideration to be made for the allocation of class nodes to input interconnections. This is because the input to hidden layer interface has been changed. As a result the respective output nodes only have a partial view of the input data. This will have a profound effect upon the manner

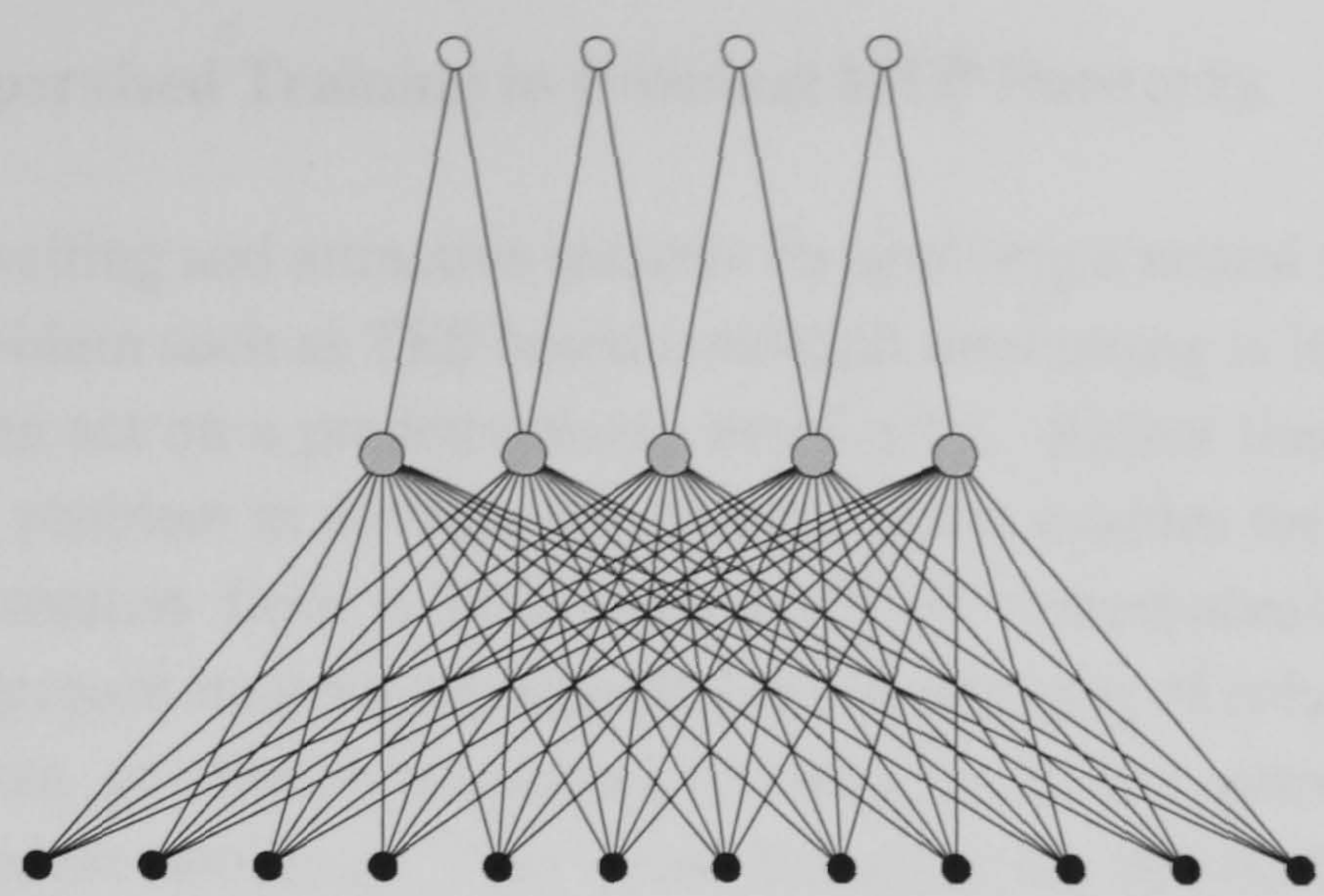

Figure 4-2 A three layer network with partial hidden-to-output layer connections

in which output nodes are assigned. Both techniques are viable but the choice is dependant upon the specific application and the way in which the raw input data is related to the classes to be assigned by the network. In the case of the TES application which is the main focus of this work the latter variation is less attractive because it requires substantially more to be known about the interaction between the mechanical gearbox faults and the corresponding acoustical effects. In contrast the pruning of connections between hidden and output layer affords more room for manoeuvre in terms of data application and thus knowledge of the acoustical effects of fault conditions is less critical.

The transition between a single large all encompassing all-class-in-one-network (ACON) and the other extreme, a series of grouped one-class-in-one-networks (OCON) in performance terms is not well defined. There are several factors which can effect the relative performances of these various network configurations foremost amongst which is the data which is presented to the networks and the level to which the classes are subdivided by reducing network interconnectivity. The practical evaluation of partially interconnected networks of the type defined in Figure 4.2 is covered in section 6.7. In this section the conclusions from a series of trials using amplitude A-matrix data as the input medium are discussed.

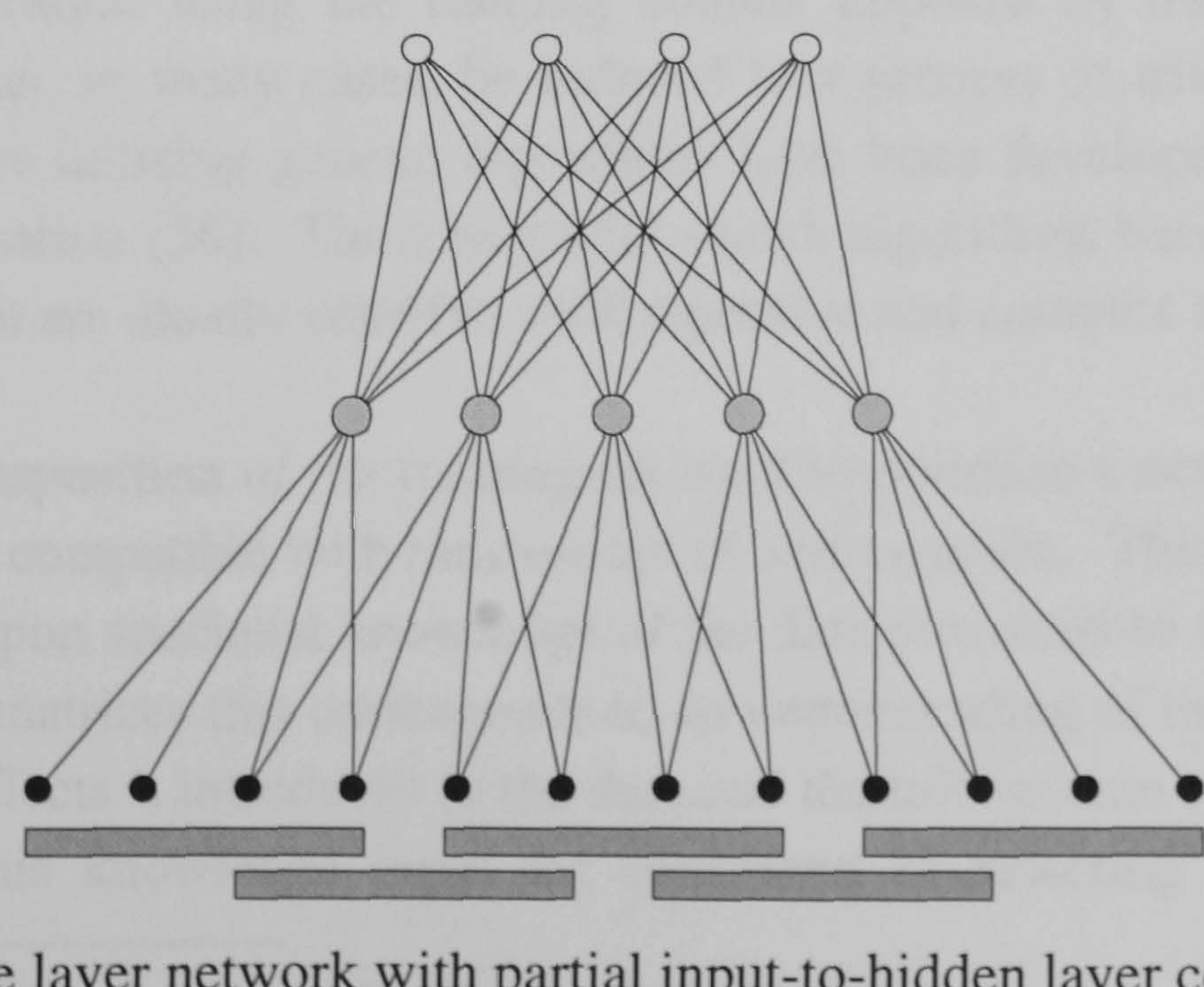

Figure 4-3 A three layer network with partial input-to-hidden layer connections

4.4 Applying Supervised Training to Practical MLP Networks

One of the most exciting and attractive reasons for applying a neural network solution to a classification problem such as TES based condition monitoring is its ability to learn by example rather than act on a predetermined set of rules. Rather than having to specify every detail of the problem in advance the training phase enables the network to extract the relevant information from a precompiled set of representative data examplars. Whilst a neural approach may not necessarily provide the type of solution afforded by an optimised algorithm or ruleset developed by experts it can provide a manageable solution of reasonable accuracy. The more complex the application the more cost effective and viable the neural approach becomes. In the case of a feed-forward multilayer perceptron architecture the effectiveness of the solution is very much dependant upon the training phase of the application. The regimes used to train the network are as important as the architectural considerations used to define its physical arrangement in the first place. The control of this regime can be separated into two clearly defined areas. The first is the physical requirements of the back-propagation algorithm used to control the weight tuning in the network (Eqs.1,2) and the second is the composition of the training set presented to the network during the weight tuning phase.

$$w_{ji}(t+1) = w_{ji} + \Delta w_{ji}(t+1) \qquad (1)$$

Where w_{ji} is the weight for the j^{th} node at time $(t+1)$. The change in weight is defined as:-

$$\Delta w_{ji} = \alpha\delta_j(t+1)o_i(t+1) + \beta\Delta w_{ji}(t) \qquad (2)$$

α and β are the learning rate and momentum parameters respectively[1].

It is the second of these two facets which presents the greatest challenge to the practical application of neural techniques to TES based condition monitoring. With the ready availability of powerful computational engines to perform training the optimisation of a particular configuration using the training bounds imposed by the back-propagation algorithm itself can, in many cases, be reduced to a process of trial and error. More recently techniques utilising genetic algorithms have been developed to automate this process of optimisation [58]. These powerful search algorithms based on a mechanism of natural selection are ideally suited to such repetitive and complex tasks.

In contrast the composition of the training set used to optimise a network for a specific application is less compatible with the concept of automisation. This aspect is generally more dependant upon specialist knowledge of the data presented to the network. In the case of TES data matrices this corresponds to an understanding of the signal conversion mechanism, the effects it introduces to the data and the information contained within the data. Without this knowledge input the likelihood of selecting a sufficient set of

[1] For further reference the reader is asked to refer to Appendix B.

relevant exemplars and training the network most effectively is reduced. The practical discussion of the selection of training data and the number of exemplars required in a typical TES application are discussed in sections 5.2.3 and 5.2.4. In sections 4.4.1 and 4.4.2 the theoretical discussion of this network optimisation is covered in more depth.

4.4.1 Optimisation of the Back-propagation Mechanism in a TES System

The back propagation algorithm requires three learning related parameters to be defined which control the process of data presentation, error calculation and weight update. The learning rate, α, limits the magnitude of the weight change for each nodal interconnection for each iteration of the back-propagation algorithm. This in turn determines both the speed of convergence and the network state at convergence. A series of practical trials was performed using live acoustically derived TES data to study the effects this parameter had on the convergence times and performance of several network configurations and selected data sets. Figure 4.4 illustrates the rate at which convergence is achieved for different training rates. The training sets contained data matrices generated using a basic amplitude TES histogram conversion scheme which was then applied to two different network architectures, one with 20 nodes in the hidden layer and the other with 8 nodes. Given the two fixed data sets and network configurations it illustrates the growth trend in the number of iterations required for convergence as the learning rate parameter is incrementally raised. Figure 4.5 demonstrates that despite the extended learning period induced by the increase in magnitude of the learning rate parameter there is little variation in the subsequent performance of the network when tested against previously unseen data.

This suggests that the learning rate should be kept as low as possible to optimise the network solution. However whilst a reduced learning rate provides enhanced learning resolution and reduces oscillation during gradient descent learning it does generally

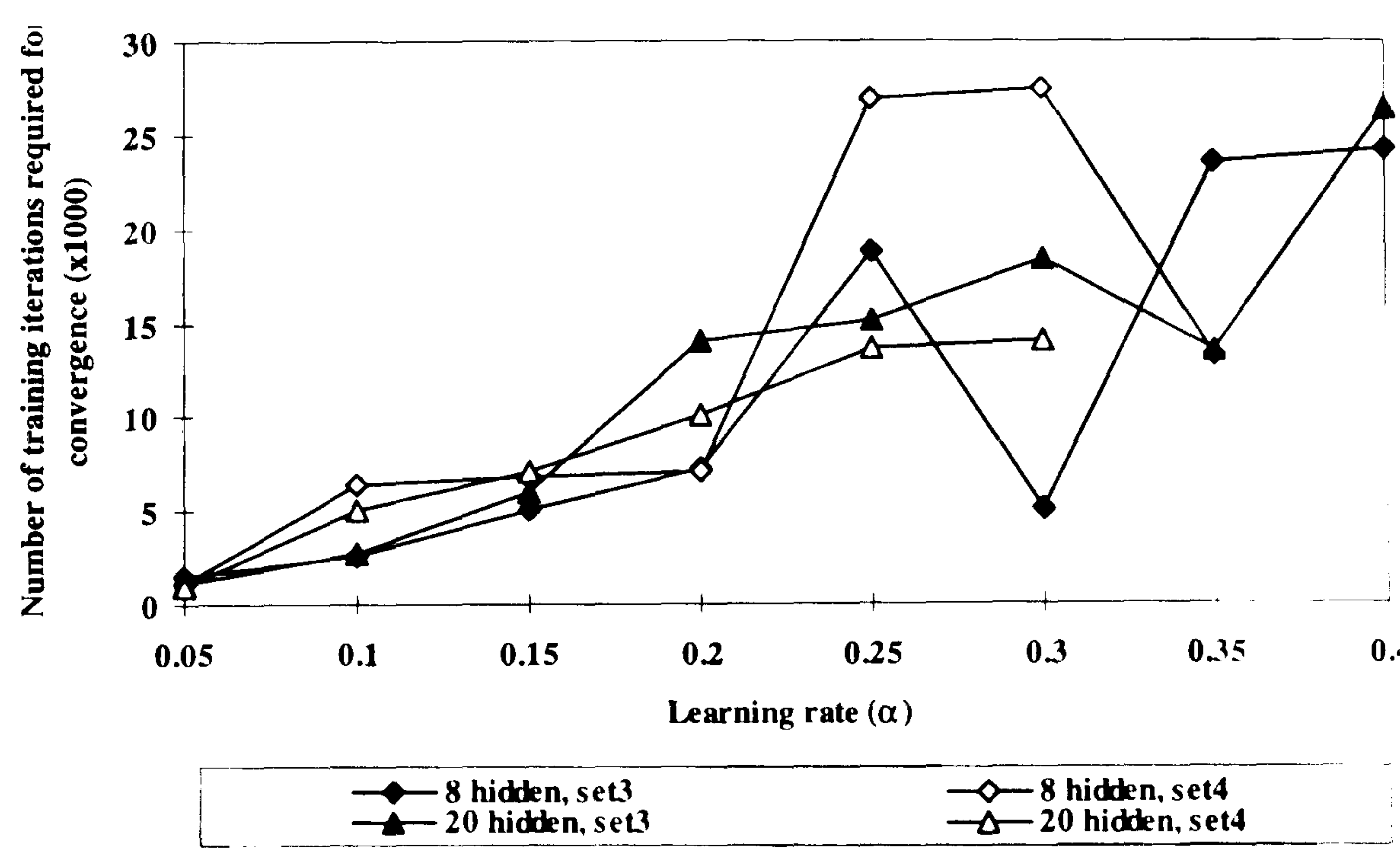

Figure 4-4 The effect of learning rates on network convergence for two different data sets and two network configurations

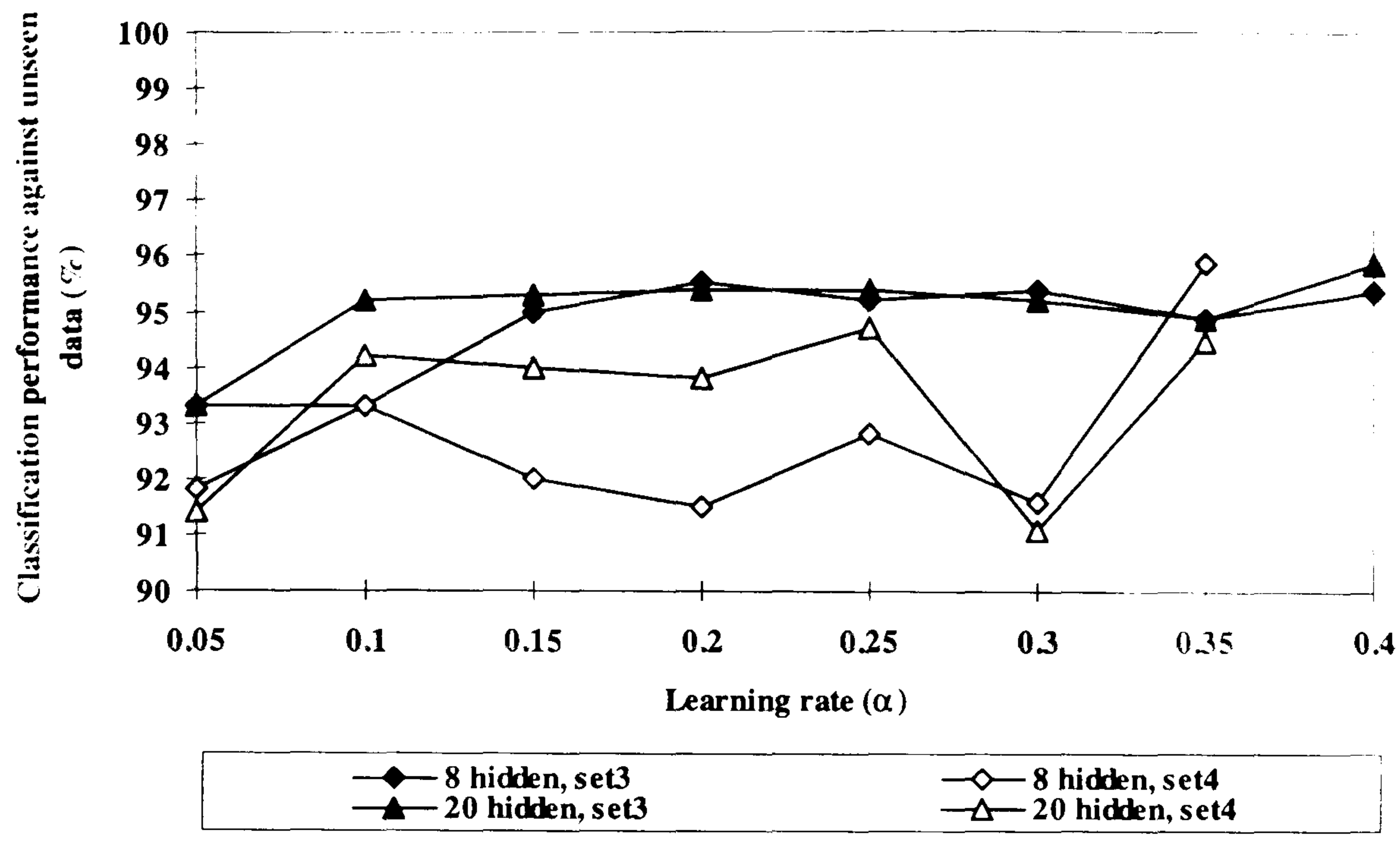

Figure 4-5 The classification performance of the networks trained using different learning rates

extend the time to converge. What is required is a means of accelerating the learning process whilst at the same time retaining the advantage accrued by a smaller learning rate. To satisfy both of these two conflicting demands a momentum factor, β, can be introduced which provides the necessary acceleration whilst retaining the fine-grained weight update resolution of a lower training rate. The β factor is used to tune the weight update algorithm throughout training depending upon the gradient at each update cycle. If the gradient of the combined network entropy is increased then the momentum factor acts to increase the step size, whilst a reduction in the gradient reduces the weight update momentum. This not only improves the speed of convergence towards a global minima but also ensures that weight oscillation about the minima caused by large step sizes is minimised as a global minima is approached. For the network training illustrated in Figure 4.4 a momentum factor of 0.95 was employed which significantly improves the convergence performance of the network during the training phase. The final training control parameter is the error bound term. This controls the point at which the network is deemed to have converged, where all output node errors are within the error bound for all the training set patterns. This term is somewhat dependent upon the activation function implemented in each of the nodes. With the sigmoidal function used in the TES networks requiring binary type target outputs values close to 0 and 1 were employed. The sigmoidal function itself is constrained to these bounds so outputs of 0 and 1 would require weights approaching infinity.

In association with these algorithm control parameters the frequency with which the corresponding weight updates are performed during training can also affect the final performance. Two types of update mechanism are commonly employed. These are respectively termed block adaptive and data adaptive weight revision. When a block adaptive mode is employed the weights in the network are updated after each complete epoch of the training data set. In contrast a data adaptive strategy updates the weights after each and every exemplar presentation and back-propagation to the network. In terms of the relative performance of the two techniques the block adaptive method is generally more robust as a result of the cumulative averaging influence exacted upon the error terms over each full data epoch. It does however impose additional overheads in processing and data storage capacity during training. In contrast the data adaptive method is generally more sensitive to noise effects on individual data patterns and as a result is generally more appealing for on-line applications. This technique also provides a more effective means of adapting to the local gradient and does not require additional data averaging or storage. It is this technique which has been employed for all practical trials involving the TES data acquired from the gearbox testbed system for fault analysis.

4.4.2 Optimisation of the TES Sourced Training Data

Whilst there are many facets to the development of a good neural condition classifier the one area which can make or break the viability is the raw data used to train the network. Whilst other areas can improve the speed of training or accuracy of performance within reasonable bounds this raw data forms the backbone of the problem definition to the network. Without sufficient initial data space definition no network will be capable of acceptable state separation. Data definition, in this context, not only refers to the basic presentation medium but also of the manner in which this information is presented to a network during the training phase.

The specific application of an amplitude based TES coding scheme to generate the primary data from an acoustic sensor prior to network application necessitates particular attention for this very reason. Intuitively this technique is likely to be susceptible to variations in the signal level of the source based as it is upon the amplitude of the individual signal components. Changes in the acquired signal level would be expected to cause fluctuations in the perceived dynamic range of the source which are likely to manifest themselves as variations in the magnitudes of individual elements within the data matrices. This is due to the predetermined nature of the TES code table responsible for the allocation of codes based upon the features of individual signal components.

The practical evaluation of the effects of such fluctuations on the performance of any neural classifier to which the data is subsequently applied was carried out by artificially varying the signal level at the conversion stage. In trials with two network configurations, one a fully interconnected three layer network with 8 nodes in the hidden layer and the other a similar network with 20 nodes in the hidden layer the disparity in performance was clear. When the TES training data was generated from a signal with a reduced dynamic range the classification performance on unseen data was significantly impaired. Compared to the correct classification of 94% of all test data with TES data generated using an optimised dynamic range the reduced dynamic range caused performance to fall to approximately 78% against the same unseen data. This reduction is caused by the "blurring" of the energy boundaries which define the amplitude codeset used for signal conversion. This results in incorrect allocation of TES codes during the conversion and subsequently manifests itself in the matrix data generated from the code stream. The reduced quality of the signal information contained in the matrices and applied to the network produces a corresponding reduction in interclass resolution.

It is clear from these trials that the application of an amplitude based TES scheme necessitates the addition of a suitable signal conditioning stage prior to conversion to minimise these adverse effects. There are further discussions of these dynamic signal effects in Chapter 5 which are based upon further practical trials and are presented in sections 5.2.1 and 5.2.5.

4.4.3 Training and Testing Data Conventions

Practical trials which were performed to evaluate the effectiveness of applying simple MLP networks to classifying gearbox condition states are presented in Chapters five and six. All these trials were performed under strict control so as to provide a reasonable opportunity for direct comparison of the differing techniques. In order to maintain this realistic perspective on the practicality of such techniques to the industrial arena certain rules were applied to the selection of data for performance evaluation of the differing architectures and coding schemes. These rules were similar in principle to those outlined by many other neural researchers regarding the data applied to each network. Essentially there must be a clear distinction between the data used to train a network and that used to evaluate its performance. In many situations the distinction is made at the most basic level. That is providing that the data used to evaluate a network has not previously been applied during the learning phase it can be used for performance evaluation.

However for the purposes of the practical trials in this work the distinction between the two data sets was separated still further. Rather than simply using data which may have been acquired within the same physical recording but not used for the training phase the distinction was made that training and testing data must be sourced from separate physical recordings. This added distinction was made so that evaluation could be performed under more "realistic" conditions. If the technique were applied in a practical situation this same physical distinction could not be avoided. Small perturbations in the local environment and possibly in the daily operating conditions encountered could be expected without a change in physical state necessarily being introduced. In some cases this resulted in training and testing data being recorded on different days. As far as was possible no special consideration was made to eliminate natural environmental variations in the day to day acquisition of the data used for practical trials.

4.4.4 Enhancement of the Network Training Phase

Whilst the back-propagation algorithm currently provides the most common means of tuning the weights of practical neural classifiers it is inherently slow to converge. It has already been shown that the rate at which the weights are modified affects the network weight solution attained and the time required to reach this solution. Many researchers have invested time in developing techniques intended to improve this phase of the application. Most result in modified training algorithms which seek to dynamically modify the way in which the network learns. Some, like the work of Hossein *et al* [59], have studied the effects upon the convergence times of increasing the magnitude of the initial weights. This technique is complicated by the average fan-in of units and the specific data set being applied which can force nodes into the saturation regions of their activation functions. In such situations the training is degraded often producing less acceptable solutions with little or no improvement in convergence times.

More often than not the element which provides the greatest headroom for optimisation is the dynamic modification of the learning rate used in the back-propagation algorithm [60] as well as the particular method of error back-propagation [61]. However in the final analysis the time taken to train a specific network application is only one of a number of factors to be considered. In many cases it is the least important of the factors affecting the network definition. Alpsan *et al* [62] make a particularly pertinent comment regarding any such efforts expended upon improving the speed of network convergence which is relevant not only our particular TES application but also to many general applications. This is that whilst there are many techniques now available to optimise networks for speed of training most if not all are flawed in terms of the resultant performance they provide over routines which are not optimised simply for speed. In the specific application to condition monitoring the rate at which a network can be trained is most definitely of secondary importance to its raw classification capability. Network generalisation which is directly linked to this raw performance is often best achieved by removing complex dynamic learning modifications from the basic back-propagation algorithm and resorting to lengthier training runs. With the capability of modern computational processors developing so rapidly the extended timing overheads imposed by these non-optimised techniques will be reduced significantly, providing better training without necessarily imposing unreasonable training demands.

4.5 Chapter Summary

This chapter has dealt with the fundamental considerations of the application of artificial neural techniques to automated acoustic monitoring of a testbed gearbox system using TES. Theoretically at least the networks themselves should be capable of providing a means of reducing the role of skilled operators in this process. This potential is imparted by the ability of networks to identify patterns within presented data without the need for specific rules or guidance to be provided during the decision making process. The selection specifically of a multilayer perceptron implementation from amongst the wide range of neural networks which have been developed since the late 1950's has been discussed. This simple network architecture brings with it certain characteristics which require careful consideration in terms of safety. These characteristics are all the more relevant considering the intended application. This aspect of the implementation has been discussed together with some potential solutions in section 4.2.1.

The selection of a software simulation package to evaluate the application of these network techniques has been discussed together with a brief description of the flexibility that the chosen package provides both in terms of data and network configurability. For the purposes of this particular research such non-optimised software implementations provide a good balance between performance and flexibility. The potential for future enhancement of the applications through the development of hardware implementations is briefly discussed together with some examples of the type of packages which are currently commercially available for this purpose.

The main focus of the Chapter however is upon the specific problems which are associated with the MLP implementation selected. Clearly the configuration of the network in terms of the number of layers and the number of individual processing elements within these layers will affect the performance. The selection of number and size of each of these layers is discussed together with some of the general rules used by other researchers to identify good starting points from which to iteratively optimise the architecture for specific classification requirements. As a result of the discussions made networks containing a single hidden layer are thought to provide a good starting point for further evaluation. Within this selected architectural model a hidden layer size of between 10-30 nodes would seem to provide adequate performance. The problems associated with pruning the links between individual layers within the networks is also considered.

Having discussed the network itself the next stage in the application which has been evaluated is the training phase. The quantity, type and selection of data have all been singled out as areas which require practical evaluation. A few of the potential pitfalls which may be encountered at this stage are also discussed, particularly the problems associated with optimising the generalisation and minimising the memorisation capability of networks as a result of architecture or training schemes.

Two other important aspects of the implementation have been considered one of which has been discussed in association with some results acquired during network evaluation. The first is related to the importance of data set selection. This is discussed with reference to the expected impact that specific TES data presentation mechanisms will have upon the subsequent network capability as well as the additional effects of template selection within a presentation type. The second aspect of the application procedure to be selected for evaluation is the modification of parameters used to direct the network training phase during which data exemplars are presented to the network. Since these parameters are used to update the interconnecting weights which impart the learning ability to the network they were expected to impact upon the manner in which the back-propagation algorithm traverses the weight space for a specific classification space. During a series of simple evaluations the effects upon classification of varying both the learning rate, α, and the momentum factor, β, were studied. The contrasting needs of an optimum step resolution and enhanced weight velocity have been considered. These early trials highlighted the necessity for a low α value and high β value to optimise the performance.

Whilst this Chapter has provided a brief discussion of the key elements of applying TES data matrices to neural networks for the purposes of monitoring the mechanical state of a testbed gearbox system it has also highlighted the potential weaknesses. The feasibility of providing adequate training for the networks must yet be more fully evaluated whilst the safety considerations must not be overlooked. Despite these factors it is clear that the potential does exist for enhancing the monitoring of complex devices through the application of simple MLP networks.

Chapter 5

Page

5. Practical Trials of TES Based Monitoring Carried out on a Gearbox Testbed

In application terms the most important question which must be answered is whether TES really can provide the condition monitoring fraternity with a viable alternative to higher cost, knowledge intensive system state classification. Already it is clear that if it is to be effective as the primary signal preconditioning component in an integrated neural based monitoring system, as outlined earlier in Chapter 3, it must be capable of retaining sufficient signal data within the symbols generated. This is, of course, the primary prerequisite of any signal conversion mechanism employed for the purposes of monitoring systems not simply TES. Assuming that the prerequisite can be met by TES, consideration must then be made for the subsequent classification strategies this data is applied to. As has been detailed in Chapter 4 all the strategies which are to be considered within this thesis centre upon the application of neural classification techniques. The ultimate goal being to ascertain whether a system utilising this core TES acoustic data to drive a neural classifier is capable of deriving physical information about the target. To provide a viable alternative in condition monitoring applications this must be accomplished with a minimum of effort and in a manner which does not require highly trained operators. Both this Chapter and the next are devoted to the discussion of a series of practical trials performed using the gearbox testbed system to evaluate this capability.

The focus of the trials is upon the different elements within the classification process which potentially affect overall performance. Specific areas here include the interaction between the neural classifier and the raw TES data as well as the architectural considerations of the neural classifiers themselves. The intention of the trials being to attempt to identify key elements of an optimal system and to identify specific limitations arising as a result of these configurations. At this stage of the investigation it is entirely appropriate to perform the trials on a representative, or simplified, system rather than on a production equivalent. The added complexity associated with a production based system would only serve to increase the potential for ambiguity in the ensuing comparative examinations of particular configurations. The intention, at this time, is to identify whether or not the principle of applying such methods is acceptable and whether they are likely to be considered sufficiently accurate or financially acceptable. In doing this it was important to attempt to recreate, in the representative system, a range of illustrative faults so that the techniques studied may reasonably be expected to carry over into a more complex production based system at a later date.

A simple mechanically configurable gearbox testbed system was developed as the representative system for the trials. Its aim was not to replicate specific faults which occur in a particular gearbox system but be capable of mimicking a set of illustrative faults which may be controlled within a test environment. Primarily then the key to judging the effectiveness of the TES techniques is to measure the accuracy with which each of these illustrative faults can be identified and the simplicity with which the TES data can be presented to the neural networks to achieve this. Another question which arises in terms of the potential future adaptability of the technique is the degradation in performance associated with an unconstrained acoustic environment. Most practical situations in which systems would be expected to operate are noisy. As such, the classification engine will almost certainly be subjected to a diverse range of both internal and external sources of noise. Internally the conversion of the acquired signal itself will introduce distortion whilst externally the surroundings in which the target system is situated will typically consist of other additional acoustic sources. The susceptibility of the classification mechanism to such additive acoustic noise is important. Attempts to quantify these effects are an essential part of the evaluation process. Clearly a practical classification engine should be capable of guaranteeing sufficient accuracy whilst operating in the presence of reasonable levels of additive background noise.

The employment of neural classification to the TES data also brings with it further constraints which must be considered. Since the type of network proposed in this work requires a training sequence, the demands of the training program required to achieve adequate performance must be examined. If training of the classification system is dependant upon a rigid data collection and selection mechanism then cost savings made through the reduced requirement for an operator may become absorbed by the requirement for highly skilled personnel to oversee the system training program. For this reason the training program is a primary target for performance evaluation. Ideally a balance should be sought between the need for skilled personnel and the necessity to minimise the time required to complete a training sequence. Ideally this would consist of a semi-automated system which required only periodic attention by a operator.

The final consideration in terms of the effectiveness of the technique as a whole is the expected response time. For the most part this is academic as far as the operational phase of the classification is concerned. With the rapid development of computer technology it is unlikely that under operating conditions the response of a reasonably sized neural network could become an overriding issue. However this is not necessarily the case as far as the training phase is concerned. The mathematically intensive nature of a neural network which relies on a separate training program inevitably brings with it the need for a powerful mathematical engine to perform the training within reasonable time scales. Since network training times are directly proportional to the volume of data and network complexity required there is a basic need for rationalisation in both of these areas. It is important therefore to establish at an early stage in trials the architectural and data requirements for an operationally adequate classification system.

The remainder of the Chapter contains a description of the basic test system used to investigate the areas identified so far as well as the application of the more basic TES data types to the definition of its fault states. The first section covers the technical description of the testbed system used throughout the trails, outlining the various fault configurations which can be introduced. The remaining sections focus on the application of the simple 300 element amplitude and minima symbol histogram matrices which were defined earlier in section 3.5.1 of Chapter 3. Application of the more complex A-matrix data along with discussion of some of the more demanding fault states will be covered in Chapter 6.

5.1 Gearbox Testbed System

The selection of a gearbox fault simulator for the purposes of analysing a TES based conditioning system was made because of their universal use in mechanical systems. Wherever there is a need to alter the rotational velocity between input and output drives a gearbox may be used. They are most commonly used in situations requiring the transfer of high loads between shafts where friction dependant systems would prove inadequate. The test platform which was developed for the monitoring trials, detailed in this and the next Chapter, is intended to be a much simplified version of these common types of load transfer gearbox. Its reduced complexity enables the rapid and secure acquisition of acoustic data essential for trials by minimising the number of degrees of mechanical freedom which must be monitored during recordings. This in turn focuses the attention of the trials on the monitoring techniques and the affects of specific variations in data presentation rather than the mechanical stability of the gearbox.

As well as being designed for the necessary mechanical simplicity two other key requirements were placed on the testbed design. Firstly, the ease of access to the internal parts of the system was required to facilitate direct acoustic coupling and thus more easily perform measurements of acoustic sensor sensitivity. Second and most importantly the system had to be mechanically configurable to provide the means to carry out trials on the various classification techniques. This configurability needed to provide sufficient movement of internal parts to mimic several faults which commonly occur in generalised gearbox systems. Figure 5.1 shows diagramatically the unit which was used.

Having described some of these basic design concepts it is necessary to become acquainted with the physical implementation of these requirements in so far as this

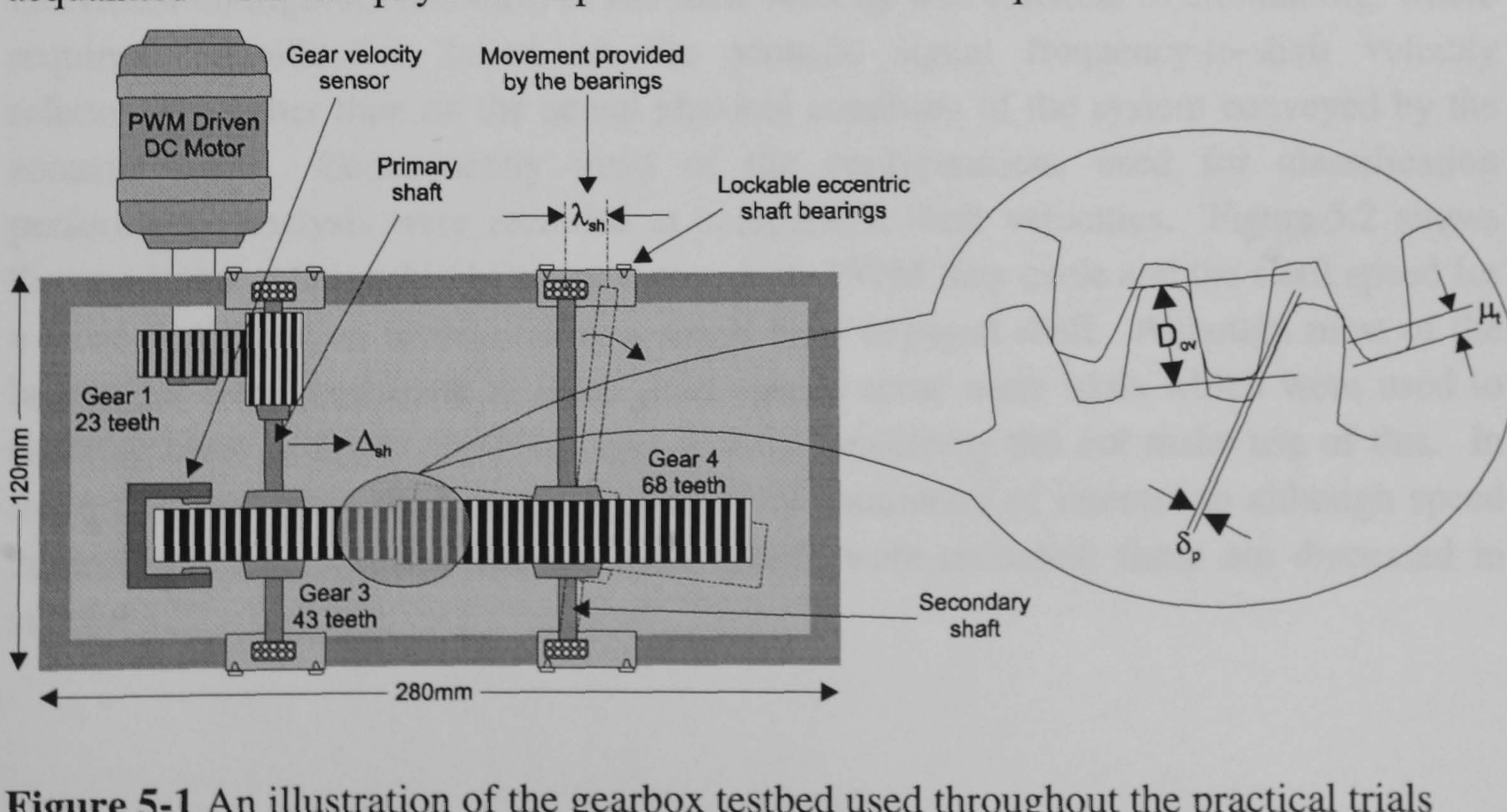

Figure 5-1 An illustration of the gearbox testbed used throughout the practical trials

affects the acoustical properties of the emissions. From Figure 5.1 it will be noted that the unit consisted of a solid metal plate box casing with an exposed top which provides a direct path to maximise acoustic coupling from the source to the microphone sensor used to record the emissions. The box contains two gear shafts, the first being an intermediate load transfer shaft and the second a final drive shaft. The unit is powered by an externally mounted electric motor which drives the intermediate shaft via a gear wheel with 23 teeth at speeds up to 4000rpm. The motor is driven using a pulse modulated (PWM) source which provides the speed control necessary during data acquisition. The rotational velocity is directly measured using a photo-detector cell arrangement housed in the gearbox, the LED source and detector units being mounted on opposing sides of the driven gear on the first shaft. The coupling between source and detector is through a series of holes drilled through the gear wheel. Velocity measurements are made by converting the LED detector signal frequency into a corresponding rotational velocity using knowledge about the number of holes located in the gear wheel. The PWM motor drive signal is derived from a digital signal processing board which drives a power amplifier circuit powering the motor. This same DSP simultaneously performs the shaft velocity measurement described previously by decoding the LED signal.

The motor itself drives the first shaft with a step up ratio of 2.6:1, which in turn drives the second with a step down ratio of 0.6:1. Both of these shafts are secured into the external casing with screw adjustable eccentric bearings at each end. These are able to provide each shaft with 3° of freedom relative to the shaft axis permiting a range of simple alignment like defects to be simulated between the primary drive and output shafts. In all configurations the relationship between actual and simulated faults is important in terms of the confidence level which can be placed on any conclusions obtained regarding TES conversion and neural classification of the mechanical processes involved.

Since the speed is not independent of but dependant upon the driven load of each individual configuration control of the shaft velocity was essential to eliminating, where required, classification based on the acoustic signal frequency-to-shaft velocity relationship rather than on the actual physical condition of the system conveyed by the acoustic data. Consequently most of the configurations used for classification performance analysis were recorded at comparable shaft velocities. Figure 5.2 shows the non-linear relationship between motor drive PWM duty cycle and the shaft speed for a fixed configuration incorporating a single fully engaged shaft. Although most of the later trials were performed at fixed shaft speeds some early trials which were used to estimate basic network capability and acoustic sensitivity did not make use of this. In these first few trials shaft speed itself was the parameter of interest so although speed control was still required several shaft speeds were recorded; these are discussed in section 5.2.

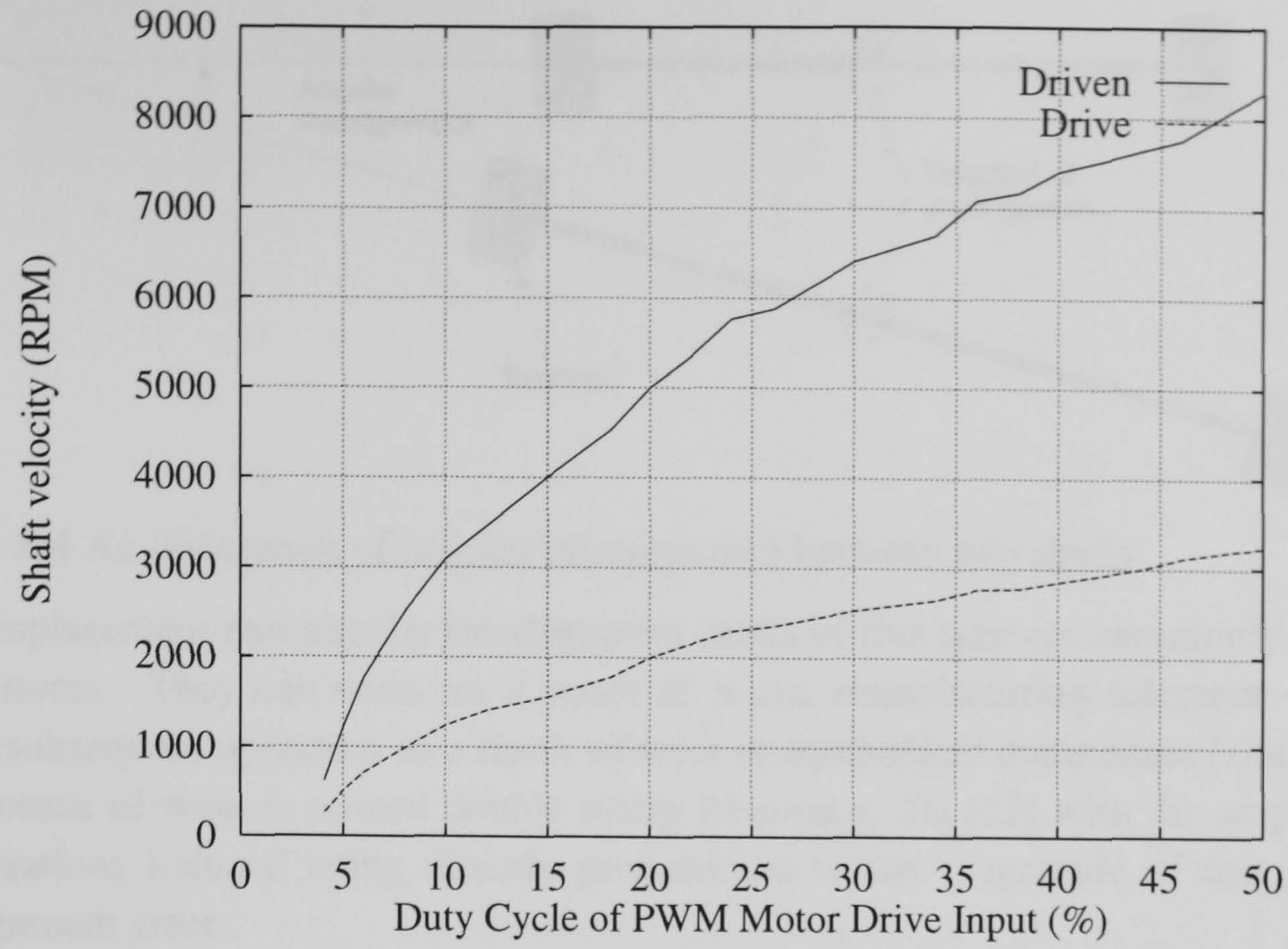

Figure 5-2 Relationship between motor drive PWM signal and the primary shaft velocity for a fixed testbed configuration

5.1.1 Simulation of Specific Fault States on the Test System

Simple displacement misalignment was simulated by rotating both anchor bearings on a shaft by equal amounts whilst locking the second shaft in a fixed position. This effectively varies the distance between shafts and consequently varies the amount of displacement misalignment which is induced. Figure 5.3 illustrates this degree of freedom graphically. Angular misalignment is achieved either by rotating both the anchor bearings on one shaft in opposing directions, or by fixing one bearing and rotating the other. Since each bearing can provide 3° of freedom, the maximum angular misalignment which may be simulated per shaft pair is 6°. This type of misalignment is detailed graphically in Figure 5.4. It will be noted from the system diagram in Figure 5.1 that neither of these errors can be simulated, with the bearing mechanism, in exactly the same way they would be expected to occur in the production equivalent. The eccentric bearings will introduce additional offsets which cannot themselves be eliminated from the fault states. However for the purposes of a test system used to simulate representative faults these additional effects are considered to be acceptable.

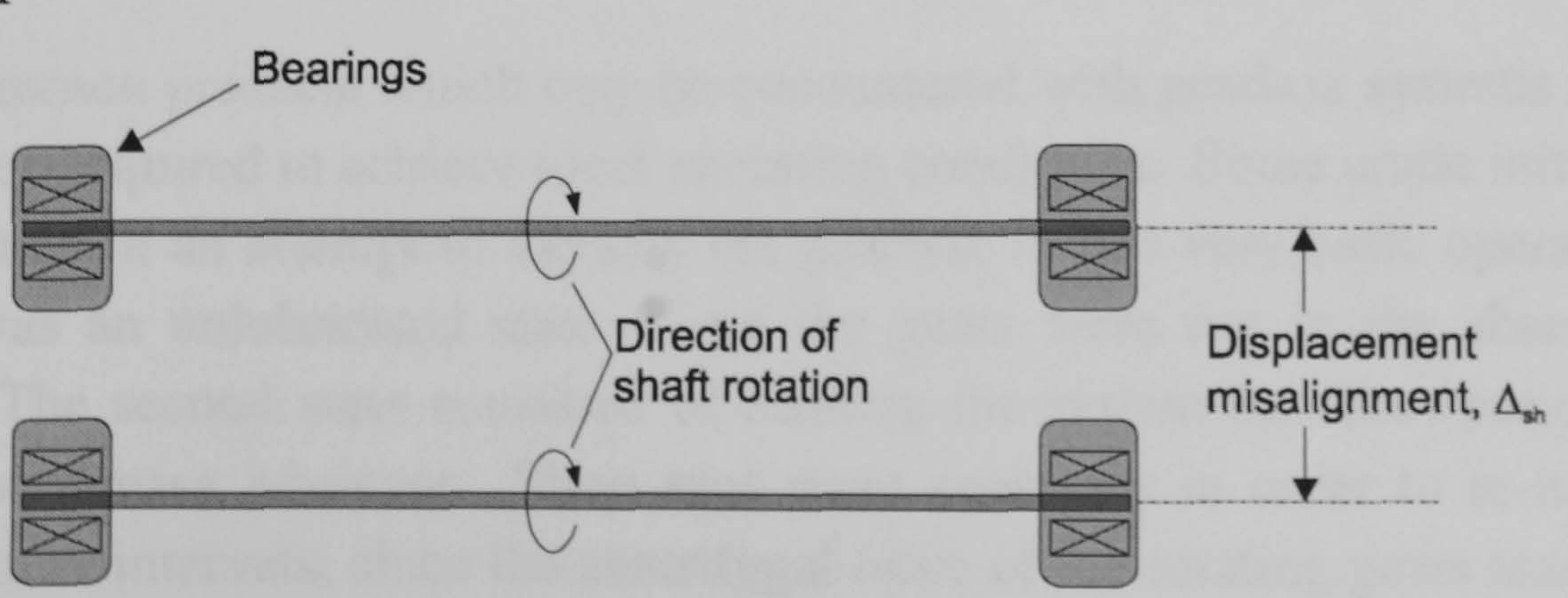

Figure 5-3 An illustration of displacement misalignment between two shafts

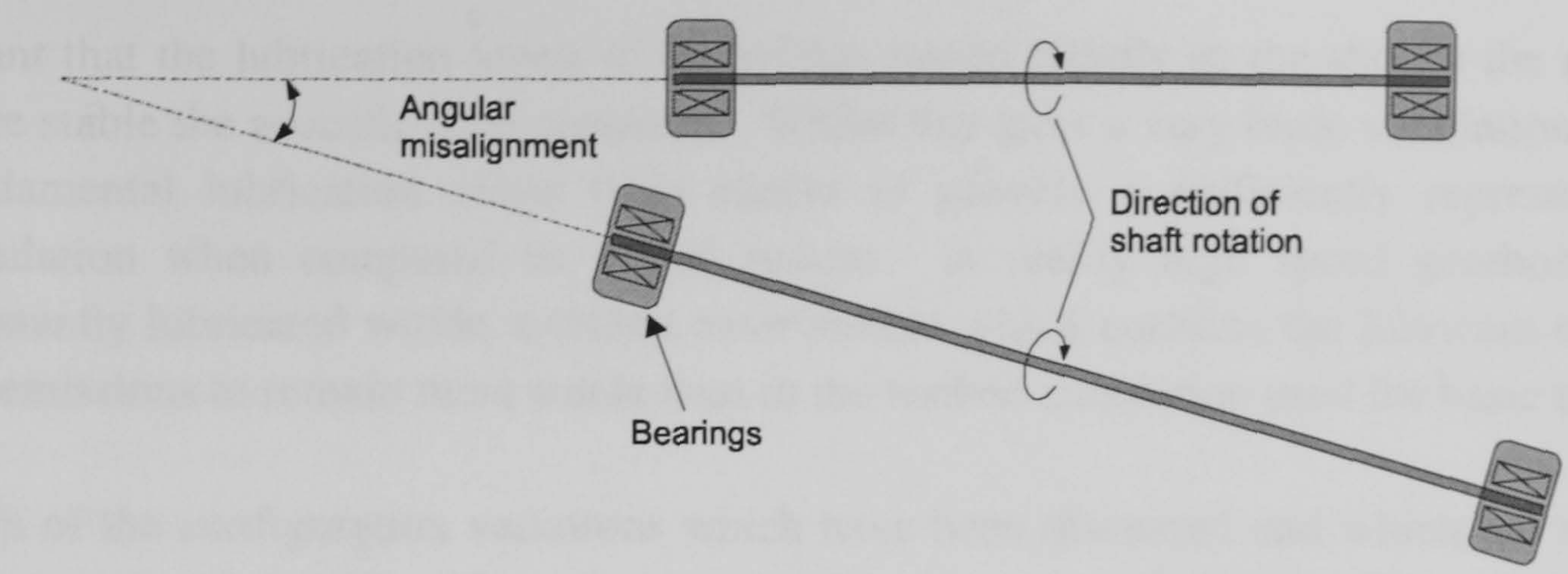

Figure 5-4 An illustration of angular misalignment between two shafts

Both displacement and angular misalignment faults of this type are commonly found in real systems. They can occur as a result of initial manufacturing tolerance errors or during subsequent operation as a result of wear or mechanical component failure. Both are a source of what is termed double rotary frequency, $2f_0$ [63] with the amplitude of the vibrations induced being directly proportional to the magnitude of the particular misalignment error.

