Optical Method and Neural Network for Surface Roughness Measurement and Surface Pattern Classification

by

Zahide Yilbas, B.Sc.(Eng.), M.Sc.(Eng.)

Ph.D. 1998

OPTICAL METHOD AND NEURAL NETWORK FOR SURFACE ROUGHNESS MEASUREMENT AND SURFACE PATTERN CLASSIFICATION

By

Zahide Yilbas, B.Sc.(Eng.), M.Sc.(Eng.)

SUPERVISOR: Professor M.S.J. Hashmi Ph.D., D.Sc.

This Thesis is submitted to Dublin City University
as the fulfilment of the
requirement for the award of the degree of

Doctor of Philosophy

School of Mechanical and Manufacturing Engineering
Dublin City University
April, 1998

DECLARATION

I hereby certify that this material which I now submitted for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed ____________________
Candidate

I.D. No.: 96971614

Date: 9th July, 1998

ACKNOWLEDGEMENT

I am very grateful to Professor M.S.J. Hashmi, Head of School of Mechanical and Manufacturing Engineering, for his thesis supervision, support and many constructive discussions during the course of this thesis. To the technical staff of the school who dedicated many hours in assisting the experimental work.

Finally, I would like to thank to my family for their love, understanding, tolerance and continuous support throughout this research.

Zahide Yilbas

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION, LITERATURE SURVEY AND SCOPE OF THE WORK

CHAPTER 2: SURFACE CHARACTERISTICS AND GENERATION OF PATTERNS ASSOCIATED TO SURFACE ROUGHNESS

CHAPTER 3: EXPERIMENTAL WORK

CHAPTER 4:NEURAL NETWORK: SIMPLIFICATION AND APPLICATION TO SURFACE PATTERN CLASSIFICATION

OPTICAL METHOD AND NEURAL NETWORK FOR SURFACE ROUGHNESS MEASUREMENT AND SURFACE PATTERN CLASSIFICATION

Zahide Yilbas B.Sc. Eng., M. Sc. Eng.

ABSTARCT

In this present study, two optical methods employing diffusive and specular reflections from the steel surfaces are considered to measure the surface roughness value (Ra). The first method is introduced for Ra ≥ 2 μm while the second method is employed for 0.1μm $\leq$ Ra ≤ 2 μm. The peak intensity of the reflected beam, and a Gaussian curve parameter of a Gaussian function, approximating the peak intensity of the reflected beam, are measured for the first and second methods, respectively.

Since a unique Ra value exists for a surface, the data collected for each profile were combined to produce a profile representing the Ra value for that particular surface, which is Gaussian in nature. The relationship between Ra and the standard deviation of Gaussian function (SDGF) was developed.

An experimental set up associated with both methods has been designed and built. In this case, a He-Ne laser beam was used to scan the workpiece surface while fiber optic probes were employed to collect the reflected beam. To calibrate the fiber optic probes, Ra is measured initially using a Bendix surface proficoder.

A back-propagation neural network classifying the surface patterns resulting from the first method was developed. A network simplification based on the self-pruning of the weights was employed. Control chart patterns resembling the possible surface profiles were developed when training the network.

It is found that, the resolution of the surface texture measurement improved considerably in the case of presently employed optical method. The neural network developed for this purpose could classify the resulting surface patterns successfully. Newly introduced self-pruning method results in an improvement in the network performance and minimisation of the network structure and computing time.

The first scheme used in the second method gives an improved standard estimate of error. The linear relationship was found between the Ra values and SDGF of the reflected beam intensity. Higher the SDGF values result in higher surface roughness. However, the measurement is limited to a certain range of Ra values; in this case, the accuracy of the measurement drops considerably as the Ra value reduces below 0.1 μm.

NOMENCLATURE

A	Constant of Gaussian Function
a	Amplitude of cyclic variation (taken as 51 or less)
A_j	Activation Level of unit j
B^2	Broadness of Gaussian Function
B	Gaussian Curve Parameter
DRMS Error	Change of RMS Error in Consecutive Training Iteration
DELTARMS	Change of RMS Error during Weight Deleting and No-Deleting Operations
$f(S_j)$	Sigmoid Function