In addition to alignment faults both tooth damage and wear effects can be exhibited when gear based systems are exposed to high levels of stress for continuous periods of operation. Such faults were simulated in the test system by machining of the tooth tip itself. Whilst in reality this type of tooth damage is rarely perfectly symmetrical over the width of the tooth or teeth the faults simulated in the test system do result in similar periodical fluctuations to those experienced under operating conditions. For the purposes of estimating the sensitivity of the various classification techniques to tooth damage this particular fault condition was limited to the introduction of a single defect in a tooth on one gear. During the acoustic acquisition stage and prior to system state evaluation trials this fault was progressively worsened over three separate recording sessions by additional machining of the tooth. Initially the fault was introduced to the system with a 1mm section of the tooth tip removed. Following acoustic data recordings, this was increased initially to 2mm and then to 3.5mm by further machining of the tip. As with the misalignment errors these fault states represent somewhat artificial failure modes since tooth damage is often accompanied by a change in the physical properties of the material caused by the continuous impact stress. Furthermore the primary failure can often cause additional damage as a result of the ingestion of the tooth fragment into the system. Such effects are difficult to simulate within the current test environment.

Another common problem which may be encountered with gearbox systems is the level of lubrication required to achieve ideal operating conditions. Some crude initial analysis was performed in an attempt to identify the gearbox in two very basic operating states. The first was an unlubricated state where the gears were run in the absence of any lubricant. The second state consisted of running the system for short periods of time with a heavy grease lubricant. Short runs were necessary in order to re-lubricate the gears at regular intervals, since the centrifugal force of the rotating gears tended to strip the grease and as the system is not self-lubricating this lubricant is lost. Essentially this meant that the lubrication levels of recordings varied rapidly so the shorter the run the more stable the acoustic state remained. Whilst this gave a very basic simulation of the fundamental lubrication states it is unable to provide a sufficiently representative simulation when compared to a real system. In reality high speed gearboxes are constantly lubricated within a closed environment which contains the lubricant causing the emissions to remain more stable than in the testbed simulation used for basic trials.

Each of the configuration variations which have been discussed and which are used in the analysis of classification techniques in the remainder of this Chapter and the next can be described using a set of basic physical measures. These are not intended to be considered an absolute definition of the configuration but as with the states themselves a representative description. The first two states were the crudely simulated lubricant state configurations in which the acoustic emissions of gears 1 and 2 meshing in two distinct lubrication states were analysed. In these two recordings gears 3 and 4 were disengaged and did not contribute to the group emissions. The remaining ten states concentrated on shaft and tooth variations. The three parameters which have been used to define each of these test conditions are illustrated in Figure 5.1. These are the linear shaft displacement, Δ_{sh}, the relative offset of the two shafts, λ_{sh} and the quantity of the material removed from a single tooth on the 3rd gear, μ_t. Variations in these parameters will introduce variations in the tooth overlap, D_{ov}, and the pitch variation, δ_p, between the third and fourth gears which are used to describe the various condition states. In all these remaining ten states the gear and bearing lubrication was kept at a constant minimal level so as to eliminate interference caused by fluctuations during classification. For reference purposes each of the configurations are allocated a particular state descriptor. The ordering of these states does not necessarily correspond to the physical differences between states and is only intended as a reference for the trial evaluations.

State descriptor	Gear status	Physical description
State 1	3 & 4 completely disengaged	Ratios 1 & 2 lubricated
State 2	3 & 4 completely disengaged	Ratios 1 & 2 unlubricated
State 3	3 & 4 completely disengaged	$D_{ov}= 0, \delta_p=$ N/A.
State 4	3 & 4 partially engaged	$D_{ov}= 1.5$mm, $\delta_p= 2.5$mm
State 5	3 & 4 partially engaged	$D_{ov}= 4$mm, $\delta_p= 0.5$mm
State 6	3 & 4 fully engaged	$D_{ov}< 5$mm, $\delta_p< 0.5$mm
State 7	3 & 4 fully engaged, with offset	$\lambda_{sh}= 2$mm.
State 8	3 & 4 fully engaged, with offset	$\lambda_{sh}= 3$mm
State 9	3 & 4 fully engaged, with offset	$\lambda_{sh}= 5$mm
State 10	3 & 4 fully engaged, tooth wear	$\mu_t = 1$mm
State 11	3 & 4 fully engaged, tooth wear	$\mu_t = 2$mm
State 12	3 & 4 fully engaged, tooth wear	$\mu_t = 3.5$mm.

TABLE 5.1 A description of the fault states employed during practical trials

It may be noted by closer inspection that these defining physical measures are not necessarily fully independent. For instance states 7, 8 and 9 are defined only in terms of the shaft offset, λ_{sh}, whereas in fact both D_{ov} and δ_p will also vary along the width of the gear teeth. The same will be true of D_{ov} in the case of the simulated tooth fault in states 10, 11 and 12. However the measures outlined in table 5.1 provide sufficient indication of the differences between particular states and are intended only as such. Each of the states have been selected so as to provide not only a means of determining the capabilities of the various classification techniques but also as a means of estimating the sensitivity of particular techniques to physical changes. Both of these factors are important in terms of the industrial applicability of TES techniques.

5.1.2 Effect on Group Emissions of Simulated Mechanical Faults

Each of the states discussed in the previous section will affect the acoustic properties of the group emissions in different and for identification purposes hopefully sufficiently unique ways. Before moving on to the discussion of practical identification techniques it is worth considering some of the acoustic effects which each of the predefined physical states is likely to introduce to the TES data matrices used for neural classification. Whilst an ideal gear pair would transmit power with no change in shaft angular velocities and with zero loss of power in practice transmission systems are not ideal and energy is dissipated acoustically as a result. Practical gear systems will generate natural levels of acoustic noise as a result of these factors, but imperfections, damage and wear will result in additional variations in the levels and spectral content of this noise emission. Discrepancies in tooth spacing or profile, shaft alignment or bending due either to production inaccuracies, damage or wear will produce periodic accelerations and decelerations in the gear pairs which will contribute towards the acoustic emissions. The casing design and gear material composition will also vary the level and spectral content of the emissions. In addition to these mechanical effects factors such as lubrication level and system temperature will also introduce additional variations to the group emissions. It is these unwanted additional factors that each of the simulated states is intended to mimic.

Taking firstly the example of the simulated displacement states 3-6. The modification in relative base pitch, δ_p, simulated in these four states will have two important effects upon the transfer of rotational energy taking place in the system. An increase in δ_p, caused in this case by relative shaft displacement will introduce a proportional increase in tooth central impact energy transfer as a result of load variations. These are caused by fluctuations in drive shaft acceleration of the drive gear relative to the driven gear. However there will also be a reduction in the frictional energy caused by the relative motion of the gear teeth against one another since the effective tooth contact patch is reduced by the displacement. As δ_p is reduced the situation is reversed. These variations should appear on amplitude TES matrices as energy related contour movements.

The shaft offsets simulated in states 7-9 will also result in energy shifts due to the variable contact force exerted on the meshing gears by a skew shaft. In the simulated faults defined by these states the shift will be constant in nature. The more complex non-static energy shifts caused by unbalance in a bent shaft were not simulated during the course of this work. States 10 to 12 which mimicked catastrophic failure of a single tooth induce acoustic variations as a result of variations in shaft acceleration similar to if generally more severe than shaft displacement, states 3-6. Whilst the previously described displacement states will result in a constant increase in emissions over a full shaft cycle, tooth failure introduces periodic fluctuations which are dependant upon the extent of the tooth damage and on the rotational velocity and loading of the associated parts

5.1.3 Acoustic Data Acquisition

Having described now in some detail the design of the gearbox testbed and each of the specific configuration states it is necessary also to give consideration to the means by which the emissions be acquired. For the practical work described in this and the next Chapter a single uni-directional condenser microphone located approximately 50mm from the open casing was used to record the acoustic emissions. The physical positioning of the microphone could be freely controlled in order to perform measurements on the sensitivity of the TES data generated from various locations. As the group acoustic emissions from the gearbox are made up from many simultaneous point sources so it is reasonable to assume that the acoustic emissions, and thus the TES data, will be position sensitive just as is the case with the more commonly used contact type sensors such as accelerometers or velocity transducers. Without practical trials the extent of this acoustic sensitivity is difficult to predict. It is therefore crucial to gauge these effects so that the means by which TES techniques may be applied and also possibly their effectiveness in noisy environments may be estimated. The need for complex arrangements of directional microphones to produce acceptable results, for example, would militate against the use of acoustic TES in a cheap monitoring system.

For each of the predefined states a set of acoustic samples were recorded onto a high quality audio media under controlled conditions. The control consisted of monitoring the shaft velocities during the recording of each unique acoustical state. In cases where the shaft velocity was not the identifiable parameter this monitoring and associated control was loosely constrained. This was because whilst it was important to keep the shaft velocity relatively stable in such cases it was also considered necessary to maintain the systems usability by minimising the number of complex constraints placed on the basic data capture. Consequently during the acquisition of data for the practical trials not involving shaft speed state identification shaft velocity fluctuations of approximately 200rpm were deemed acceptable. This corresponded to variations in shaft velocity during acquisition of approximately 7%. Recordings which remained within these bounds were included in the acoustical archive, whilst those which did not were discarded.

Another key to building a stable archive of emissions records was the periodic inspection of the testbed to determine the current mechanical status. Between each of the recorded segments included in the archive the mechanical configuration of the system was checked for stability. This added safety precaution was mainly intended to ensure that the bearing locating mechanisms had not become loose and caused an unwanted change in mechanical configuration during the recording process.

5.2 Application of Simple Amplitude Histogram Matrices for Classification of Basic Shaft Velocity States

There are two major advantages to using the simple amplitude TES symbol histogram matrices defined in 3.5.1. The first is the ease of generation of the matrix data, requiring only summation of the symbols in the converted TES stream over short fixed time periods. The second is the physical size of neural network which is required for the application of the generated data. In trials the TES conversion symbol table consisted of only 300 elements which equates to a similar number of input layer nodes for the classification network, rather less than the A-matrix scheme. Whilst the main focus of the application of TES to the archive of acoustical data acquired from the gearbox testbed was the identification of specific physical fault states some initial suitability trials were performed using these very basic histogram matrices to identify a series of four unique gearbox velocity states. These four unique velocity modes are not contained in the earlier state Table 5.1 but are instead defined below in Table 5.2.

State descriptor	Gear ratio status	Physical description
State A	3 & 4 completely disengaged	Primary shaft velocity of 500rpm
State B	3 & 4 completely disengaged	Primary shaft velocity of 1200rpm
State C	3 & 4 completely disengaged	Primary shaft velocity of 1500rpm
State D	3 & 4 completely disengaged	Primary shaft velocity of 2000rpm

TABLE 5.2 A description of the four basic velocity states used during initial trials of TES application to neural networks

In these preliminary trials the 300 element neural network input layer was connected to a 10 node hidden layer and a 4 node output layer (one per velocity state). The simplicity of the primary data generation combined with a moderate network size, and thus small number of interconnections, results in a classifier configuration requiring a relatively low computational overhead. Each histogram matrix applied to the network was constructed from a one second segment, or token, of the raw TES stream generated from the recorded acoustic archive and converted into the amplitude TES symbol format. Four sets of data were produced in total from the recorded archive for network training and evaluation, each containing ten histogram matrices corresponding to each of the four states defined in Table 5.2. Thus each data set contained 40 matrices, or 40 seconds of acoustic data, clustered into unique state groups of ten matrices. Ten separate train and test sessions were performed with the neural classifier, each using different permutations of training and testing data from the four available sets. In each trial one or two of the four data sets were selected for training purposes with the remainder dedicated to the testing of the network. Thus two different sizes of training and testing data sets were used. In training 40 and 80 matrix template sets were used whilst for performance evaluation 80 and 120 element matrix sets were employed. The use of these differing data set ordering and size permutations enabled early indications of the effects of the data sensitivity discussed previously in Chapter 4 to be assessed for this basic TES presentation format. The results of these trials are given in Figure 5.5.

The error rates encountered during classification testing ranged from 0-17.5% depending upon the quantity and composition of the data used during training and testing. This disparity in error performance between similar sized data sets illustrates graphically the dependence between the data presented to the network and the global network error minima which is achieved during convergence. Whilst trial 3 produced an acceptable 2.5% (3 errors in 120) error rate, trial 1 produced considerably more errors with 17.5% (21 errors in 120) of the test set classified incorrectly. When the training data set size is doubled in size the disparity between best and worst performing networks is reduced from 15% to 10%. This improvement is most noticeable in trial 10 (rightmost plot in Figure 5.5) where the extended data set (containing sets 3 & 4) eliminated classification errors in the test set altogether.

In addition, some fluctuation in network performance was noted during the trials when the data set presentation ordering was modified during training runs. This was undoubtedly the result of the effect the data presentation has upon the training of network weights. Effectively the path taken by each weight over the error surface of the network during training is sensitive to the contents of each training exemplar and thereby the order in which they are presented during the training phase. As an example when the data sets used for training in trial 10 were reversed a classification error was introduced where previously there had been none. These stability effects filter on

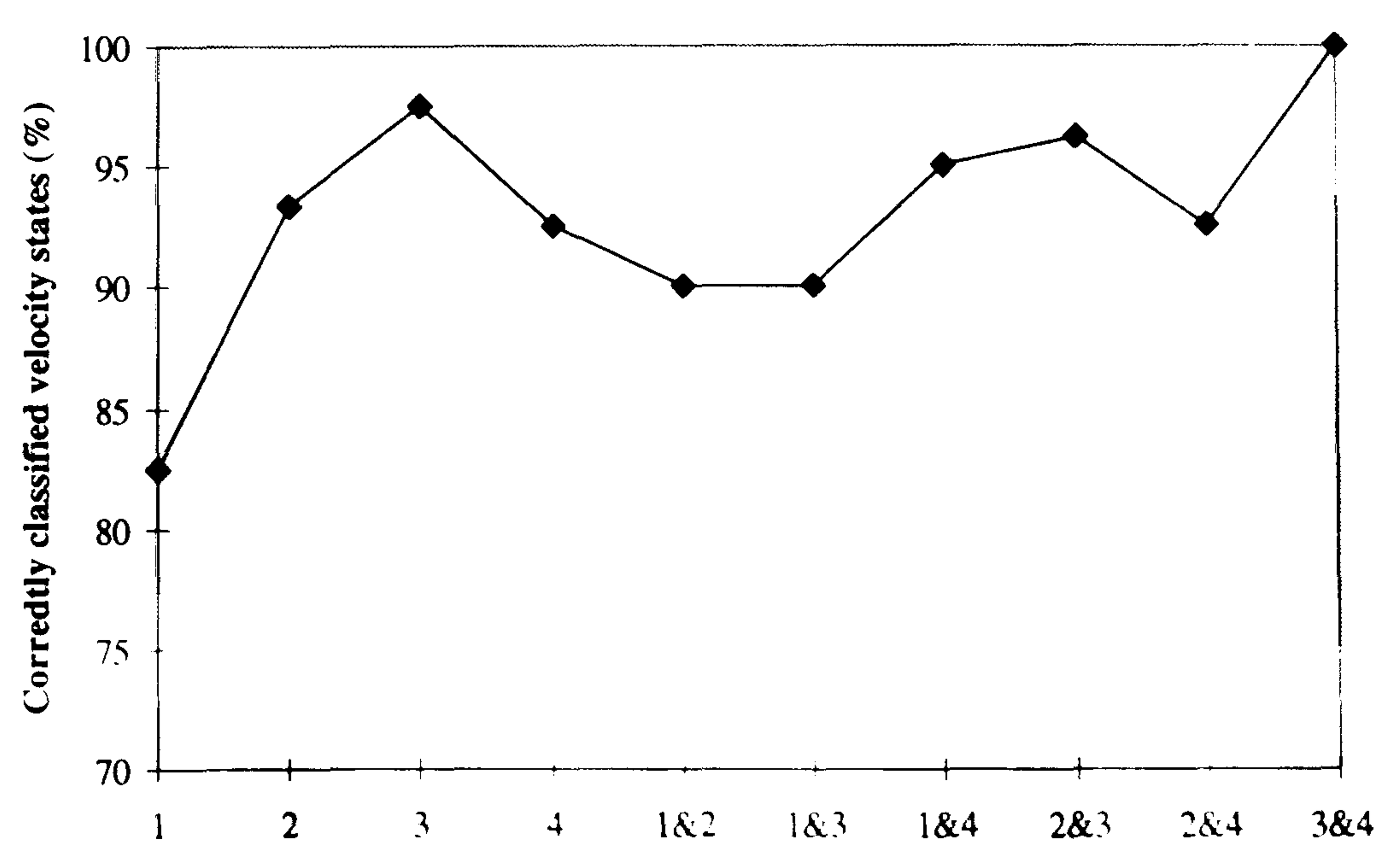

Figure 5-5 Results of the trials employing amplitude-frequency data matrices

through to the various parameters of the training control mechanism itself. Again in tests with the trial 10 data it was found that the error performance may be improved by testing for convergence less frequently. This would indicate that the original test sequence was probably terminating training at a point where network convergence was acceptable but not necessarily optimised for the particular data set. The improvement in performance introduced by lengthening the time between successive convergence tests can be accounted for by a reduction in the subsequent global network entropy caused by the inclusion of additional training iterations. In subsequent trials this problem is uncovered as a common theme where neural classifiers are employed.

It is clear from these trials that amplitude-frequency histogram matrices provide sufficient raw acoustic information to separate simple shaft velocity states, even when training runs are limited to relatively short, 10 second segments of acoustic tokens for each state. Whilst the subject of this classification trial, shaft velocity identification, is not in itself particularly demanding it is still encouraging to see that a basic feed-forward network had been able to separate the state space with reasonable accuracy. It is also evident that the combination of network, data and training regime can impact measurably upon classification performance as a result of variations between specific training and testing data sets.

5.2.1 Positional Sensitivity Associated with Amplitude Histogram Matrices

Having performed some initial work and ascertained the basic suitability of simple amplitude histogram matrices to the identification of a few rudimentary gearbox shaft velocity states consideration was made as to the microphone positional sensitivity of this particular presentation technique. Microphone sensitivity had earlier been identified as being one area which could adversely complicate the application of the neural acoustic technique. For a series of initial trials, sensitivity was defined as the effect upon classification performance of variations in microphone position between the training and testing phases of between 100-200mm. Figure 5.6 illustrates the three different microphone positions which were used to acquire the acoustic data used for amplitude TES histogram generation for evaluating a similar neural classifier configuration to that applied previously in section 5.2. This time however a fifth output node was added to the network to enable detection of a gear lubricant.

A network was trained on histogram matrices generated from testbed emissions acquired at a constant speed in both unlubricated and artificially lubricated states from the three microphone positions. Baseline sensitivity of the configuration to these sensor positions when classifying data into lubricated and unlubricated states was estimated by using data acquired from the same physical microphone location for both training and testing. In tests carried out on the three longitudinal positions (1, 2, 3) using training and test sets of 20 matrices, lubrication classification accuracy ranged from 55% (2), to 80% (3). Having ascertained the baseline performance from each of the separate microphone locations a second trial was performed to estimate the location sensitivity of the data. In

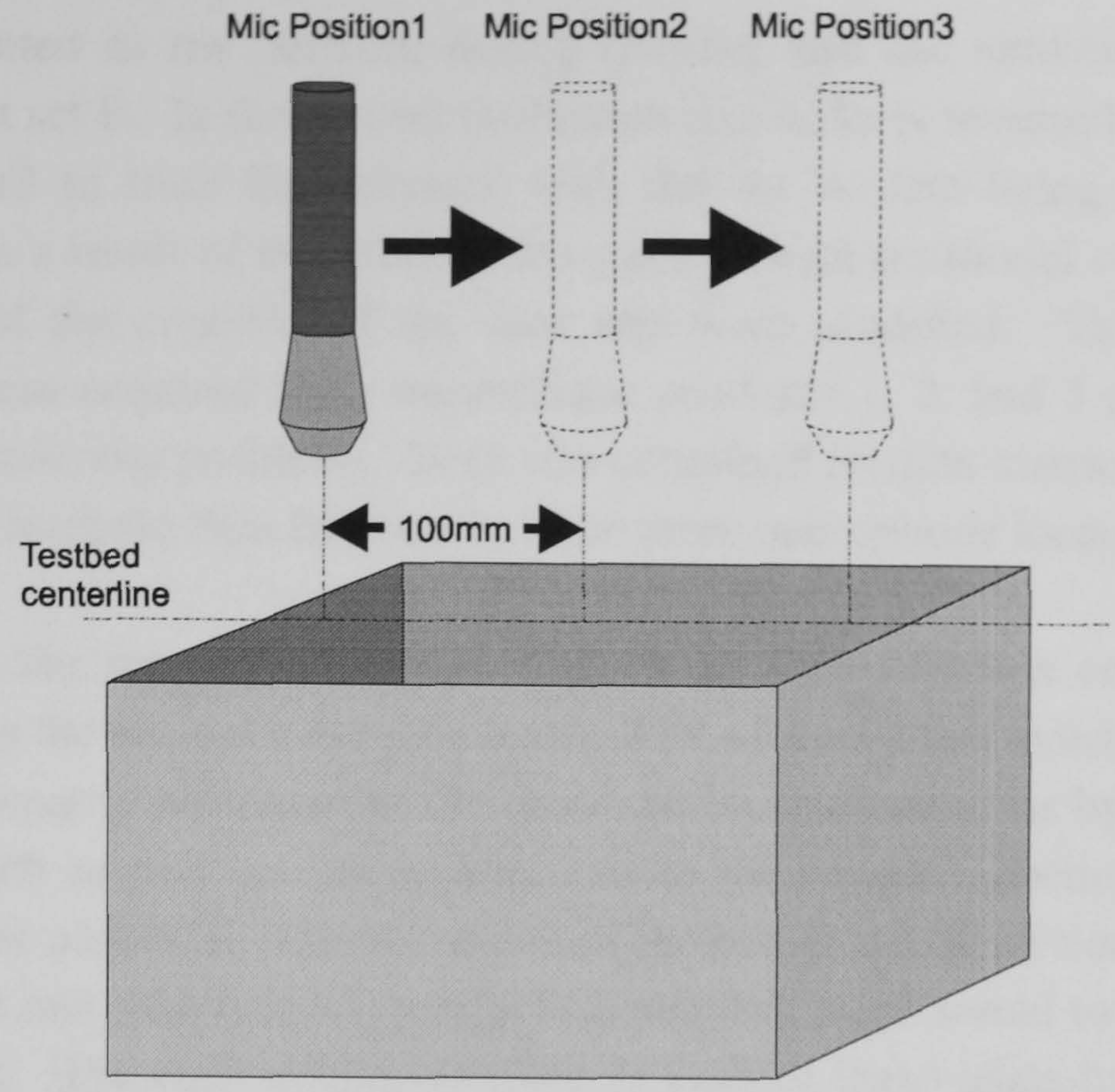

Figure 5-6 The microphone positions used to measure the sensitivity of acoustic data sets

this trial the network was trained on histogram matrices generated from acoustic recordings taken with the microphone in one location and tested on data generated from the remaining two positions. In this way the positional diversity of the respective microphone positions could be estimated. The results of this are provided in Table 5.3. Whilst not wholly conclusive in this instance, the results indicate that there is a relationship between the microphone location and the network performance. Location 3 performance was eroded from 80% to only 50% when train and test locations were separated.

Acoustic acquisition positions Training set : Testing sets	Classification of lubricated/unlubricated states
1:[2,3]	54%
2:[1,3]	60%
3:[1,2]	50%

TABLE 5.3 Results of the trials using amplitude histogram data acquired from different microphone locations to train and test a three layer network

5.2.2 The Effect of Histogram Matrix Presentation on Network Performance

The sensitivity of neural classifiers upon the data sets applied during the training phase has already been noted for the shaft velocity classification problem. A further trial was completed using the network defined in section 5.2.1 to quantify the additional effects of data ordering on the tokens used in the tests discussed so far. This trial consisted of two separate train and test phases, both using data from two unique data sets (A and B) but applied to the network in different orders. For the first evaluation, matrix tokens in data set A are presented to the network during training and the network performance is measured against set B. In the second evaluation this order is reversed and instead set B matrices are used to train the network with the set A data being used to measure performance. As a result of the information gained about positional sensitivity made in the previous trial the contents of the data sets were modified. This time both sets contain data tokens acquired from microphone positions 1, 2, and 3 aimed at reducing some of these sensitivity problems. Both sets contained 60 data matrices, corresponding to 20 seconds of acoustic data from each of the three microphone locations.

In the first trial the network identified 95% of all data matrices correctly following training, whilst in the second it classified only 80% of the 60 test matrices correctly after training. The disparity between the two runs can be accounted for by variations in the two data sets with respect to one another and to the acoustic phenomena which they represent. In this particular instance the training data in set A obviously enhances the training program and subsequently produces a network more suited to the data matrices in the testing set. The performance attained in both of these trials by using data from multiple sensor positions highlights the necessity for adequate network generalisation if the effects of sensor position, seen previously, are to be minimised. The ability to reduce the effects of the sensor selectivity are of particular relevance to industrial implementations of a classification system where restrictions on this flexibility should be discouraged. The results here provide proof that given adequate training these sensitivity effects can be significantly reduced.

5.2.3 Varying the Internal Structure of the Neural Network

Whilst some of the physical data acquisition and network training constraints had been considered in the initial trials the architectural requirements of the neural classifier itself were not. From the point of view of implementing a realisable and cost effective monitoring system, in this instance with histogram data matrices, these architectural considerations are similarly important. As discussed in Chapter 4 the architectural definition of any supervised network applied to data defining a state space relies upon a balance being sought between network size and the corresponding computational overhead required during both training and operation. On one hand the fewer nodes there are in the different network layers the fewer the number of interconnections and thus the lower the computational overhead required. However the fewer nodes there are the more restricted the dimensions of freedom of the network become and with it its classification capability. In contrast, as network size is increased so too does the capability, training time and the data set required for adequate generalisation. Since the input and output layers are effectively fixed by the data format and state space definitions respectively the only significant room for flexibility in network size and configuration is within the hidden layer.

To study the effects of these architectural facets performance evaluation was carried out using the shaft velocity state matrix tokens utilised earlier in 5.2. Two evaluations were performed, one including and one excluding the lubrication status of the gears. Initially the data matrices taken from the eight state classification problem discussed previously in 5.2.1 were evaluated against a range of different internal network configurations. The hidden layer was varied in size between six and fifty nodes, whilst other architectural and training parameters for the networks remained stable. Most important in this respect was the stability of all training and test data sets. Each contained 80 data matrices, 10 per state corresponding to 80 seconds of amplitude histogram TES data in total acquired whilst the microphone was fixed in position 2. The subsequent training rates and classification performances of each of these network configurations are graphically represented in Figure 5.7 and Figure 5.8.

Somewhat expectedly the number of training iterations and the time required for network convergence in these tests is related to the number of nodes contained in the hidden layer. Below six nodes in the hidden layer the number of iterations required for convergence produced training times considered to be unreasonable and additionally did not provide quantifiable improvements in state separation. Beyond 50 nodes, the training times became similarly excessive despite a reduction in the number of iterations required to reach convergence. The increase here in the time to convergence, despite a reduction in the number of iterations, is caused by a corresponding increase in the computational time taken to perform each cycle of the matrix presentation and error back-propagation as a result of the additional number of nodal interconnections.

Despite the differences in the time taken to complete the training phases for the various configurations correct identification of the shaft velocity states during the testing phase remained relatively stable. Over the range of configurations examined it varied by only approximately 4%. Identification of the lubrication status during the same trials showed a little more variation, approximately 13%. Given the results obtained from these trials there seem to be few indications that any specific configuration is more suited than another to the separation of these particular states although networks containing 10, 16

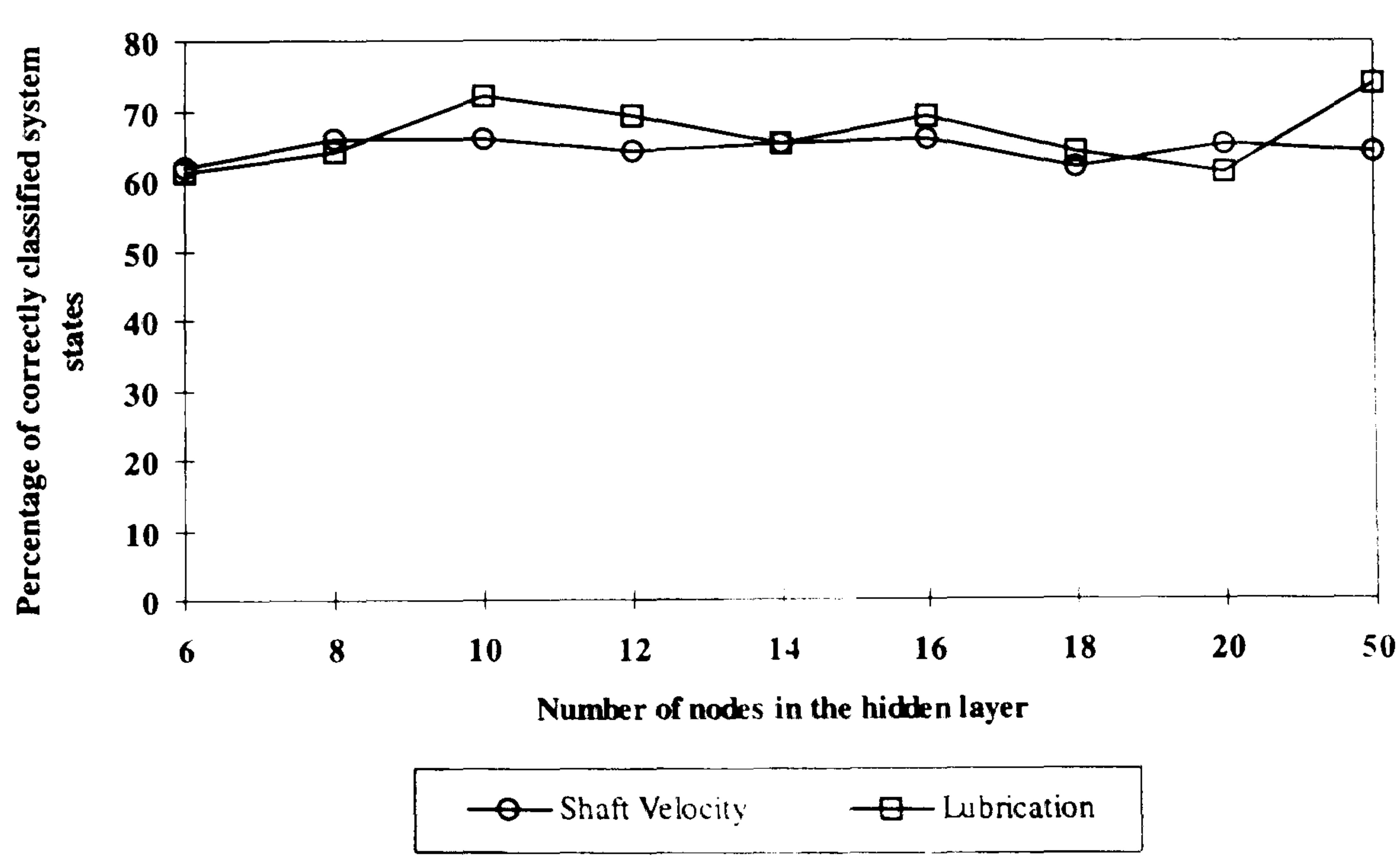

Figure 5-7 Network performance for velocity and lubrication status classification with selected hidden layer configurations in an eight state data space

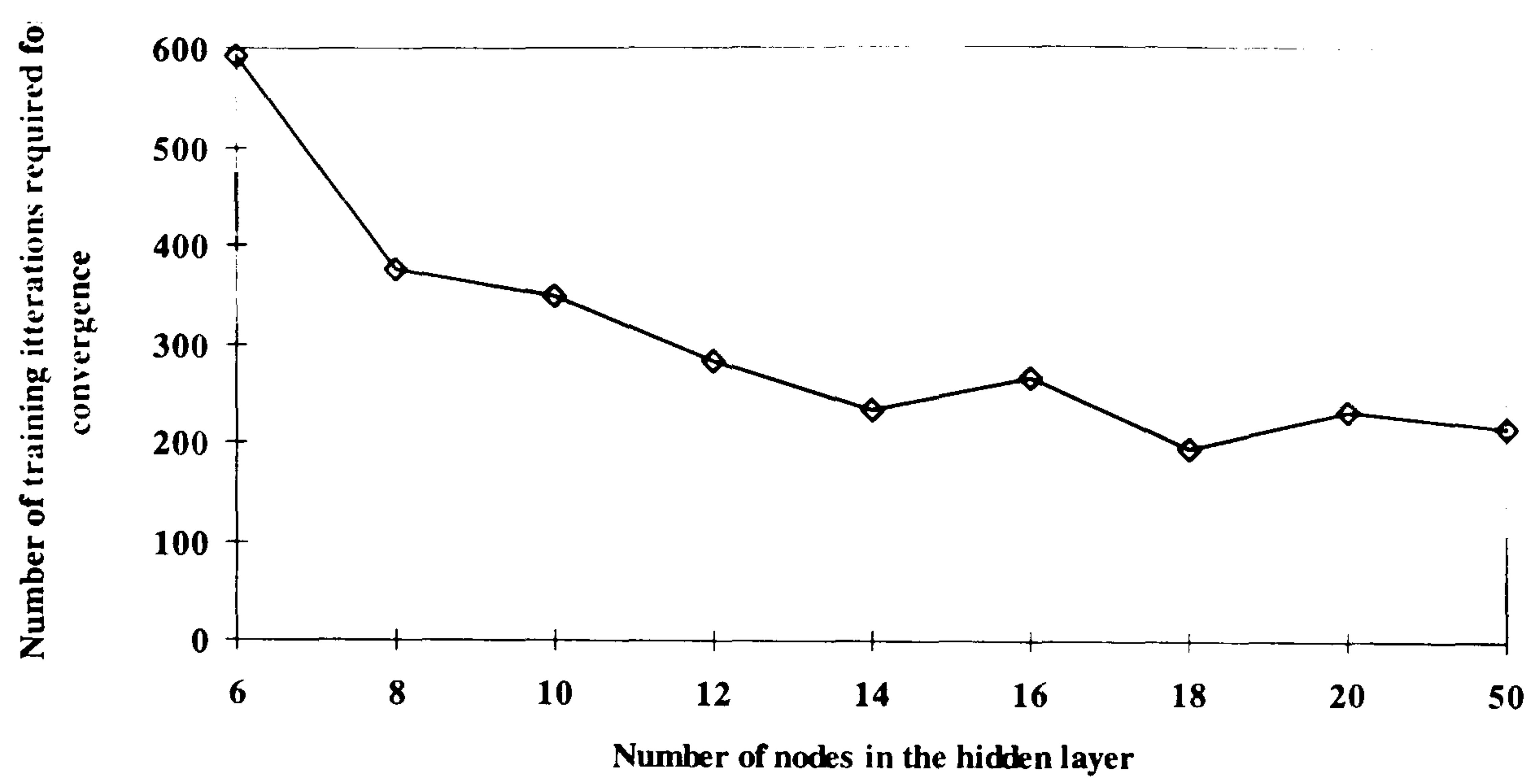

Figure 5-8 Network convergence for various hidden layer configurations in an eight state data space

and 50 nodes provided marginally better separation of the lubrication state in this evaluation. However, there is likely to be some degree of data sensitivity to be accounted in these findings as a result both of initial training weights and the path followed subsequently by each weight during training. Given this premise and the results obtained it would seem reasonable to select a particular architecture(s) from this range primarily on the computational requirements of the network during the training and testing phases rather than on point performance with individual data sets.

5.2.4 Detection of Four Shaft Displacement Misalignment States

One area of concern in terms of the acoustic data used in the eight state network used in previous trials was the use of an artificially simulated lubrication status. The dynamics of the gearbox were such that the state of lubrication could, and did, change rapidly during the course of each acquisition recording. This effectively introduced additional dynamic states into the data space which combined to degrade the networks perceived performance. Consequently these particular states were removed from the decision space to increase the confidence in results attained in relation to the performance requirements of the hidden layer in the network architecture. However, at the same time reducing the output space to the four remaining shaft velocity states would reduce the classification problem to a considerably less demanding frequency focused exercise. Thus rather than continue the discussion of network architectures with such comparatively simple states, a further four, more demanding states (3, 4, 5, 6) were selected. They correspond to four shaft misalignment states each of varying degrees of severity and were each recorded at a controlled constant shaft velocity.

Since the dimensions of the input matrix required for misalignment detection remain unchanged a similar 300 element input layer was required to apply the TES data from the amplitude histogram matrices providing the raw acoustic information to the network.

This was coupled to a four node output layer with each node corresponding to one of the four unique shaft position states. These two external layers were connected via a configurable hidden layer containing between two and one hundred nodes. As with previous trials the training regime was strictly controlled to ensure that the effects of a range of hidden layer configurations could be evaluated in isolation. Control consisted of maintaining a fixed set of network training parameters including learning rate (α), momentum (β) and convergence bound. Each network configuration was trained using a fixed training set containing a series of the 300 element amplitude TES histogram matrices. Each of the data sets within this fixed training set corresponded to approximately 14 minutes of acoustic data. Following successful completion of the training phase for each configuration, the networks were subjected to an identical set of acoustic test patterns corresponding to approximately 7 minutes of recorded data. The results of the training and testing phases of these more demanding shaft position detection trials are given in Figure 5.9 and Figure 5.10.

The use of four more complex positional states placed an increased burden upon the network to classify the data adequately when compared to the earlier shaft velocity trials which had been carried out. The demands of this increased burden were partially offset by the application of a considerably extended training data set to further improve the networks learning capability and generalisation properties. The performance of these networks when compared with those from previous trials is particularly interesting. Despite the added complexity of the requirement to separate the more acoustically similar shaft positional states the general performance of the networks at identifying each of these states was better than that achieved with the eight states described in the previous section. This improvement, although partially attributed to the elimination of the dynamic variations in the lubrication state included in previous trials, was mainly due to the tenfold expansion of the training data set.

As with earlier evaluations the deviation in network performance over the range of network architectures tested appears inconsistent as a result of the combined affect on the networks of data dependencies and weight updates during training. As an example a network with a six element hidden layer correctly identified 82% of the test patterns whilst a 32 node layer only identified 76.2% of the test patterns correctly. Again since it seems reasonable to surmise that both the training path taken by individual weights and the diversity in content of the acquired data sets themselves will effect the performance of each network selection of an "ideal" layer size again becomes a non-trivial exercise.

As with the trials on velocity and lubrication detection it is better to assume that a range of hidden layer architectures will provide adequate performance and make the selection of the hidden layer attributes based on the computational demands of each configuration. Thus providing the hidden layer contains a reasonable number of nodes for a particular classification problem enough network flexibility should be available to separate each class and it is unnecessary to expend further effort on identifying a notionally ideal configuration. For this particular shaft misalignment problem a hidden layer containing 10-20 nodes would seem to offer the best compromise between

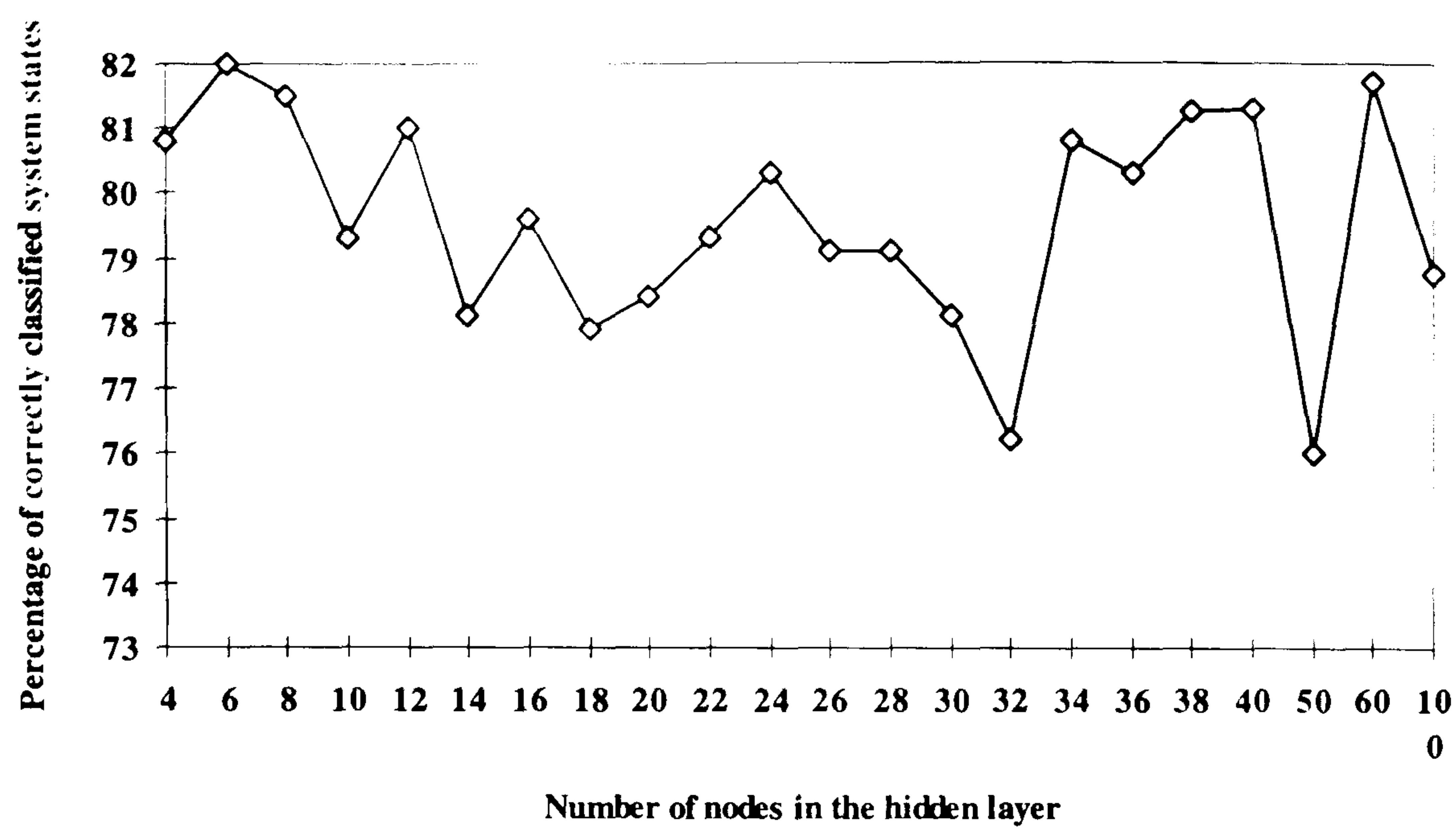

Figure 5-9 Network performance for shaft position detection with selected hidden layer configurations in a four state data space

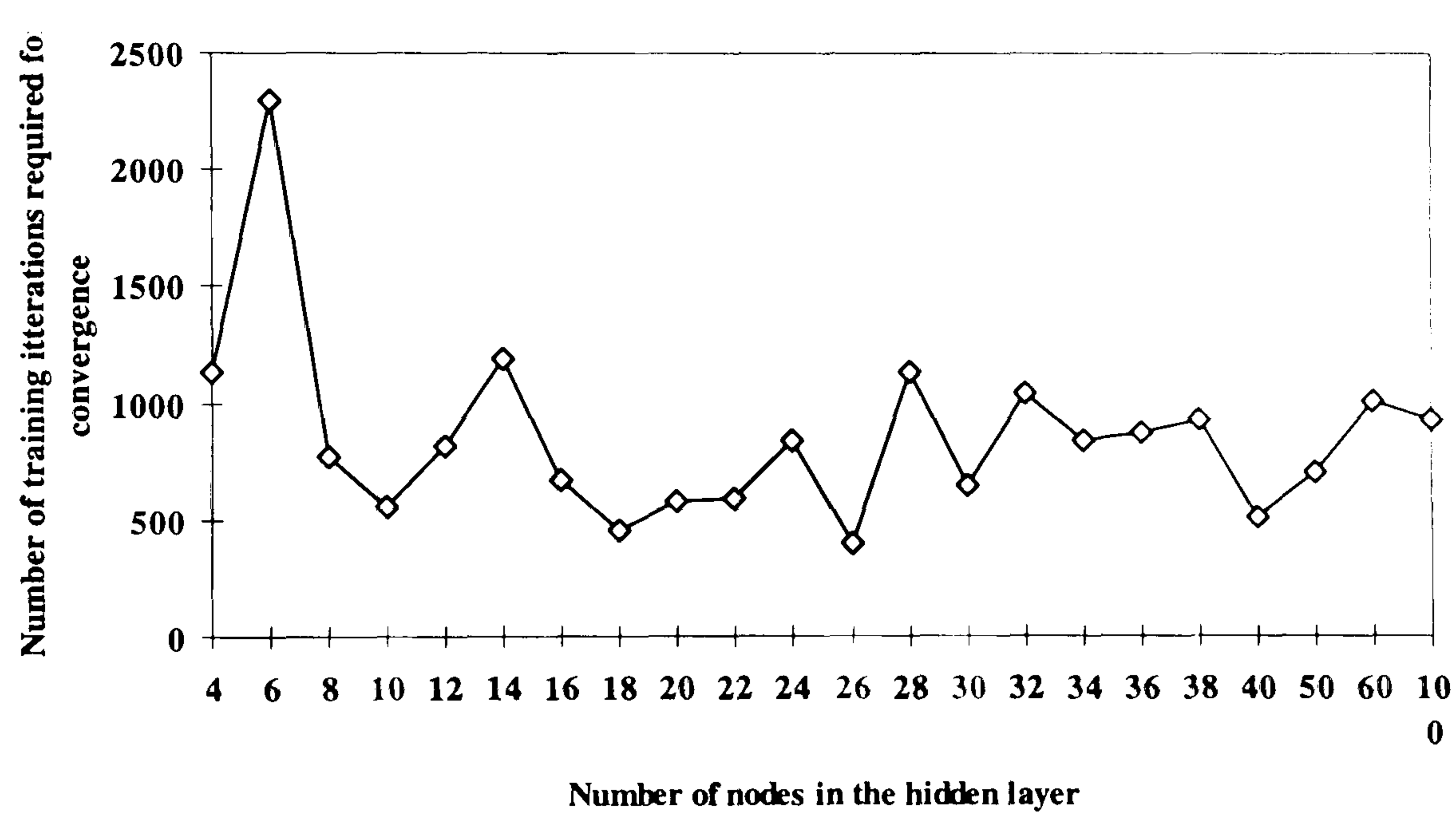

Figure 5-10 Network convergence requirements for various hidden layer configurations when detecting the shaft position during operation

adequate performance and optimised computational requirements during training and operation.

5.2.5 Improving Generalisation by Extending the Network Training

The initial results drawn from the evaluation performed on the configuration of the hidden layer in networks detecting shaft status provides a stable baseline from which to investigate further the effects of varying some of the other parameters associated with neural classifiers. Extending the quantity of data applied during the network training phase is one such parameter. From the performance achieved during the preliminary trials with shaft alignment the importance of this aspect of the training cannot be overlooked. For further comparative evaluation of this aspect of the implementation it is essential once again to maintain not only the same basic network platform and training regime but also to use training and test sets based upon the data used previously.

To maintain the necessary data stability each of the training data sets used in the earlier evaluation was retained to form the cores of the training sets generated for the additional testing. These core sets were complemented by selecting a further seven minutes of TES histogram matrix data to extend each of the training sets to 21 minutes of data in total. As in the case of earlier trials, network convergence testing during this phase was performed after each and every iteration, or epoch, of the extended training set. In reality this meant that convergence testing was not performed after an identical number of network data presentations to the previous trials due to the extended size of the set. However it was assumed for the purposes of the evaluation that this change, taken in isolation, would not cause a significant distortion to the subsequent performance.

At this stage no analysis had been conducted into defining methods for selecting a balanced range of data sets with which to most effectively and rapidly train the system. It was therefore reasonable to, once again, assume that data dependency amongst the four training data sets would provide a variety of network solutions, some of which may be more suited to the test patterns than others. Thus, as was done previously, four separate data sets were provided for the evaluation to take account of the likely effects upon the performance of randomly selected additional data sets being added to the core data sets. By using four separate data sets it is not only possible to quantify these effects but also to account for them when drawing conclusions from the trials. Ultimately it also enables comparison of the contents of the respective data sets and their physical attributes to determine the causes of such data dependencies.

Figure 5.11 and Figure 5.12 below graphically illustrate the results of the training and classification trials with the various network configurations using each of the four extended training sets. Overall performance of the network was improved with the addition of the extra TES data in the training groups. Comparing network classification performance with the reduced data set, illustrated in Figure 5.9 above, and those obtained with the extended data sets, illustrated in Figure 5.11, it should be noted that the magnitude of this performance gain is between 0-16%. As predicted the effects of data set dependencies caused the enhancements to be separated into two distinct data set sub-categories.

When sets 2 and 4 are presented during training the resultant classification is improved by between 0-6.5% whilst with sets 1 and 3 the improvement in performance is in the region 11.3-16.6%. Essentially, this indicates that training sets 1 and 3 produce better network generalisation than sets 2 and 4. In spite of this performance gap it is evident that even without specific data selection the larger the training set, the better able the network is to correctly classify the unseen test data subsequently.

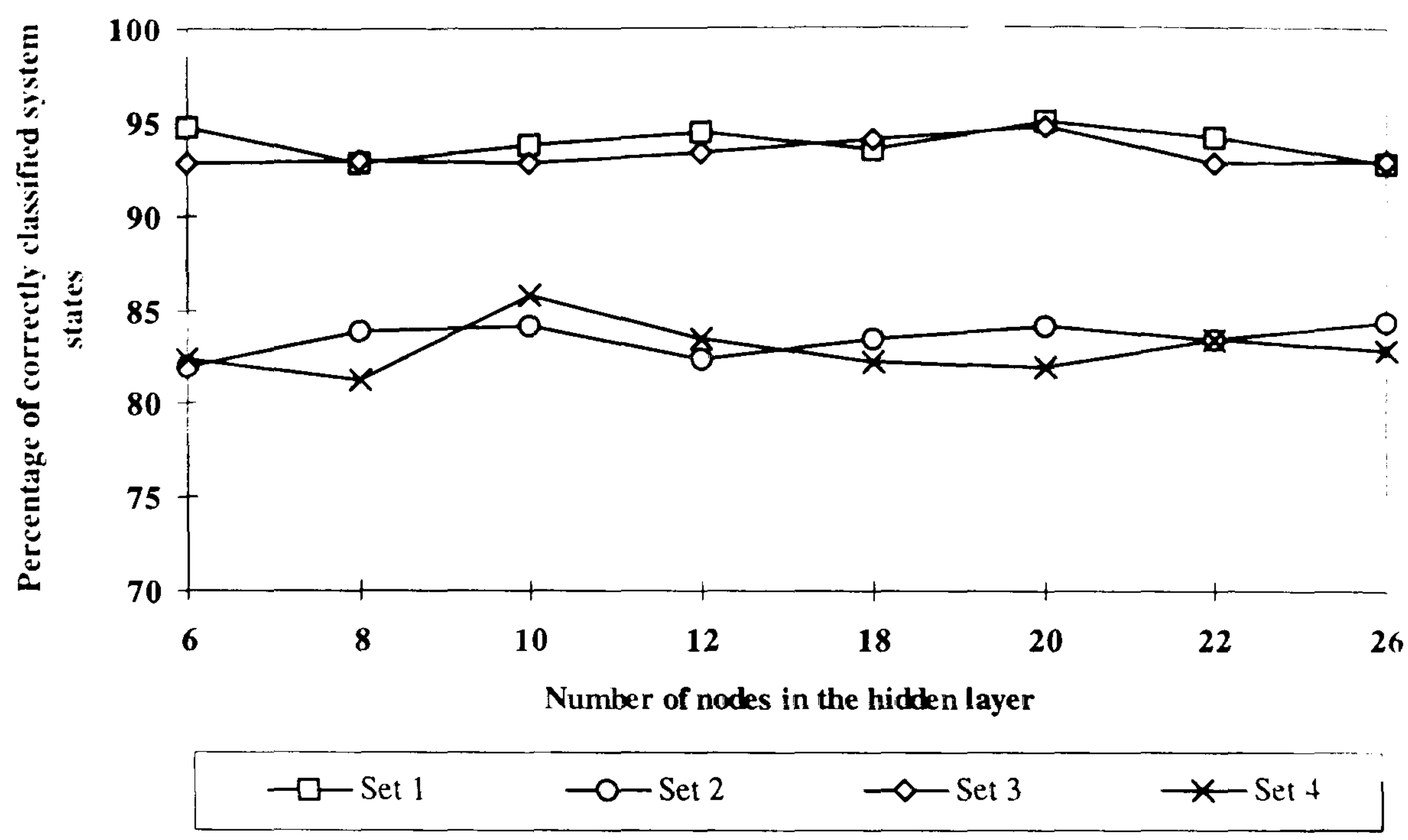

Figure 5-11 Classification of shaft alignment defects for various hidden layer network architectures following training with four extended data sets

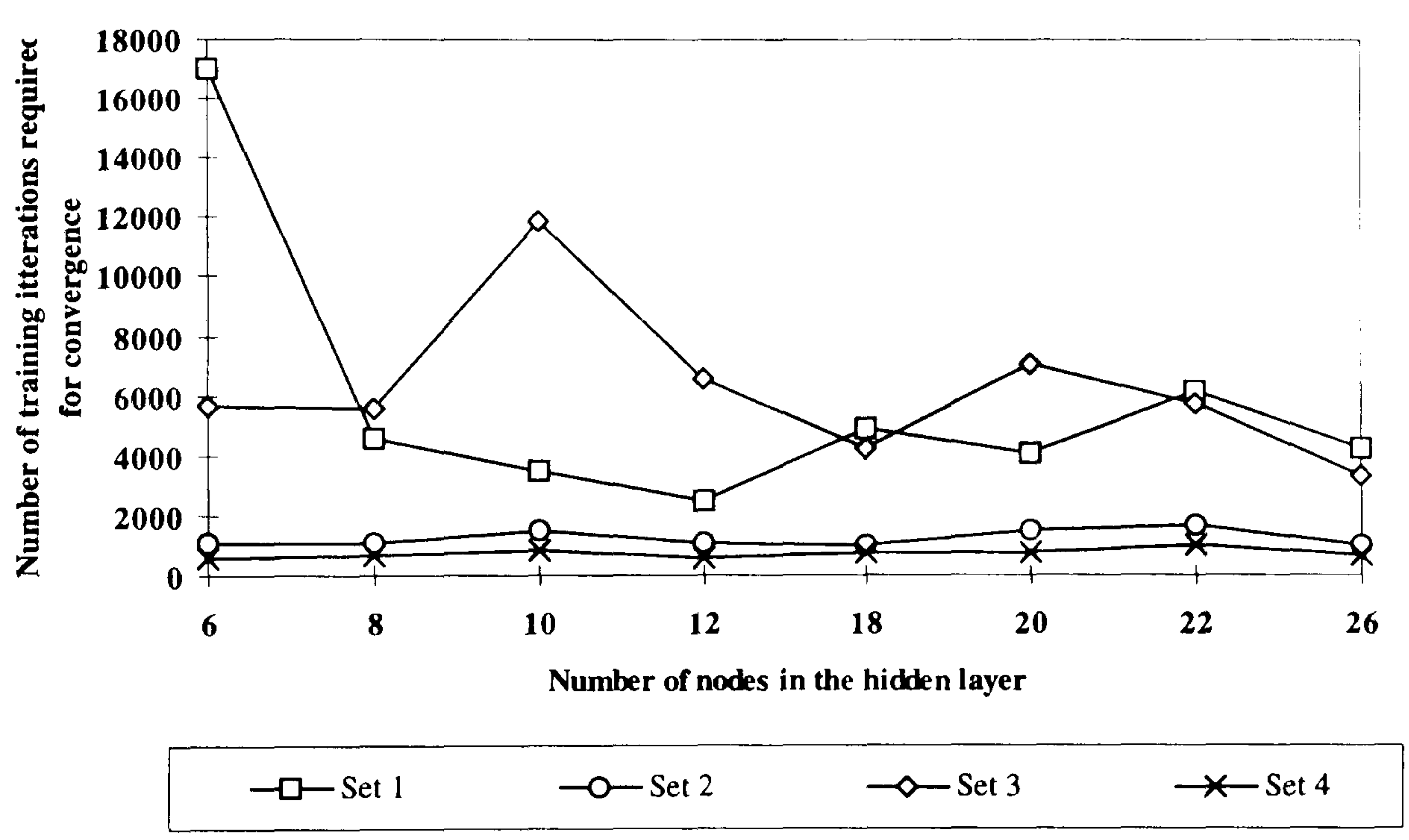

Figure 5-12 The network training characteristics for various hidden layer network architectures using the four extended training data sets

In these trials a 50% increase in training set size effectively guarantees a worst case error rate of 18.7% which is comparable to the best performance achieved with the smaller data set used earlier. However optimal performance was restricted to network configurations trained with sets 1 and 3 where approximately 95% of the unseen test patterns could be correctly identified with a 20 node hidden layer. With these two particular training sets all network configurations achieved better than 93% accuracy, a significant improvement over the core training set used during the earlier evaluations.

This improvement is however associated in all but 3 of the 40 network and data set combinations with a corresponding increase in the number of training iterations required to achieve the preset convergence bound. Since the time taken to train a network is a function not only of the number of iterations required to converge but also of the time taken for each iteration this period is extended still further by the larger data sets being used. However providing that the training periods are still reasonable, which they were in this instance, then the primary selection criteria should always remain the classification performance of the network rather than rapidity of training. Of the trials performed here only a 6 node hidden layer architecture trained using the set 1 extension and a 10 node configuration trained using the set 3 extension need necessarily be eliminated based upon their excessively demanding training requirements. As with the performance testing there are two distinctive performance sub-groups. Both sets 2 and 4, which produce a reduced classification performance take significantly less time to converge than sets 1 and 3 which take longer to converge but provide better separation when exposed to the test set. This direct correlation between the number of iterations required for convergence and the subsequent performance of the network is a further indication that the differences are caused by data dependencies.

5.2.6 Feasibility of Performance Optimisation Through Data Selection

Whilst early trials have shown that TES data presented to a neural network in amplitude-frequency histogram matrix format can provide a reasonable means of condition identification it has also emphasised the fact that the selection of suitable training data is important if performance is to be maximised. The difficulties of using randomly selected data to extend a training set have been highlighted by the discrepancies in performance achieved with the four different data extension sets applied in 5.2.5. Whilst a process of data selection should provide a more optimised solution to a given problem it also imposes additional demands upon the system operator. To achieve this solution either trials must be performed over a wide range of operating conditions or the parameters affecting performance within the data sets must be identified. It is necessary therefore to further quantify the benefits a prolonged training regime, consisting of performance analysis feedback, has when compared to an automated training scheme which employs random data of varying training benefit. To estimate this potential for improvement several additional data sets were developed from, and for comparison against, the extended sets used in trials described in the previous section.

The additional data sets which were used for the trials in section 5.2.5 were arranged into four further groupings, each containing a different mix of the original data sets. Sets 1 and 3 which boosted the performance of the core set previously were combined to determine whether the performance could be enhanced still further by combining these two "good" sets. Sets 2 and 4, which when added separately to the core set previously, had produced a less effective network solution were combined to see if performance could be improved simply by weight of data. A third set used in the trial contained all four of the extension sets, both good (1,3) and bad (2,4). Could this provide a more balanced and possibly better range of TES histogram matrices with which to train the network? The fourth training set used in this evaluation consisted of the two good extension sets combined with the core data set used previously. Would the addition of the second good extension set to one of the better sets used previously further improve the network solution?

As expected those training sets which contained the data from sets 1 and 3 once again provided the best training material for the various network configurations(see Figures 5.13-5.14). When combined the trained network was able to classify 94% of all test matrices correctly. This performance is comparable to the training data used previously in section 5.2.5 where sets 1 and 3 were used as a means of extending the core set. However this level of performance is achieved with only two thirds of the quantity of training data used previously. Allied to this the number of iterations of the back-propagation learning algorithm necessary to achieve such performance, between 2500-6000, proved to be more consistent over the range of network configurations than when they were employed simply as extension data to a core set.

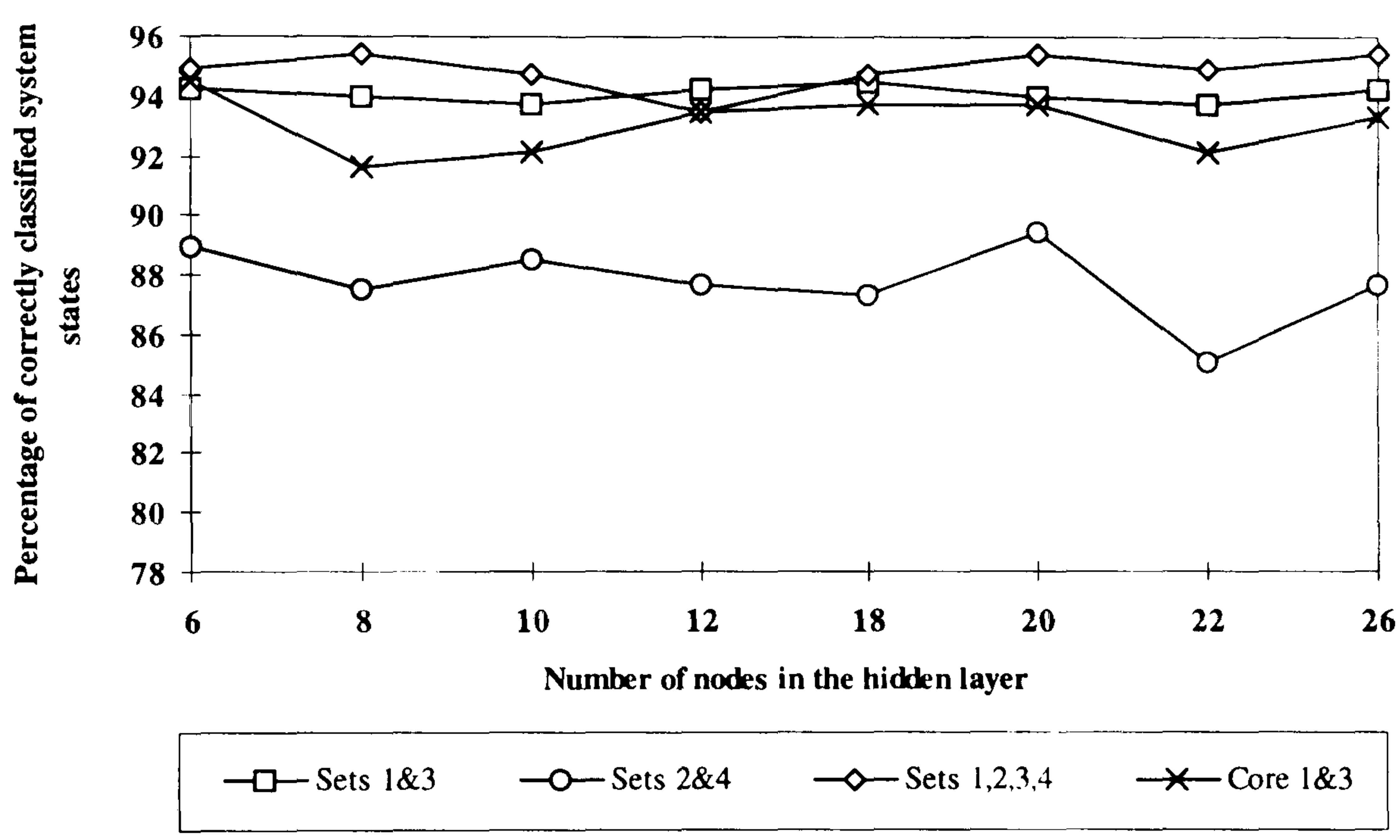

Figure 5-13 Network performance for various hidden layer architectures trained with a variety of selected training data sets

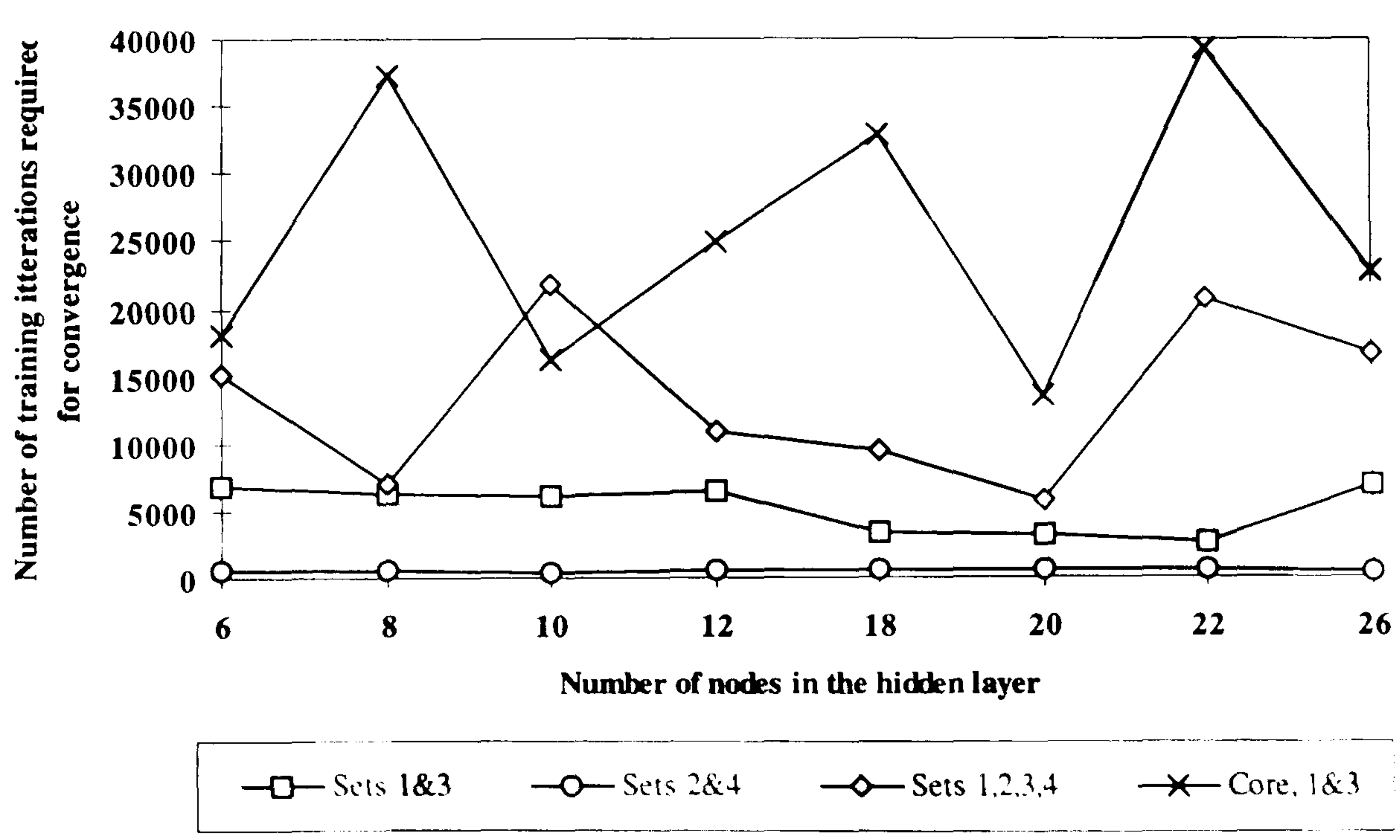

Figure 5-14 Illustration of the training requirements for each of the network configurations using the four selected training data sets

When sets 1 and 3 were simultaneously added to the core data set used in the previous trials the network showed no further improvement in state identification performance indicating that the data saturation point had probably been reached. This data set also seemed to result in the network requiring disproportionately long and also more erratic training periods to converge. The exact reason for this remains unknown. It is possible that the two data sets contained differing matrix anomalies which combine to drive the weight updates in a more erratic manner and thus necessitate additional data presentations to achieve the required convergence limit. If this had been the case though the consequent network unusually seems no more able to classify the test exemplars than either data set used in isolation which might be contrary to expectations.

When sets 2 and 4 were combined and used to train the various configurations the networks required substantially fewer training iterations, by a factor of at least 10, to achieve convergence. The trained networks which result from this however are only able to classify between 85-89% of the test set correctly. Whilst this still corresponds to an improvement of between 3-4% in classification performance and a 50-250% reduction in training requirements when compared against their use in conjunction with the core data set previously they are still less effective than sets 1 and 3. The data set which produced the optimal network solution during these trials however consisted of combining the four extension sets together. With this data set network performance varied between 93.5-95.5% depending upon the hidden layer configuration. However this 1.5% improvement, when compared with the set consisting only of 1 and 3, comes at the expense not only of a data set twice the size but also of a training program which requires up to eight times the number of data iterations to converge.

In conclusion, all four data sets provided a reasonable level of network performance when identifying the shaft displacement misalignment faults and whilst the selection of good data sets provides a more optimised solution acceptable performance can be achieved without necessarily resorting to a prolonged series of comparative testing and selection trials. This statement is backed up by the performance with the two extension sets which proved less effective previously. When combined these in fact provided reasonable performance whilst at the same time requiring less training time in which to achieve this than they had done earlier.

5.2.7 Impact of Dynamic Signal Properties on the TES Conversion Process

As discussed in Chapter 3 the conversion of an acoustic signal, acquired initially with a microphone, into an amplitude TES representation of this source requires determination of a normalisation factor. This factor is defined as being the maximum signal magnitude of the source attained over a given pre-conversion sampling period, which for practical trials was 10 seconds. After the initial pre-conversion period this normalisation factor is stored and then subsequently applied to all samples prior to TES conversion. Essentially, this causes each individual TES segment conversion associated with a single matrix to be performed with a unique dynamic range. This normalisation is essential to the conversion process if the data is to be used later for identification purposes. However utilising such a basic approach to the acquisition and conversion procedure in the interests of minimising system complexity and reducing cost has implications on the matrix generation phase. With this in mind it is essential to be aware of any adverse training effects induced as a result of the self-imposed dynamic variability which may significantly reduce the systems subsequent classification capability.

In an attempt to gauge the practical effects of this conversion technique a series of trials was performed in which the train and test group histogram matrix members were selected by analysing the normalisation factor characteristics used for each specific conversion from which they originated. Four different matrix data sets were used to train each network prior to evaluation against one of three test data sets generated. Normalisation factor based statistical representations of each of these train and test sets is illustrated in Figure 5.15 below; the horizontal axis is used only for plotting purposes. The normalisation factor axis identifies the relevant statistical coefficient for each data set component. Two neural architectures were employed to perform the tests, one containing an 8 node and the other a 20 node hidden layer. In both cases the remainder of the structure was unchanged from previous trials and consisted of a standard amplitude TES 300 element input layer allied once again to a four node output layer.

Initially the two networks were trained using training data sets 5 and 6 each containing histogram matrices corresponding to 21 minutes of acoustic record taken from the archive. Training set 5 was weighted with low biased matrix conversion coefficient components thus corresponding to a TES stream with a narrower dynamic range. In contrast set 6 was weighted with matrix components having higher biased conversion characteristics and thus a wider dynamic range. The primary test set, set *a*, which had high biased TES conversion characteristics and contained 7 minutes of histogram matrices was used initially to determine the baseline performance characteristics prior to evaluation of the remaining configurations classification performance. Whilst some network architecture and data dependency based variations in the respective convergence and classification performances is to be expected what Table 5.4 clearly identifies is what appears to be a wider divergence of classifications rates for the different training sets. Training set 5 seems to provide the back-propagation learning

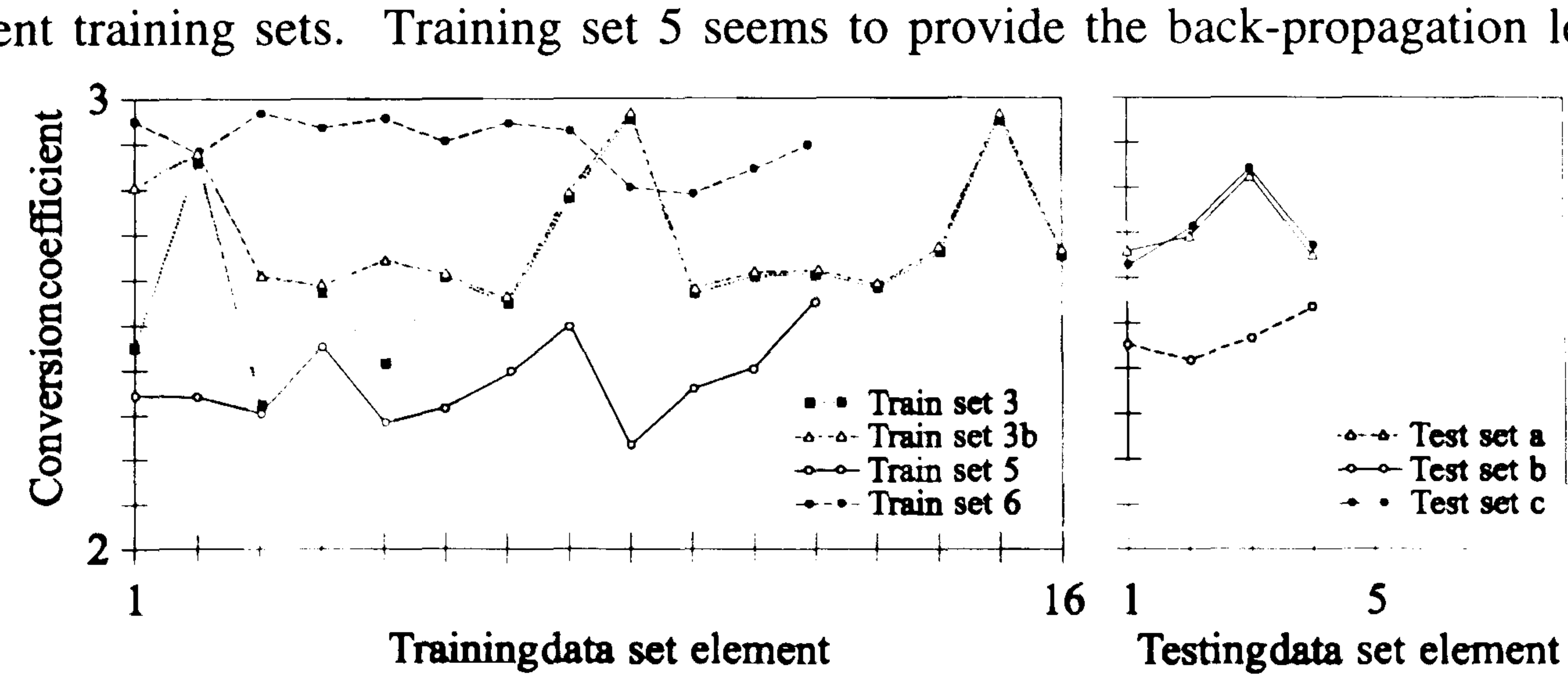

Figure 5-15 Normalisation factor characteristics for the four training and three test amplitude TES histogram matrix data sets

process with TES data which is statistically more dissimilar to that of set *a* characteristics, resulting in a poorly trained classification engine. In contrast training set 6 produces a network capable of correctly identifying approximately 93% of all test patterns presented from set *a*, albeit also requiring significantly longer to converge during the training phase.

TES histogram training set	Number of nodes in the hidden layer	Training iterations required for convergence	Correct classification of system state (%)
5	8	2732	75.7
5	20	984	74.8
6	8	8502	94.2
6	20	20592	93

TABLE 5.4 Results of network classification given two opposing normalisation bias based training sets analysed with test set *a*.

To ascertain whether this apparent performance mismatch can be partially rectified by statistical selection of data a second test set, set *b*, was generated to test the trained networks. The statistical composition of this test set, seen in Figure 5.15, resembles more closely the lower biased statistical distributions of training set 5. Somewhat expectedly, the performance, detailed in Table 5.5, is improved when this statistically more similar test set is employed for network evaluation. However in reality both training sets show improvement with training set 5 still apparently at a distinct disadvantage relative to the training data acquired using a wider dynamic range and contained in set 6.

TES histogram training set	Number of nodes in the hidden layer	Training iterations required for convergence	Correct classification of system state (%)
5	8	2732	78.6
5	20	984	76.7
6	8	8502	98.8
6	20	20592	99.5

TABLE 5.5 Results of network classification given two opposing normalisation bias based training sets analysed with the lower biased test set, set *b*.

Assuming for the moment that there is a correlation between the statistical balance of the data sets and subsequent performance in terms of the signal normalisation performed prior to TES conversion it is essential to attempt to quantify these effects. Consequently further comparative tests were performed against test set *a* using two additional training sets, set 3 and set 3*b* each containing 28 minutes of acoustic matrix tokens, slightly more than either set 5 or set 6 previously. The matrix data for set 3, the baseline set, was selected at random without regard to the statistical characteristics of the individual histogram matrix elements. This data was used to train two baseline classifiers, again one with an 8 node hidden layer and the other with a 20 node hidden layer. The second new training set, set $3b$, whilst based on the data contained in set 3, had some low biased components replaced by equivalent data components with higher bias characteristics. In this way if the normalisation factor is responsible for the performance anomalies then an improvement in classification performance would be expected when the controlled data of set $3b$ is used during the training phase rather than the randomly selected data in set 3. The results of this comparative test are illustrated in Figure 5.16. In each case networks were trained at three different α rates to minimise the likelihood of results being distorted by local network minima effects during training.

Considering the relatively minor alterations made to the composition of training set 3 any variation in performance induced by the altered characteristics of training set $3b$ data was expected to be small. However whilst the eight hidden node architecture does show signs of the type of improvement expected the 20 hidden node network displays characteristics which are contrary to those expected. During the course of further trials

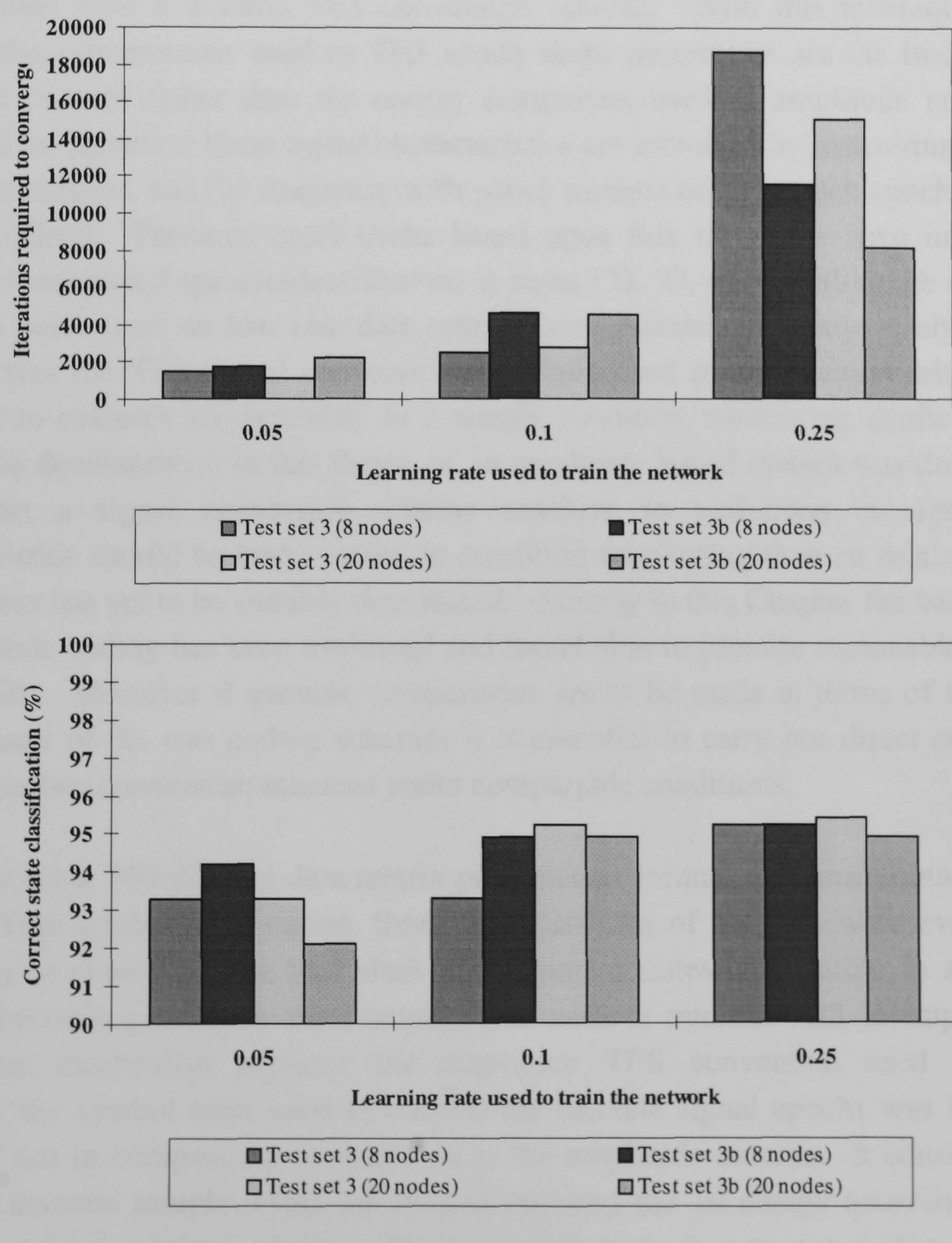

Figure 5-16 Results of network classification analysis with test set *a* using a randomly selected data set and an enhanced data set to train two networks at three different learning rates

to back these findings up similar small performance differences were noted. It would seem from these trials therefore that provided the training sets are not generated with statistically low conversion bias characteristics and contain a reasonable range of characteristics within the TES histogram matrices the classification performance will remain relatively unaffected by small fluctuations in subsequent test data characteristics. However the acoustic acquisition must ensure that the type of discrepancy seen between training sets 5 and 6 is eliminated since this will introduce additional unnecessary performance degradation.

5.3 An Evaluation of the Application of Minima TES Histogram Matrices to Shaft Alignment Classification

Prior to the development and application of an amplitude TES data format in this thesis all research into signal classification, with or without neural classifiers, had been implemented with a minima TES conversion scheme. With this technique the two primary data components used as TES epoch shape descriptors are the frequency and harmonic content rather than the energy component used in amplitude coding. As described in Chapter 3 these signal characteristics are extracted by measuring the epoch length, in samples, and the frequency with which minima occur in each epoch within the sampled signal. Previous applications based upon this technique have mainly been limited to automated speech identification systems [32, 37, 64, 65] although some work has been performed on low rate data transmission systems [33, 66]. Only relatively recently was the TES signal conversion technique used in conjunction with a neural classifier to evaluate its capability in a simple condition monitoring application [38]. Whilst the development, in this thesis, of an amplitude based system was driven by the belief that a signal conversion scheme sensitive to variations in signal energy characteristics should be better suited to condition monitoring than an minima focused scheme this has yet to be suitably determined. Already in this Chapter the basic concept of amplitude coding has been evaluated and found able to provide reasonable condition information. However if genuine comparisons are to be made in terms of the relative performance of the two coding schemes it is essential to carry out direct comparative tests of the two conversion schemes under comparable conditions.

Using the same 300 element data matrix presentation format, this time containing basic minima TES symbol information, three configurations of network were evaluated for suitability to classifying the four shaft misalignment states used earlier in section 5.2. Whilst the table generation itself remains the same a minima TES lookup table and conversion mechanism replaces the amplitude TES conversion used previously. Similarly the symbol table used to convert the discrete signal epochs was identical in format if not in composition to that used in the amplitude scheme. It consisted of the same 30 discrete sample levels but instead replaced the 10 energy quantisation levels with 10 unique minima quanta. During initial tests it was noted that the matrix variability produced by the four gearbox states using this scheme was more limited than the earlier amplitude matrices. The number of individual symbols required in the conversion process for all states was similarly constrained, producing when visualised, matrix maps with much reduced contour variations. Even prior to performing any performance comparisons it was felt that this reduced matrix diversity between states would not only affect performance but increase the network training times.

The trials were performed using identical acoustic condition recordings taken from the archive and used previously to evaluate the amplitude histogram classification mechanism. The recordings were used to generate four training sets compiled for these trials each containing approximately 28 minutes of minima histogram matrix data. Two different network configurations, containing 10 and 20 nodes respectively in the hidden layer, were employed for the tests. As previously, each configuration was trained at a variety of different training rates to minimise the likelihood of local minima effecting performance during the evaluation phase. The momentum factor, β, and error bound used for each of the training runs was the same as used during the earlier trials with amplitude data.

As initially surmised when the minima matrices were used to train each of the network configurations it was found that training times were indeed prolonged substantially when compared to the equivalent amplitude data matrices. In fact in several cases the network training phase was halted without convergence being achieved. Two configurations which eventually did converge provided acceptable state separation with little sign of the expected degradation in performance. The results of the tests, when four 7 minute test data sets were presented to the trained networks are illustrated in Figure 5.17. The 10 node hidden layer configuration required over half a million training iterations to achieve convergence whilst the twenty node configuration required approximately 45500 iterations. Both of these examples are unacceptably long for practical applications when compared to what can be achieved using amplitude conversion. Whilst it is possible that some of the other architectures or training parameter variations which were halted prematurely would also eventually have

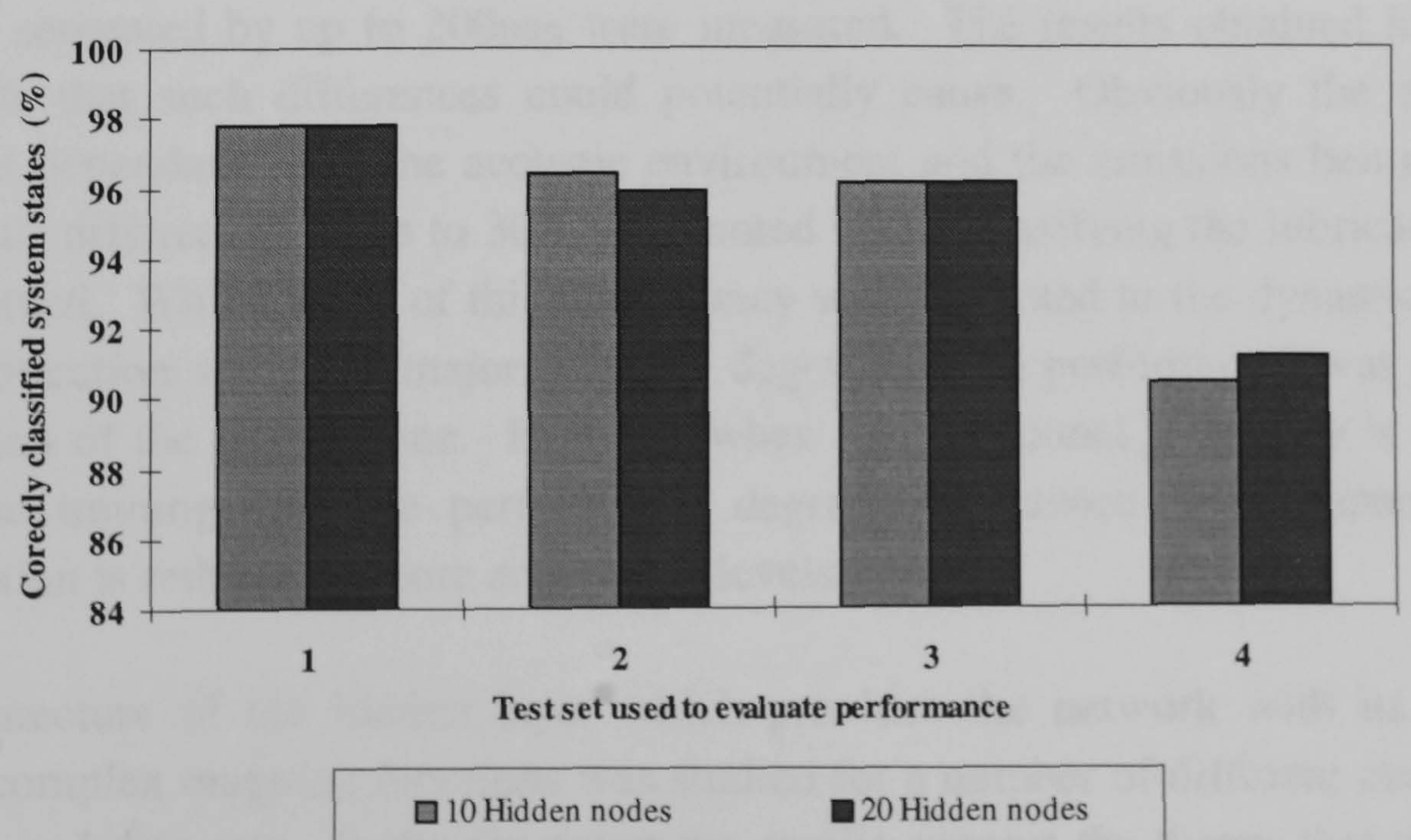

Figure 5-17 Results of classification analysis tests performed on three network configurations trained using simple minima histogram matrix data

produced converged networks they would be similarly unacceptable for a practical application. Therefore, despite the positive comparison in terms of the classification performance of the two examples which did eventually converge against a similar network trained using the amplitude histogram matrices the technique is not considered to be viable for use in a system implementation for this particular set of condition states.

5.4 Chapter Summary

This Chapter has described the concepts which were used in the design of the gearbox testbed system used to simulate the mechanical faults which are used to evaluate some TES based monitoring techniques. This covered the description of a series of representative physical states corresponding to displacement, and tooth wear as well as the crude simulation of lubrication status. The design, though simple, provided an ideal means by which acoustic data could be acquired relatively easily to provide a stable baseline for the TES matrix techniques to be evaluated.

Initial investigations centred on the application of basic histogram matrices generated using the novel amplitude TES conversion technique and were applied to three layer MLPs firstly to measure their ability to recognise four simple shaft velocity states. Even with small training data sets of the order of 40-80 seconds correct identification of the velocity status was achieved at least 82.5% of the time. Depending upon the data sets used for training this figure could be improved to 100%. These early results provided the impetus for examining the technique still further.

At an early stage of the discussion regarding the use of acoustic data the question of sensitivity was raised. Since one of the primary reasons for adopting the use of an acoustic data acquisition scheme was its potential simplicity the effects upon the classification mechanism of a loosely controlled microphone position had to be studied. In trials the effects of selecting training and test matrix tokens from microphone positions separated by up to 200mm were measured. The results obtained highlighted the effects that such differences could potentially cause. Obviously the effects are somewhat dependant upon the acoustic environment and the emissions being recorded but in trials differences of up to 30% were noted when classifying the lubrication status of the testbed. Whilst some of this discrepancy was attributed to the dynamic variation of the lubrication states the majority of the degradation in performance was caused by the location of the microphone. However when this positional flexibility is accounted for in the training data the performance degradation caused by movement during classification is reduced to more acceptable levels.

The architecture of the hidden layer which provides the network with its ability to perform complex mapping functions was studied for a number of different classification problems and data sets. In the most part the results support the theory that there is not an ideal configuration for a given problem unless the data sets are fixed. Instead, provided that the network has sufficient hidden layer nodes for the problem the addition of further nodes is unlikely to affect significantly the performance. Optimisation of the architecture of this central layer is thus better performed in terms of the computational overhead required to implement a particular network configuration.

The identification of shaft displacement status proved, as expected, to be more demanding than classification of the velocity status. Initial trails used 14 minutes of acoustic matrix tokens to train a basic network which was subsequently able to correctly identify 82% of tokens in a test set. Further extension of the training data set by one third improved the performance under test conditions to approximately 93% depending upon the specific training set used. The disparity in network performance, depending upon the training data used, varied but in the worst case amounted to a variance of approximately 10%. Further trials were carried out to evaluate the potential of data selection as a means to further enhance the performance. Whilst practical evaluation could be applied to the selection of data sets it was found that similar improvements could be achieved simply by extending further the size of the data sets used during the training phase. During the course of these trials successful classification of 95% of a test set was possible with an extended data set where 94% could be achieved with a smaller but selected set. In most applications the extension of the training data set would be more desirable and easier to implement than the addition of a further evaluation phase to the practical application of the technique.

The final aspect of the application of an amplitude TES conversion scheme which was studied was the performance of the classifier under varying conversion conditions. The effects of dynamic variations in the normalisation coefficients applied to the discrete signal prior to symbol conversion was discussed earlier as a potential difficulty. In practice the performance statistics can be affected by up to 20% when the training data sets are weighted with matrices generated with low bias normalisation characteristics. However as with the microphone sensitivity considerations, provided that the data sets are neutrally biased the performance remains acceptable.

In the final section of the Chapter the application of minima histograms was evaluated. Unlike the amplitude technique there was little evidence to suggest that this method was suitable for the identification of physical state in practice. Whilst the technique was able to perform reasonably under certain circumstances the training performance was very erratic and in many cases the networks were unable to converge. This would be totally unacceptable in a practical system implementation. Of the two simple matrix types used to provide signal condition information in this Chapter it is reasonable to conclude that the amplitude based matrix types studied provide reasonable accuracy with networks that learn at an acceptable rate to be considered for further system evaluation. It is also reasonable to conclude that the simplified minima TES data format is unable to provide an acceptable balance between classification capability and learning performance. Both of these simple techniques are however considerably more basic than the A-matrix data format employed by previous researchers for the purposes of automated signal identification.

Chapter 6

6. The Application of A-matrix Data to the Classification of System State

Having evaluated the suitability of the basic histogram matrix formats in Chapter 5 it is essential now to consider the extension of the signal conversion process to encompass the additional shape information imparted by converting a raw TES stream into an A-matrix format. The detailed mathematical description of this conversion process was presented in section 3.5.2 of Chapter 3. Whilst this conversion technique requires only modest additional computational overhead when compared to the basic histogram matrix format discussed in the previous Chapter it should in theory improve the subsequent classification. This Chapter considers not only the practical application of the technique, in both minima and amplitude conversion forms, but seeks also to identify whether or not the expected improvement can realistically be achieved. For the purposes of this thesis it is also important to identify whether the expected improvement in classification can be achieved without recourse to excessively large neural classifiers or complex data analysis and the additional computational overhead that this would necessitate.

The primary requirement prior to converting a TES symbol stream into an A-matrix format, independent of the parameters being used in the conversion, is to reduce the symbol set required to represent the signal. This is crucial to the feasibility of applying the subsequent data to a neural classifier since the number of elements in an A-matrix is the square of the number of symbols in the symbol set used for conversion. If the symbol set, in either amplitude or minima formats, were used in their initial fully populated state the matrices presented to the neural classifier would require 300^2, or 90,000, elements. which would impose an unacceptable computational burden on any implementation of network classifier. Of course any reduction in the symbol set used for conversion introduces further distortion to the converted signal. As has already been discussed in Chapter 3 the key to successful optimisation of the symbol table is to ensure that any additional distortion remains at an acceptable level so that sufficient signal information is retained whilst the computational requirements imposed on any classifier are minimised. In the practical trials discussed in this Chapter minima symbol tables containing 30 entries and amplitude symbol tables consisting of 40 entries were found to provide adequate signal information without incurring excessive additional distortion.

The symbol selection necessary to reduce the tables down to only 30-40 entries is essentially a statistical exercise necessitating individual symbols in a test TES stream to be monitored. The necessary analysis, taking as an input a raw TES symbol stream, was performed automatically using a custom utility which extracted and presented the statistical signal details. Practical application of this process, for both conversion types, necessitated a series of statistical symbol plots to be acquired over a range of condition states to eliminate symbol biasing for individual states. Once the data has been extracted from a symbol stream a global symbol allocation threshold can then be applied to the symbol set. Only if a symbol is generated by the TES converter at a rate equal to or greater than this threshold is it included in the new A-matrix symbol set. This initial stage reduces drastically the size of the required symbol table. However since each symbol within the original 300 entry table has a finite probability of occurrence it is necessary to ensure that even those symbols which do not surpass the threshold occurrence frequency are catered for. This is efficiently achieved by means of a nearest neighbour symbol allocation, the details of which were discussed in Chapter 3.

The practical application of this process to the symbol table is best explained by taking as an example the provision of an A-matrix symbol set for use in an amplitude based coding scheme. Initially a fully populated 300 element symbol table similar to the one illustrated in Figure 6.1 was used to convert a series of acoustic samples acquired from the gearbox testbed in each of four displacement misalignment configurations to produce sufficient sample data for symbol optimisation. By placing the allocation threshold limit, in this example set at 0.25% for the amplitude TES format, only those symbols shaded in Figure 6.1 surpassed the threshold and warranted immediate inclusion in the final symbol table. During trials for the displacement states this group of symbols constituted roughly 74% of all symbol types generated by the converter during operation, the remaining 26% of symbols falling below the required inclusion threshold. Having such a large minority of infrequent symbols spread over the remaining table elements however poses a secondary problem. Although each of the remaining symbols only had a likelihood of occurrence in the region of 0.1% they were still able to swamp the majority symbols had not careful consideration been given to the manner of table restructuring. It was therefore considered vital to represent them in a manner which minimised the likelihood of distortions being caused by such grouping.

In addition to ensuring that reallocation did not introduce such unwanted symbol swamping effects during this secondary optimisation stage it was also necessary to give consideration to minimising symbol warping both in the frequency and energy domains. The compromise solution eventually selected, the final table for which is presented in Figure 6.2, consisted of reducing the individual shaded codes on the left side of the table by introducing distortions of up to one energy level for those elements which surpassed the threshold. Two additional codes, 6 and 17, catered for the remaining elements on

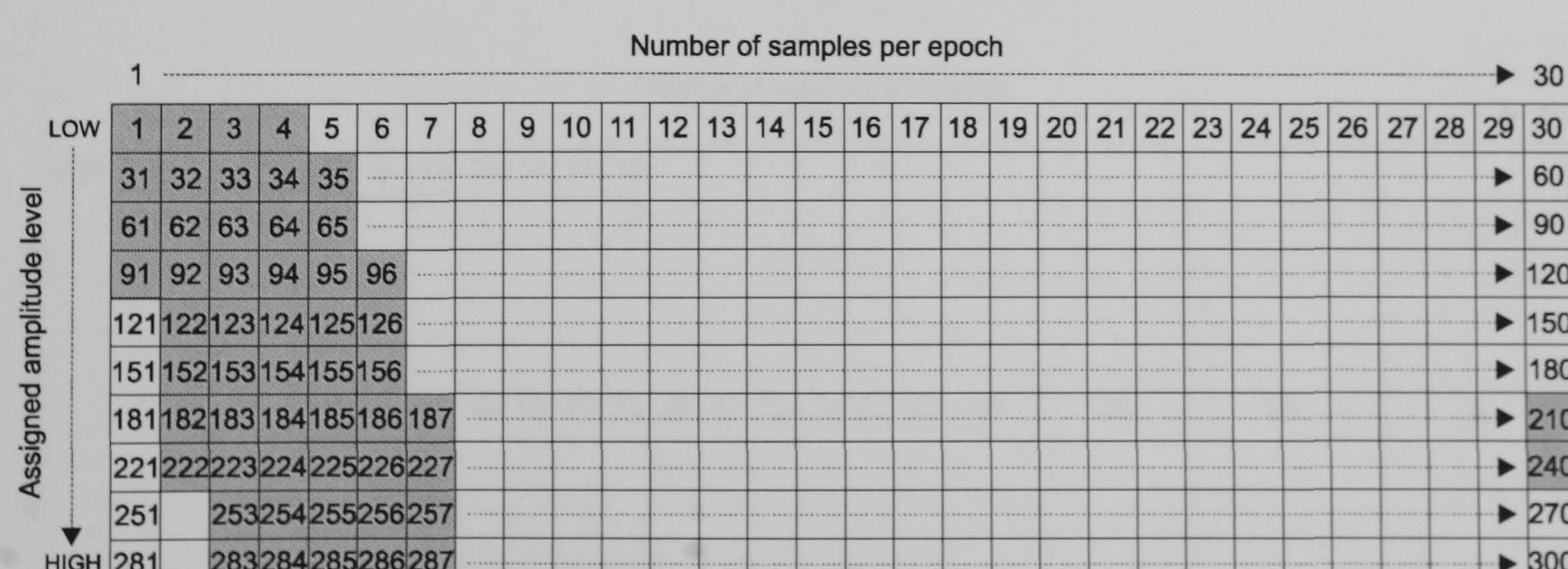

Figure 6-1 Primary symbol allocation table for an amplitude coding scheme consisting of 10 energy and 30 sample quantisation levels. Shaded elements identify all symbols having a frequency of occurence of 0.25% or greater

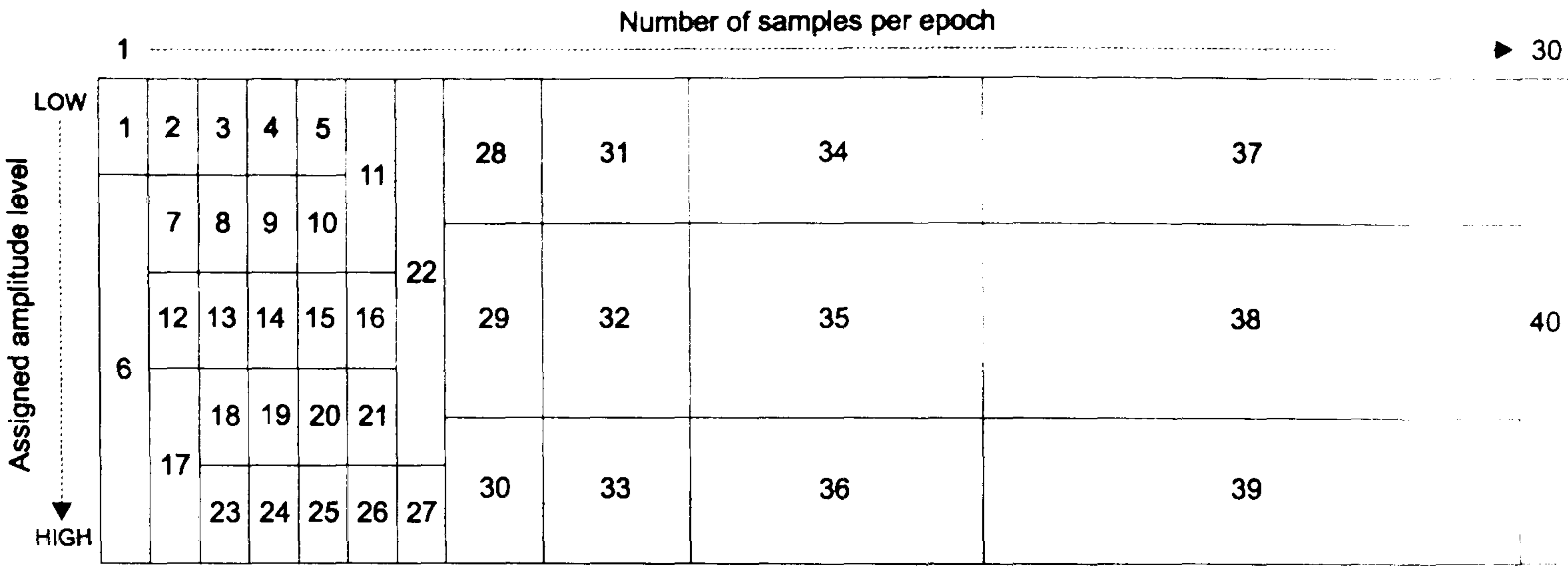

Figure 6-2 An optimised 40 element symbol allocation table for an amplitude coding scheme

the left side of the table which had not exceeded the symbol threshold. A single symbol was used to describe the right hand column of the table which was occupied, in the most part, by symbols generated as a result of low frequency sample truncation by the coder. As such this symbol in fact provides a good indication of the low frequency content of the source signal. The remaining 220 table elements were allocated to 12 symbols, the boundaries of which were selected so as to enable a crude level of energy information to be conveyed whilst at the same time minimising the frequency warping. The discrete energy levels were separated into high, medium and low quanta whilst the frequency axis was subdivided into four non-linearly separated subgroups which provide optimised correction for the effects of frequency warping.

In this final table the symbols on the left which were originally marked for immediate inclusion vary in allocation probability up to approximately 9.5%. The reallocated entries in the central region which are composed of the lower frequency elements not originally assigned have occurrence probabilities in the region of 1-2%. This 40 element symbol table was found to provide an adequate balance between minimising the symbol set and thereby the conversion distortion and maximising the information content in the symbol stream.

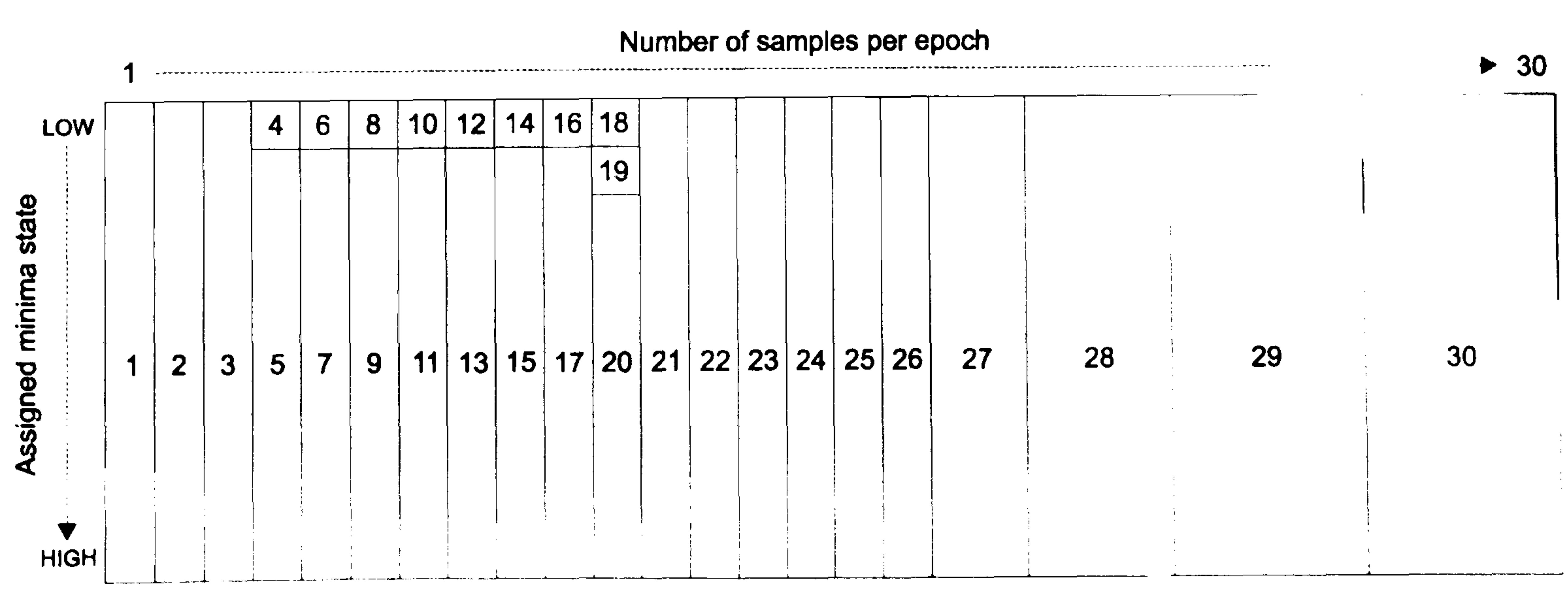

Figure 6-3 An optimised 30 element symbol allocation table for a minima coding scheme

All these same principles can be applied to the analysis of the symbol stream data in a minima based coding scheme. The optimised symbol table generated from signals converted using this coding scheme, seen in Figure 6.3, has a slightly different format due mainly to the lower probability of minima occurring in the discrete signal epochs. However the same technique of sample distortion reduction is employed at the right hand side of the table. Unlike the amplitude implementation which required a 40 element conversion table, adequate performance can be attained in a minima based TES system with only 30 symbols in the conversion table.

6.1 A Comparative Study of the Two Minima Data Presentation Types

The application of minima data, in its basic histogram matrix format, discussed previously in Chapter 5, to three different network architectures proved relatively unsuccessful. Whilst direct comparisons with amplitude TES did, in some cases, prove reasonable in terms of classification performance they did not provide adequate insurance of network convergence within acceptable time limits. This was essentially a by-product of the reduced symbol diversity in the minima symbol stream generated from the acoustic testbed for each of the simulated mechanical faults. However this reduced diversity makes simpler the generation of a suitable symbol table for an A-matrix conversion scheme. As illustrated in Figure 6.3 a compact symbol table containing only 30 unique symbols proves more than adequate. Moreover, the A-matrix presentation model, whilst more complex to derive than the earlier histogram type, conveys additional signal shape detail which in the simpler format was discarded. This additional detail may be sufficient to overcome the basic lack of symbol diversity observed previously. To identify the improvements, if any, that this additional shape detail provides over the more rudimentary format a series of comparative trials were performed. If the research carried out by Vu *et al* [38] which focused upon diesel engine state identification is used as a benchmark then a 30 code table should indeed provide sufficient source detail to enable suitable state separation. The key questions which the trials sought to answer were whether the technique can fulfil the separation requirements sufficiently and if so whether the rate of network learning improves as expected compared with the earlier histogram format.

To enable effective comparisons to be made between these two minima presentation formats several key elements of the trials already detailed in Section 5.3 were retained. The basic network architecture was retained but with an input layer extended from 300 to 900 elements to accommodate the expansion of the presentation matrices. Both the number of elements within the hidden layer and the number of unique states to be identified were retained. Further to this the acoustic state samples themselves were taken from the same archived recordings of the four gearbox states used for the histogram evaluations earlier. These recordings were used to generate two new training sets, each consisting of 28 minutes of A-matrix formatted data, with similar basic statistical acquisition characteristics to those used in the earlier tests. As with the previous evaluations the networks were trained at a range of different learning rates to minimise the likelihood of local minima disturbing the results. Four test sets were also selected, containing approximately 7 minutes of A-matrix data, with which to grade the performance of the classifiers using this enhanced conditioning technique. The results of trials on two network configurations are detailed in Figures 6.4 to 6.8. There are several observations to be made regarding the performance of each of these networks when trained and tested on the various combinations of data compiled. Neither the training rate parameter, α, nor the different data sets used with the networks radically affected their subsequent performance, although clearly in the case of training set 2 an α rate of 0.4 did not result in network convergence for either configuration within a reasonable time period.

The effect of the α parameter on the rate at which the network converges is illustrated in Figure 6.8. As expected the relationship between the training rate and convergence shows reasonable proportionality although the associated training data also plays a part

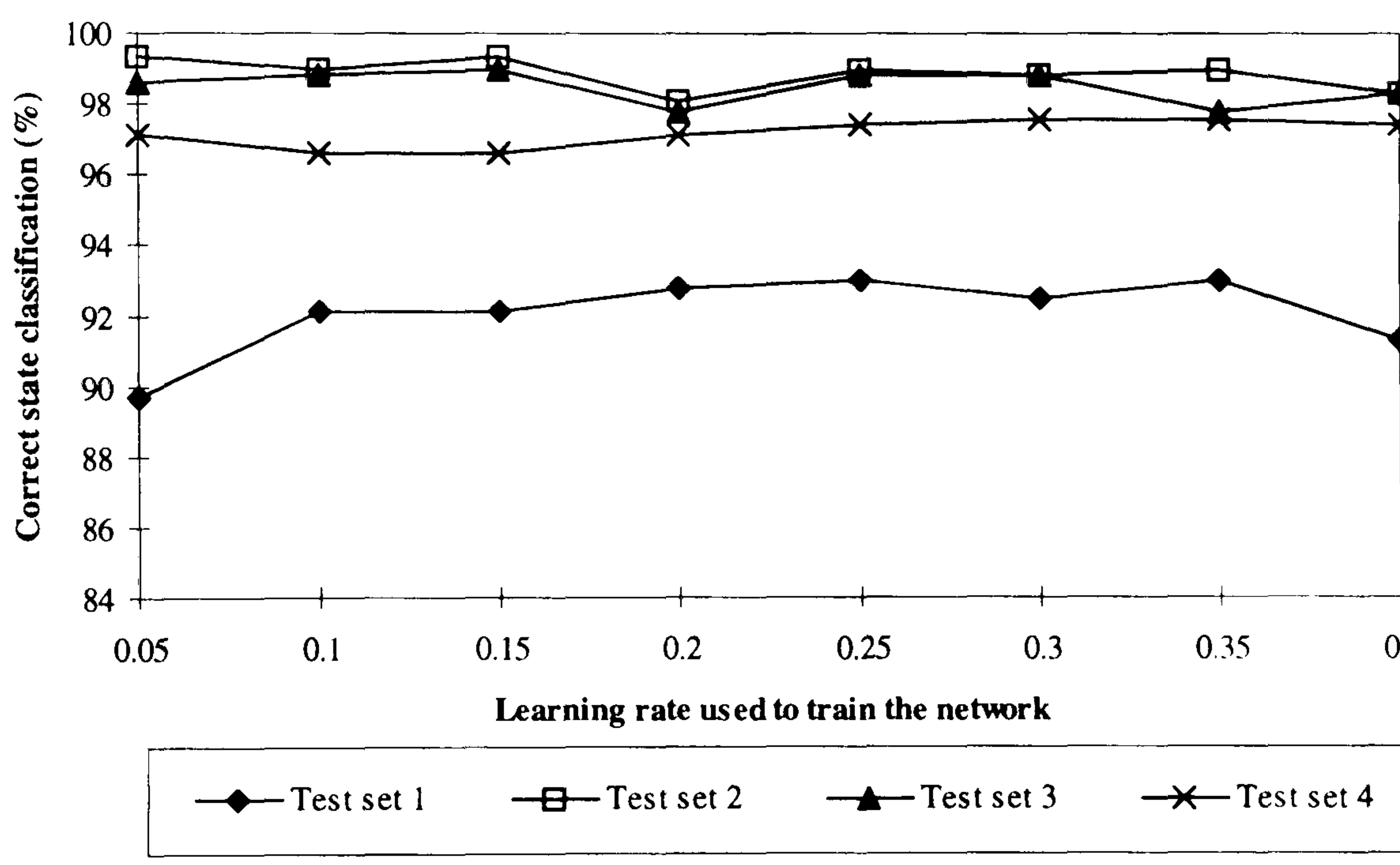

Figure 6-4 Classification for a 900-10-4 network configuration trained on data set 1 and tested against each of the four separate test sets

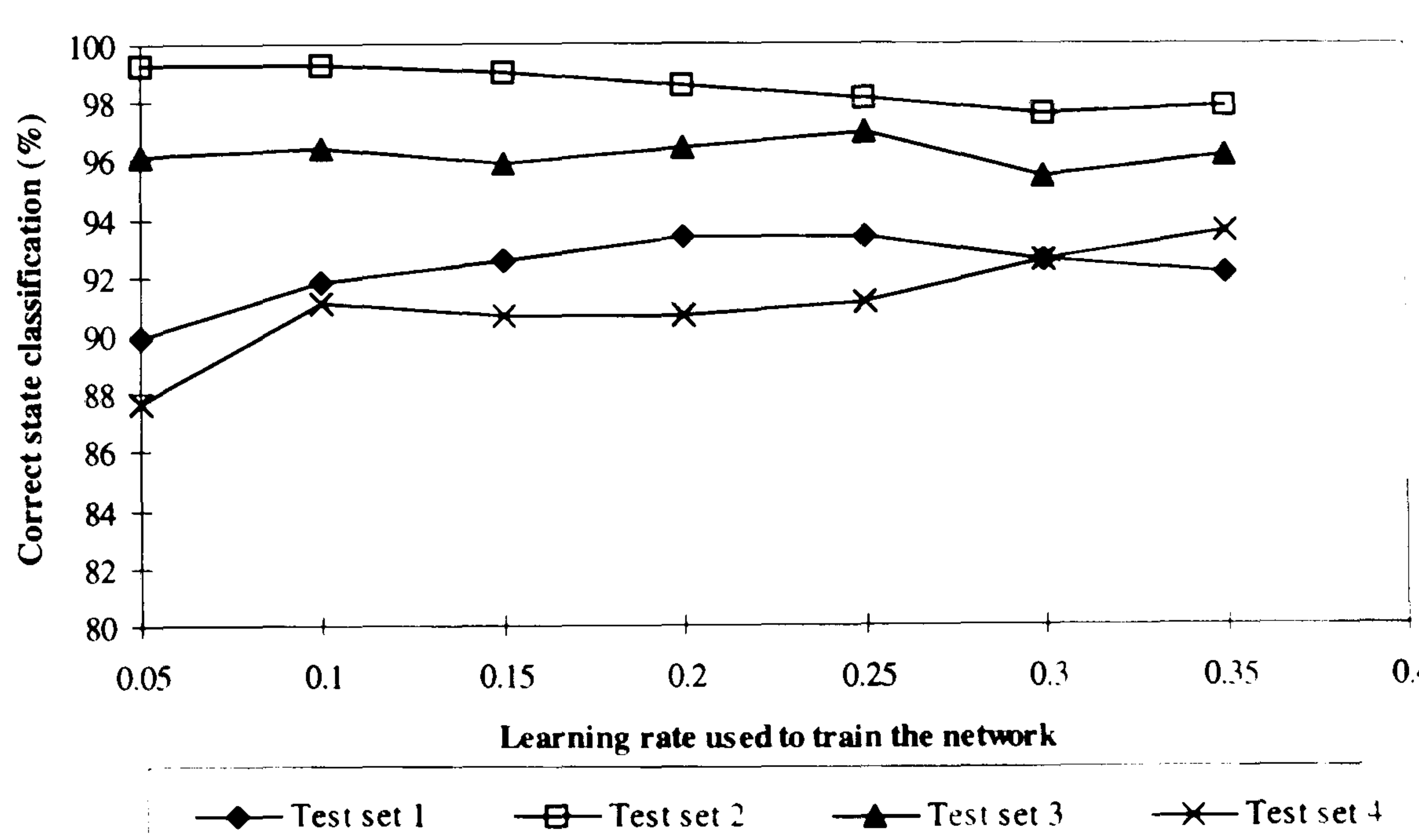

Figure 6-5 Classification for a 900-10-4 network configuration trained on data set 2 and tested against each of the four separate test sets

in defining the overall weight path taken during training. The variation in convergence times seen for $\alpha>0.25$ in some configurations (Figure 6.8) is most probably caused by oscillatory and erratic behaviour induced in the weight updates by the increased step size. A rate of between 0.05 and 0.2 provides acceptable and more predictable performance when combined with the momentum parameter, $\beta=0.95$. For α values above 0.2 the convergence becomes less predictable and often excessively prolonged. In the most extreme case identified here, a 10 hidden node layer architecture trained on test set 2 at $\alpha=0.35$ required nearly 168,000 training iterations to converge. This degraded still further to the point where at $\alpha=0.4$ the network training was halted prior to achieving convergence after 200,000 iterations.

However if the network performance, in those combinations which did converge, is compared there is little to differentiate between a network requiring 1,000 training iterations and one requiring 10,000. The most obvious visual difference in performance

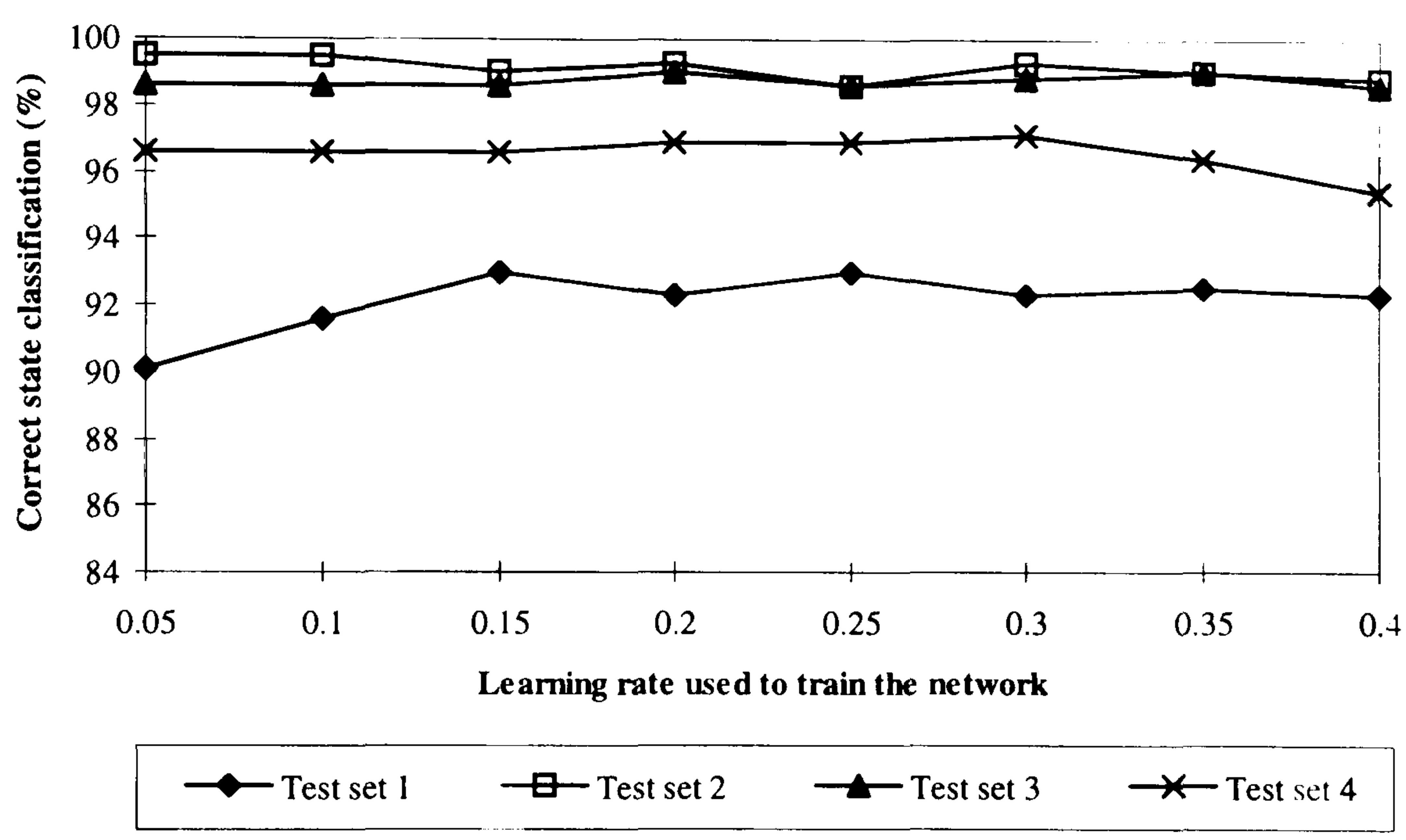

Figure 6-6 Classification for a 900-20-4 network configuration trained on data set 1and tested against each of the four separate test sets

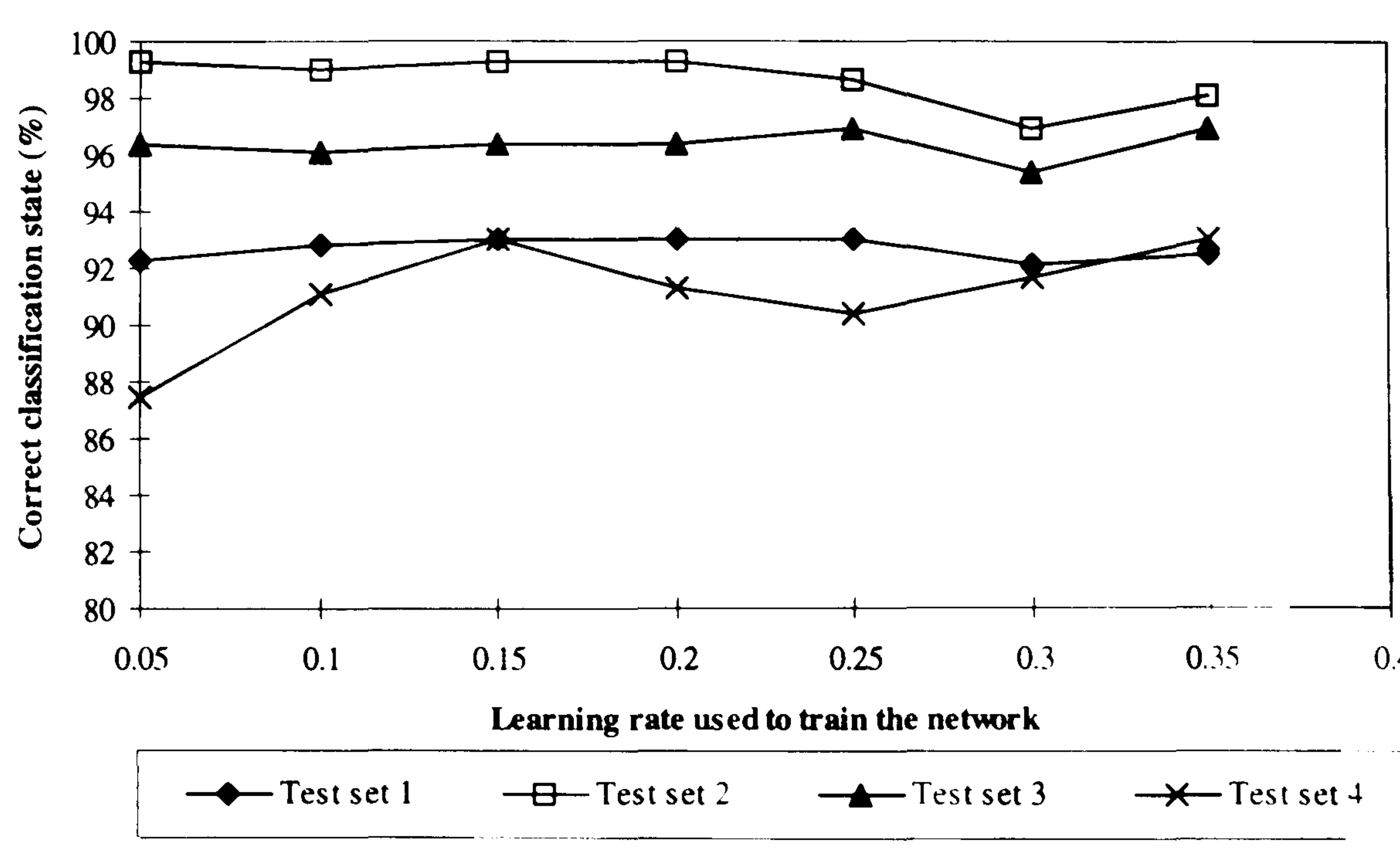

Figure 6-7 Classification for a 900-20-4 network configuration trained on data set 2 and tested against each of the four separate test sets

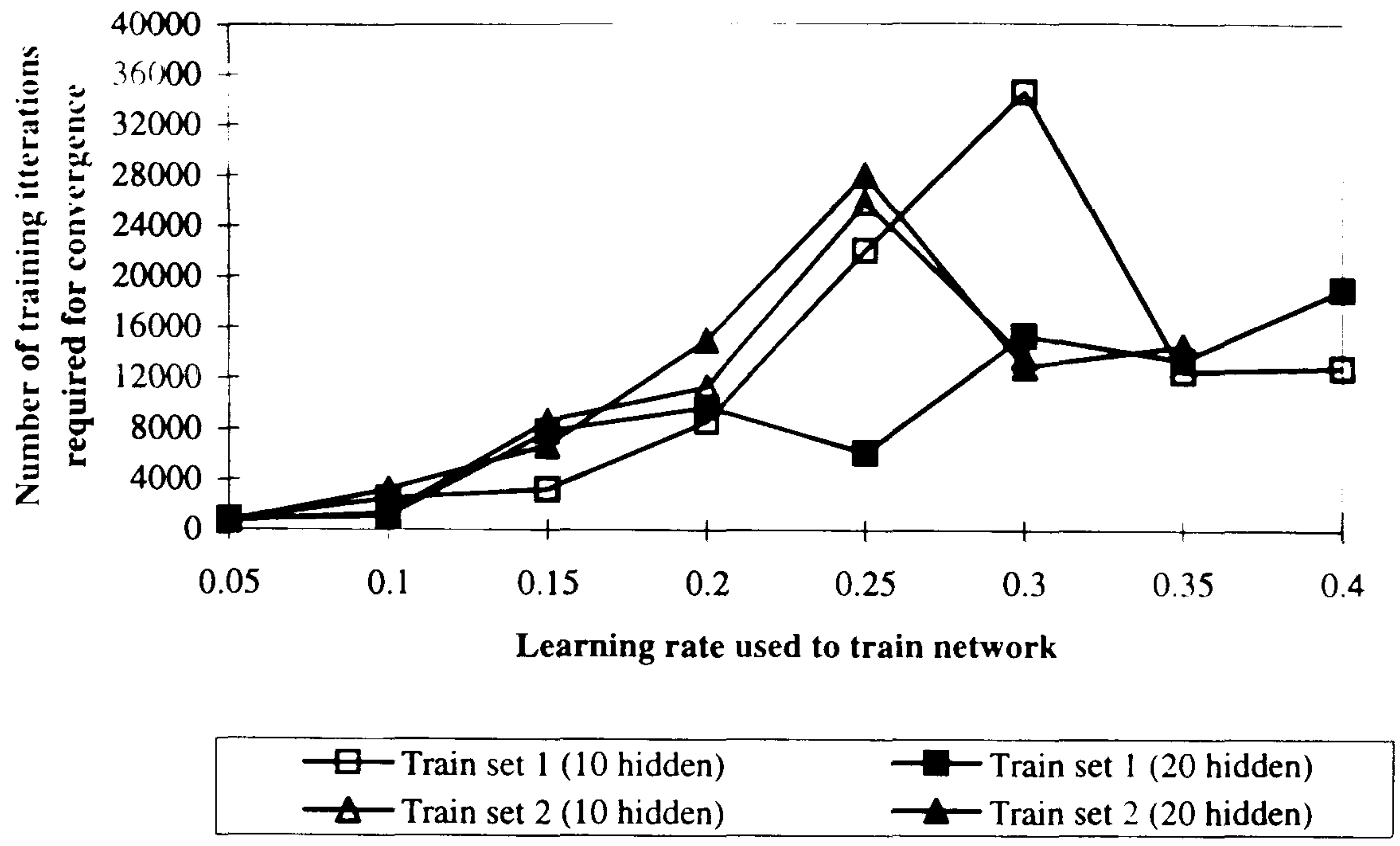

Figure 6-8 Rate of network convergence for each of the four network configuration, minima A-matrix training data set combinations used

characteristics illustrated in Figures 6.4-6.7 are the variations caused by the data diversity in the various training and test sets. Whilst test set 1 resulted in the worst performance, 90-92% over the range, test sets 2 and 3 gave the best performance with correct classification rates of between 96% and 99.5%. The best overall performance was produced by the 10 node hidden layer architecture with rates of 98-99.5%. Test set 4 highlighted the inconsistencies in training data sets, resulting in 97% performance when trained with set 1 or 88% performance when trained with set 2.

From this comparison of the two minima network data presentation types it is clear that, although ultimately classification performance remains reasonably comparable, the training requirements are considerably less onerous when the A-matrix format is used. Comparing the case of a 10 node hidden layer network presented with histogram data and one presented with A-matrix data this reduction is substantial, from 500,000 to only 1,000 iterations. Thus as expected the technique does overcome the difficulties faced by the lack of symbol diversity identified in the previously discussed histogram matrix trials. The additional shape information has also enhanced the training phase which becomes much less demanding of processor overhead than the earlier histogram driven networks proved to be.

6.2 An Evaluation of Amplitude A-matrix Conditioning on Misalignment Classification

Whilst the basic minima histogram matrices produced reasonable performance in instances where network convergence was achieved the convergence proved erratic. The erratic behaviour was reduced through the introduction of a reduced symbol set and with it the inclusion of additional signal shape information using the A-matrix conversion algorithm. This presentation scheme provides comparable classification performance coupled with the added advantage of more predictable and less onerous training requirements. By comparison amplitude TES data, even in its basic histogram format, provides reasonable performance coupled with acceptable training requirements. The possibility of further improvements in both the training and classification phases through the introduction of an amplitude A-matrix format seemed not only feasible but very likely.

The overriding drawback of this technique is the comparative increase in the network size required to implement the classification phase of the scheme. Whilst both of the basic histogram formats require a 300 element input layer and minima A-matrices require a 900 element structure the amplitude A-matrix format requires 1,600 elements. This five fold increase over histogram data requirements is a direct result of the square relationship between matrix dimensions and symbol table size. However given the availability of relatively low cost computational power even this obstacle would be acceptable if performance can be improved. Assuming that the presumption concerning computational resources can be met the single most important aspect of the techniques evaluation should be the training requirements necessary for its practical implementation and the subsequent capability of the classifier so produced.

As with all the other comparative tests performed so far this evaluation was carried out using acoustic source samples of the four gearbox displacement misalignment states. The training sets each contained 28 minutes of matrix data converted using the statistically generated 40 element code table described at the start of this Chapter. Four test sets, each containing 7 minutes of matrix data corresponding to the gearbox states was used for evaluation purposes. Both the 10 and 20 element hidden layer, four output state network configurations were employed once again, this time with a 1,600 element input layer rather than the 900 element configuration described in 6.1. As a result of these previous evaluations a narrower range of training rates, α, were employed in an attempt to reduce the likelihood of unpredictable weight behaviour during training which had earlier impacted upon results. All the remaining training and testing

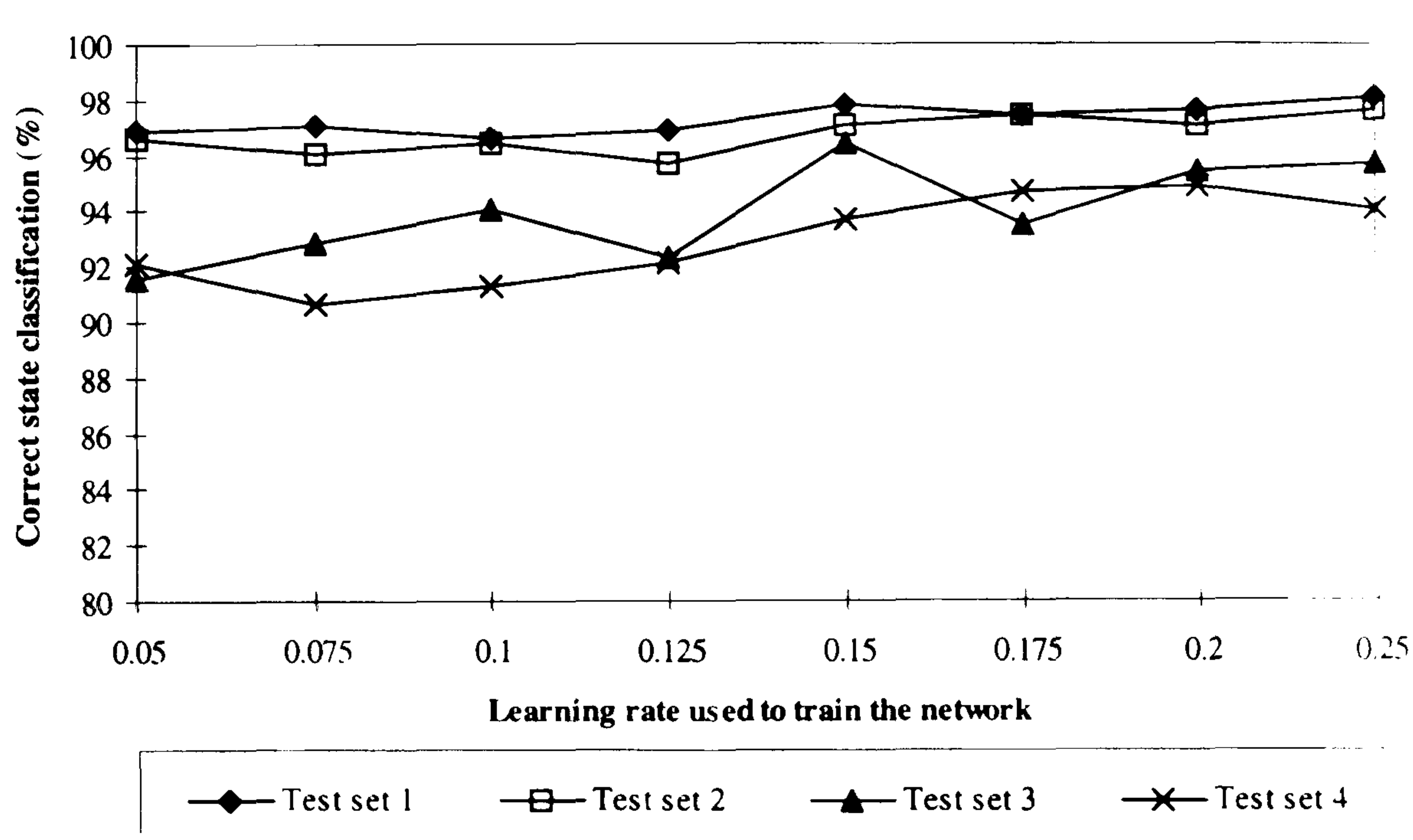

Figure 6-9 Classification for a 1,600-10-4 network configuration trained on data set 1 and tested against each of four separate test sets

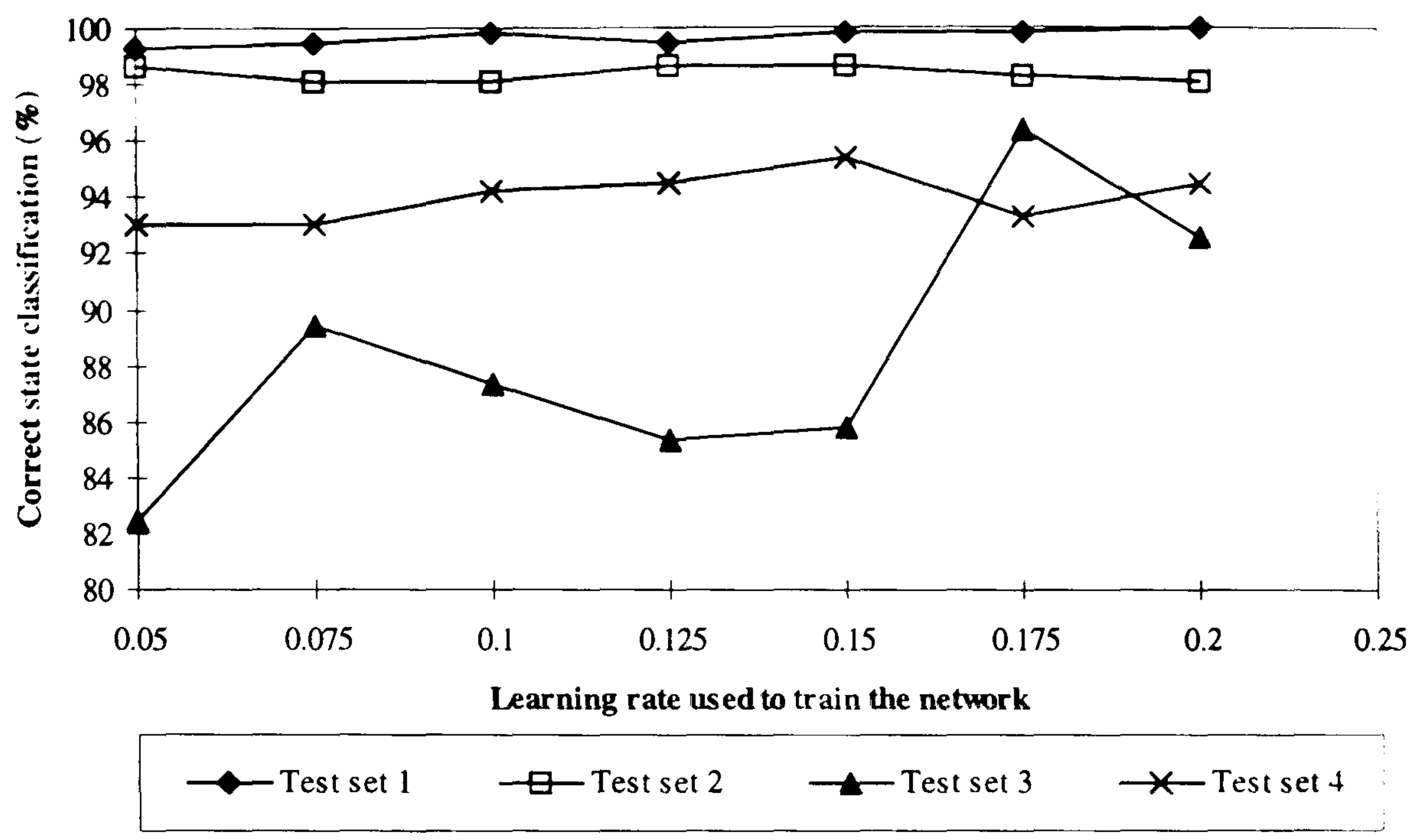

Figure 6-10 Classification for a 1,600-10-4 network configuration trained on data set 2 and tested against each of four separate test sets

parameters were fixed to maintain the integrity of any direct performance comparisons made with respect to the other techniques described so far.

From the trials performed using the amplitude derived A-matrix presentation format it is clear that the technique does indeed provide many of the predicted gains in performance both during the training and classification phases. A sample of these trials results are graphically presented in Figures 6.9 to 6.13. Whilst worst case separation proved to be below that expected at 82% the average rate of correctly classified states was in the region 93-98%. The worst case performance identified during this evaluation was the result of network training with data set 2 followed by testing with set 3. Three data set and network combinations produced better than average performances by classifying between 98-100% of all test data correctly. In addition to this general performance enhancement, training of the various networks involved required significantly fewer data passes than all of the presentation types evaluated so far. In most cases 40-400

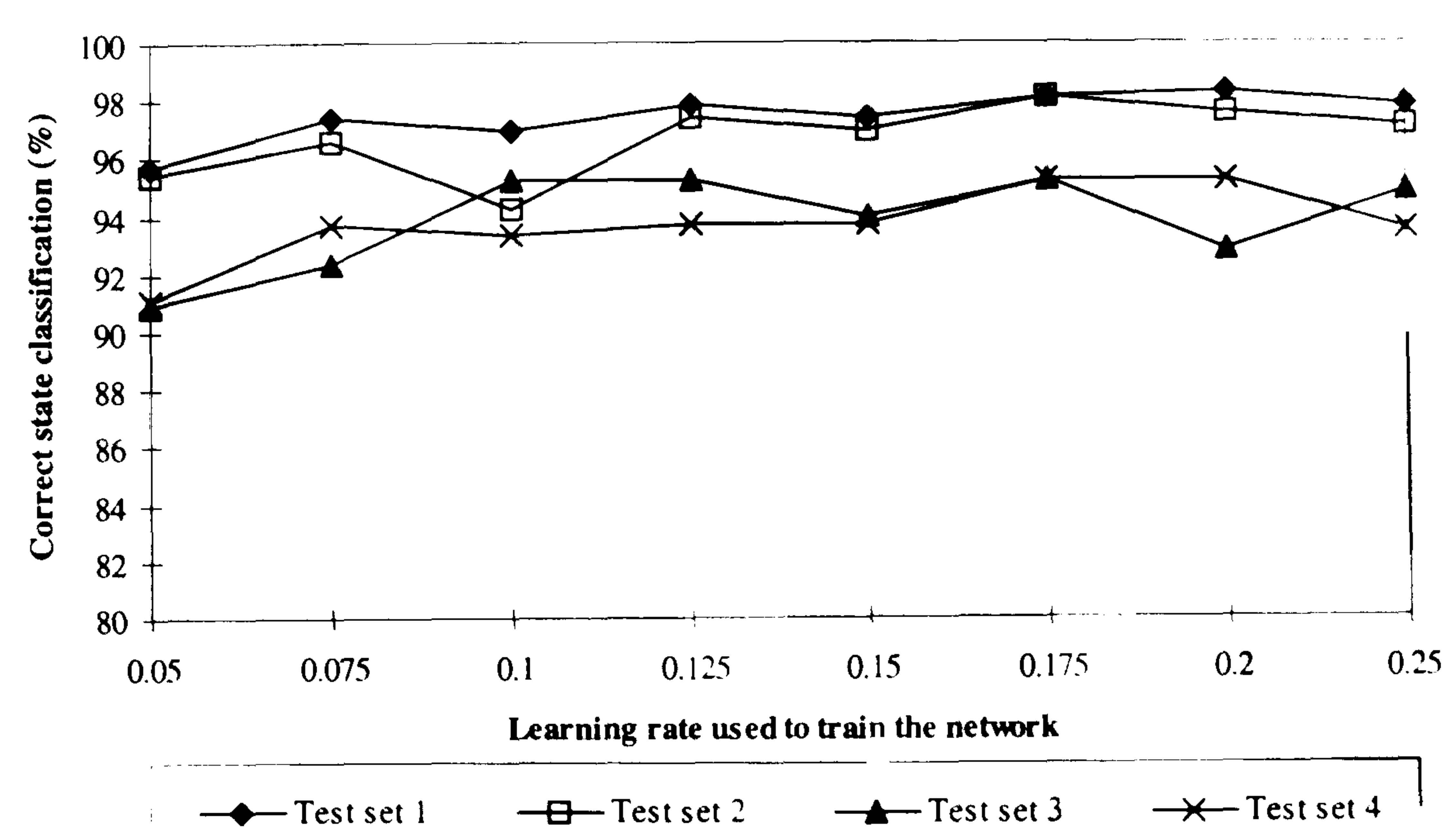

Figure 6-11 Classification for a 1,600-20-4 network configuration trained on data set 1 and tested against each of four separate test sets

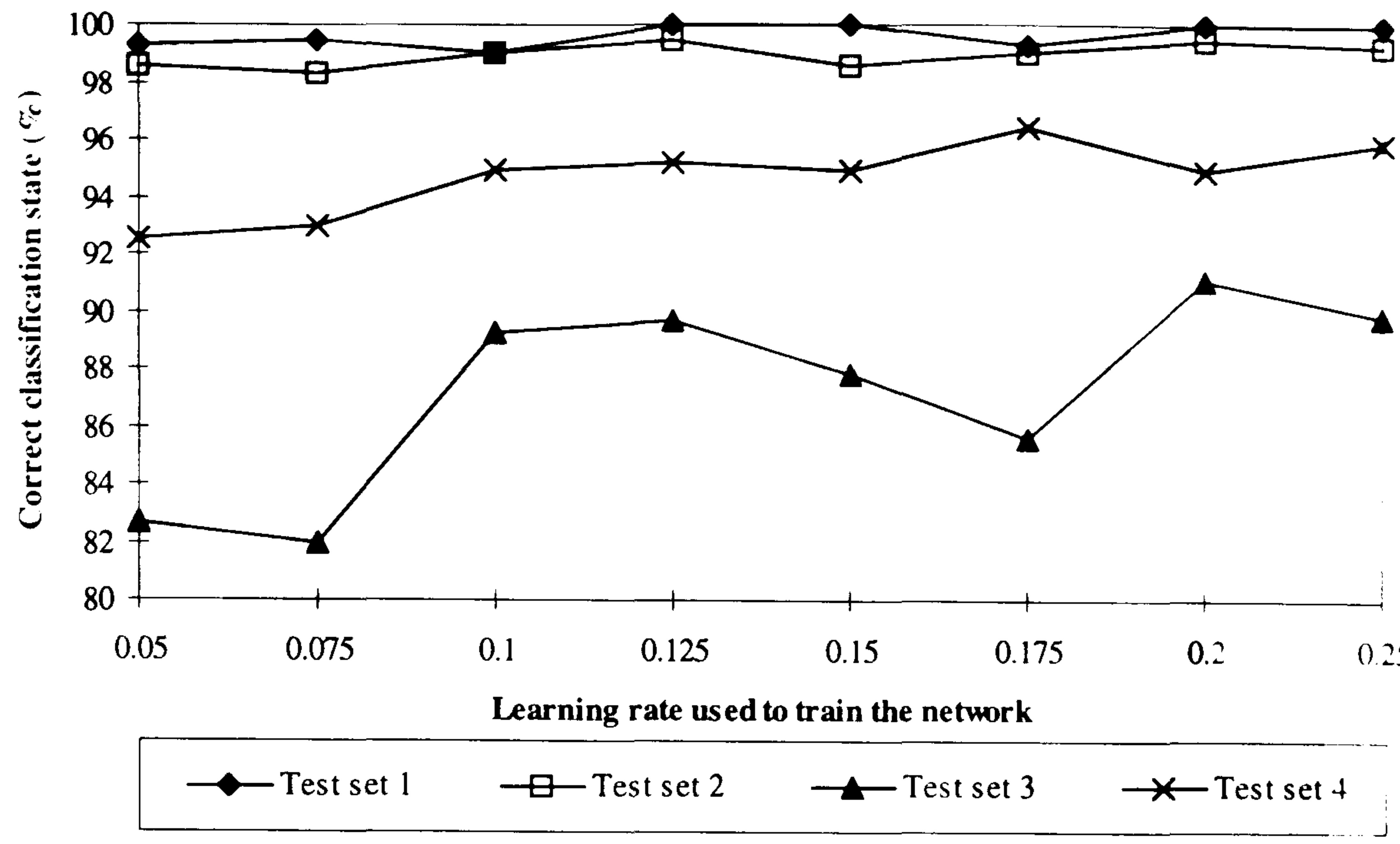

Figure 6-12 Classification for a 1,600-20-4 network configuration trained on data set 2 and tested against each of four separate test sets

iterations were sufficient for the network to converge over the range of training rates applied. This represents a significant improvement upon previously evaluated signal conditioning and presentation strategies. In some cases halving the training times.

Once again, as expected, the relationship between the training rate, α, and the number of training iterations subsequently required to achieve convergence over the α range employed is approximately linear. However, from the results illustrated there would seem to be little to indicate that fewer training iterations, as a result of a reduced α, result in an inadequately generalised network or that performance is impaired as a consequence. This is a good indication that the error surface for the weight training with this classification problem is relatively smooth and is not significantly affected by local weight space minima.

These early trials indicate clearly that, as predicted, the combination of amplitude TES

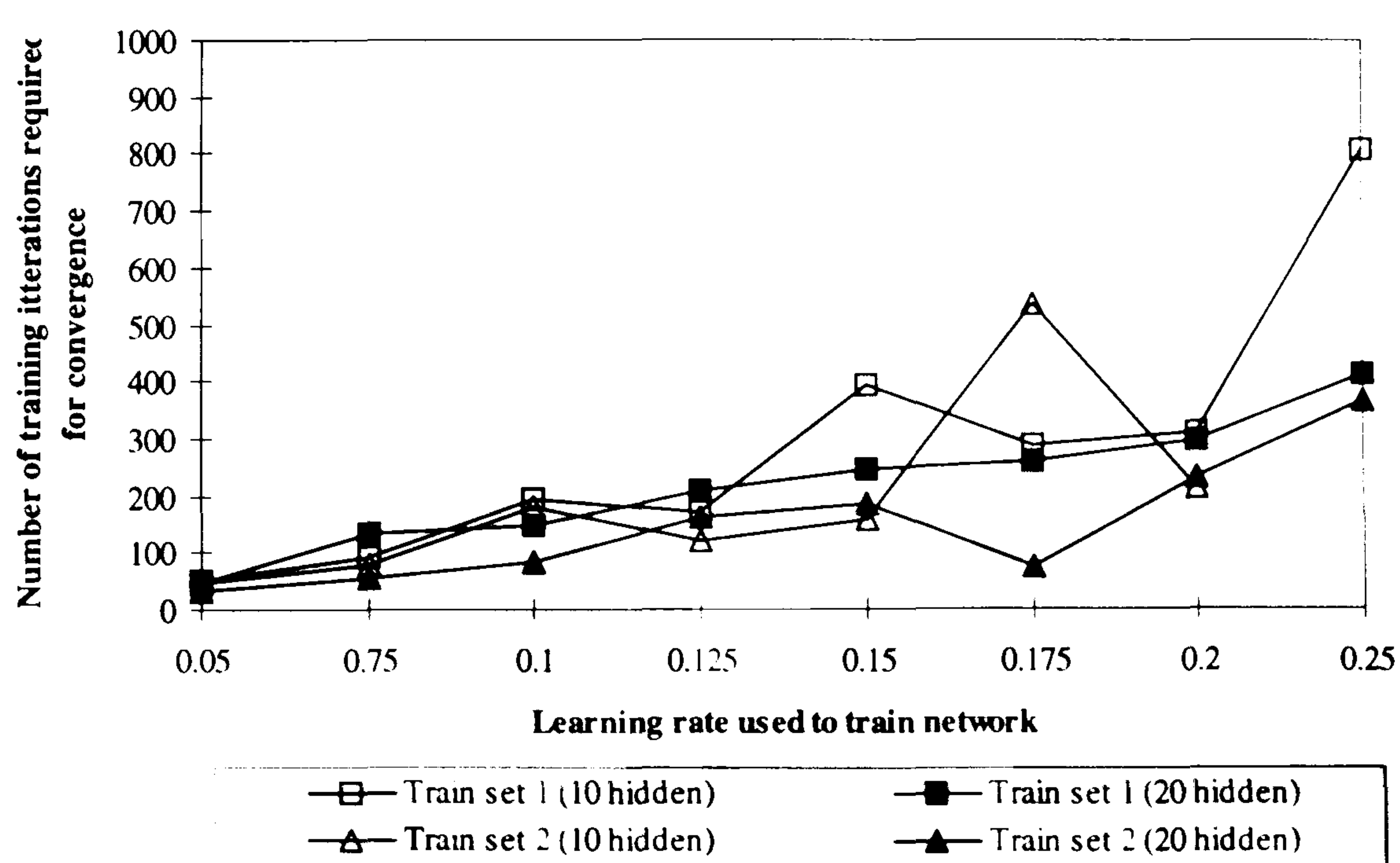

Figure 6-13 Rate of network convergence for each of the four network configuration, amplitude A-matrix training data set combinations used

coding and A-matrix compression provides the best compromise between network size, learning ability and classification capability. These results highlight the potential of the technique as a powerful and effective tool with which to classify the acoustic signals emitted by the gearbox. The remainder of the Chapter is dedicated to studying in greater depth the techniques applicability to additional mechanical faults which may be simulated using the gearbox testbed system. In addition evaluation was performed on potential restrictions which may be associated with a practical implementation as well as the physical characteristics of a variety of network architectures used to classify the data presented.

6.3 Restrictions Imposed on the Acquisition and Application of Amplitude TES by Neural Techniques

Whilst the majority of the evidence from trials with amplitude A-matrices provides good encouragement for the further development of this technique there was one specific data combination which produced a network solution with significantly reduced classification capability. This discrepancy, illustrated in Figure 6.10 and Figure 6.12, and associated with the combination of the second training and third test sets resulted in more erratic classification. Between 82% and 90% of the test matrices were correctly identified independent of the network configuration applied. As with the other presentation formats some variation in perceived performance is to be expected due to the random nature of the data selection. However, a disparity of this magnitude was unlikely to have been caused solely as a consequence of random data selection from a reasonable set and thus it warranted further investigation.

The initial focus of attention was on the characteristics of the smaller test sets rather than on the training sets simply because of the number of data permutations involved. Seven further test sets were generated each based upon the contents of test set 3 which

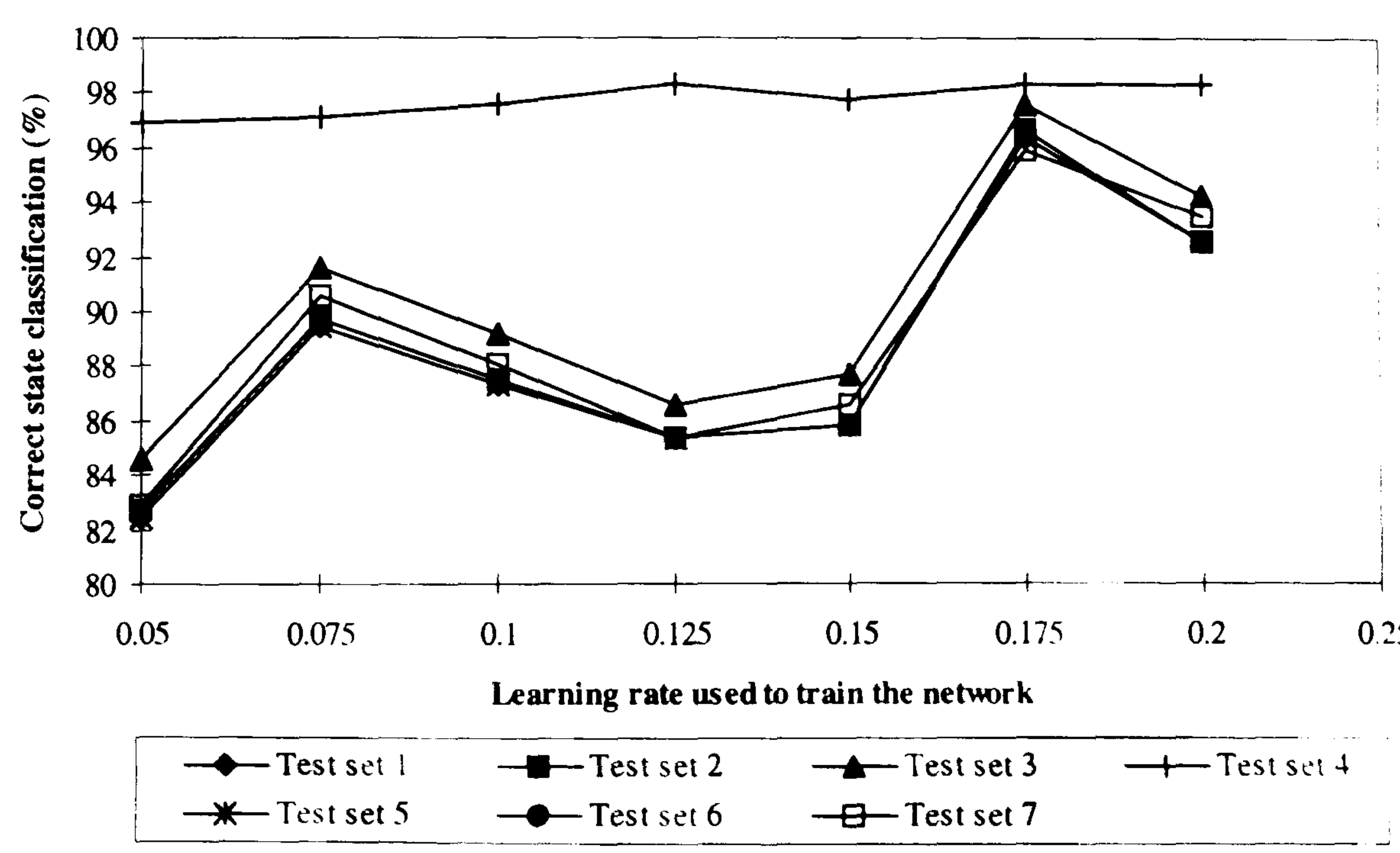

Figure 6-14 Classification for a 1,600-10-4 network configuration trained on data set 2 and tested against each of the seven test set variants

had previously highlighted the disparity. Four of the new sets were constructed by replacing, in turn, all matrix tokens pertaining to one of the four gearbox states (3, 4, 5, 6). In each case the tokens were replaced by corresponding state data from test set 1. The remaining three sets, again based on set 3, had the tokens for states 3, 4 and 5 respectively replaced by equivalent data matrices not previously used but still corresponding to the state which had been removed. Another series of trials was then performed using these test sets applied to both original network configurations again trained using data set 2, the other element associated with the problematical performance results. The intention of the trials was to identify whether or not any specific element within the original set could be attributed to the degradation in performance.

Clearly from the results obtained in this evaluation, graphically illustrated in Figures 6.14 and 6.15, the fourth test set produces similar perceived performance to the previous data combinations covered in section 6.2 the remainder still result in poor classification. The distinguishing feature between the newly generated test set, set 4, and the remainder of the new sets are the A-matrix tokens corresponding to state 6. When this particular token sub-set is replaced in the test set by another sub-set corresponding to the same state, as it was in set 4, the perceived network performance is improved by between 7% and 14% to approximately 98%. Having identified the apparent cause of the degradation in perceived performance it was then necessary to consider why the discrepancy had occurred. Three possible causes were examined.

- An additional physical fault, configuration change, or excessive shaft velocity variation had occurred during the original archive recording period.

- Data bias, whereby the configuration of the training set had in some way become weighted so as to produce a network which was insensitive to these particular state 6 data tokens.

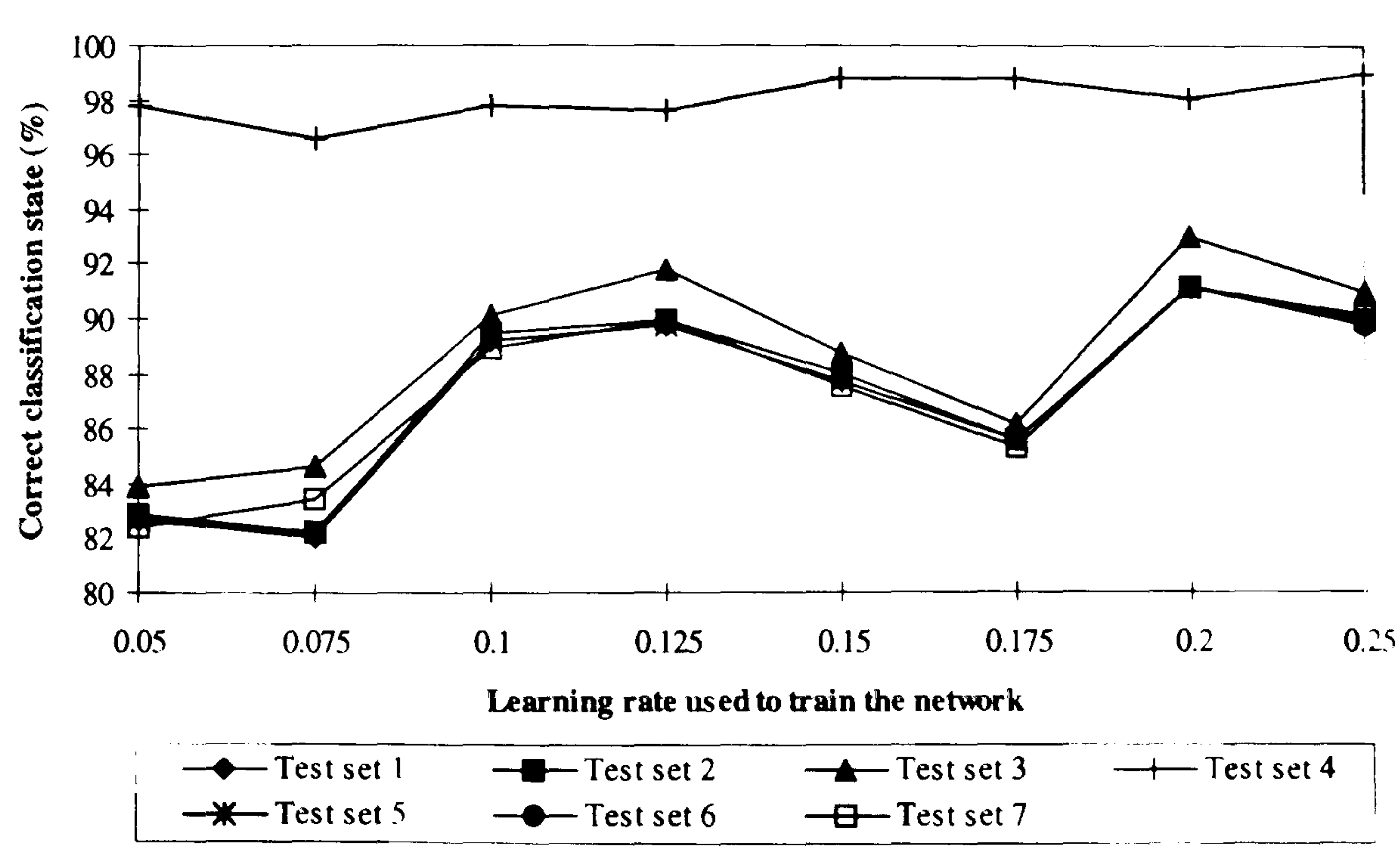

Figure 6-15 Classification for a 1600-20-4 network configuration trained on data set 2 and tested against each of the seven test set variants

- That a fault had been introduced during the TES signal conversion stage through the selection of an unrepresentative normalisation coefficient prior to commencement of the conversion.

All three possibilities may have been sufficient to introduce perturbations to the matrix tokens which led to the degradation in perceived performance. In order to clarify the cause or causes of this discrepancy all three of the hypotheses were evaluated in turn.

To identify whether or not a physical discrepancy had caused the fluctuation five further test sets were constructed. In each of these five sets those matrix tokens associated with state 6 were replaced by additional state 6 matrix tokens generated from the same archived recording retrieved from the taped database. The TES symbol stream for four of the new sets was produced from the same physical segment of the tape, in each case

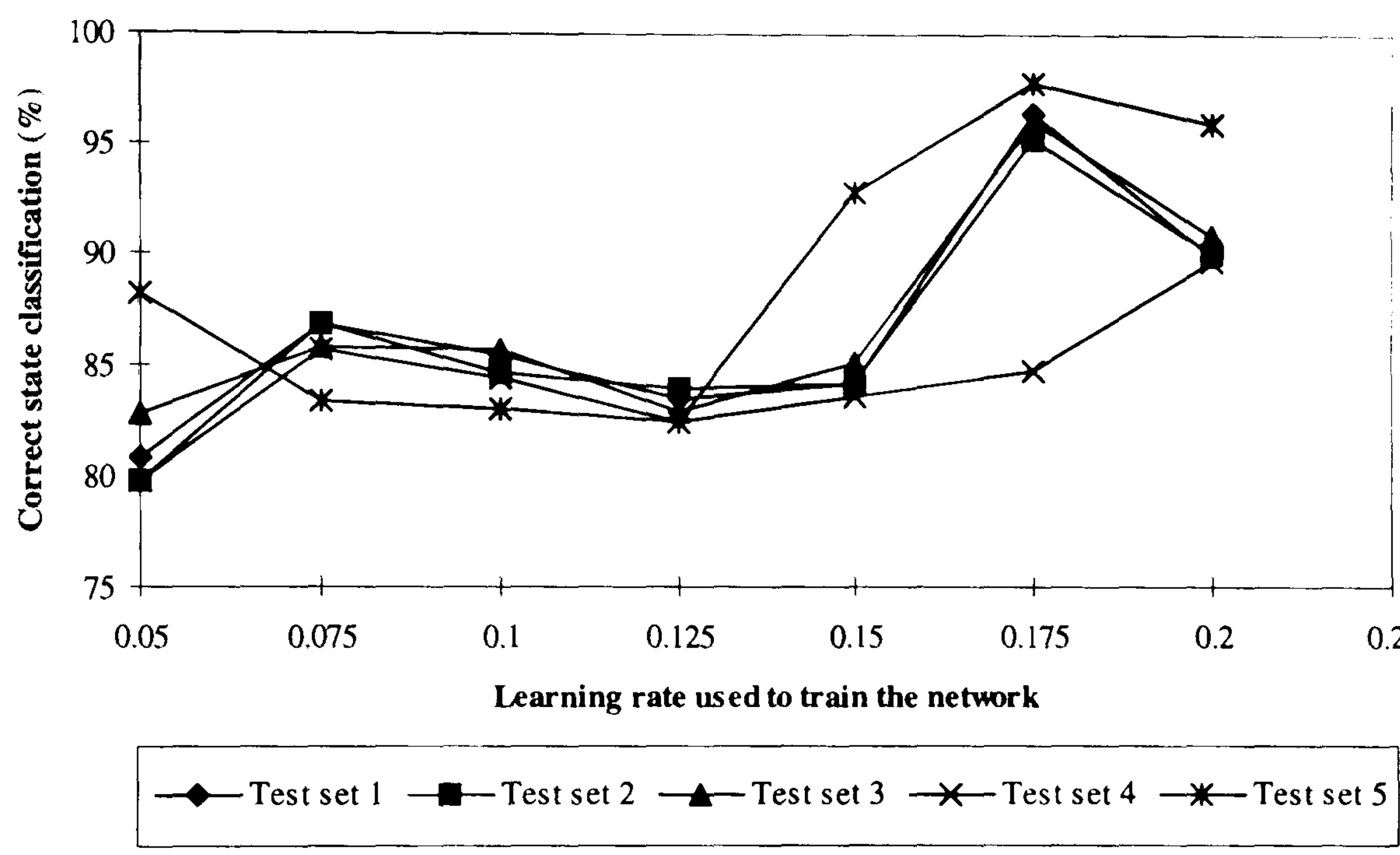

Figure 6-16 Classification for a 1,600-10-4 network configuration trained on data set 2 and tested against each of the five test sets containing replaced state 6 data

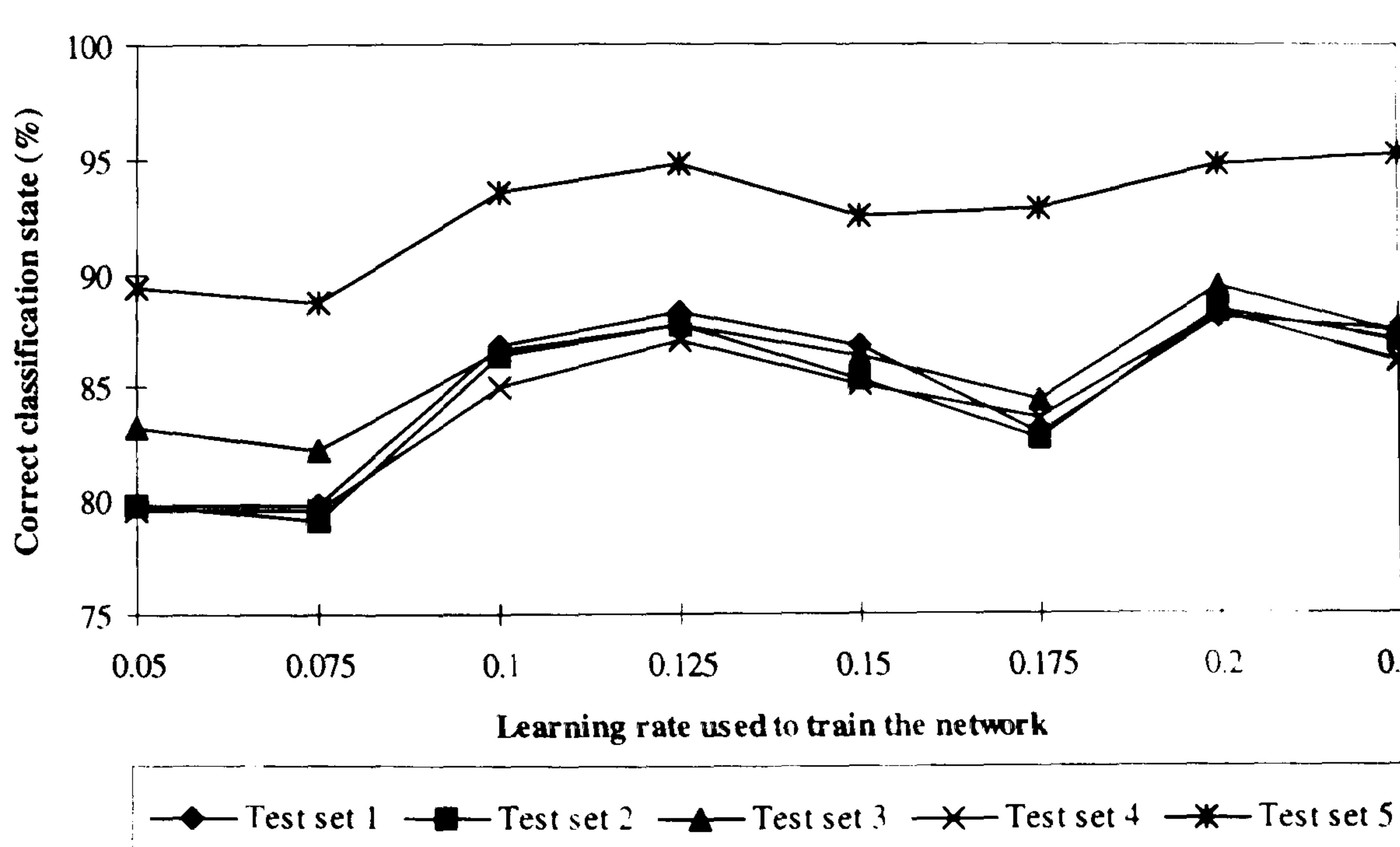

Figure 6-17 Classification for a 1,600-20-4 network configuration trained on data set 2 and tested against each of the five test sets containing replaced state 6 data

using a different normalisation coefficient, whilst the fifth was generated from a segment later in the recording and had similar normalisation properties to the original data. Both of the trained network configurations were evaluated using these five test sets. The results of these trials are illustrated in Figures 6.16 and 6.17. The set composing matrix tokens generated from a different physical section of the archived recording, set 5, produces a tangible improvement in the perceived performance over those containing tokens generated from the original segment.

Whilst this is more clearly defined in the 20 node hidden layer configuration it is also seen in four of the seven trained network configurations having 10 hidden nodes. The remaining four new test sets produce results comparable with earlier network performance. This provided an early indication that the most probable cause of the disparity in relative system performance was the signal characteristics of a particular physical segment of the acoustic recording used to generate the original test data set. However there had still been a finite possibility that the inconsistency was simply the result of an inadequately formed training set rather than a physical perturbation. Since essentially the network performance is controlled by information the network extracts from the training set any mismatch, or bias, is likely to be distinguished by a disparity in the subsequent performance evaluation.

To eliminate data bias as a cause a further three training sets were created, once again based on set 2. Each of these examined specific areas of the data which could feasibly have produced the variation in performance. All three focused on the order in which data is presented to a network during a training run based upon specific physical attributes and the conversion characteristics of the data. Of particular interest in this respect were the shaft velocity and the amplitude TES normalisation coefficient attributes of each data set within the training group. Already in previous trials the effect of the conversion coefficient on performance has been highlighted so there is good reason for believing that the effectiveness of the network training with A-matrix data

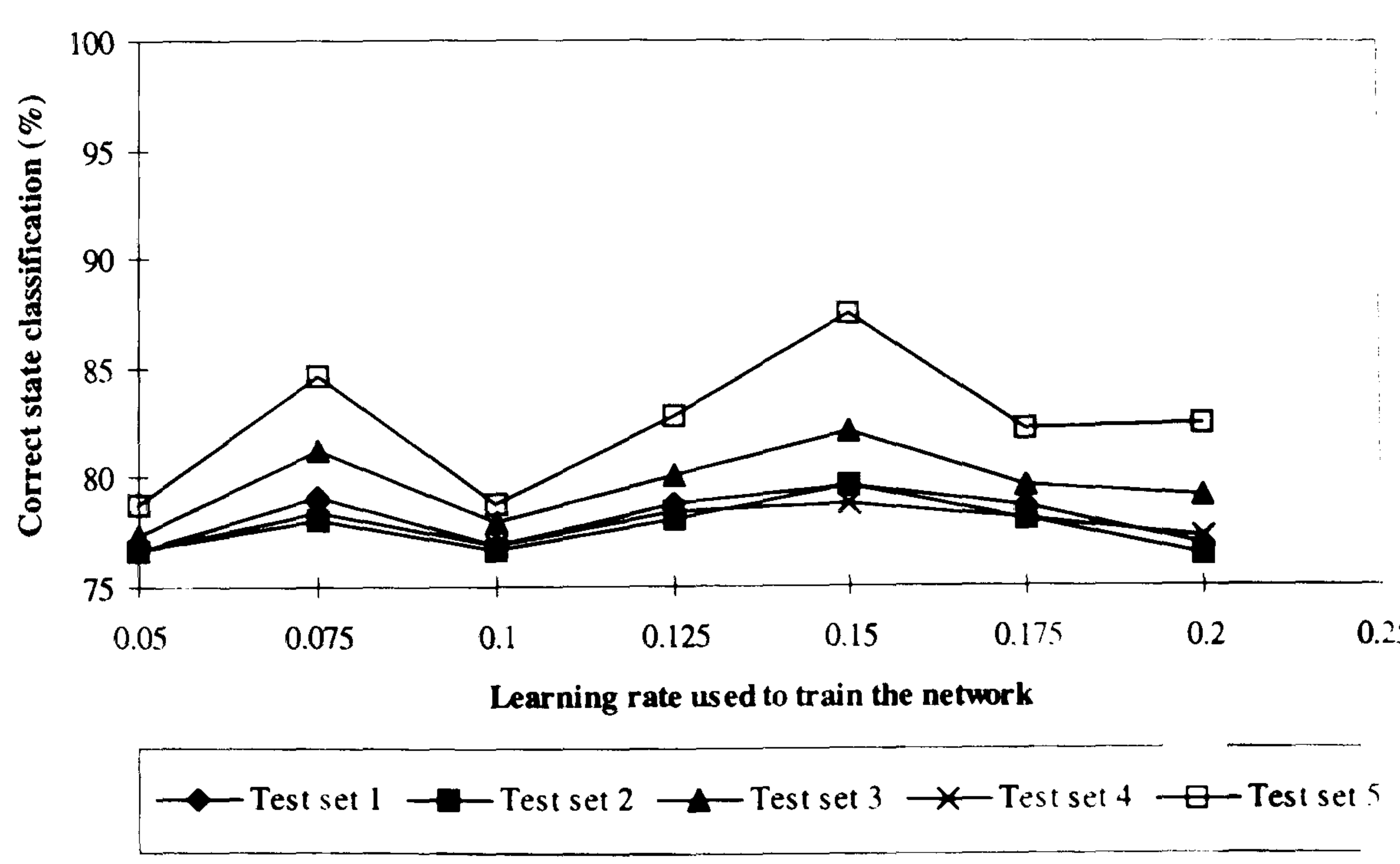

Figure 6-18 Classification for a 1,600-10-4 network configuration trained using a data set containing TES state data clustered into different shaft velocity biases

may be similarly affected. However initial investigations centred upon the effects of the shaft velocity component of the acoustic samples at the time they were acquired and prior to TES conversion.

Two alternative training sets were produced using the data from set 2 as a template to gauge the effects that this velocity attribute exacts upon the network. One set, set 3, was subdivided into four smaller subgroups each corresponding to 7 minutes worth of matrix data. Two of these subgroups contained matrix token data generated from archived recordings with a high biased velocity attribute whilst two contained data generated from recordings with low biased velocity characteristics. This technique produces a training set comprising of TES A-matrix tokens with four subgroups of alternate velocity attributes, 50% high biased split into two equally sized subgroups and 50% low biased, again split into two subgroups. The absolute difference in shaft velocity between low and high bias subgroups is relatively small, only in the region of 50rpm or

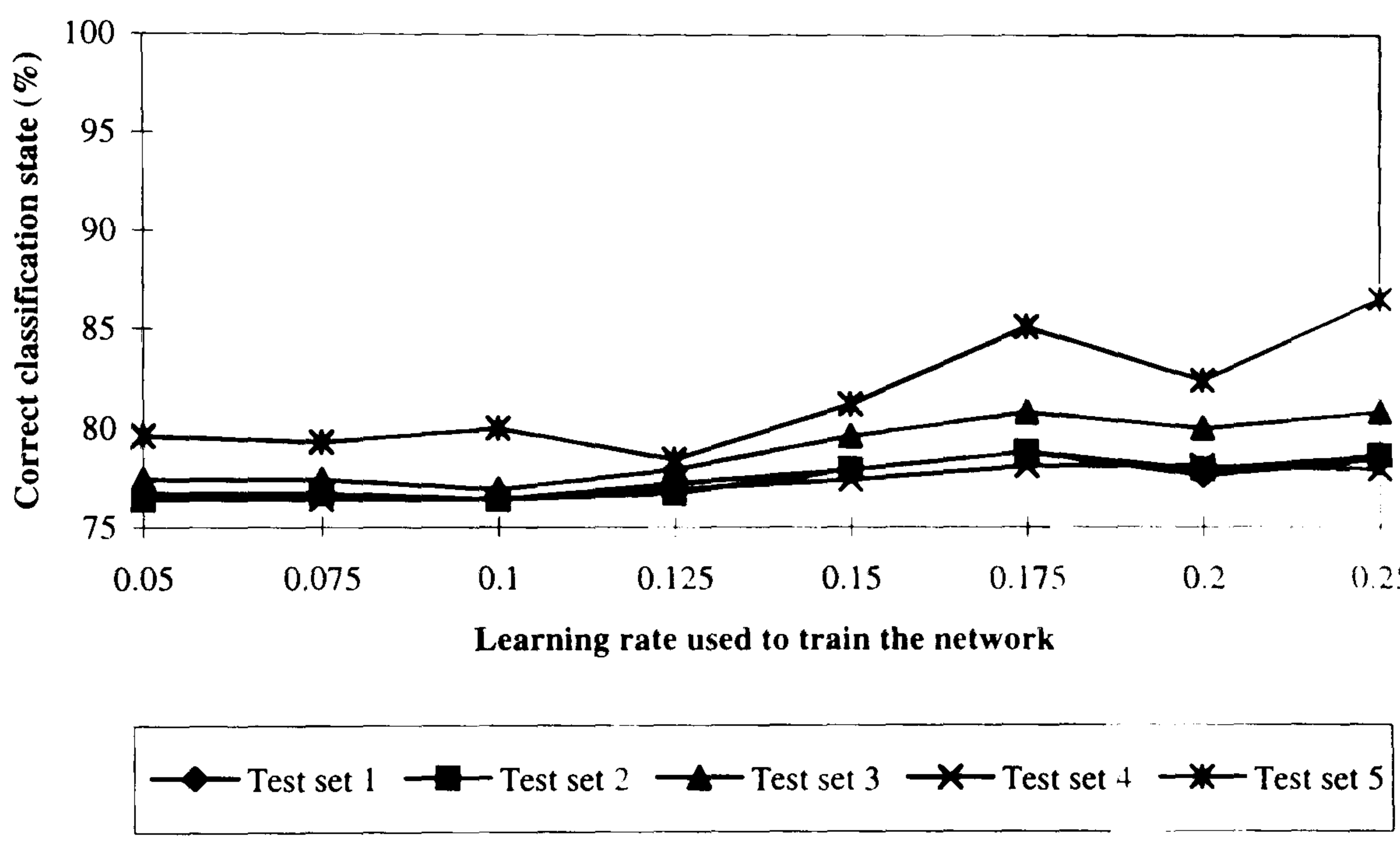

Figure 6-19 Classification for a 1,600-20-4 network configuration trained using a data set containing TES state data clustered into different shaft velocity biases

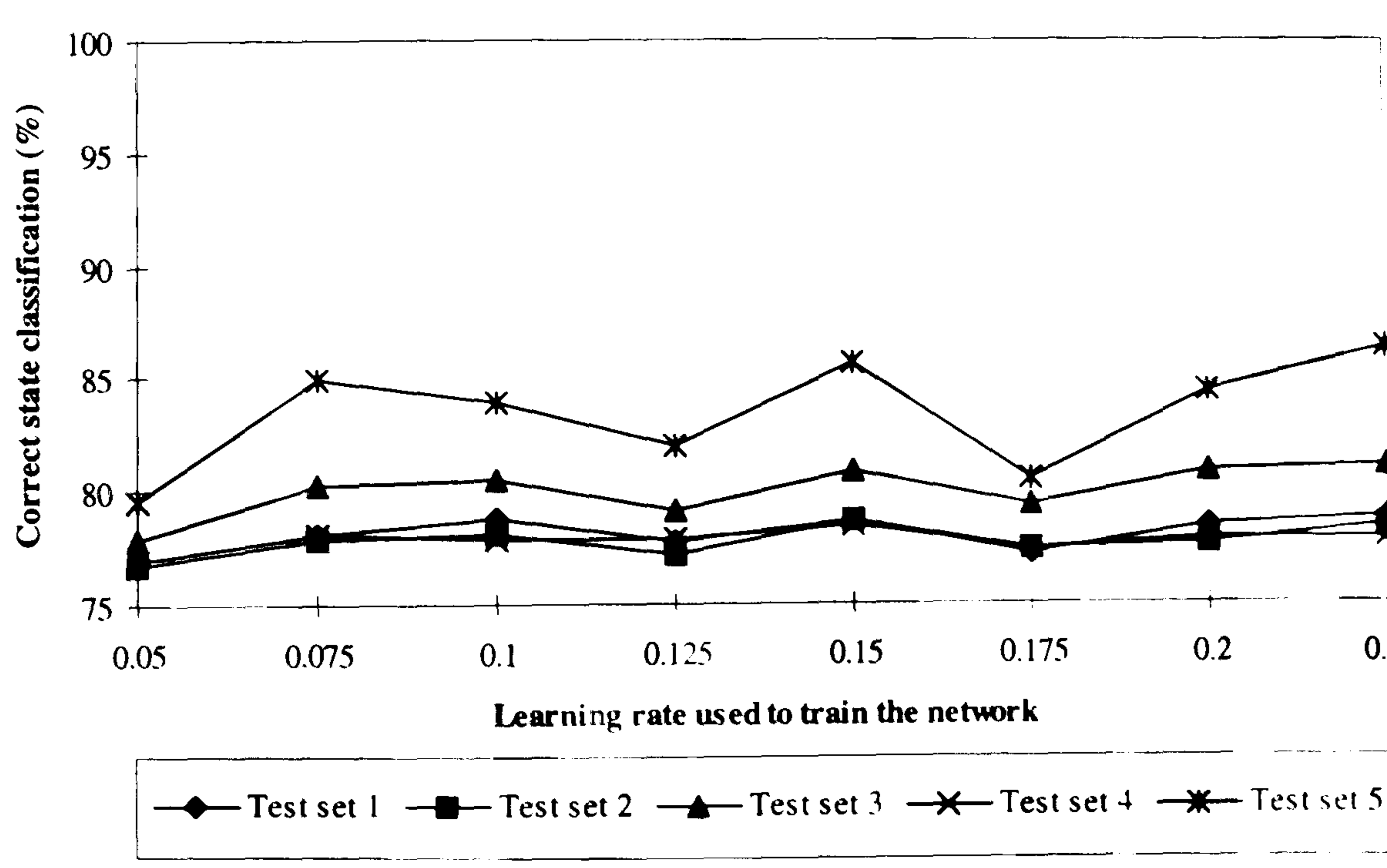

Figure 6-20 Classification for a 1,600-10-4 network configuration trained using a set containing TES state data with evenly distributed shaft velocity attributes

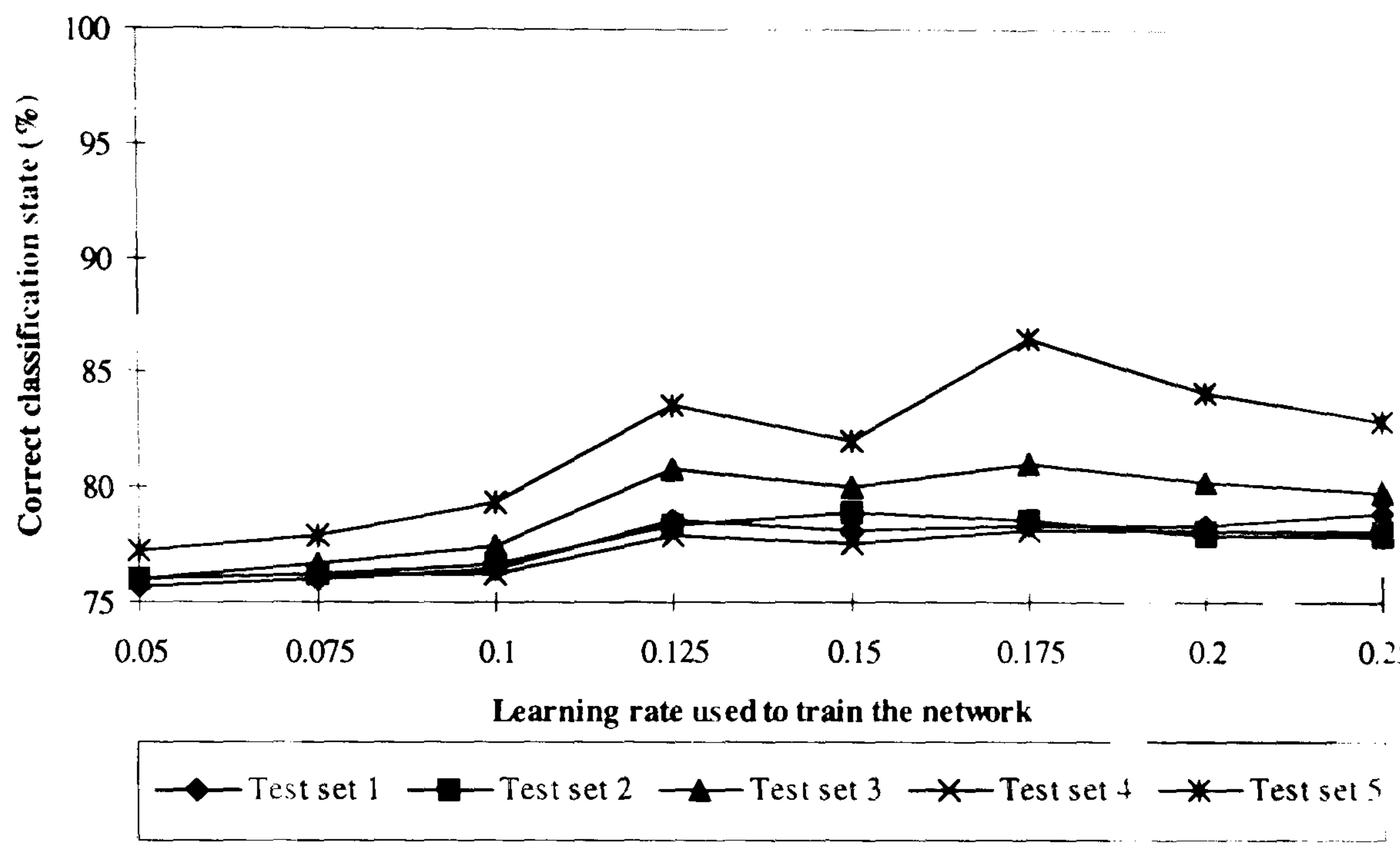

Figure 6-21 Classification for a 1,600-20-4 network configuration trained using a set containing TES state data with evenly distributed shaft velocity attributes

1.5% of shaft speed. It is though conceivable that even such a relatively minor variation may be sufficient to affect the matrix tokens in such a way as to inadequately train the network and so produce the disparity in results which were noted. The second of the two new training sets, set 4, contained exactly the same matrix data as the first, set 3, but reordered so that the velocity attributes were uniformly distributed throughout the set rather than concentrated into four sub-clusters.

From the evaluation results presented in Figures 6.18-6.21 the indication is that the velocity attribute was not the cause of the original performance degradation which was identified. The perceived performance against each of the five test sets actually worsens and has certainly not been enhanced through the selection of matrix tokens based specifically upon their velocity attributes. Despite the degradation in performance set 5, with its modified state 6 data tokens is still associated with the best perceived classification performance as it had been previously.

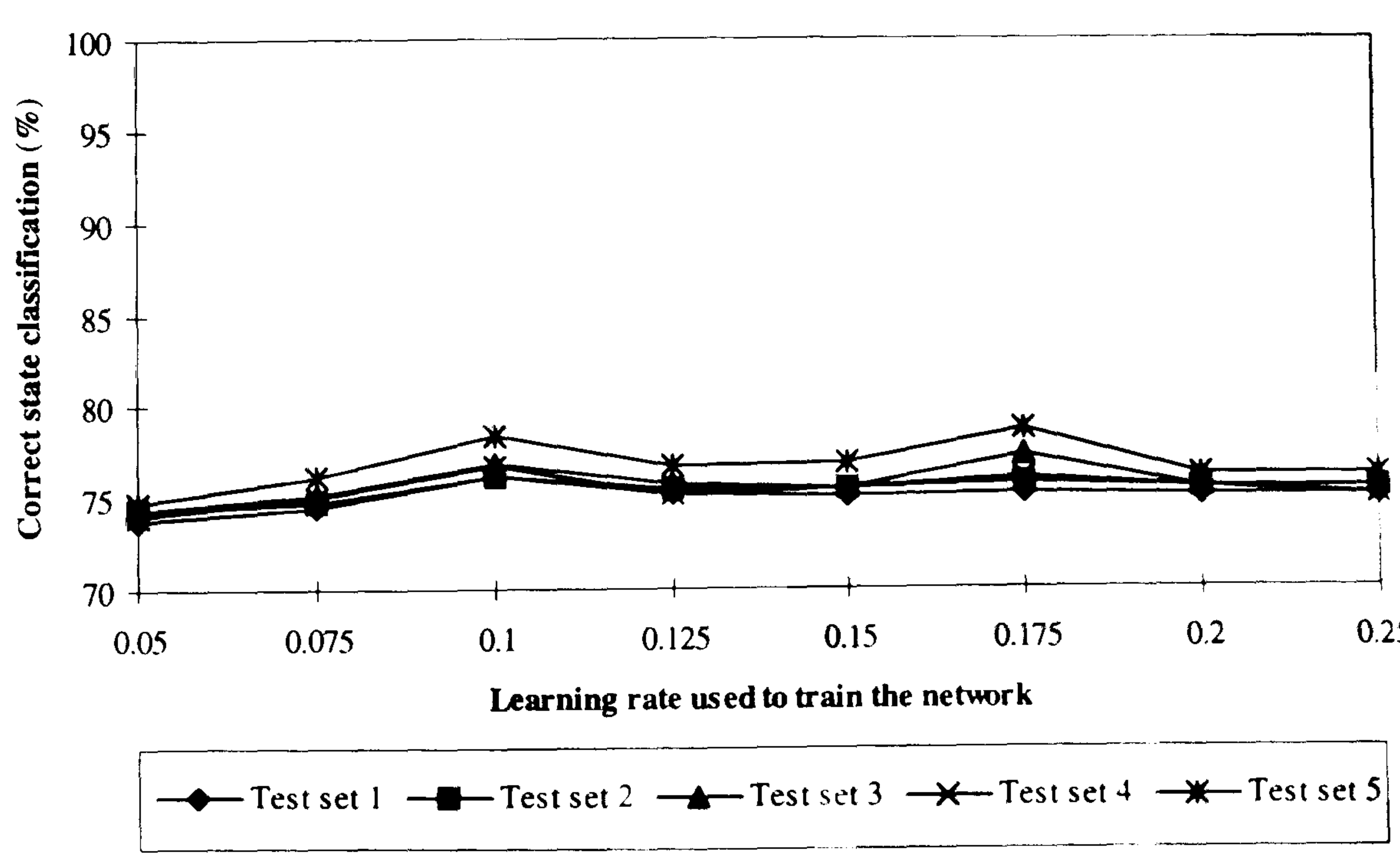

Figure 6-22 Classification for a 1,600-10-4 network configuration trained using a set containing TES state data with evenly distributed normalisation attributes

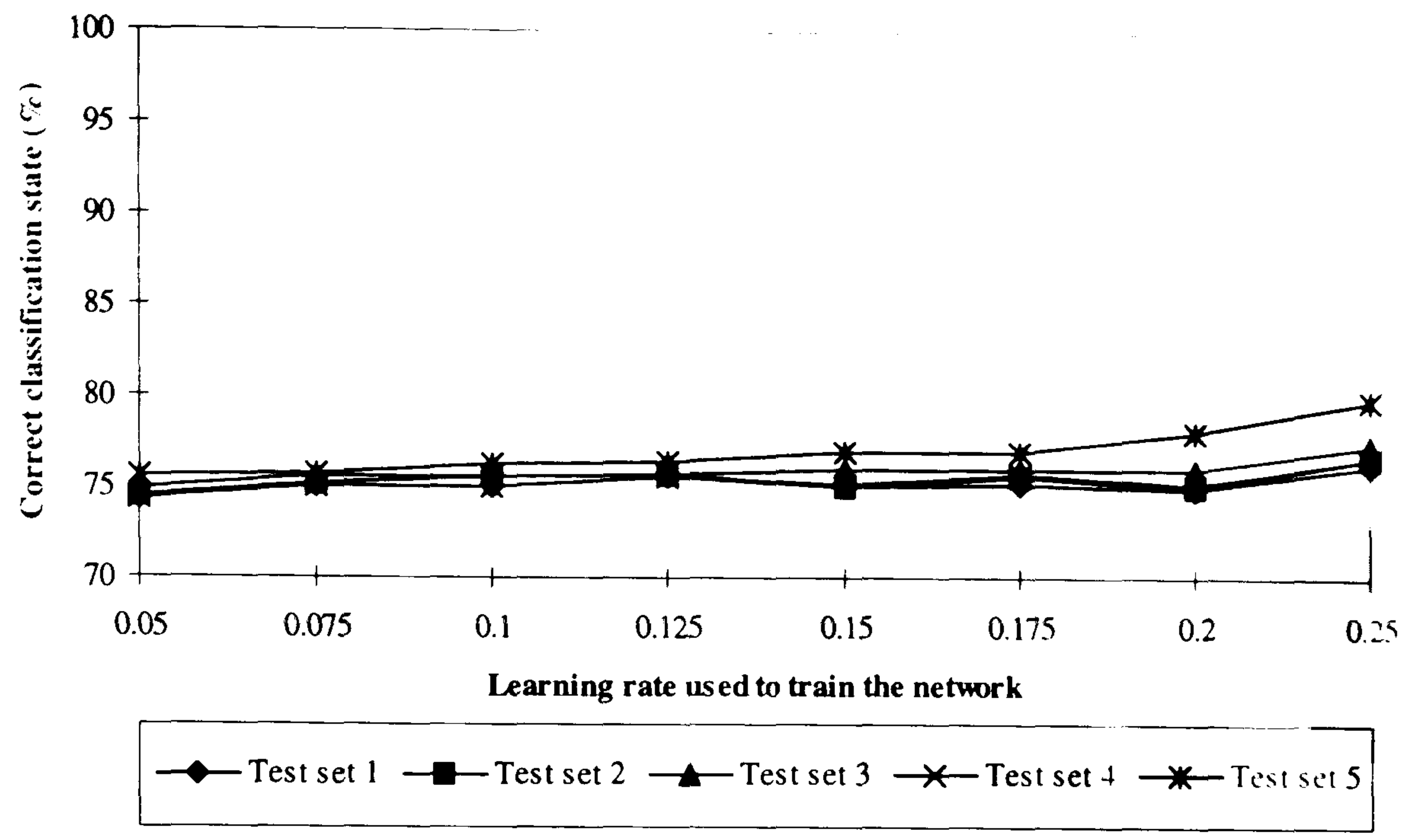

Figure 6-23 Classification for a 1,600-20-4 network configuration trained using a set containing TES state data with evenly distributed normalisation attributes

The second physical attribute of the TES data, other than the velocity attribute, which should be considered as a source of potential performance degradation is the normalisation coefficient. The effect of this factor acquired prior to and then used during the conversion to amplitude TES codes which provide the indication of the source status has already been discussed in relation to the histogram matrices. To evaluate the effect of this conversion parameter on the efficiency of the A-matrix training data and thus the efficiency of the subsequent network learning a further control set was developed which specifically targeted these normalisation attributes. Rather than the spread of coefficient attributes found in the randomly selected training set a further set was generated which contained data with an evenly distributed range of normalisation coefficients. The results of this test are presented in Figures 6.22 and 6.23. Again they show clearly that selection of specific training data matrices in terms

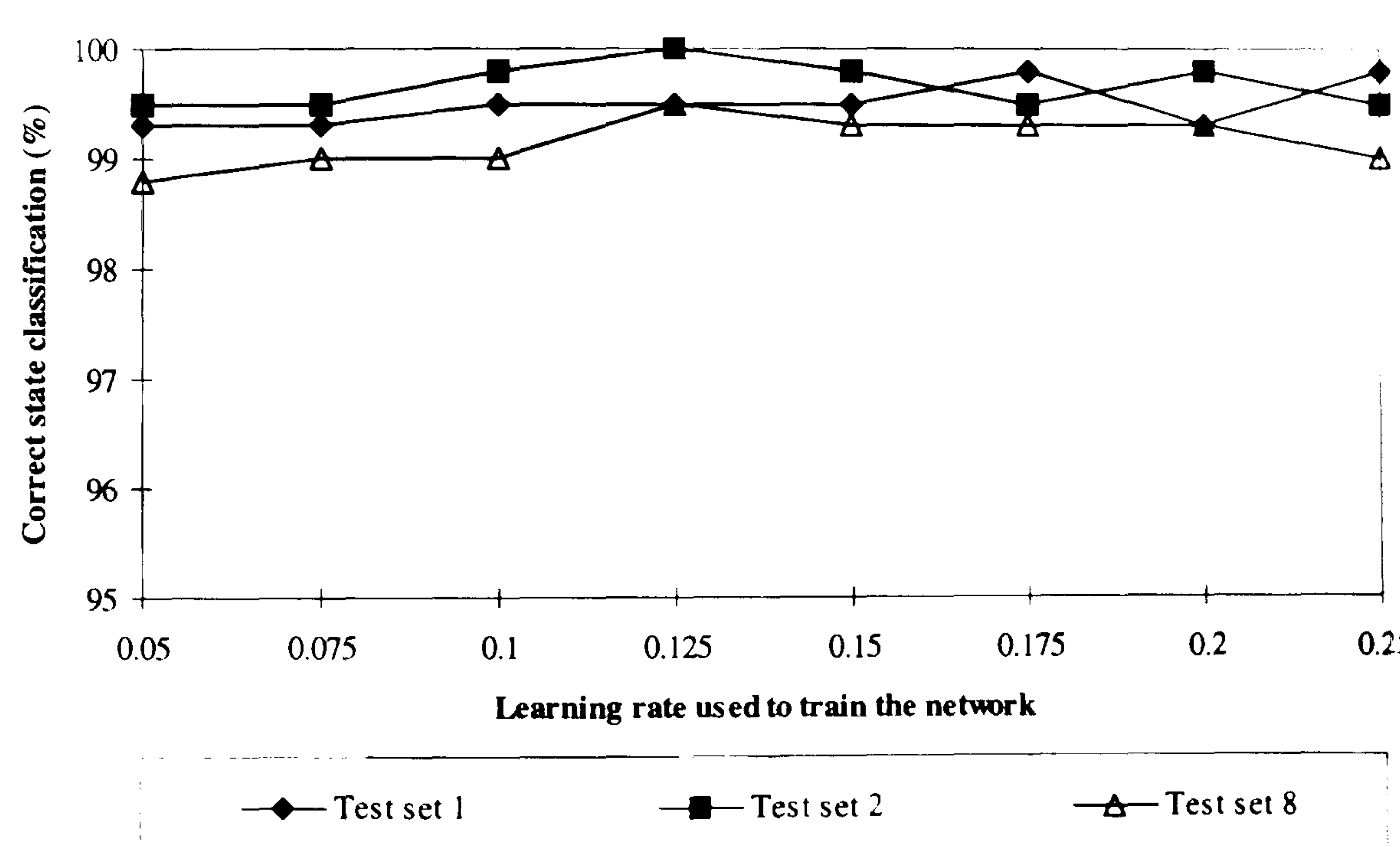

Figure 6-24 Performance of a 1,600-20-4 network configuration trained using TES state data with grouped shaft velocity attributes (TR3) and tested on data sets not containing the state 6 tokens identified as causing the drop in perceived performance.

of particular normalisation characteristics does not improve the perceived performance of the network when presented with those test sets which proved difficult previously. In fact the performance against set 5 which previously was affected less severely is degraded still further by selecting a training set in this manner.

The results of the evaluations on both the training and test sets appears to suggest that in fact the most likely cause of the under performance of the network seen originally is associated with a particular segment of the original recording corresponding to state 6. This was backed up further during a series of three retrospective trials employing some

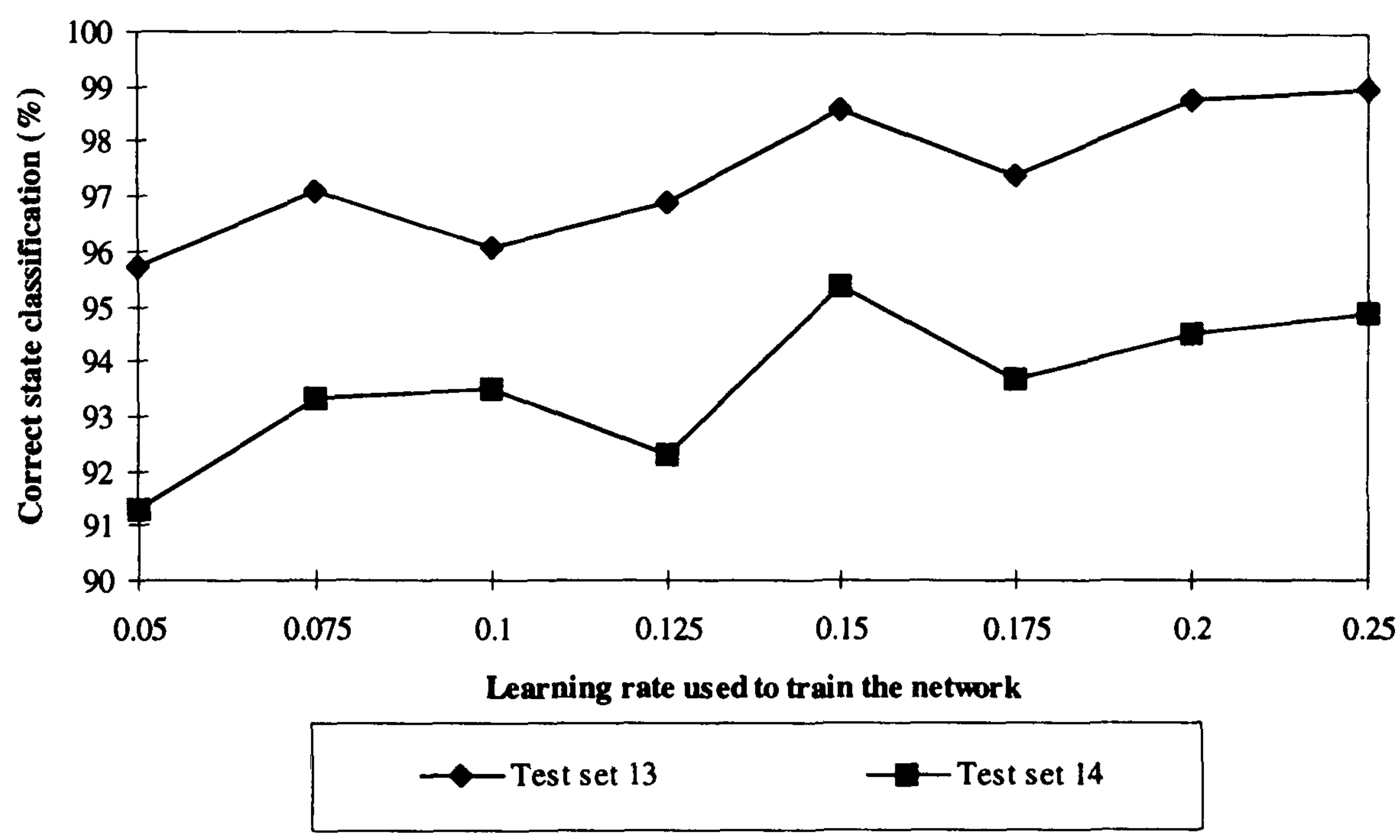

Figure 6-25 Performance of a 1,600-10-4 network trained using the original TES data tokens (TR1) and tested using sets containing state 6 data generated from the suspect recording (set 14) and the corrected state 6 data (set 13)

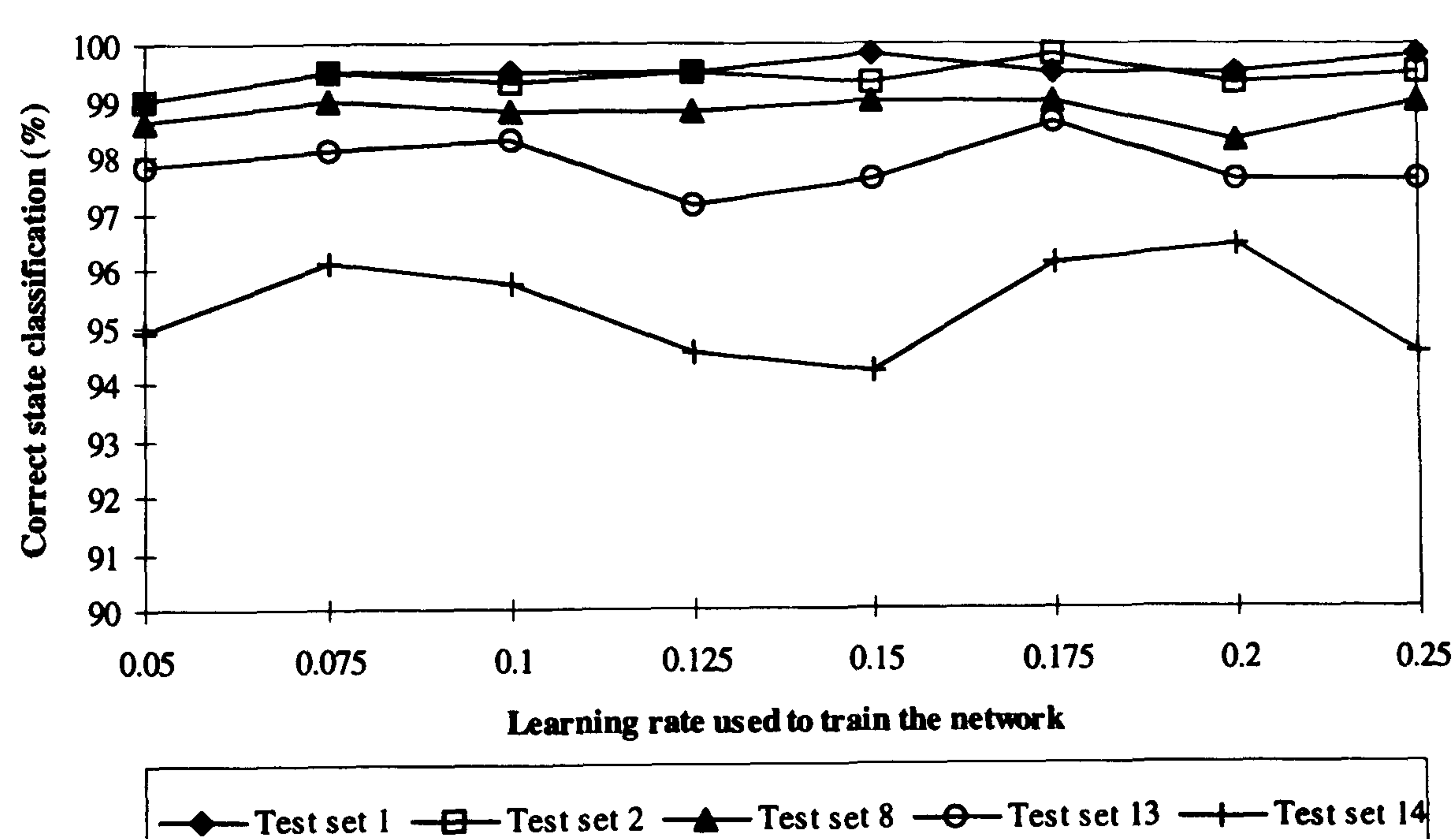

Figure 6-26 Performance of a 1,600-10-4 network trained using a data set containing tokens from the acoustic anomaly (TR6) and evaluated on sets containing state 6 data produced from recording set 14 and from other recordings (1, 2, 8, 13).

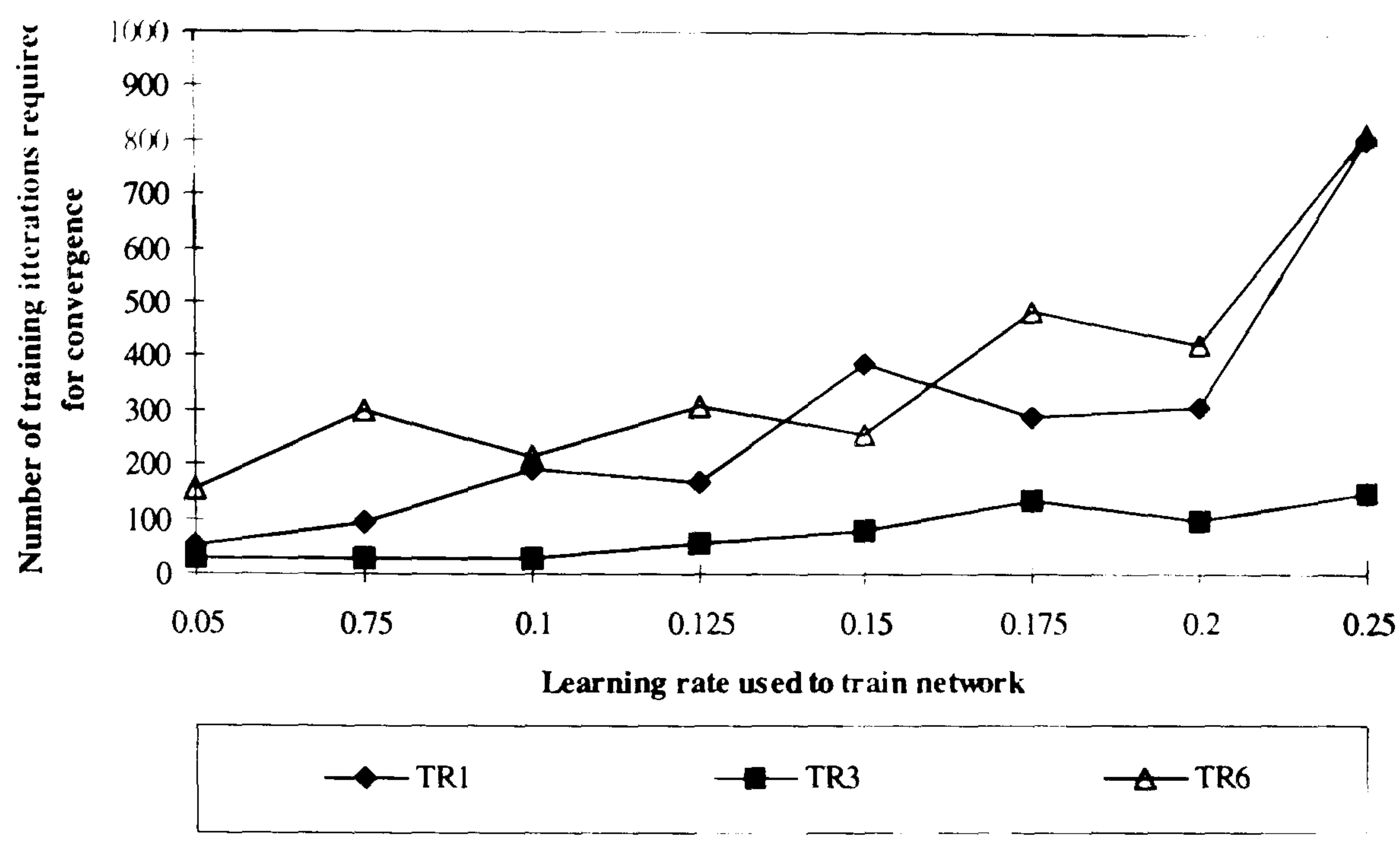

Figure 6-27 Training statistics for each of the three trials illustrated in Figures 6-24-26.

of the newly generated training and test sets used to evaluate the data selection and set construction discussed so far in this section. The training and testing statistics for these trials are illustrated in Figures 6.24-6.27. They further substantiated the notion of an abnormal acoustic record by producing similar perceived performance to those used originally in Section 6.2 when the suspect data tokens were removed. Likewise when the suspect record was included (Figure 6.25) the performance dropped once again. The trial depicted in Figure 6.26 illustrates this. It demonstrates that whilst performance against the test set containing the disparity can be enhanced by modifying the training set the performance against the remaining sets is not affected. Whether this unusual data set was caused by an unexpected fault in the system or extraneous acoustic noise during the early part of the recorded state sample from which this data was extracted is not clear.

6.4 Detecting Angular Misalignment through the Application of Amplitude A-matrices

Angular misalignment is another fault which commonly occurs in systems in which rotational energy is transferred between close coupled shafts. It is caused when the two opposing shafts between which rotational energy is being transferred become non-uniformly offset as was illustrated graphically in Figure 5.3 (see Chapter 5). Such faults induce a vibration component which if measured from opposing bearings on a shaft produce components out of phase by π^0. More traditional schemes for identifying such faults rely on measuring the vibrational phase offsets between sensors attached to opposing bearings. Attempts to detect such faults using a single acoustic source represent a novel if more challenging approach. Initial trials centred on three gearbox configurations which mimicked crudely the shaft offset states. Simulated using the eccentric bearings these three configurations introduced relative bearing offsets of between 1-2mm over the length of the 100mm shaft. The magnitude of these offsets was made intentionally small so as to provide an indication of the likely sensitivity and thus

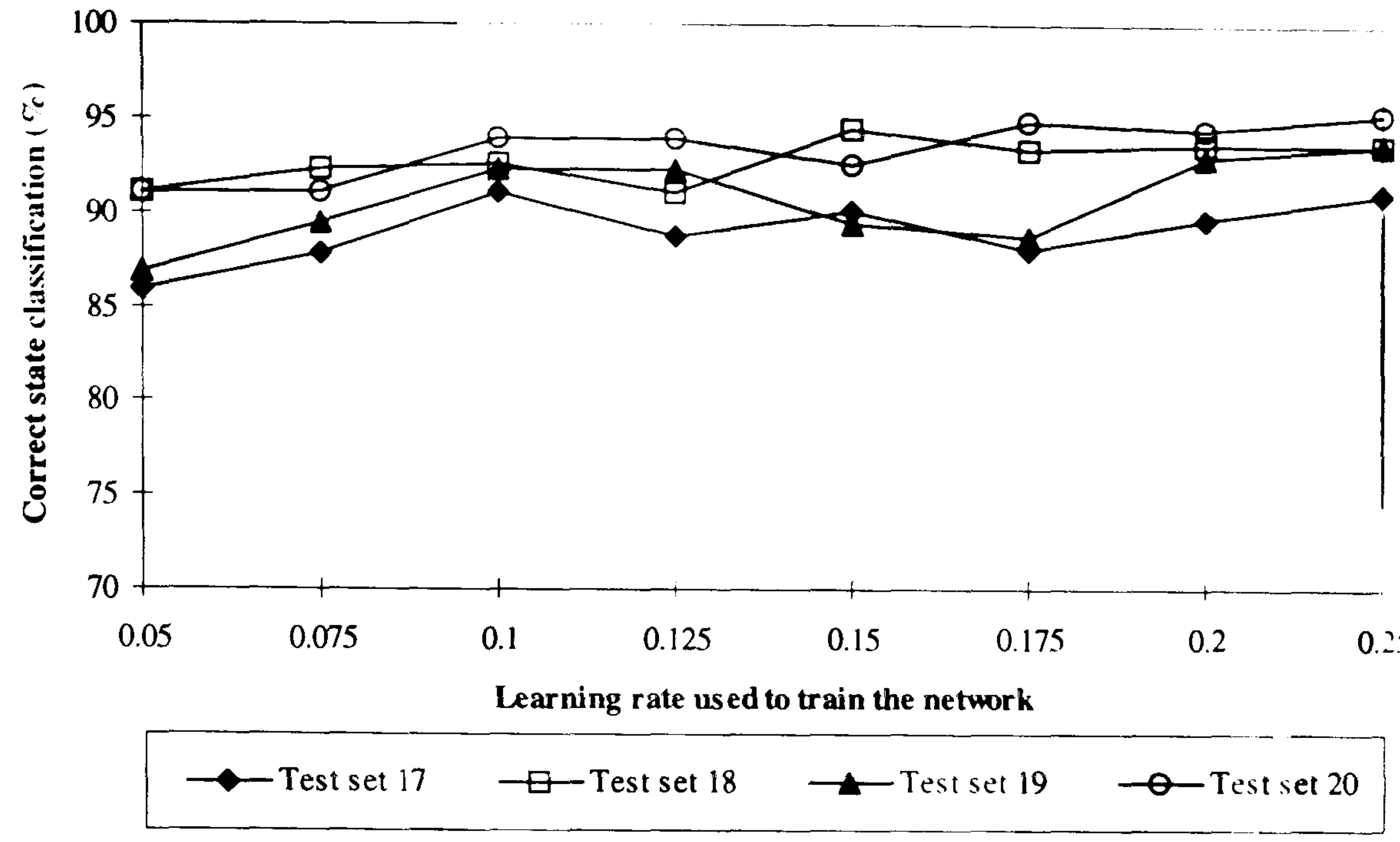

Figure 6-28 Performance of a 1,600-20-3 network configuration trained using the set 8 TES state data information and tested on randomly selected sets containing the three angular misalignment states

the capability of an acoustic identification scheme.

As with earlier trials differing network configurations were evaluated to minimise specific configuration peculiarities when measuring performance. In these trials two configurations were employed the first with 10 and the second with 20 hidden nodes. Both had a common 3 element output configuration with each node corresponding to one of the three unique angular misalignment states. The networks were each presented with identical amplitude A-matrix tokens corresponding to the three states generated from the taped archives. For the performance evaluation a series of training and test sets, each with a range of randomly selected train and test data were generated. Whilst the training sets each contained 16 minutes of acoustic matrix tokens the test sets contained five minutes of token data. The results are presented in Figures 6.28-6.30.

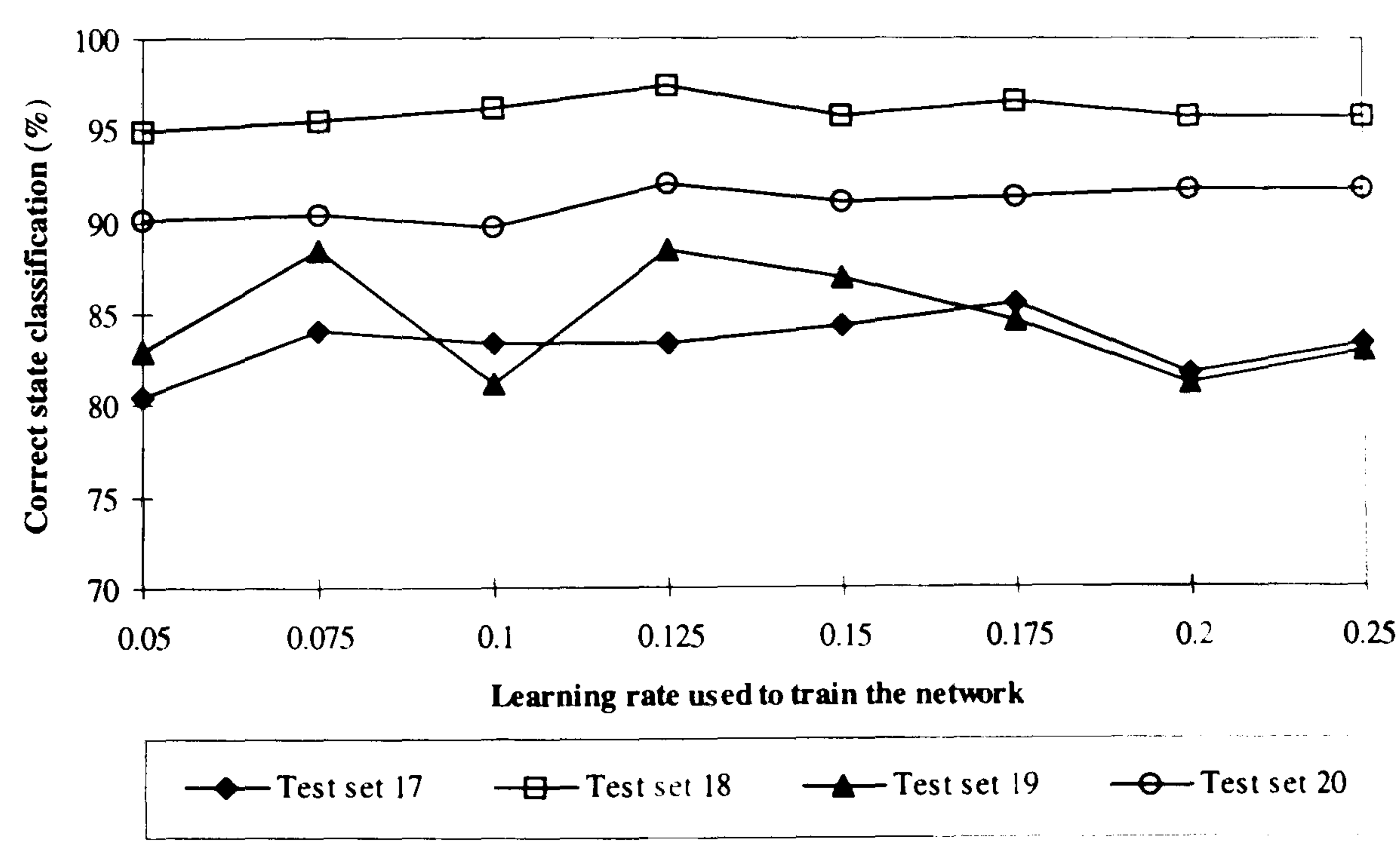

Figure 6-29 Performance of a 1,600-20-3 network configuration trained using the set 9 TES state data information and tested on randomly selected sets containing the three angular misalignment states

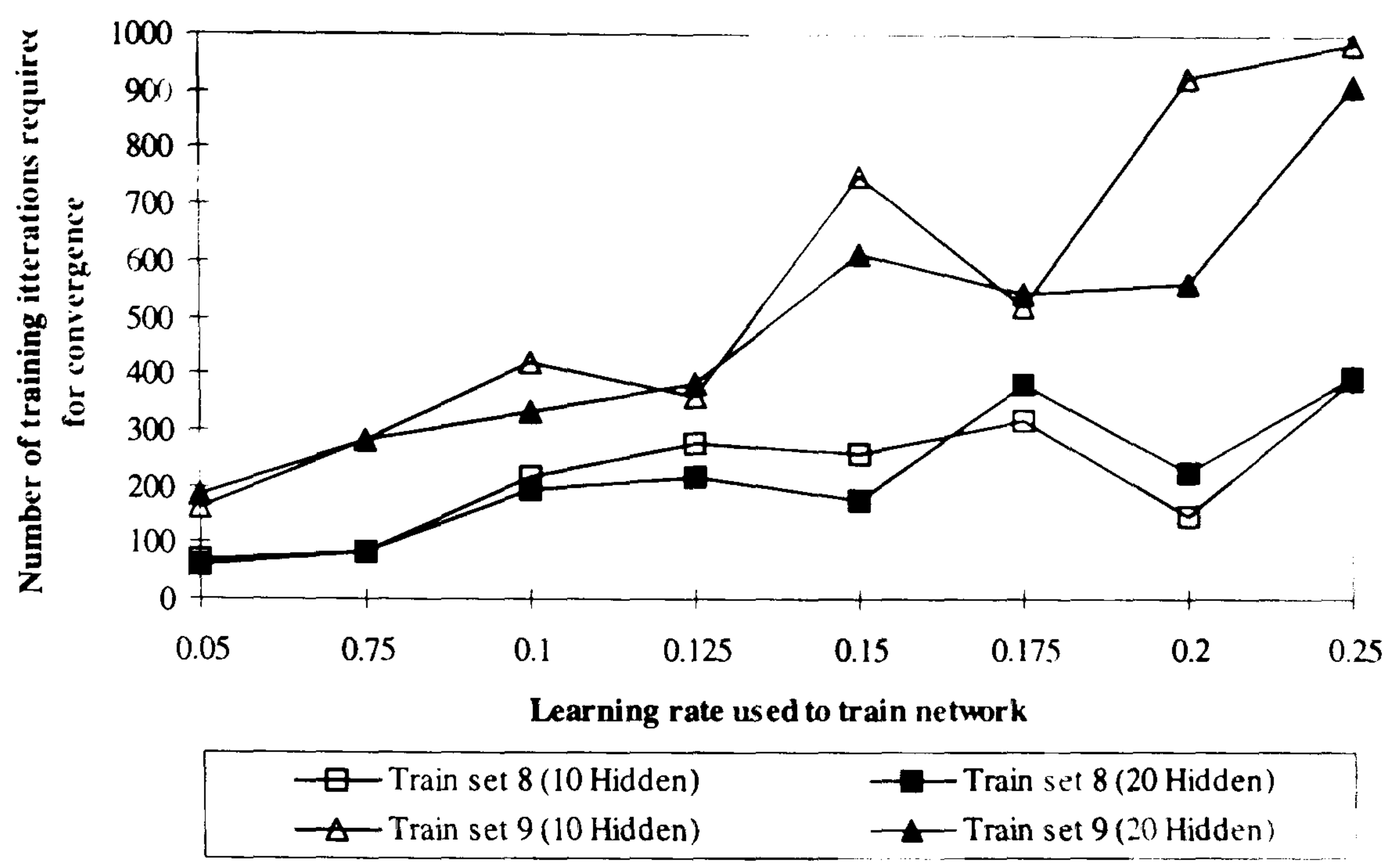

Figure 6-30 Training statistics for both of the networks evaluated

Whilst both network configurations were able to separate each of the states under evaluation the 10 node configuration produced generally inferior performance to that of the 20 node network. As in previous studies the time required for each network to converge is dependant upon the configuration and the contents of the randomly selected training sets. This is graphically illustrated in Figure 6.30. In this case set 8 provided the more consistent and rapid convergence, requiring between 60-400 iterations of the training data prior to completion when compared to that required by set 9.

However when the training and testing performances are compared it is clear that the extended convergence requirements of set 9 do not result in the generation of a more capable classifier. In fact the more erratic training behaviour seen in Figure 6.30 is carried through in those networks to the testing phase illustrated in Figure 6.29. Whilst this training set produced the best perceived performance, 97% of states correctly identified when measured against set 18, it also produces the worst perceived performance, at 80% when evaluated against set 17. Training set eight by comparison produced a more consistent classifier still capable of correctly identifying up to 95% of the test patterns presented without similar fluctuations in performance and for a much reduced training requirement. Given the network performances achieved in this simple series of tests with randomly selected data it is reasonable to conclude that amplitude A-matrix data provides sufficient acoustic information to separate even the relatively small mechanical angular misalignment errors in shafts simulated on the testbed gearbox system.

6.5 Detecting Gearbox Tooth Failure using Amplitude A-matrix Data

Toothed gears are commonly used to transfer rotational loads between close coupled shafts particularly in situations where large torsional loads are involved. As a consequence the gear teeth are subjected to rigorous operating conditions which introduce high levels of stress. This in turn causes wear to the teeth which can ultimately lead to failure. The identification of wear and failure patterns is therefore classed as a high priority in condition monitoring systems particularly when applied to safety critical systems. In the description of gearbox states in Chapter 5 (see table 5.1) three configurations were identified as simulating a point tooth failure on a single gear wheel in the testbed system. This particular tooth fault was simulated in three stages by progressive removal of material from the tip of the gear tooth. At each stage a symmetrical section of the tooth tip was removed across the entire width of the gear.

For a series of practical trials four gear tooth states ranging from healthy to worn were used to generate A-matrix data subsequently used to evaluate a series of detection networks. The healthy state, state 6, was allied to three progressively more severe failure modes. The first identifiable failure state used for these practical evaluations corresponded to a 1mm cross section of tip having been removed (state 10), the second corresponded to a 2mm section (state 11) and the third to a 3.5mm section (state 12). The mechanically driven acoustic components which combine to produce the group emission for each of these states are based on the shaft input frequency, f_i, and the shaft output frequency, f_o. In addition to these there are also harmonic meshing components based on the physical characteristics of the system which contribute to this group emission. The harmonic meshing components f_{mi} and f_{mo}, which can reach $5f$ in the spectrum, are defined below in terms of the number of teeth t_i and t_o on the respective input and output drive wheels.

$$f_{mi} = f_i \cdot t_i \qquad -(1)$$
$$f_{mo} = f_o \cdot t_o \qquad -(2)$$

Faults in the condition of the intermeshing teeth will impact upon the spectral components causing further elements as a result of the disturbance of the rotational properties of the meshing gears. In addition to these spectral effects there will also be cyclic energy disturbances caused by the rapid acceleration/deceleration cycle of the drive and driven gears relative motions to one another as the damaged tooth meshes with opposing healthy teeth. Whilst this is undeniably a simplified view of the combined acoustic effects on a system with a single tooth fault they represent a definition of the generalised type of effects introduced. Other factors such as shaft loading can also further complicate the group emission. It is clear, even from the simplified model, that identifying such relatively small faults introduced into the testbed system is by no means a trivial exercise.

The results from the network evaluations employing amplitude TES A-matrix data representing the four described states are illustrated in Figures 6.31-6.33 below. They show clearly that whilst the network is able to identify each of the four tooth fault states from the TES acoustic data the basic system performance against these tokens is below that attained in previous trials with more simple faults. The main cause of this relative performance deficit when compared to previous evaluations is the reduced diversity of acoustic data caused by the reduction in the resolution of the physical variations. As an indication of the narrowing of state boundaries and the increasing similarity between the acoustic tokens of adjacent states the training phases for both networks become lengthier. In the case of the network comprising a 20 node hidden layer this training requirement has increased by up to 30 times in some cases. Subsequent performance evaluation for the 20 node configuration against three selected test sets indicates that the best achievable performance is in the region of 75% whilst in some cases only 53% of test patterns could be correctly identified. A 10 node configuration whilst not illustrated here had similar performance characteristics.

When the same four acoustical states were retrospectively applied to a network using the amplitude histogram matrices successful network convergence became erratic and classification performance was reduced still further to below 50%. Whilst the A-matrix

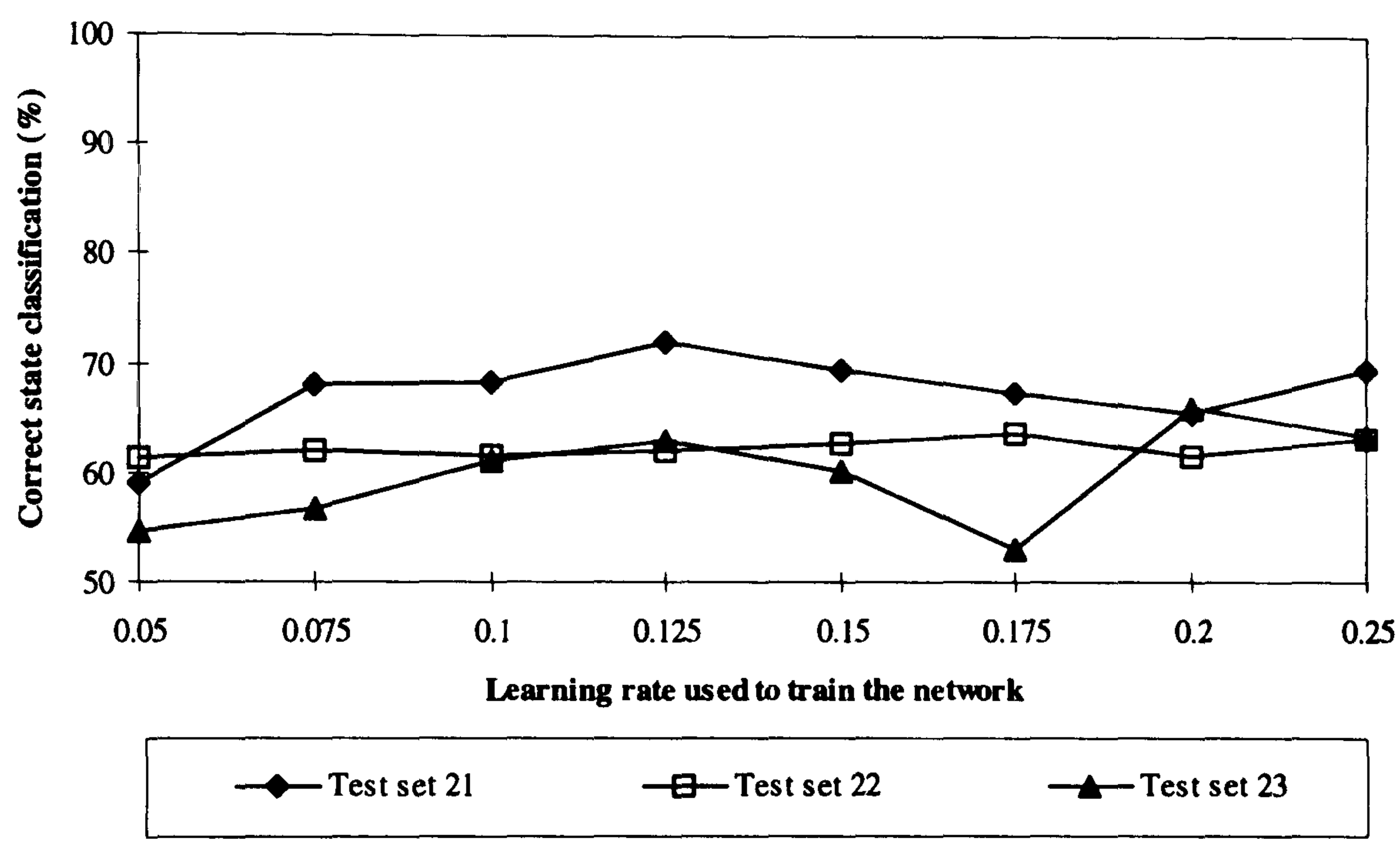

Figure 6-31 Performance of a 1,600-20-4 network configuration trained using the set 11 TES state data information and tested on randomly selected sets containing the four tooth wear states

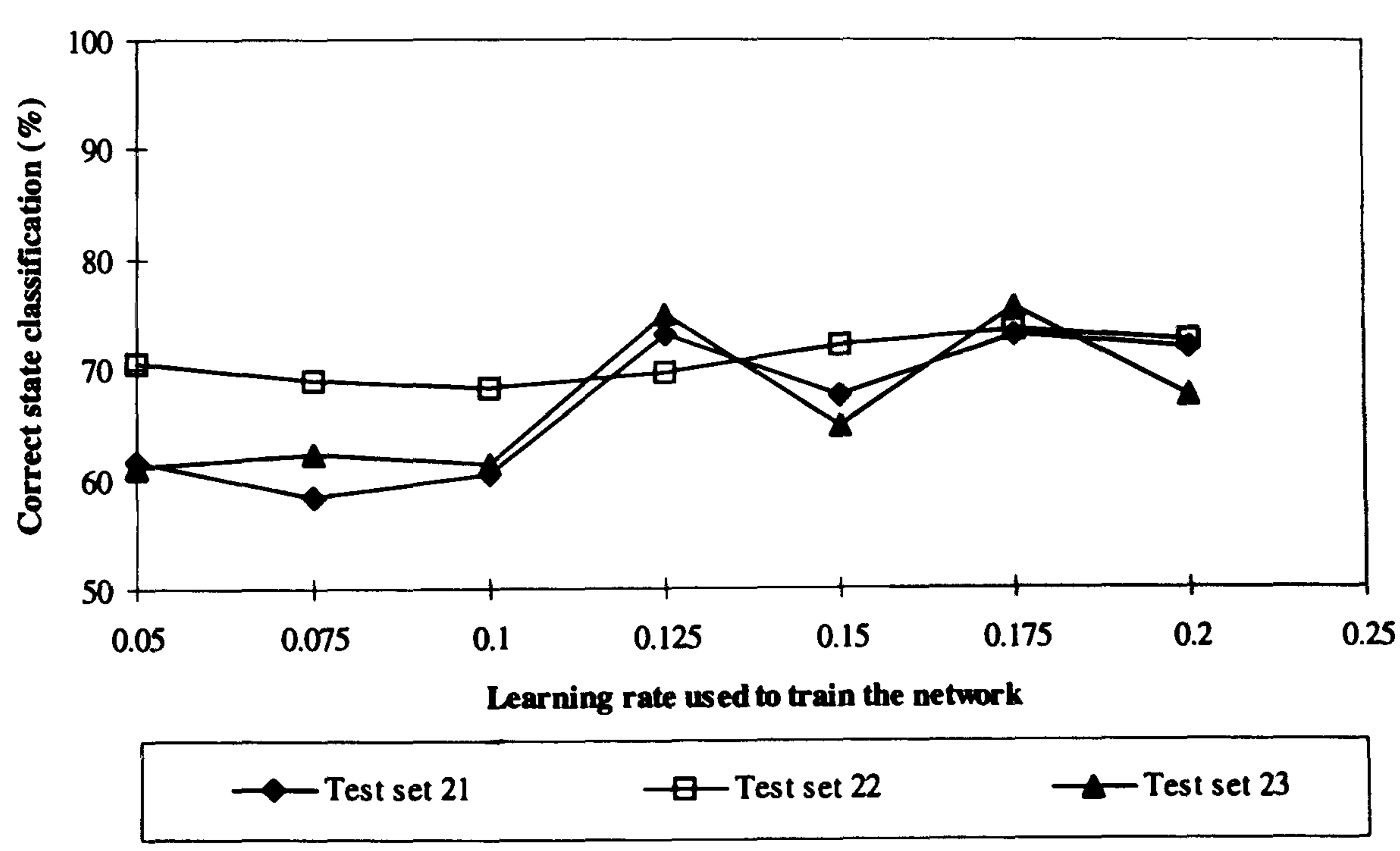

Figure 6-32 Performance of a 1,600-20-4 network configuration trained using the set 12 TES state data information and tested on randomly selected sets containing the four tooth wear states

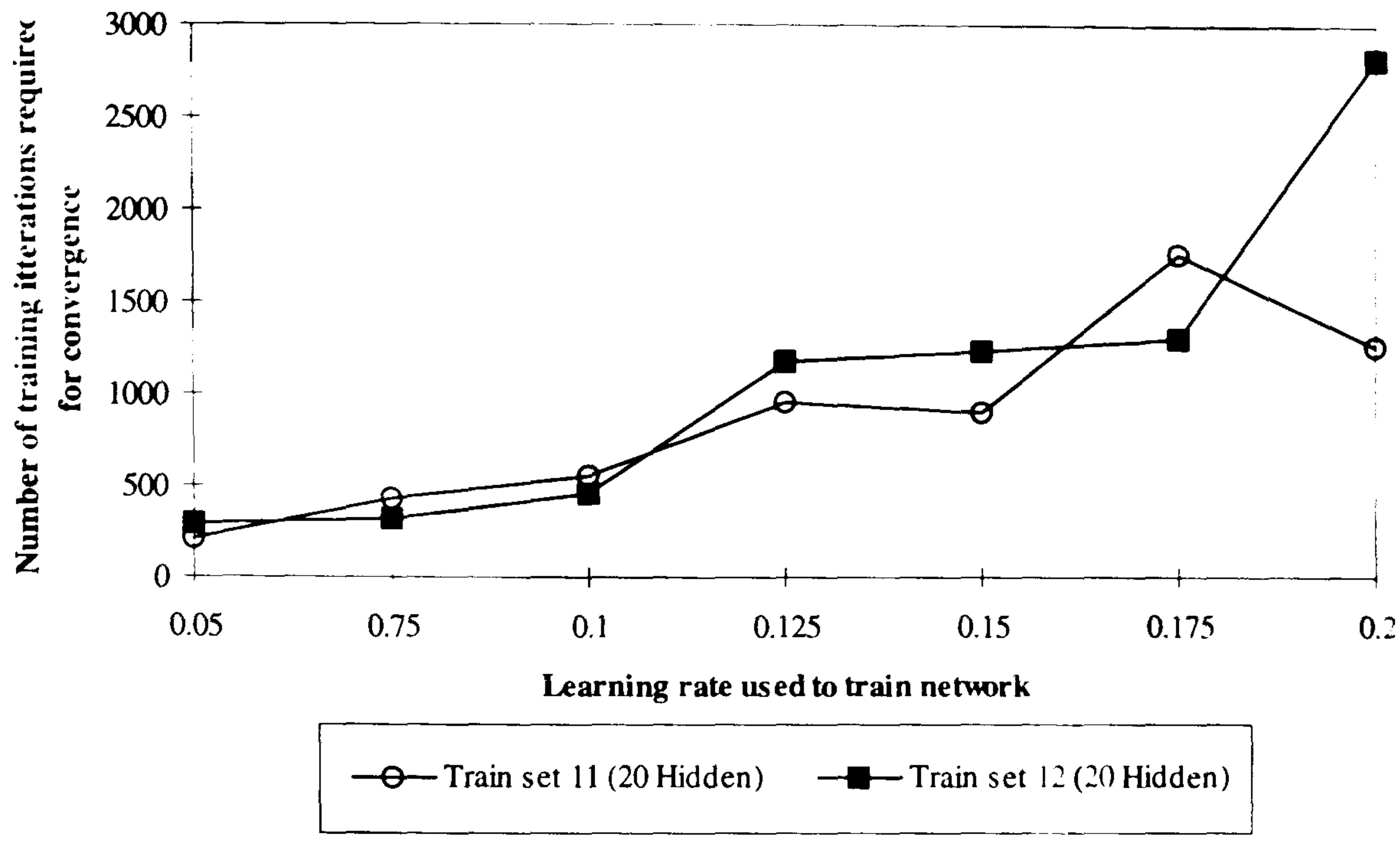

Figure 6-33 Training statistics for both of the networks evaluated

still appears to provide the best means of network presentation when compared to those used previously it is unlikely that without further processing this technique could be used to effectively monitor a system for such failure modes. Certainly from the performance attained with the data sets used in trials it is clear that some further data selection testing would be required to produce a system with even a reasonable level of classification capability.

6.5.1 Comparison of Inter-State A-Matrix Tokens by Partial Network Training

Although it was assumed that the degradation in performance noted in 6.5 was wholly due to the demands placed upon the classifier network as a result of the similarity between states this could not be proved without further examination. If indeed the degradation was associated with inadequate network sensitivity then it was reasonable to assume that the performance could be enhanced by artificially reducing the state resolution required by the classifier. During the earlier trials each of the four applied states had incremental physical differences corresponding to not more than 1.5mm of tooth section removed from a single tooth on a gear containing 43 teeth. In order to evaluate a reduction in the classification resolution an additional series of trials were performed by applying the four states in pairs rather than as a single set. In this way the networks perceived acoustic similarity between pairs of states may be estimated in isolation.

The trials were performed using the same 20 element hidden layer, 4 output state configuration used previously to enable direct comparisons to be made with the earlier findings. In addition training during each of the pair tests was carried out using the same quantity of acoustic data for each specific state as had been used in the earlier trials. Thus each of the new training sets contained exactly half of the data used for previous four state training, or seven minutes per state in total. To maintain an adequate level of confidence for later performance correlation all the remaining parameters associated with training were unchanged.

If the premise regarding state resolution was accurate then classification performance figures attained from the pair trials and illustrated in Figure 6.35 would be expected to increase the greater the physical separation in status became. The first three columns starting from the left of the graph (in Figure 6.35) correspond to adjacent tooth state boundaries and can be expected to result in relatively more errors than the remaining three columns which correspond to boundaries of more than a single physical state. The network performance measured for these first three state pairs, each corresponding to tooth wear disparities of between 1-1.5mm, ranges from 73% to 84%. The remaining three state pairs correspond to network classification against states differing in physical terms by between 2-3.5mm. Whilst two of the three tests did show improvements in classification to 94-96% consistent with the earlier predictions the third comparison actually resulted in reduced rather than enhanced performance. States 10 and 12 however could only be separated by the network correctly during the evaluation stage in 53% of cases. This contradicted the earlier expectations and raised doubts regarding the premise that the earlier evaluation results discussed in 6.5 had been the direct result of

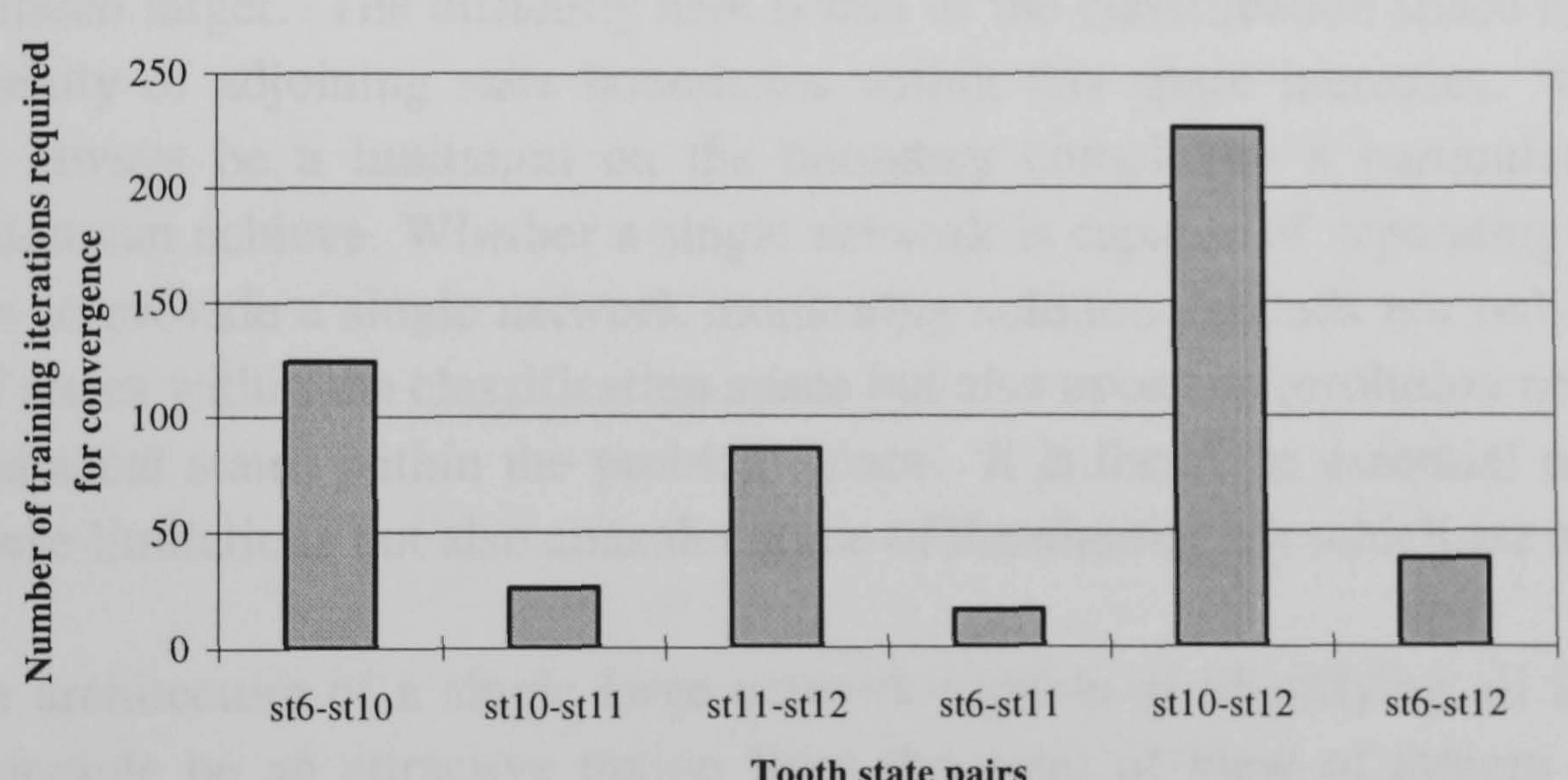

Figure 6-34 Training statistics for each of the different state pairs with the network

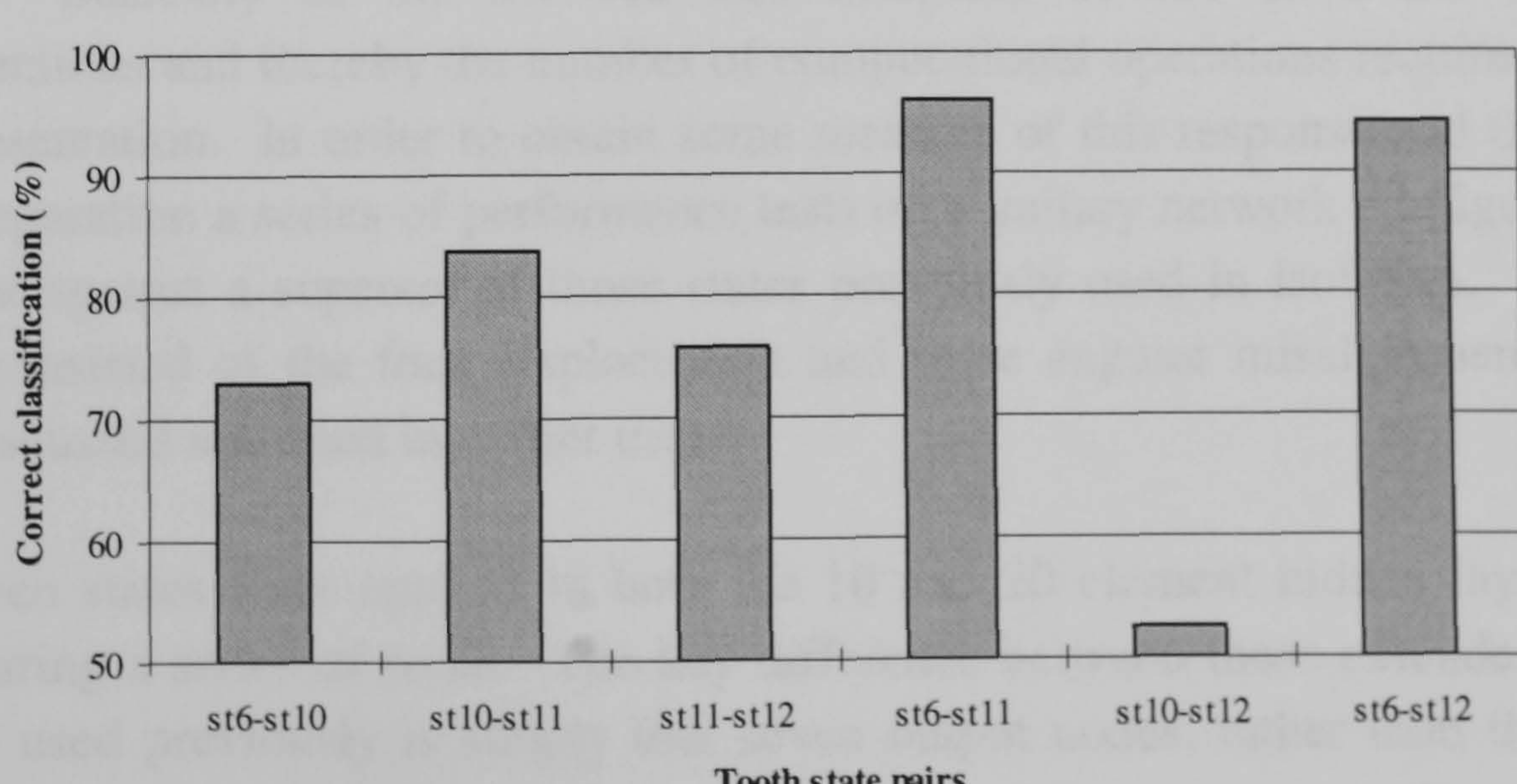

Figure 6-35 Performance of the network under evaluation with amplitude A-matrix data sets corresponding to each state pair

unreasonable demands being placed upon the network.

However what has not been discussed up to this point are the non-linear effects elicited upon the A-matrices by changes in the TES data stream caused by state changes. In actuality the relationship between the gearbox state, the acoustic emissions and the A-matrices subsequently generated is complex. In this instance it has produced matrix templates for two gearbox states which despite being relatively more physically dissimilar than other state pairs have very similar matrix properties when presented to the network for classification.

6.6 Evaluation of the Effects of Expanding the Size of the Classification Space in a Fully Interconnected Network Architecture

Up to now the classification space for each of the network configurations subjected to performance evaluation has been limited to only three or four unique states. If a TES based neural classification scheme is to be considered for inclusion in a practical monitoring system it is reasonable to expect that the number of potential fault states would be much larger. The difficulty here is that as the classification space expands so the complexity of adjoining state boundaries within this space increases. Ultimately there will always be a limitation on the boundary complexity a particular network configuration can achieve. Whether a single network is capable of separating sufficient fault states to provide a single network monitoring solution depends not only upon the number of states within the classification space but also upon the resolution necessary to separate adjacent states within the problem space. It is therefore essential not only to explore these limitations but also consider some of the alternatives which are available.

Whilst the architecture of a single large network capable of identifying all fault states may in principle be an attractive option from the point of view of system simplicity there are some potential drawbacks. Primary amongst these are the computational limitations associated with a single large fully interconnected network in terms of response. Basically as the network size increases so too does the number of interconnections and thereby the number of computational operations required per input vector presentation. In order to obtain some measure of this response and the capacity for data separation a series of performance tests on a unitary network configuration was carried out against a superset of those states previously used in isolation. In all, this superset consisted of the four displacement and three angular misalignment positions already discussed and used in earlier trials.

These seven states were applied to both the 10 and 20 element hidden layer network models during a series of trials. The key difference between these extended networks and those used previously is simply that seven output nodes, rather than the previous three or four, are required to cater for the expansion of the state space. The elements of network operation which were monitored during the trials were the basic error performance and the number of training iterations required to facilitate this. The time taken to present all the training data matrices during a single back propagation training iteration will obviously be increased as a result of the data expansion necessary to cater for each of the states in the extended space. However the increase or otherwise in the number of these iterations required to achieve convergence provides a better indication of the relative complexity of the data space expansion. The raw classification performance on unseen test data also provides further indication of the sensitivity of the extended network to the increase in diversity of the state space.

The training itself was performed by applying A-matrix data generated from 35 minutes of recorded acoustic samples, five minutes per state for each of the seven states. The findings attributed to this particular trial provide encouragement for the modest expansion, from four to seven, of the state space allocated to a unitary fully

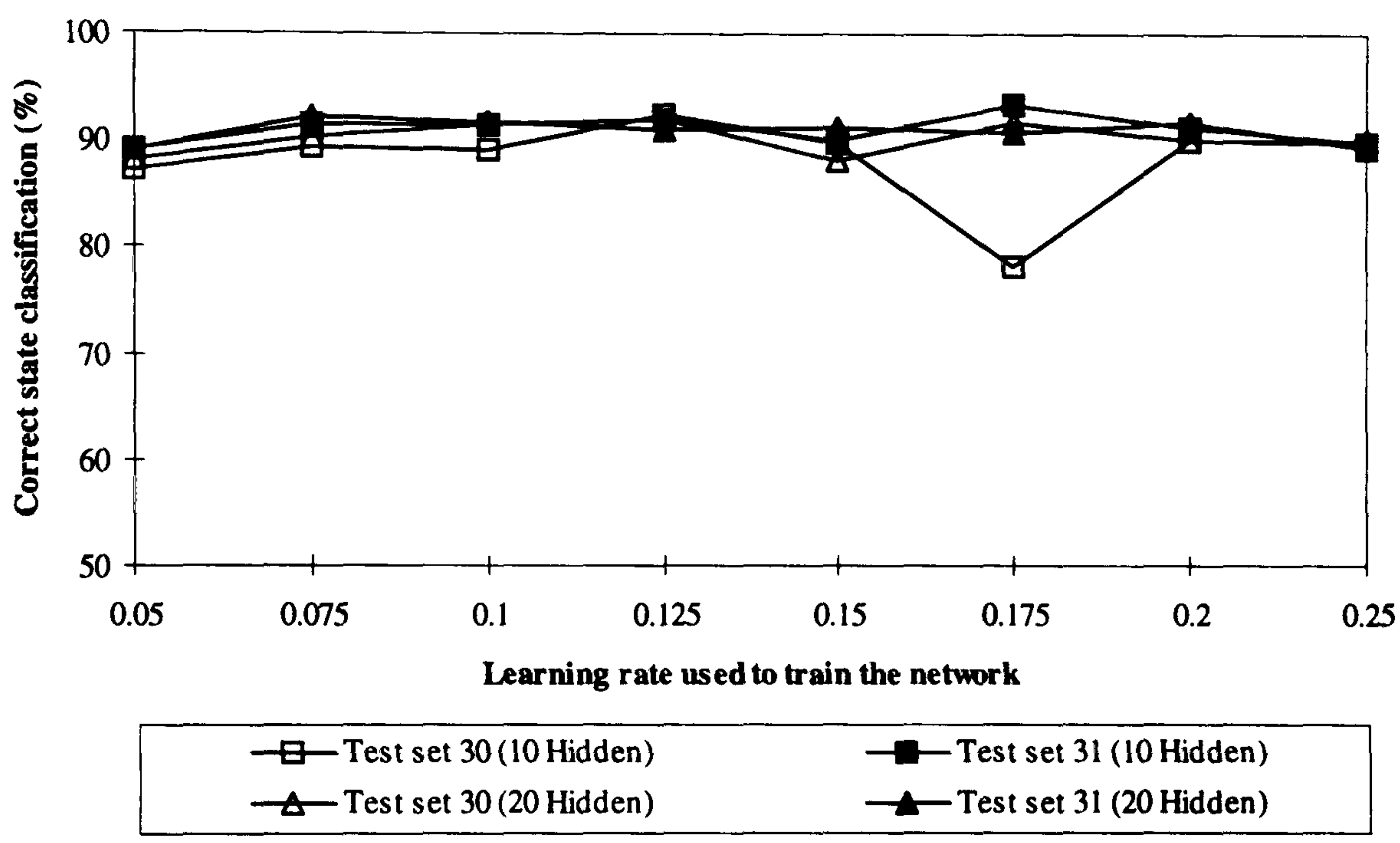

Figure 6-36 Performance of the extended network configurations trained using set 20 containing amplitude A-matrix data relating to the seven states and tested using two randomly selected sets

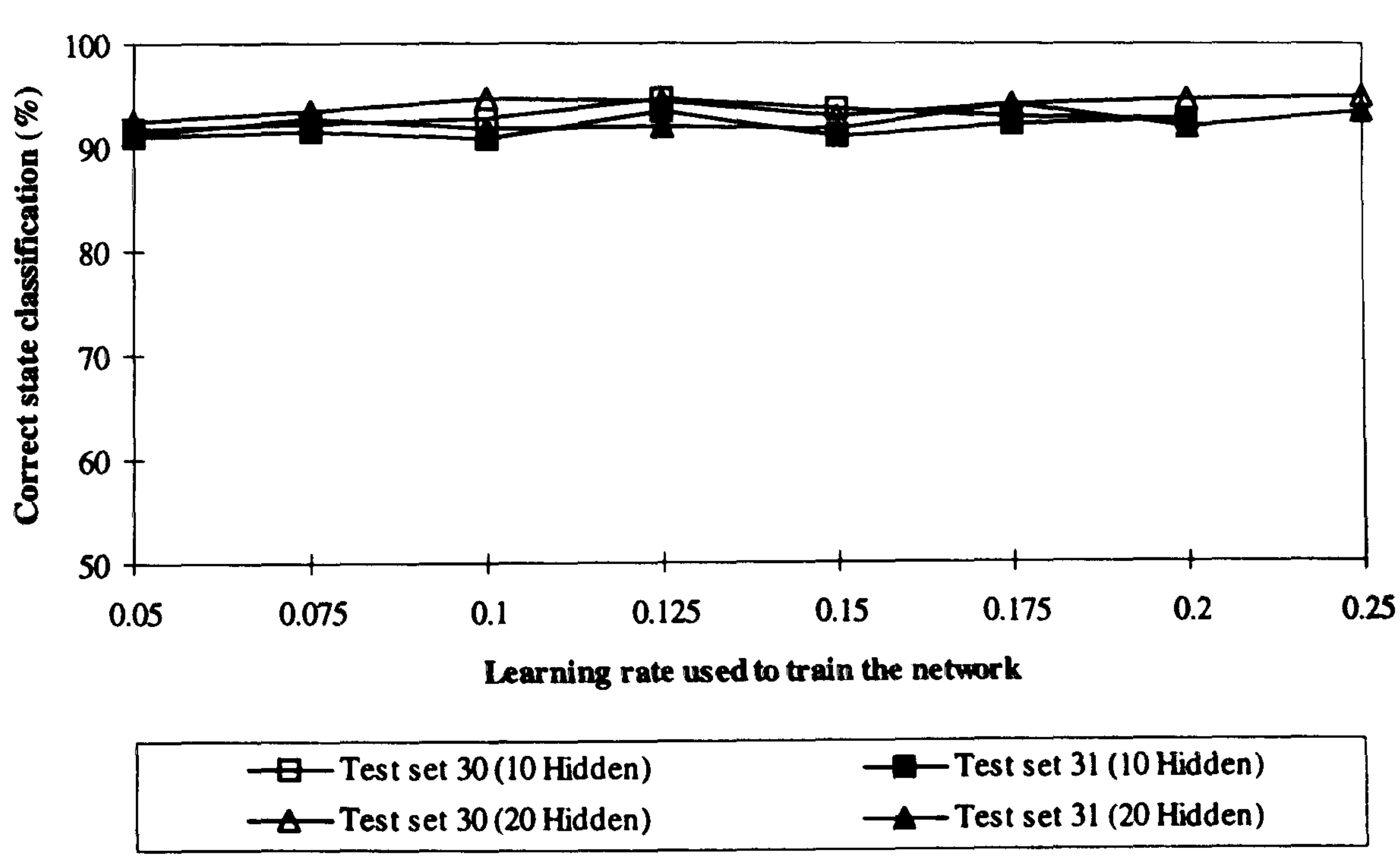

Figure 6-37 Performance of the extended network configurations trained using set 21 containing amplitude A-matrix data relating to the seven states and tested using two randomly selected sets

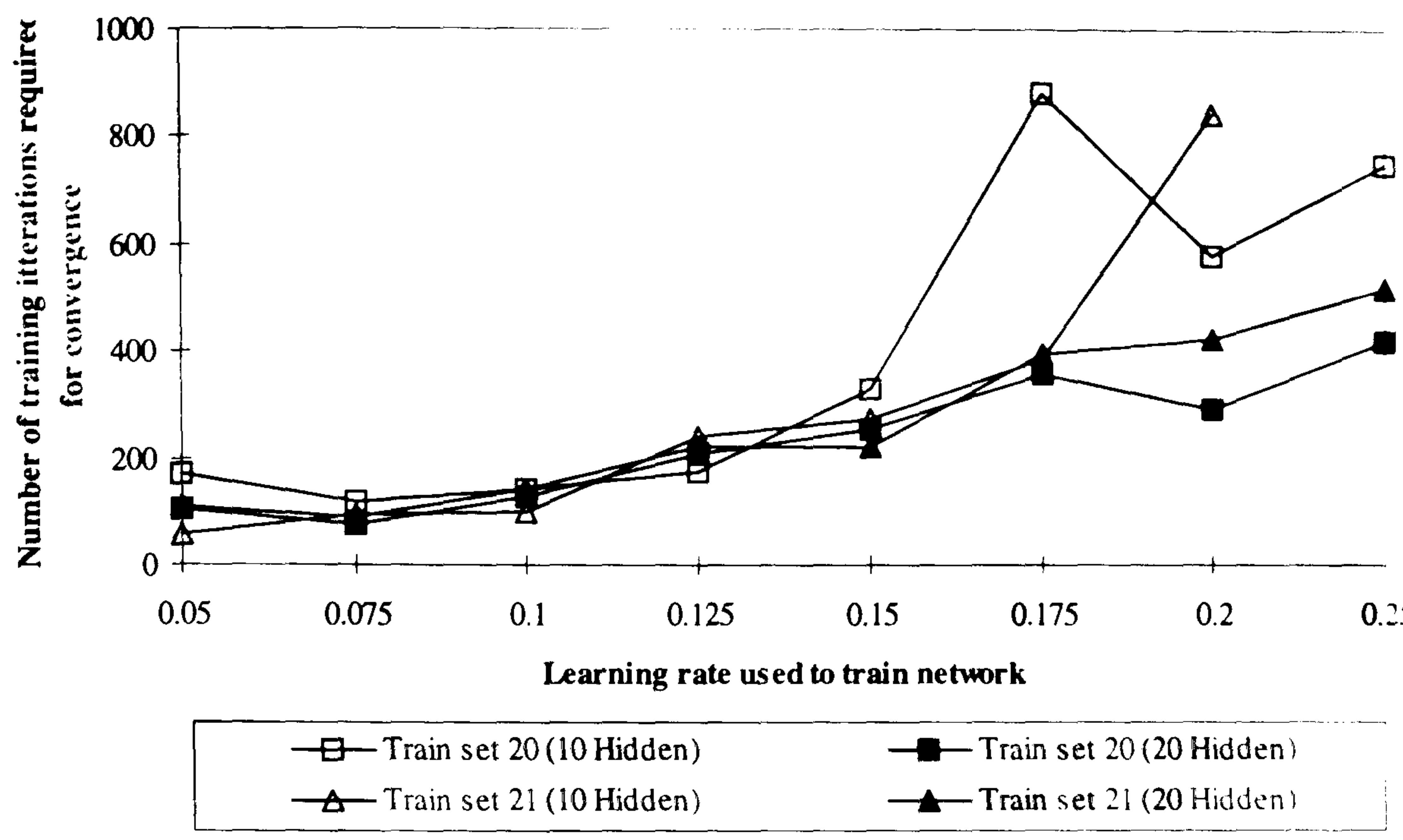

Figure 6-38 Training statistics for both of the networks evaluated

interconnected network. Whilst a reduction of between 5-8% in overall performance was noted when compared with the earlier classification of displacement alignment states, a general improvement was noted over previous angular alignment classification. Similarly performance fell below the best achieved previously for these states (97%) but showed consistent improvement over the range. More notable was that the general consistency of classification of each of these states was improved as a result of the expansion. Where previously several of the train and test set combinations introduced discrepancies in performance of up to 15% making the definition of acceptable training procedures more complex the expansion of the data space reduced this. With the exception of the α=0.175 point in Figure 6.36 the performance between data set combinations within the extended state space did not vary by more than 4%.

In comparisons of the relative network performance the 20 node hidden layer configuration proved to be marginally better in terms of both rapidity and consistency of network convergence and classification. This network was able to outperform the 10 node configuration by up to 2.3% for 80% of all train and test set combinations used in the trials. It is reasonable to conclude from this modest state expansion trial that increasing the state space does not necessarily incur penalties in terms of training requirements or classification performance. However without further extension of the space it was difficult to estimate were the limitations of such an expansion lie.

6.7 Evaluation of the Internal Network Architecture on Classification

So far the effects of the network architecture on performance have been limited to studies on the effects of the number of nodes within the hidden layer under fixed conditions. All of these networks have been fully interconnected with all input nodes connected to all hidden layer nodes and all hidden layer nodes connected to all output, or state, nodes. However the networks internal connectivity may be considered flexible in the same way as the number of nodes or the training parameters used to configure them have been. Both of these other two properties have been shown to affect the convergence and classification capabilities when tested against real data. Having already covered in some detail now the capabilities of a fully interconnected architecture on a number of different acoustic state data matrices it is time now to study the interconnectivity of the networks themselves. In this section consideration will be given to pruning of the internal nodal connections and the effects this has upon state identification by the network. Investigations focus upon two specific architectural variations within the hidden-to-output layer connections of the networks. Whilst both concern the implementation of a partially interconnected interface between the two layers, the first has no overlapping of the associated interconnections and the second permits partial overlapping of the interconnections to be applied. For further details regarding these configuration parameters the reader should refer to Section 4.3 in Chapter 4

6.7.1 Applying the Seven State Space to a Partially Interconnected Network Model

Initial evaluation of partially interconnected structures was carried out by developing two networks with non-overlapping interconnections between the hidden and output layers. The first of these networks was developed with a 21 node hidden layer so that performance comparisons could be made against the 20 node fully interconnected configuration used in section 6.6. The expansion to 21 nodes from the 20 used in the fully interconnected configuration was necessary due to the necessity for a seven node output layer with no hidden layer overlaps. For the purposes of this comparison any enhancement or degradation in performance caused by the addition of this single extra node will be considered as negligible. In this configuration nodes in the hidden layer are connected to each of the seven output nodes, or classes, via only three internal interconnections. This represents a seven fold reduction in output layer connectivity over the fully connected model which in turn reduces the number of computational operations necessary during both learning and classification. In addition to the 21 node model a second configuration containing 35 hidden nodes was developed for evaluation. The 14 extra nodes in this configuration provide a further two hidden nodes per output class and are used to study the significance upon classification of varying the internodal tasking demands. Both networks were used to classify the seven gearbox shaft states, previously used in Section 6.6, to evaluate the fully interconnected configuration.

In the trials training requirements for both configurations varied between 60 and 920 iterations depending upon the data set, the specific configuration and the α value used during training. When the 35 node configuration is presented with identical training data it generally converged more rapidly than the 21 node configuration, on average requiring 10% and 30% fewer data iterations. When the training statistics for these partially interconnected configurations are compared to similar fully interconnected models an increase in the number of iterations required for convergence is noted. In fact, from observations made during trials with identical training and test data the fully interconnected architectures converge on average after 30% fewer iterations than their partially interconnected counterparts despite having more interconnections and equivalent numbers of internal nodes.

The performance of both networks, illustrated in Figures 6.39-6.41, again exhibits good consistency over the full range of training parameters as the fully interconnected system had done. The 35 node architecture provides only marginally better classification, between 0.3-0.7%, than the 21 node architecture classifying on average 90.6% of all test patterns correctly. However the general performance of the partially interconnected networks is marginally below that of similar fully interconnected networks. When the performance of the 21 node variant is compared to previous work with a 20 node fully interconnected network the degradation in performance is of the order of 1-2%.

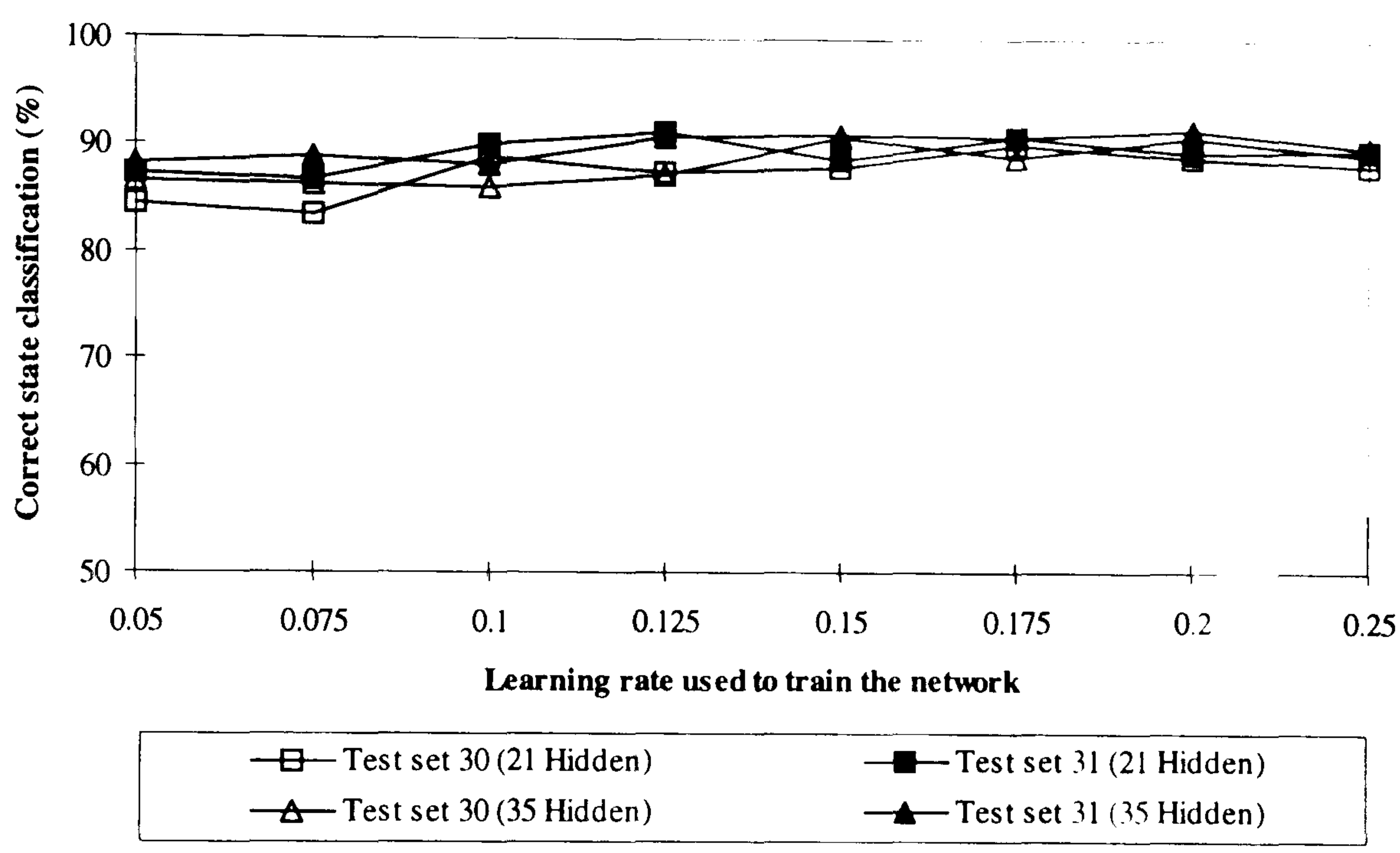

Figure 6-39 Performance of the 21 and 35 hidden node network configurations trained using set 20 containing amplitude A-matrix data relating to the seven shaft states and tested using two randomly selected sets.

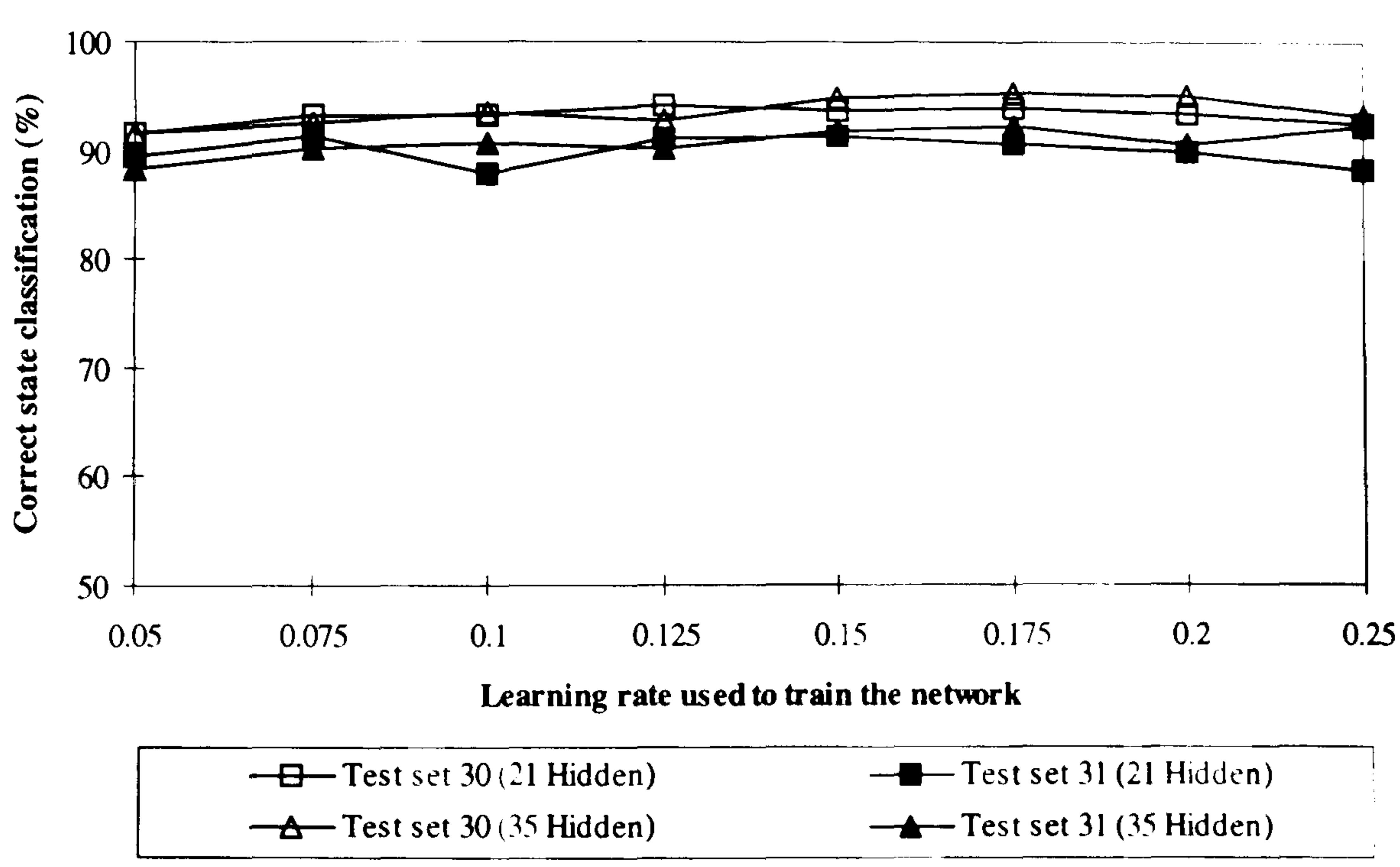

Figure 6-40 Performance of the 21 and 35 hidden node network configurations trained using set 21 containing amplitude A-matrix data relating to the seven shaft states and tested using two randomly selected sets.

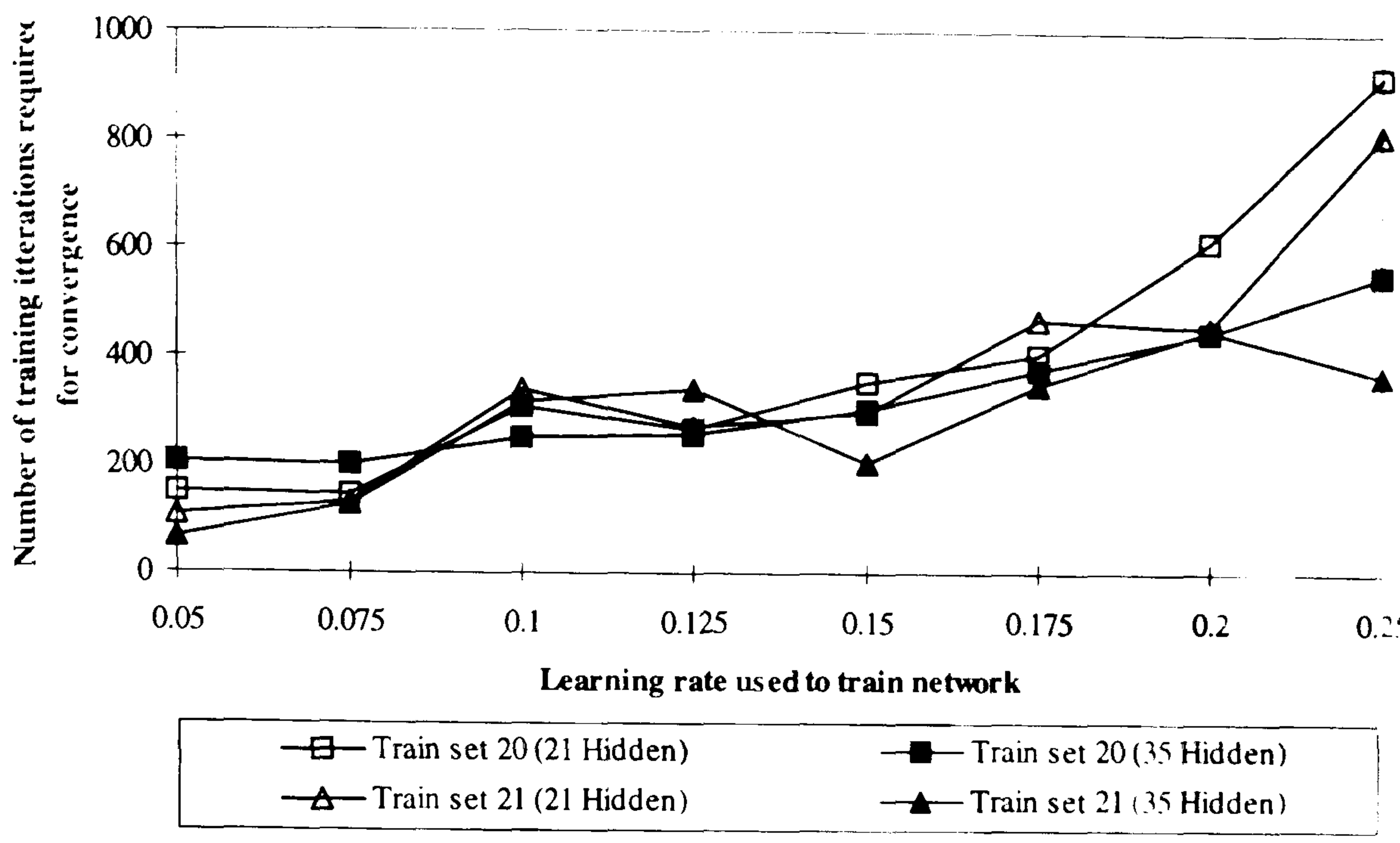

Figure 6-41 Training statistics for both partially interconnected networks.

6.7.2 Applying the Partially Interconnected Network Model to Tooth Fault Analysis

The reason for selecting the tooth fault detection problem for further testing with a variety of partially interconnected networks was to identify whether or not the decision boundaries can be further optimised through the rationalisation of the network interconnections. Whilst in earlier trials the general performance of fully interconnected network configurations was adequate for most states one specific inter-state boundary proved difficult to define. In the initial trials the separation of states 10 and 12 caused significantly more errors, in the region of 47%, as a result of this ill defined boundary. By pruning the networks internal connectivity the number of conflicting signals received by each of the output, or state, nodes from the hidden layer is reduced. This in turn reduces the burden on each output class to simultaneously satisfy all the requirements of the matrix vector applied at the input stage. The intention of the trials is to determine whether such rationalisation is sufficient to enhance the performance of the network when presented with the conflicting requirements of states 10 and 12.

The internal pruning takes two basic forms both of which have previously been discussed in Chapter 4. The first is complete compartmentalisation of output classes whereby each output class is connected to a unique subset of the hidden nodes with no overlapping, or interaction, between nodal subsets. Likewise the second method attempts to compartmentalise the classes but allows a degree of overlapping, or crosstalk between the internal interconnections. For the purposes of this trial four partially interconnected architectures were considered and assessed against the baseline performance of a fully interconnected network model. The internal structure and nodal dimensions of the four separate configurations are listed below.

1. A 20 node network with partial interconnections between the hidden and output layers. All the feed-through connections are distributed symmetrically

with groups of five allocated per output node with no overlap between the adjacent groups.

2. A 32 node network with a comparable architecture to the first model. However in this instance each output node has a group of 8 unique connections from hidden layer nodes, again with no overlap.

3. A 20 node network with 8 connections from the hidden layer to each output node. Since only 20 nodes are in the hidden layer this configuration employs symmetrical overlapping of the feed-through groups. The overlap is applied at each boundary so that four hidden nodes are shared between adjacent output classes as is described in Chapter 4.

4. A 19 node network similar to (3) again with a symmetrical overlap but this time with 7 hidden layer nodes being allocated to each output class node.

From the results acquired during evaluation of the four networks described in Figure 6.42 it will be noted that in all but one pair state, st10-st12, the partially interconnected networks take longer to converge than the benchmark 20 node fully interconnected network. In the case of the st6-st10 pair up to four times as many training iterations may be necessary. In this respect they follow the pattern set in the earlier evaluation of partially interconnected structures outlined in Section 6.7.1. However, in contrast to the previous trials with partially connected networks most of the test evaluations performed with the gear tooth states indicate that classification proficiency is actually improved, in one particular instance by 6.6%, rather than reduced as they had been earlier in Section 6.5. However despite generally acceptable class separation performance of 75-95% the st10-st12 boundary still remained difficult to define with sufficient accurately. Despite an improvement in separation of approximately 3.5% on the benchmark figure produced by the fully interconnected network for these states only about 56% of all test matrices were correctly identified.

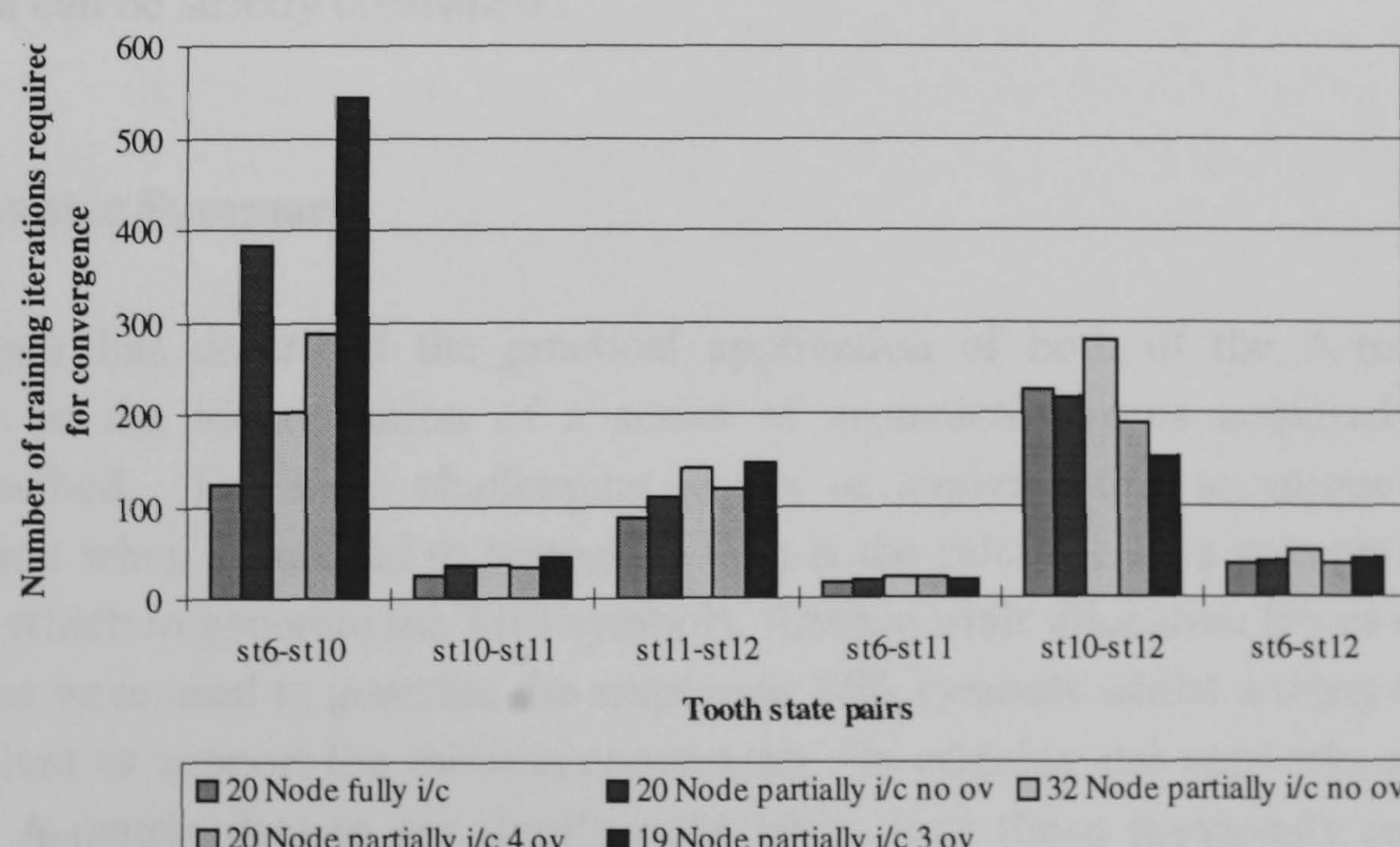

Figure 6-42 Training statistics for the five different network architectures when trained to recognise each of the four system gear tooth state pairs

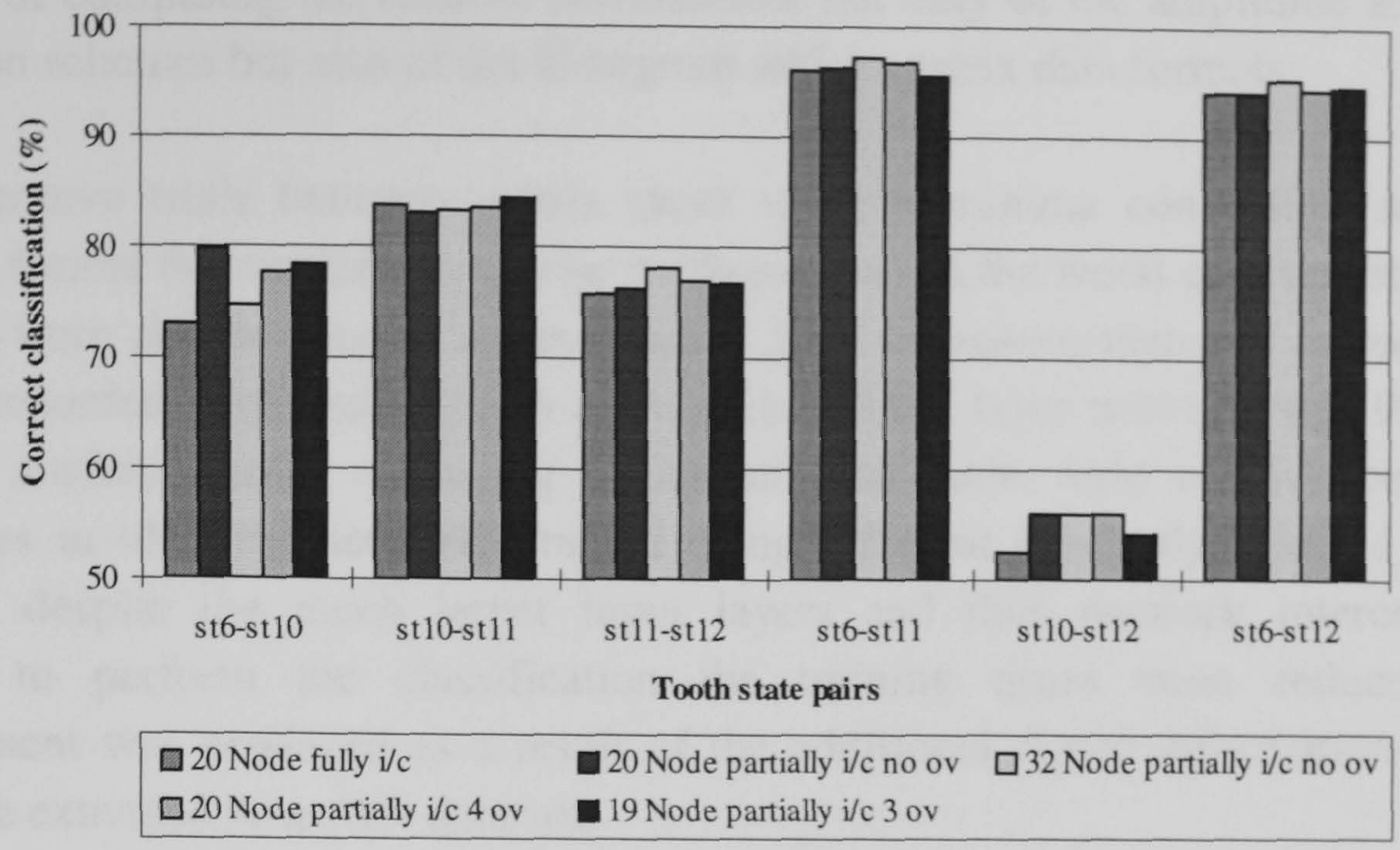

Figure 6-43 Performance of all five network configurations when evaluated against a test set containing amplitude A-matrix data corresponding to state pairs

In general these trials have proved inconclusive in terms of identifying fully the effects of the internal network architecture. It does however seem clear from the tests completed that there are, with these acoustic tokens at least, no significant gains to be made in terms of improved training and only relatively small enhancements to be made in certain cases in terms of subsequent classification. Reductions in performance, where they exist, are not offset by comparable improvements in computational overhead attained by the rationalisation of the network interconnections. Only significant pruning of the input to hidden layer connections is likely to produce any meaningful savings in the time taken to perform each matrix training and classification presentation. The drawbacks in pruning this layer however are the likely complications introduced to the classification process itself by outputs nodes relying upon partial information for decisions. Such a scheme is only really applicable in situations in which the failure modes can be clearly defined at the outset and the input vector generation for network application can be strictly controlled .

6.8 Chapter Summary

This Chapter has described the practical application of both of the A-matrix TES techniques to the identification of a series of acoustical tokens acquired from the gearbox testbed. The most challenging aspect of applying this technique to neural identification when compared to histogram data is the selection of a suitable allocation table with which to generate the TES symbols. During trials allocation tables containing forty entries were used to generate the amplitude TES symbols whilst a thirty entry table was sufficient to support the minima conversion. In addition the networks required to apply this A-matrix data to are significantly larger than those previously required for histogram application. The trials detailed in this Chapter were performed with the intention of comparing the relative performance not only of the amplitude and minima conversion schemes but also of the histogram and A-matrix data formats.

In comparative trials between matrix types using a minima conversion scheme the A-matrix format demonstrated superior performance. In the worst case test scenario the networks were able to classify approximately 90% of unseen matrices correctly whilst the best recorded performance with a ten node hidden layer network was 99.5%. As with the studies carried out using histogram data there were relative performance differences in identical networks trained using different randomly selected data sets. However despite the much larger input layers and thus network interconnections required to perform the classification the training times were reduced. This improvement was produced as a result of the additional signal information contained within the extended A-matrix data sets.

A similar, if less dramatic, improvement was identified during early trials using the amplitude conversion scheme in conjunction with the A-matrix data format. Classification performance with this data format was generally in the 93-98% region with the worst case being 82% and the best achieving 100%. Once again despite the requirement for an even greater expansion of the input layer, to 1,600 nodes, the training times were reduced significantly. In the early trials with amplitude TES A-matrices some discrepancies were identified between specific data sets which were not readily attributable to simple data differences. In an attempt to identify the cause of this discrepancy a series of additional tests were performed to compare the performance with a modified data set. These tests focused on both specific physical and general bias characteristics of the data set identified as the cause of the performance discrepancy. The conclusion after these tests was that the discrepancy had been caused by specific physical characteristics of a particular acoustic record within the data set.

The amplitude TES A-matrix data format, having been identified as the most desirable means of generating condition information, was applied to two other identification problems which were simulated with the testbed gearbox. When applied to angular misalignment conditions it was capable of identifying between 80-95% of all unseen data correctly. Whilst this was classed as acceptable the technique proved less so against acoustic data representing tooth failures. In trials using acoustic tokens representing four failure stages in a single tooth of between 1-3.5mm the network performance degraded to between 53-75%. Given this perceived capability it would seem that even for a reasonable level of performance to be achieved some additional data selection would be necessary to optimise the classification. Some work was performed using networks trained on subsets of these tooth failure states to identify whether the performance was attributable specifically to the resolution required to distinguish the individual states or not. There was some evidence to support this notion. When states which differed physically by 1-1.5mm were compared correct identification was achieved in 73-84% of cases. In contrast when the physical discrepancy was widened to between 2-3.5mm perceived performance improved to between 94-96%. However in the case of states 10 and 12 which correspond to physical differences of

2.5mm the performance was drastically reduced to 53%. This highlighted one of the difficulties associated with classification of this type. That is that there is not necessarily a linear relationship between the physical aspects of a particular condition and the acoustic emissions generated as a result of it. When this is combined with a learning based classification mechanism as opposed to the more conventional programmed approach relationships of this sort can result in variable system performance. In the case of the tooth failure identification the performance of the network is acceptable in most instances. However the combined performance, which was found previously to be inadequate, was mainly caused by the difficulties surrounding the separation of states 10 and 12.

The expansion of the state space associated with a particular network configuration was investigated by applying four displacement and three misalignment states to a series of network configurations. Whilst there will ultimately be a physical limitation attributable to a particular network configuration in terms of the complexity of the boundaries it can define, the seven state system which was implemented for trials did not surpass this. During the trials a reduction in the performance against the displacement misalignment states of between five and eight percent was noted. However in contrast a small improvement in the general performance with the angular misalignment states was observed. In general though the consistency of the networks over a range of training runs was much improved with a performance against unseen data of approximately 90%.

The final aspect of the application of neural networks to the classification of TES data which was considered was the connectivity of the networks themselves. The investigations performed on this element of the application concerned the removal of some of the hidden-to-output layer connections. The results of the investigations indicated that whilst in most cases the operational accuracy is reduced by a small percentage, in the case of the tooth failure identification improvements of up to 6.6% could be achieved. However in all cases the training requirements were extended as a result of the elimination of nodal connections. In conclusion the small gains achieved in processor overhead as a result of the rationalisation of the network architecture did not provide sufficient performance benefits to offset the additional training requirements.

Chapter 7

Page

7. Final Summary

This Chapter summarises the research undertaken during the period of study and already detailed in the main body of this thesis. It outlines a number of objectives, defined at the outset of the study period, as well as a discussion of the key features of the monitoring techniques developed to satisfy these objectives during the research undertaken. The discussion of these techniques will be made with reference to practical trials performed upon a testbed gearbox system which was constructed to simulate a series of identifiable physical faults. The last two sections in this Chapter highlight the key achievements of the work and seek to identify a number of areas within the scope of the work which the author feels to be worthy of further study.

The work contained within this thesis describes the development and evaluation of a TES based neural network condition monitoring system applied to a custom built simplified gearbox fault simulator. The work consists of a discussion of two TES coding schemes and two selected neural application strategies. A series of practical trials concentrated on the identification of a number of simple faults seeded into the simulation testbed and recorded acoustically. The trials were used to carry out a fundamental evaluation of the problems associated with these novel techniques as well as determining the capabilities of such a system using the various different data formats.

7.1 The Original Objectives

Many of the traditional condition monitoring techniques adopt empirical analysis or theoretical modelling as a means of identifying state indicators. These more conventional techniques can be relatively simple to apply in well defined problems where localised environmental conditions either do not pose a threat to the success of the monitoring or can be strictly controlled. The aim of this work has been to study the potential of novel methods for identifying the fault modes which occur in machinery on-line, in real-time and without necessarily imposing some of the more severe restrictions upon the manner in which this processing is performed. The self-adaptive learning techniques of neural networks can potentially make simpler the application of condition monitoring to a range of machines by removing the necessity for specifically defining condition indicators or controlling localised environmental additive noise. However at the outset there were several aspects of the proposed solution which necessitated careful evaluation to determine whether in practice these methods were capable of producing the kind of enhancements envisaged. Clearly the potential exists for removing certain undesirable aspects of the more conventional techniques only to replace them with different ones associated with the new techniques.

The impetus for this study is provided by the increasing demand for greater control of the production process within the industrial environment. The monitoring of specific elements within this process is an integral part of the ability to control them in order to both improve efficiency and enhance safety. Traditionally, monitoring has been a human intensive operation requiring a combination of both skill and experience. As the demands placed upon equipment intensify, the necessity to operate more efficiently and within tighter physical tolerances has led to ever higher operational stresses and thus indirectly increased operational costs. As a result of these demands the research into automated methods of condition monitoring has accelerated considerably in recent years. Many methods now exist which are capable of reducing the operational costs attributed to the maintenance of a range of complex machine types. They employ both a variety of sensor types and condition extraction algorithms to identify the state of machines either during operation or periodic maintenance. There are however a number of problems which can reduce the effectiveness of some or all of the techniques currently available. A few are outlined below:

1. System cost.
2. Implementational complexity.
3. Overheads associated with the skilled operatives.
4. Sensitivity of the identification techniques to localised conditions.

The intention of this research was to evaluate some novel techniques which seek to overcome some or all of these concerns. The key focus at the outset of the research period was the identification of techniques which answered two of these concerns. These related specifically to the operational overheads associated with the necessity for skilled personnel and the simplicity and ease of use of the systems developed.

Neural networks were identified at an early stage as a potential means of reducing the necessity for skilled personnel without necessarily eliminating the human strengths of experience and adaptability. The acceleration in their use for a wide range of tasks in recent years owes much to their ability to learn rather than be programmed. This makes them particularly useful in situations where complex problems are difficult to fully define. Their distributed architecture also imparts them with the ability to perform these complex tasks in the presence of additive noise or where the data presented is incomplete.

In conjunction with the selection of neural networks as a tool to perform the identification a simple discrete signal coding scheme, TES, was chosen as a means of presenting the condition data to the networks to enable the classification. TES itself has previously been applied to human speech both for the purposes of automated identification and for low data rate radio transmission. The extension of this technique to acoustic emissions from machines for the purposes of condition identification was a natural progression. In fact some early evaluations of this type had already been performed by Vu *et al* [38] with reasonable success.

In summary the general direction of the research from the outset was to evaluate TES as a means of condition data application and neural networks as a robust yet flexible means of condition identification. A PC environment was selected as the basis for the system implementation together with the use of digital signal processing technology for the more computationally intensive signal acquisition and conversion stages of the processing. These two elements not only fulfil the basic computational requirements to achieve reasonable response but also fit in well with the desire to develop cost effective means of condition monitoring.

7.2 System Performance Under Trial Conditions

The main focus of the thesis has been the evaluation of an amplitude TES conversion scheme whereas previously all the speech work had been limited to the application of a minima based technique. The early work of Vu *et al* also employed a minima conversion to present a neural classifier with acoustic condition cues for emissions from diesel engines. Amplitude TES, in contrast, is a novel approach which conveys information about energy and frequency variations in a source. In conjunction with this change of signal conversion strategy initial trials employed a more basic neural network presentation format to that employed by earlier researchers. Rather than post processing the TES stream acquired from the signal source to generate an A-matrix the raw statistical information about symbol generation over short periods is applied to the neural network. During practical trials each of these so called histogram matrices presents the symbol allocation statistics over a one second period for the source in a 300 element neural network data vector format.

In early trials on gearbox shaft velocity identification using the amplitude data format the neural classifier was, depending upon the adequacy of the training data applied, able to correctly identify between 82-100% of a pool of 120 test vectors. The quantity of data required by the training phase for this level of performance was relatively short at only 40-80 seconds. When the classification problem was made more difficult by introducing the lubrication status of the system the performance of the classifier worsened. Under these circumstances the classifier was only able to correctly identify the shaft velocity state in approximately 65% of test cases and lubricant status in 60-70% of cases. Considering the inherent uncertainty of the lubrication simulation data this represented a reasonable rate of success.

During trials in which the classifier was presented with acoustic tokens acquired from simulations of the more complex shaft displacement problem the training requirements for the amplitude based system had to be extended considerably to offset the increase in complexity of the boundaries in the corresponding data state space. During these trials the extension of the training phase to include TES matrix tokens amounting to 14 minutes of acoustic emissions was sufficient to achieve identification rates of between 76-82% from a test pool of 420 unseen data sets. By further enhancement of the data presented during the training phase this could be improved to approximately 93%. Network convergence in these trials required the presentation of data to the network for back-propagation learning between 500-2,300 times.

In back to back comparative trials the minima TES technique was applied to identical shaft displacement acoustic emissions and was able to achieve similar and in some instances slightly better separation of test data vectors than the amplitude format had. In some tests 98% of all matrix tokens presented to the neural network were correctly identified. However, the single overriding drawback encountered with this particular data presentation format was the erratic nature of the network training when presented with this classification problem. Whilst in some instances it was necessary to present the training data sets approximately 45,000 times in others 500,000 iterations were required. In many instances the training was aborted prior to network convergence due to the inability of the back-propagation algorithm to identify a network weight configuration which satisfied all the training requirements. Thus despite the apparent capability of the networks in certain circumstances it was apparent that this signal coding and presentation scheme was not suitable in the majority of trials involving shaft misalignment detection. The variation in acoustic emission properties caused by these faults could not be as suitably conveyed through the application of minima TES descriptors as they had already proved to have been with the amplitude scheme.

When the minima TES data was subjected to a further transformation using the A-matrix algorithm prior to network application both the learning ability and classification capability were improved. In trials with the same misalignment faults used to evaluate histogram data matrices results showed that between 90-99.5% of all test vectors could be correctly classified following a training program comprising 28 minutes of acoustic token data. Whilst ultimately the performance proved only marginally better in some cases the most obvious advantage of this presentation format was the improvement in the likelihood of success during the weight training phase. Not only did the networks converge more readily but the time taken to reach a stable weight state was reduced significantly. Where 45,000 iterations was considered reasonable earlier the introduction of additional acoustic shape cues, associated with the A-matrix format, reduced this to as few as 1,000 iterations in some instances. The additional overhead incurred in this instance as a result of the necessary A-matrix post-processing and the extension of the network input layer from 300 to 900 elements was more than offset by the improvement in system performance.

The classification enhancements identified as a result of employing the A-matrix symbol stream presentation format led to trials of the fourth network presentation format. This combined the novel amplitude TES coding scheme with the more advanced A-matrix data conversion algorithm. The expectation prior to practical trials was that improvements in system capability similar to those attained in the case of minima TES would be achieved. During subsequent trails, again using identical misalignment faults, the networks were able to identify between 82-100% of unseen acoustic matrix tokens depending upon the training data sets employed. The wide variation in perceived performance described here was due in the most part to the type of training data selection performed during the trail. With training sets of comparative size to those used in the minima trials correct identification of test vectors was more usually in the range 93-98%. The main drawback of this coding type is the necessity for a larger code table to implement the scheme which in turn produces larger matrices containing 1600 elements, considerably more than for any previous application scheme. However this is once again offset by further improvements in the number of data iterations required to achieve convergence. In trials between 40-400 iterations were usually sufficient to converge the network weights.

Because of these encouraging trials the amplitude A-matrix technique was singled out for further studies at this stage. Overlooking the network expansion aspect of the scheme this method had already been identified as being the most suitable technique for the presentation of acoustic cues associated with displacement misalignment of the gearbox shaft. Three further acoustic states corresponding to various stages of shaft angular misalignment and four states representing tooth wear and failure were employed during these later tests. In the case of angular misalignment trials 16 minutes of matrix data was sufficient to adequately train a classifier. During the training phases between 60 and 1,000 data iterations were required to achieve a satisfactory weight configuration and in subsequent trials 80-97% of test vectors were correctly classified. The tooth wear/failure states proved to be somewhat more problematic. Whilst training seemed to identify suitable weight configurations reasonably quickly, usually within 250-3,000 iterations, the subsequent test results were relatively poor. In some cases only just over half of the test vectors could be correctly classified.

Despite the knowledge that these simulated tooth fault states represented a non-trivial classification exercise it was felt that the performance should have been superior to that observed. Thus a series of additional trials were performed in an attempt to identify the cause of this reduced capability. They highlighted a single state-pair which proved most difficult to separate during the trials. Where other states could be identified with relative ease (94-96%) this particular pair, although not adjacent in terms of physical likeness, could only be identified correctly in approximately 53% of cases. Difficulties of this sort are caused by the complexity of the relationship between physical state and the acoustic emissions which are converted into TES format signatures. This is the result of the training phase not having identified the perimeters of the state boundaries with sufficient accuracy to be applied to the previously unseen test vectors. Identifying when such failures are caused by insufficient or unrepresentative training data exemplars and when they are caused by insufficient raw data is undoubtedly one of the keys to the successful application of neural techniques to such complex data types. Such separation is usually empirically derived using a process of trial and error.

7.2.1 Physical Restrictions of the Implementation Scheme

Whilst evaluation of the fundamental capabilities of TES based monitoring techniques has been the primary aim of this work the appraisal of the methods by which these techniques are applied is also important. Several of the trials therefore focused on the application process with the intention of identifying and evaluating particular elements within the procedures where restrictions affecting performance might be encountered. Four key elements where it was felt such effects may be experienced were studied.

The first two areas centred on the initial acquisition of the acoustic signals which form the basis of the monitoring mechanism. The first of these was the positional sensitivity of the acoustic data which impacts upon the requirements imposed upon implementers in terms of the rigidity of the rules surrounding data acquisition. Often in monitoring systems relying upon contact sensors the position of these sensors is a primary factor in the success or otherwise of the system. The intention of employing a microphone was to reduce the effects caused by complex contact transmission paths so making the application of monitoring a more straightforward and non-intrusive procedure. However, during trials the effects of varying transmission paths remained identifiable when the microphone position was varied by between 100-200mm. Despite this it was found that providing the data presented to the network classifier during training took account of this positional diversity the capability of the classifier on subsequent unseen data taken from a variety of positions was not significantly impaired. In practical terms this necessitates the acquisition and inclusion of acoustic emissions from a range of positions in the data used for the learning phase of the implementation.

The second aspect of the acoustic acquisition monitored during trials was the effect on performance of localised acoustic conditions. Although this aspect was not specifically singled out for analysis the effects were assessed indirectly by studying system performance using data recorded without the imposition of rigorous localised restrictions. By capturing the acoustic data in this manner it effectively guaranteed the inclusion of some additive noise. However one of the strengths of neural networks is their ability to perform input-output mappings in the presence of incomplete or noisy data. During the course of the practical trials the performances of networks remained acceptable despite not having imposed acoustic controls over data acquisition. It is therefore reasonable to suppose that providing the levels of external localised noise do not become excessive there is no necessity to provide an additional filtering stage within the TES coder.

The remaining two areas in which it was felt application restrictions could arise concerned the handling of the acoustic signal following initial acquisition. At this stage the signal is converted into a representative TES stream and is used to produce condition signature matrix data. Following compilation this data is used to train a neural classifier. Intuitively the characteristics of these data sets will affect the "knowledge" gained by the network about the problem being "taught". Thus the characteristics of the converted signal and the subsequent matrix data will affect the performance of the classifier when presented with unseen data. During the course of the trials two aspects of this process were singled out for examination in order to define their impact upon the TES data and thereby the network performance.

The first of these was the stability of rotational velocity necessary to reliably perform condition identification of gearbox fault conditions independent of the instantaneous operational state of the machine. In terms of implementation what was seen as undesirable was the necessity either to maintain a steady velocity whilst monitoring or to train the monitoring system to identify the fault set under many different machine operating configurations. The work carried out on this aspect of the conditioning procedure was fairly limited. As with the case of microphone positional sensitivity once again the philosophy was one of training using loosely constrained data sets thereby reducing the sensitivity of the network to parameters which are unrelated to the fault state. In trials using the gearbox shaft velocity was constrained within a 200rpm band during acoustic fault state data acquisition. This constraint, which represented approximately 7% of the average velocity, did not significantly affect the system performance under test conditions using data recorded at shaft operating velocities within this band.

The second aspect of the signal processing identified as a potential problem area for practical applications was associated with the dynamic properties of the raw acoustic signal. These signal properties are particularly important in amplitude TES coding which relies upon a normalisation coefficient for symbol allocation. As a result variations in the signal periodicity must be adequately catered for by the pre-coding normalisation coefficient search phase if the subsequent TES symbol stream is to provide a reasonable signal representation for fault characterisation. Trials did reveal some network sensitivity associated with this aspect of the coding process. However these effects could be minimised by careful control of the input conditioning and conversion stage of the TES symbol generation.

The conclusion from these experimentative trials was that providing the neural classifier is trained with these effects in mind and is able to extract the relevant condition cues from within the data then the classifier will be relatively unaffected by acceptable variations in the prevailing acoustic environment. The tradeoff however is that the more loosely constrained the conditions surrounding the acquisition of acoustic data are the larger the data set required to adequately describe the problem to the network.

7.2.2 Performance Optimisation

Following the initial examinations with each of the four signal coding and presentation techniques investigations were carried out on methods of further enhancing their basic capabilities. There are a number of elements of the classification mechanism which can potentially be modified to improve the performance but the areas which were focused upon during the course of this work were limited almost entirely to the neural network application phase. Within this context areas such as network configuration and architecture, classification requirements, data set size and composition, and data set ordering all affect the overall system performance against previously unseen data.

The backpropagation learning concept used in the network training phase is, by definition, a weight search algorithm which seeks to reduce the error between an input vector and a desired output pattern by backpropagating the error from output to input nodes. This is accomplished by successively calculating the contribution to the measured group error made by individual nodes within the network when an input-output vector pair are applied. The weight configuration path taken over the network's "error surface" by applying the gradient descent search algorithm is therefore affected not only by the training vectors applied at each backpropagation iteration but also by the network architecture and the chosen starting point on the error surface.

The ordering of the data was found during the trials to affect both the length of time which the network required to reach convergence and the final weight configuration attained. Whilst the actual configuration of the weights is immaterial, providing that classification is not affected, the time required to achieve convergence is important. The observed lengthening in training times can be caused by the varying demands made upon the network of individual training vectors. Whilst some vectors may drive the weights in one direction others may have differing requirements. Thus the manner in which these various vectors are applied during training affects the path taken across the error surface and thus the weight path taken by individual nodal interconnections. During the course of practical trials it was discovered that the best method of applying vectors was to not separate the individual classes too much thereby making the weight path more erratic nor by grouping the states too much causing convergence oscillation. Best results were achieved by applying the various state vectors in evenly sized and orderly subsets. In reality this technique is only able to increase the likelihood of training success because the contours of the error surface for a given problem are unknown at the outset of training. Thus the effect of each unique training vector on the direction taken along this error surface is difficult to determine in advance.

Another area which was examined for performance optimisation was the content of the data set applied during training. In many instances the performance of individual networks could be measurably improved in one of two ways. Empirical data selection generally provided the best means of generating data sets for optimised network performance and usually resulted in less training vector exemplars being required to achieve a particular performance level. However the disadvantage with this method is its implicit reliance upon an intelligent means of comparing data sets during an evaluation phase. Whether this is performed by an operator or by automated combinatory trials of data components it requires an extra level of complexity to achieve success. In contrast it was noted during trails that in the most part the extension of the data set would produce a similar improvement in performance without the necessity for the additional trial period. This mechanism though will still extend the training period simply because the quantity of training data which must be applied to the network has been increased.

From the point of view of the potential for industrial monitoring systems the effect upon a classifiers capability as the state space is extended is important. During the course of the studies relating to the application of amplitude TES A-matrix data matrices the effect of a moderate increase from four to seven states was examined. Whilst this single study is insufficient to provide conclusive evidence of the general effects of such extensions it does at least provide an indication of the type of effects which are present in this application. The results indicated that whilst the general performance was slightly reduced the consistency of classification over a range of training regimes was improved. Correct separation of the seven condition states was achieved in approximately 90% of all unseen test matrices.

The scope for architectural performance optimisation was examined in a series of trials which experimented with the interconnectivity of the network classifiers. These modified networks were evaluated against both the seven state problem which has just been discussed and with the tooth fault analysis problem discussed in Chapter 6. The intention of both studies was to identify whether or not a reduction in the competing demands of individual nodal weights could be used to enhance both the training requirements and the separation of states in problems composed of more complex state boundaries. Although the trials were not sufficiently conclusive it was apparent that in those problems it was evaluated upon no significant improvements could be identified. In the seven state problem the performance was in fact reduced whereas in the tooth fault problem some improvement was noted. In both these problems the partially interconnected networks required longer to train than their fully interconnected counterparts.

7.2.3 TES System Implementation

In many respects the application of TES data, particularly the novel amplitude coding technique, and neural networks to the classification of acoustic signals has proved to be successful. Although the research is still in its infancy there is reason to believe that these techniques represent a realistic alternative to the more conventional monitoring systems currently available. They offer the potential not only for a trained rather than a programmed solution but also one which is robust enough to operate acceptably within an industrial environment. The implementation of a significant part of this evaluation system within the PC environment also highlights the potential for low cost monitoring systems which may be tailored for a wide range of applications. Those elements of the system which were not developed under the umbrella of the PC environment, namely the neural network simulator, fall well within the capabilities of modern PC technology and could feasibly be incorporated at a later date into a combined PC based package.

DSP implementation of the combined signal capture and TES symbol generation algorithms enables significant improvements to be made in the generation of raw condition signatures over a similar PC solution. If the system were to be developed still further by implementing the neural classifier on a second DSP hosted by the PC the potential exists for a monitoring system which is able to respond in real-time to changes in the condition of a monitored system.

The most problematic aspect of implementing the type of condition monitoring system outlined in this study has been the difficulty encountered in evaluating the effects upon classification of alterations made in the training data sets. This is a direct result of the complexity with which the network encodes the learning process.

7.3 Conclusions

The author feels that the following conclusions can been drawn from the studies performed during the research period:-

- TES coding of the acoustic emissions provides a good means of data compression without reducing the information content significantly.

- The amplitude A-matrix data presentation format offers the best compromise between the necessary training requirements and the subsequent classification performance.

- The physical restrictions of the data acquisition phase can be overcome by additional training data set presentation.

- Training data is best applied to a selected network in evenly sized ordered subsets to minimise the path length taken over the network error surface during the training phase.

- Empirical data set selection can improve system performance whilst limiting the size of the training data set. However, data set extension can be used to minimise the requirements for operator intervention during the training phase.

- The application of a nodal interconnection pruning phase adds additional complexity without providing significant classification performance enhancements.

- A single network is able to classify a seven fault state system with reasonable accuracy.

- The implementations studied offer marked improvements over the capabilities afforded by a human operator at low cost and without the necessity for constant operator supervision.

7.4 Recommendations for Future Work

Despite the basic potential surrounding TES based neural classifiers there are still a number of areas which could benefit from further evaluation. This section identifies areas which the author feels were not suitably examined or hold the most potential for further improvement.

The classifier implementation described in this thesis has remained focused upon a "one application, one classifier" solution to the identification of gearbox condition state. Whilst the concept of expanding the number of states which may be handled by a single classification network has been touched upon it is felt there is much room for further work to be carried out in this area. It is highly unlikely that most "real world" applications will require only seven unique fault states to be identified. As such both the examination of the limits of data space expansion within a single network and the concept of a multi-network monitoring system should prove to be worthwhile.

Observation of the effect additive noise has upon the classification process has up to now been limited. Whilst the studies carried out during the course of this work made no attempt to eliminate potential sources of additive noise during the recording of acoustic state emissions no specific trials were performed on the degenerative effects such noise could introduce. The application of neural networks will no doubt make the classification process more robust but without practical trials the effects are difficult to quantify accurately. This particular study would help to define the environmental restrictions which would be required of an industrial application of such a monitoring system.

The neural network applied to this study is effectively only capable of making time independent decisions. In this respect it is unable to use knowledge about previous classification decisions to modify its current decision. This simplistic behaviour eliminates the advantages to be accrued by evaluating the decisions made by the classifier over a short but finite time period. In contrast a human operator can use additional knowledge about the problem under observation to come to a decision on a machines current status. One area in which the author feels significant improvements could be made is in post decision analysis of the network classifier output. In particular knowledge about the manner in which physical faults occur could be used to extend the capabilities of the classification system described in this thesis. For example by applying the basic rule that faults are not self correcting to the stream of time independent fault status decisions a more representative indication of the status might be attained. For instance if a classifier indicated that at time, T-n (where n represents the number of decisions made by the classifier) the machine status was healthy and at time T+n the machine state was healthy, but that at time T the machine status was unhealthy, then a post decision analyser could calculate the probability of the fault actually being present based upon knowledge about the indicated fault.

Whilst the trials evaluated the characterisation techniques on a series of basic fault states simulated on a simplistic gearbox testbed the ability of the acoustic TES system to identify the complex fault states associated with a real system are as yet not known. There are two areas which would benefit from further studies using more representative acoustic emissions. Firstly the evaluation of the technique using a more realistic testbed model would provide additional feedback regarding the potential of the system for industrial applications. The most important aspect of such a study should be an evaluation of the sensitivity required for acceptable fault state analysis. The second area which would benefit from further analysis is the ease with which on-board monitoring systems may be applied to mechanical devices. Of particular interest in this respect should be the evaluation of the likely success of a model based rather than a unit based training regime for high volume production type equipment. The necessity, in such applications, for a per unit training regime would significantly limit the areas in which this type of monitoring could be applied. In contrast if the training can be achieved off-line before subsequently being applied to many identical units fewer limitations are present.

References

1. R. A. King, and W. Gosling, "Time encoded speech", Electronics letters, 14, No. 15, July 1978, pp. 456-457

2. H. J. Rose, "Vibration signature and fatigue crack growth analysis of a gear tooth pending fatigue failure", 44th Meeting of the mechanical failure prevention group, April 1990, pp. 235-245

3. A. F. Stronach, C. J. Cudworth, and A. B. Johnston, "Condition monitoring of rolling element bearings", International Conference on Condition Monitoring, University College of Swansea, 1984, pp. 162-177

4. W. Dong, Y. H. Joe, and A. Mardapittas, "Monitoring chip congestion in drilling by acoustic emission", International Conference on Condition Monitoring and Diagnostic Engineering Management,1991, pp. 119-126

5. C. Cempel, *Vibroacoustic condition monitoring*, Ellis Horwood Series in mechanical engineering, 1991

6. L. G. Martins, and S. N. Y. Gerges, "Comparison between signal analyses for detecting incipient bearing damage", International Conference on Condition Monitoring, University College of Swansea, 1984, pp. 191-204

7. T. Koizumi, M. Kiso, and R. Taniguchi, "Preventative maintenance for roller and journal bearings of induction motor based on the diagnostic signature analysis", Journal of vibration, acoustics, stress, and reliability in design, Transactions of ASME, Vol. 108, January 1986, pp. 26-31

8. D. Dyer, and R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis", Journal of mechanical design transactions of the ASME, paper No. 77-DET-83

9. L. M. Rogers, "The application of vibration analysis and AE sourced location to on line condition monitoring of antifriction bearings", Tribology International, Vol. 12, No. 2, pp. 51-59

10. P. D. McFadden, "Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration", Journal of vibration, acoustics, stress, and reliability in design, Transactions of ASME, Vol. 108, April 1986, pp. 165-170

11. P. Gadd, and P. J. Mitchell, "Condition monitoring of helicopter gearboxes using automatic vibration analysis techniques", AGARD Proceedings, No. 369, 1984, pp. 29.1-29.10

12. R. Wojciechowski, "Utilisation diagnostics of rolling element bearings in electric motors", Masters thesis IMT, Poznan University

13. D. G. Astridge, "HUM- Health and usage monitoring of helicopter mechanical systems", British Journal of Non-destructive Testing, 1988, Vol. 30, No. 6, pp. 403-408

14. J. T. Broch, *Mechanical vibrations and shock measurements*, Bruel & Kjaer, 1984

15. Hewlett Packard, "Effective machinery maintenance using vibration analysis", Application note 243-1, Dynamic signal analyser applications, 1983

16. J. E. Berry, "How to specify machinery vibration spectral alarm bands", Proceedings of the EPRI V predictive maintenance conference, Knoxville, Tennessee, September 1992

17. M. Angelo, *Bruel & Kjaer technical review: Vibration monitoring of machines*, No. 1, Naerum Offset, Denmark, 1987

18. S. Biswas, T. Chander, and S.K. Bhave, "An integrated approach to diagnose gear coupling failure through condition monitoring", International Conference on Condition Monitoring, University College of Swansea, 1984, pp. 315-327

19. M. A-S. Mohamed, G. Ahmadi, and F. T. C. Loo, "A study on the mechanics of fatigue-dominated friction noise", Journal of vibration, acoustics, stress, and reliability in design, Transactions of ASME, Vol. 112, April 1990, pp. 222-229

20. R. F. Beatty, "Differentiating rotor response due to radial rubbing", Journal of vibration, acoustics, stress, and reliability in design, Transactions of ASME, Vol. 107, April 1985, pp. 151-160

21. L. Zhaoqian, I. A. Craighead, S. B. Bell, "Monitoring tool wear using vibration signals", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 114-118

22. R. B. Randall, "Computer aided vibration spectrum trend analysis for condition monitoring", Maintenance Management International, 1982, Vol. 3, No. 3, pp. 183-208

23. I. Imam, S. H. Azzaro, R. J. Bankert, and J. Scheibel, "Development of an on-line rotor crack detection and monitoring system", Journal of vibration, acoustics,

stress, and reliability in design, Transactions of ASME, Vol. 111, July 1989, pp. 241-250

24. R. G. Herbert, "Computer techniques applied to the routine analysis of rundown vibration data for condition monitoring of turbine alternators", British Journal of Non-destructive Testing, 1986, Vol. 28, No. 6, pp. 371-375

25. M. Serridge, "Cyclic analysis: An on-line automated technique for the vibration monitoring of reciprocating machines", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 459-468

26. R. Chaudhuri, "Development and application of machine health monitoring techniques for reciprocating machines", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 474-480

27. A. F. Stronach, A. B. Johnston, and C. J. Cudworth, "Detection of valve faults on reciprocating compressors", Maintenance Management International, 1985, Vol. 5, No. 4, pp. 225-232

28. R. Firoozian, J. Wu, "Model based condition monitoring of industrial machines", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 370-376

29. W. Bartelmus, "Application of some statistical estimators of the vibration signal as criteria of assessment of mesh state", International Conference on Condition Monitoring, pp. 256-265

30. P. D. McFadden, and J. D. Smith, "Information from the vibration of rolling bearings", International Conference on Condition Monitoring, University College of Swansea, 1984, pp. 178-190

31. J. C. R. Licklider, and I. Pollack, "Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech", Journal of the acoustical society of America, 20, No. 1, January 1948, pp. 42-51

32. A. G. Warner, R. D. Hughes, and R. A. King, "A direct voice input man-machine interface strategy to provide voice access for severely impaired speakers", IEE Conference publication No. 316, 1990, pp. 279-285

33. T. C. Phipps, and E. D. Chesmore, "Silence signalling in a time encoded speech (TES) system", IEE 3rd Communications symposium, Bangor, May 1991

34. F. E. Bond, and C. R. Cahn, "On sampling the zeros of bandwidth limited signals", IRE Transactions, IT6, 1960, p. 51

35. H. B. Voelcker, "Toward a unified theory of modulation. Parts I and II", Proceedings of the IEE, Vol. 54, 1966, pp. 340-353, 735-755

36. R. C. Power, R. D. Hughes, and R. A. King, "Verification, archetype updating, and automatic token set selection as a means of improving the performance of menu driven isolated word recognition systems using time encoded speech descriptions in high acoustic noise backgrounds", IEE Conference on speech input/output; Techniques and applications, pp. 144-151

37. R. C. Power, R. D. Hughes, and R. A. King, "Performance and evaluation criteria for simple TES isolated word recognition (IWR) systems in hostile tactical military acoustic environments", European Conference on speech technology, September 1987, pp. 249-252

38. V. V. Vu, P. J. Moss, A. N. Edmonds, and R. A. King, "Time encoded matrices as input data to artificial neural networks for condition monitoring applications", International Conference on Condition Monitoring, 1991, pp. 31-35

39. F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organisation in the brain", Psychological review, 65, 1958, pp. 386-408

40. F. Rosenblatt, "Two theorems of statistical separability in the perceptron", Mechanisation of thought processes: Proceedings of a symposium held at the national physical laboratory, London, November 1958, pp. 421-456

41. F. Rosenblatt, *Principles of neurodynamics*, New York, 1962, Spartan

42. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation", *Parallel distributed processing: Explorations in* the *microstructure of cognition*, Cambridge, MIT Press, 1986, pp. 318-362

43. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *Neurocomputing: Foundations of research*, Cambridge, MIT Press, 1988

44. J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the theory of neural computation*, Santa Fe Institute, Addison Wesley, 1993

45. P. Gadd, "The capability of built in equipment for VA (Vibration Analysis) monitoring of military helicopters"

46. T. G. Jellison, N. S. Pratt, and R. L. De Hoff, "XMAN a tool for automated jet engine diagnostics", Proceedings of the International Conference on Condition Monitoring, Swansea, 1987, pp. 60-72

47. P. J. Skitt, and R. C. Witcomb, "The analysis of the acoustic emission of jet engines using artificial neural networks", Condition Monitoring and Diagnostic Technology, Vol. 1, No. 1, June 1990, pp. 7-12

48. R. R. Leighton, *The Aspirin/MIGRAINES neural network software users manual V6.0*, October, 1992

49. N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, "Optical implementation of the hopfield model", *Applied optics*, 24, 1985, pp. 1469-1475

50. M. A. Sivilatti, M. A. Mahowald, and C. A. Mead, "Real-time visual computations using analogue CMOS processing arrays", *Advanced research in VLSI: Proceedings of the 1987 Stanford Conference*, Cambridge MA, MIT press, 1987, pp. 295-312

51. G. Cybenko, "Continuous valued neural networks with two hidden layers are sufficient", Technical report, Department of computer science, Tufts University, Medford MA, 1988

52. G. Cybenko, "Approximation by superpositions of a sigmoidal function", Mathematics of control, *Signals and systems 2*, 1989, pp. 303-314

53. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks 2*, 1989, pp. 359-366

54. L. Fausett, *Fundamentals of neural networks, architectures, algorithms and applications*, Prentice Hall International, 1994

55. M. Gutierrez, J. Wanng, and R. O. Grondin, "Estimating hidden units for two layer perceptrons", First IEE International Conference on Artificial Neural Networks, No. 313, 1989, pp. 120-124

56. B. Widrow, "ADALINE and MADALINE - 1963", Proceeding of the IEEE 1st International conference on neural networks, Vol. 1, 1987, pp. 143-157

57. E. B. Baum, and D. Haussler, "What size net gives valid generalisation?", Neural computation, Vol. 1, No. 1, 1989, pp. 151-160

58. S. J. Marshall, and R. F. Harrison, "Optimisation and training of feed-forward neural networks by genetic algorithms", IEE Conference Publication No. 349, June 1991, pp. 39-43

59. H. Lari-Najafi, M. Nasiruddin, and T. Samad, "Effect of initial weights on back-propagation and its variations", Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, Vol. 1, 1989, pp. 218-219

60. S. Roy, "Factors influencing the choice of learning rate for a back-propagation neural network", Proceedings of the IEEE International Conference on Neural Networks, Vol. 1, No. 1, June 1994, pp. 503-507

61. W. El Deredy, and N. M. Branston, "An update function that speeds up back-propagation learning", Proceedings of the IEEE International Conference on Neural Networks, Vol. 1, No. 1, June 1994, pp. 477-482

62. D. Alpsan, M. Towsey, O. Ozdamar, T. Tsoi, and D. N. Ghista, "Are modified back-propagation algorithms worth the effort", Proceedings of the IEEE International Conference on Neural Networks, Vol. 1, No. 1, June 1994, pp. 567-571

63. K. M. Raguliskis, *Vibrations of turbo-systems*, MOKSLFS, Vilna, Chapter 3, 1976

64. V. V. Vu, and R. A. King, "Automatic diagnostic and assessment procedures for the comparison and optimisation of time encoded speech (TES) direct voice input (DVI) systems", European Conference on speech technology, Paris, September 1989, pp. 412-416

65. R. D. Hughes, and R. A. King, "A comparison of the performance of normal and whispered speech with simple time encoded speech (TES) direct voice input (DVI) systems in a tactical military environment", European Conference on speech technology, Paris, September 1989, pp. 417-421

66. T. C. Phipps, and E. D. Chesmore, "Low rate speech encoding: New algorithms and results", International symposium on communications theory and applications, Crieff hydro hotel, Sept 1991, Paper No. 14

67. A. Sturm, H. Bittner, R. Wirth, "New aspects of digital signal processing in technical diagnosis", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 346-353

68. A. J. Penter, "A diagnostic monitoring system for machine operators", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 354-359

69. H. Zhengjia, Q. Liangsheng, Z. Hanfu, and R. B. K. N. Rao, "A valid approach to cope with uncertainty in condition monitoring and diagnostic management of machinery", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 481-484

70. M. H. C. Thomas, "Efficient interpretation of fault indicators by data fusion including machine learning", International Conference on Condition Monitoring and Diagnostic Engineering Management, 1991, pp. 80-84

71. J. Butterworth, "Vibration condition monitoring as part of a predictive maintenance programme"

72. M. F. White, M. Rasmussen, "Vibration diagnosis data bank for machinery maintenance", International Conference on Condition Monitoring, University College of Swansea, 1984, pp. 98-112

73. E. O. Sohoel, "Shock pulses as a measure of the lubricant film thickness in rolling element bearings", International Conference on Condition Monitoring, University College of Swansea, 1984, pp. 148-161

74. B. T. Kuhnell, and J. S. Stecki, "Correlation of vibration, wear debris analysis in rolling element bearing condition monitoring", Maintenance Management International, 1985, Vol. 5, No. 2, pp. 105-115

75. Z. Savic, and M. Ognjanovic, "Vibration and noise due to gear meshing", International Conference on Condition Monitoring, 1985, pp. 243-255

76. M. P. Boyce, "On-line condition monitoring of turbomachinery", Proceedings of the International Conference on Condition Monitoring, Swansea, 1987, pp. 90-111

77. K. Pipe, "Application of advanced pattern recognition techniques inn machinery failure prognosis for turbomachinery", Proceedings of the International Conference on Condition Monitoring, Swansea, 1987, pp. 73-89

78. R. M. Stewart, "Machinery management based on the application of modern diagnostics, control theory and artificial intelligence", Proceedings of the International Conference on Condition Monitoring, Swansea, 1984, pp. 32-54

79. I. C. Cheeseman, "Advanced diagnostics for helicopter machinery management", pp. 163-188

80. M. Dexing, "Fast algorithms for pitch and energy determination", pp. 316-318

81. M. T. Whitaker, and J. A. S. Angus, "A low cost continuous speech recogniser", pp. 119-123

82. A. P. Seneviratne, "Comparative study of TES waveshape approximation techniques", IREECON International, 1985, Vol. 2, pp. 1029-1032

83. A. P. Seneviratne, and J. Martin, "Time encoded speech; An assesment", Colloquium on the digital processing of speech , London. 1983

84. J. Holbeche, R. D. Hughes, and R. A. King, "Time encoded speech (TES) descriptors as a symbol feature set for voice recognition systems", IEE Conference Publication No. 258, 1986, pp. 310-315

85. A. V. Oppenheim, A. S. Willsky, I. T. Young, *Signals and systems*, Prentice-Hall International Editions, 1983, pp. 527

86. D. R. Hush, and B. G. Horne, "Progress in supervised neural networks, whats new since Lippmann", IEEE Signal Processing Magazine, January 1993, pp. 8-39

87. D. Hammerstrom, "Neural networks at work", IEEE Spectrum, June 1993, pp. 26-32

88. D. Hammerstrom, "Working with neural networks", IEEE Spectrum, July 1993, pp. 46-53

89. S. Y. Kung, *Digital neural networks*, Prentice Hall International, 1993

90. Yoh-Han Pao, *Adaptive pattern recognition and neural networks*, Addison-Wesley Publishing, 1989

91. J. C. Bezdek, and S. K. Pal, *Fuzzy Models for Pattern Recognition*, IEEE Press, 1992

92. P. Lynwander, *Gear drive systems - design and application*, Mechanical engineering series, Pub. Marcel Dekker, New York, 1983

Appendices

A. Practical TES Evaluation System

A.1 Acquisition of the Condition Signal

The first stage in the condition monitoring process is to capture the signal from the library of acoustic emissions whilst at the same time preventing it from being corrupted by the effects of aliasing. To guard against this the source is band limited using a filter prior to being digitally sampled using a digital signal processor. The characteristics of the bandpass filter used to pre-condition the signal are dependant upon the dynamic range of emissions required for classification and upon the sampling rate used for conversion. Basic sampling theory [58] states that for the converted source to be an accurate digital representation of the analogue original, the sampling rate must meet or exceed the Nyquist frequency. For the purposes of the trials detailed in this work a dynamic range of approximately 15 kHz was used by applying the necessary pre-filter, in this case a 5th order Chebyshev type, to the input stage. The characteristics of this filter are plotted illustrated graphically in Fig. A.1.

Once filtered the signal is sampled at 40kHz. Oversampling at this rate provides a margin of safety in the signal acquisition. For a commercial system implementation there may be some scope for further reduction in the frequency range employed during this initial signal acquisition. However any such reduction would be dependant upon the machinery under observation and the associated condition states which require identification. Unless these bandwidth restrictions are applied with a full understanding of the specific fault modes further sensitivity problems may be unintentionally introduced. To minimise complexity at this stage in the evaluation process no such additional bandwidth limitations were imposed. Instead the greater flexibility provided

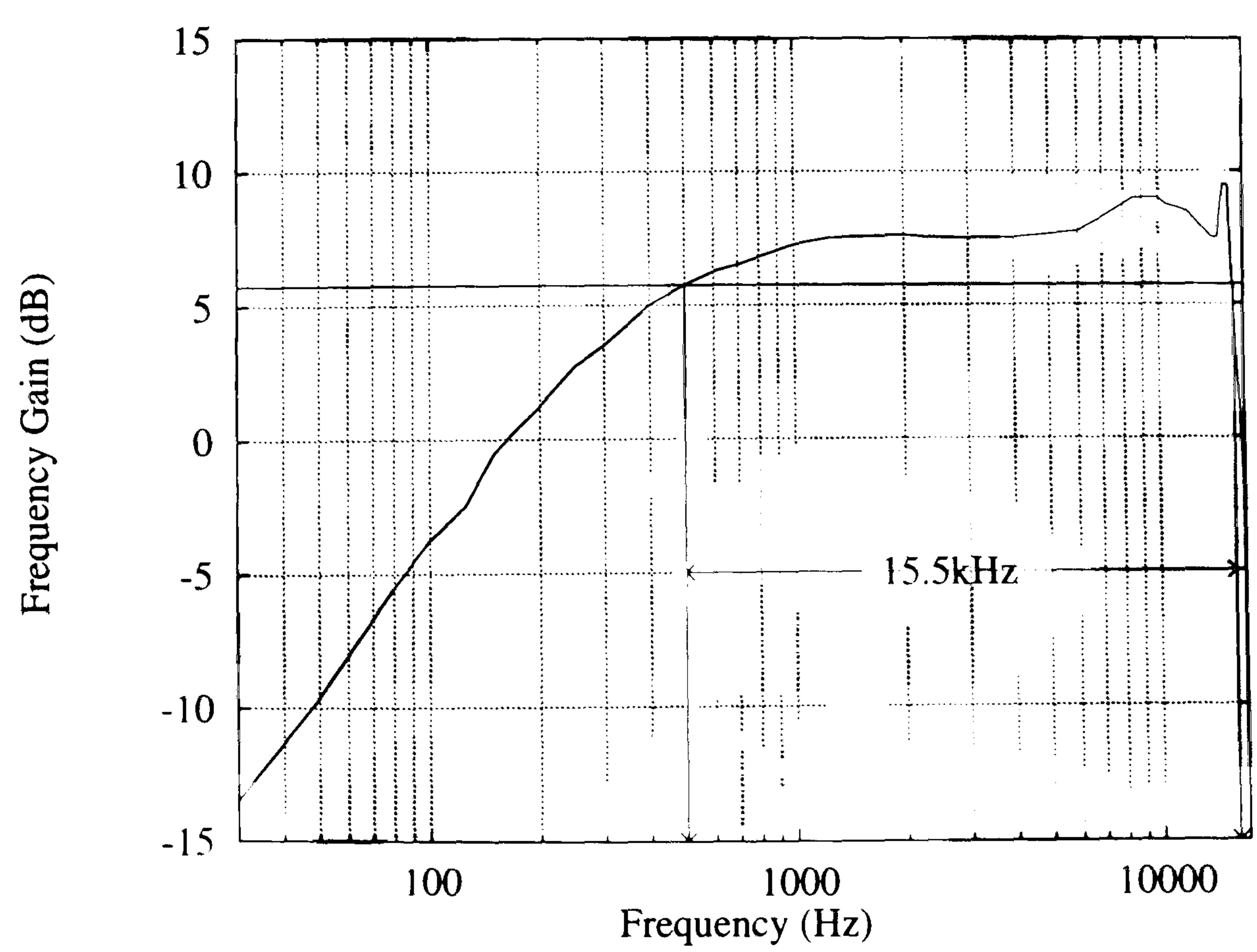

Figure A-1 Characteristics of the bandpass filter used to condition the acoustic signal prior to sampling and TES conversion by a less restricted frequency range was employed at the expense of an increased data bandwidth.

A.1.1 Real-time Conversion of Acoustic Emissions into a TES Format

Once an acoustic signal has been acquired it must be converted into a series of TES symbols before applying one of the matrix data encoding techniques discussed previously in Chapter 3. To perform this task in real-time an acquisition and conversion system was developed around a PC platform. The core of the system is based on a dedicated Digital Signal Processor (DSP) which resides in the PC. The AT&T DSP32C is, as a result of its 80ns instruction cycle time and pipelined instruction capability, able to perform the signal capture and TES conversion algorithms in real-time. This provides a significant improvement over comparative PC conversion performance. The PC itself is used to control this capture and conversion process and store the subsequent data for further processing.

The TES data is exchanged between PC and DSP using a protected, shared memory area. This memory area is divided into two blocks. Each of the blocks has a semaphore flag to indicate the state of the data contained within it which both the PC and DSP processors can read and write to. During operation the DSP converts the condition signal and writes TES symbol data to the first "empty" block and uses the semaphore to indicate that the block is currently active. As this block becomes full the DSP sets the semaphore associated with it to "full", and begins using the second memory block. Meanwhile the PC monitors the status of each of the block semaphores. When the PC sees a "full" flag it downloads the associated TES data in the block and resets the semaphore to indicate an "empty" status. Providing the data blocks are sufficiently long for the PC to complete a "full" block transfer to a ramdisk during the time taken for the DSP to acquire a block of TES data then the DSP will always have an "empty" block available to write data to. This data transfer procedure is known as a two block shuffle.

A.1.2 Associated Signal Effects

There is one stage in the conversion process yet to be discussed which can adversely affect the characteristics of the condition signal. If left uncontrolled it can cause distortions in the TES symbols assigned to the raw signal epochs. This refers specifically to signal level variations caused as a result of microphone position and amplifier gain characteristics during signal recording or playback. Fluctuations of this type at the input stage of the DSP are of particular concern in an amplitude coding scheme. To minimise these effects normalisation is applied to the signal prior to coding. This reduces the potential for inter-recording perturbations and essentially reduces the analysis to one of statistical distributions of signal energies rather than specific magnitudes. This is perfectly adequate since for a given signal type the energy distributions remain relatively stable despite any fluctuations which may occur in absolute signal levels as a result of the acoustic sensor stage. However the variations in statistical amplitude distributions between signals of different types should remain. In the discussions of the work of Martins [16] in Chapter 2 these variations in statistical distribution have already been used to monitor the development of defects. The shift of emphasis from signal magnitude to statistical distribution is also of particular interest since it provides a method of monitoring a signal source without the necessity for strict acquisition controls being imposed.

The practical application of normalisation to a signal requires a pre-determined normalisation coefficient to be applied to each discreet sample in turn. This coefficient should accurately represent the maximum magnitude of the signal over a finite coding period. In order to acquire this normalisation factor the source is sampled over a finite measurement period prior to TES coding. During this precoding measurement interval the DSP device acquires, compares and then discards successive samples in its search for the magnitude of the absolute maxima. At the end of the precoding period a normalisation coefficient is calculated and stored. This coefficient is subsequently used during coding to condition each sample prior to analysing the epoch characteristics and generating a symbol stream. Providing the initial measurement period is of sufficient duration to represent the short term average signal level during a coding run then the normalisation coefficient will accurately condition the signal. For the purposes of the work contained in this thesis a precoding period of 10 seconds was considered adequate to acquire a normalisation coefficient which could be used for single conversion runs of 60-90 seconds.

A.2 A TES Analysis Package

Having discussed the basic configuration of the PC based processing package the manner in which it may subsequently be operated should be considered. As discussed in Chapter 3 the conversion necessitates the generation of TES symbol allocation tables which themselves require statistical analysis tools to make the relevant symbol selections. To cater for this, a combined PC/DSP package was developed to perform in two operational modes. Initially it is used in analysis mode to generate the relevant symbol allocation tables and subsequently it is used in an operational conversion mode.

The analysis mode is itself separated into two phases. In the first phase the DSP acquires the signal, and measures the relevant epoch parameters such as duration, amplitude and minima frequency. This epoch data is then presented for statistical evaluation at the PC using the two block shuffle for data transfer. A graphical user interface was developed for the PC to enable the results of this evaluation to be displayed. From this graphical feedback the user is able to generate the initial, or first-stage, allocation table. The second phase of the allocation table generation is the statistical table optimisation detailed previously in Chapter 3. This requires the first-stage table to be downloaded to the DSP device and used to perform a full TES conversion of a sample of source. Once again statistical analysis is then performed at the PC on the TES symbol data generated. At the end of this cycle, an optimised, or second-stage, symbol allocation table is generated.

This manual modular approach, although slower than an automated technique, provides complete freedom in terms of symbol selection and allocation table generation which for the development system was essential. All of the off-line analysis and data transfer software for the PC was written as standard ANSI 'C' routines. Whilst 'C' does not afford ultimate algorithm performance it does provide a simple and easy method of module reconfiguration. In a commercial system implementation much of this data analysis could be performed automatically on the DSP device which would significantly improve the response.

A.3 Conversion of the TES Symbol Stream into a Condition Matrix

The system which has been described up to now is able to capture the signal, generate statistical data relating to it and with manual intervention produce an allocation table optimised for a particular coding strategy. This optimised allocation table is then downloaded from the PC to the DSP for use in an on-line real-time conversion without further PC intervention. The PC then becomes a receptacle for symbol stream data acquired from the stable taped source library. The final PC based step in the processing chain is the conversion of a symbol stream into a matrix format suitable for use in a condition identification mechanism. This requires the subdivision of the symbol stream into a series of sub-frames. Each of these frames is then used to produce a unique condition matrix containing acoustic information corresponding to that segment of the original signal recovered from the tape. The selection of the length of each frame in symbols, previously detailed in Chapter 3, is important to the subsequent analysis and recognition stages. The number of symbols contained in each frame is primarily dependant upon the aspect of the signal under observation which is of interest.

Taking the example of a gear wheel rotating at a constant 60 rpm with a single damaged tooth the effects of varying the frame length can be described. Each time the faulty tooth in the gear meshes with a secondary gear it will emit a characteristic peak of energy emission due to a rapid acceleration-deceleration action. This characteristic is expected to appear in the emitted signal once every second. Suppose then a frame length significantly shorter than 1s, for arguments sake assume this to be 0.1s is selected. In this case the characteristic energy burst due to the fault will be captured only every tenth frame. If this is the case, one of two scenarios is possible. The first is that the fault may be identified as being intermittent because of its infrequent occurrence. The second is that the fault may be missed altogether because the frame containing the characteristic to classify the machine state by has not been selected.

The solution is either to ensure that the frame length is sufficiently long to incorporate a single machine cycle or secondly to ensure that each state decision requires several frames which together make up a single machine cycle. For the practical trials discussed in this thesis the frame length was selected so as to ensure that each matrix contains a symbol set made up of at least one cycle of the lowest frequency machine event. To accommodate this requirement a frame length of 5000 symbols was employed which corresponded to approximately one second of acoustic source signature.

A.4 Configurable Neural Classification Module

The culmination of this combination of PC and DSP hosted processing is the series of condition matrices which contain the information used to characterise the fault status of the target system. These matrices however do not in themselves constitute a condition detector. What is required to perform the condition detection is a pattern classifier module. For the reasons discussed in section 3.5.3 this classifier was implemented using a simple multi-layer perceptron network. As with all the other PC and DSP based tools discussed so far this module required sufficient flexibility to be incorporated as to enable modifications to be introduced to the architecture and presentation format. Rather than develop a specific package for this purpose it was decided to make use of a UNIX hosted package, aspirin/MIGRAINES, which was readily available. This package contains a script language which is able to generate network simulations containing all the relevant flexibility in architecture to simulate large non-trivial networks without the need for a custom design. In a more commercialised system implementation this final neural based signal classification would be performed on the DSP, on the PC host, or using dedicated neural network hardware.

However for the evaluation work detailed in this thesis a SUN Sparc 10 workstation was used to run the network simulations. This platform provides significant improvements in processing power over the available PC hardware. This ensured that the learning and evaluation periods required for a simple back propagation neural network were reduced. On the matrix based condition patterns employed this varied from between a few seconds to a few minutes for each matrix during the course of the more computationally intensive training phase. The aspirin/MIGRAINES network simulations were able to import the matrix data generated at the PC in space delimited format. Each data file consisted of 104 data matrices, each with an appropriate system state identifier attached. All neural network configurations were trained using the matrix data as input stimuli and each state identifier tag as an output pattern. Further details regarding the structured training and architecture of this package can be found in Chapter 4.

B. Derivation of the Generalised Delta Rule for Network Training

The back-propagation learning algorithm is a central component of the application of neural networks to specific problems. Without a mechanism for modifying the weights which constitute the means of learning a neural network becomes useless. The derivation of the generalised delta rule for the updating of weights in multilayer networks during training is presented in this section. Fig. B-1 identifies each of the components of a simple network presented with an input vector, X_{Pi}, and the corresponding output training vector, T_{Vi}.

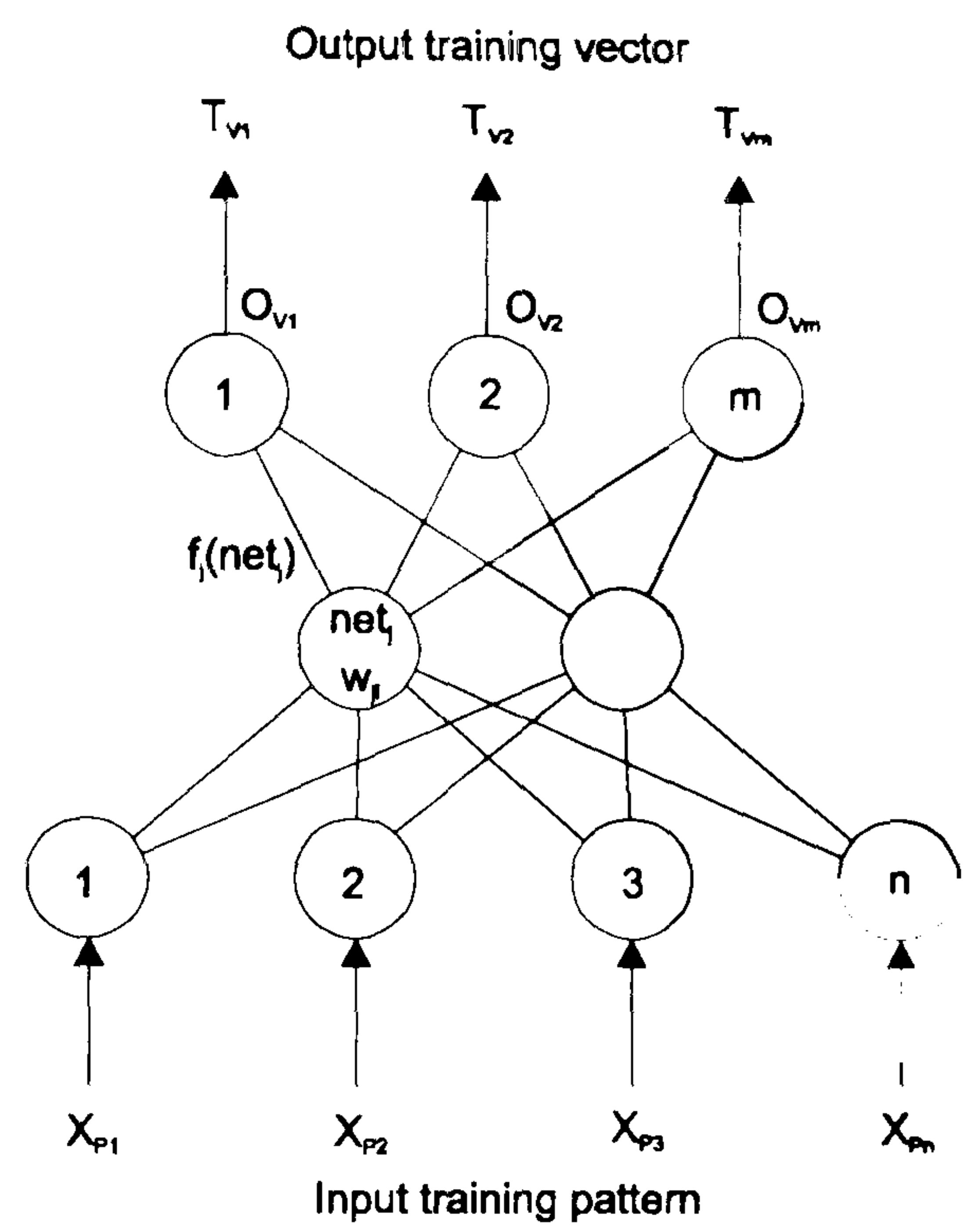

Figure B-1 A simple three layer perceptron network training with an input vector, X_p

The weighted sum of inputs to the jth node in a simple feedforward perceptron network from n nodes in the networks preceding layer is

$$net_j = w_{jo} + \sum_{i}^{n} w_{ji}x_i \qquad \text{(B-1)}$$

This input is then passed through a nodal activation, or transfer function[1], f_j, to produce the output for the jth node, o_j

$$o_j = f_j(net_j) \qquad \text{(B-2)}$$

If the jth node is an output unit then the error, E, between the actual output, o_j, and the target output, T_j, defined during the application of a training vector to the network is

$$E = \frac{1}{2}(T_j - o_j)^2 \qquad \text{(B-3)}$$

[1] During all practical trials the nodal activation was performed by a sigmoid function

$$f_j(net_j) = \frac{1}{1 + e^{-net_j}}$$

The weight change, Δw_{ji}, to help correct this error should be proportional to the contribution of that weight to the total error, E.

$$\Delta w_{ji} = -\alpha \frac{\partial E}{\partial w_{ji}} \qquad \text{(B-4)}$$

or

$$\Delta w_{ji} = \alpha \delta_{ji} o_{ji} \qquad \text{(B-5)}$$

Where δ_j is the error information term for the j^{th} node and is defined by

$$\delta_j = -\frac{\partial E}{\partial net_j} \qquad \text{(B-6)}$$

Converting this using the chain rule we obtain

$$\delta_j = -\frac{\partial E}{\partial o_j}\frac{\partial o_j}{\partial net_j} \qquad \text{(B-7)}$$

However from (B-2)

$$\frac{\partial o_j}{\partial net_j} = f'(net_j) \qquad \text{(B-8)}$$

Thus,

$$\delta_j = -\frac{\partial E}{\partial o_j} f'(net_j) \qquad \text{(B-9)}$$

and from (B-3)

$$\frac{\partial E}{\partial o_j} = -(t_j - o_j) \qquad \text{(B-10)}$$

Therefore by substitution

$$\delta_j = (t_j - o_j) f'(net_j) \qquad \text{(B-11)}$$

Likewise for a hidden layer node within the network δ_j can be derived as

$$\delta_j = f'(net_j) \sum_k \delta_k w_{kj} \qquad \text{(B-12)}$$

Thus the weight at time $(t+1)$ is defined as

$$w_{ji}(t+1) = w_{ji} + \Delta w_{ji}(t+1) \tag{B-13}$$

where

$$\Delta w_{ji} = \alpha\delta_j(t+1)o_i(t+1) + \beta\Delta w_{ji}(t) \tag{B-14}$$

and β is an optional momentum term used to accelerate the learning process.

C. Published Material

During the period of research the following papers were published:

W. Lucking, E. D. Chesmore, and M. Grayson, "Amplitude TES pre-conditioning for characterisation with neural networks", Proc. IEE International symposium on communications theory and applications, July 1993, pp. 111-113

W. Lucking, E. D. Chesmore, and M. Darnell "Acoustical condition monitoring of a mechanical gearbox using artificial neural networks", IEEE International conference on neural networks, WCCI, Orlando, June 1994, pp. 3307-3311

W. Lucking, and E. D. Chesmore, "Acoustical condition monitoring of a mechanical gearbox using artificial neural networks- further results", 10th International conference on systems engineering, ICSE '94, Coventry, September 1994, pp. 749-753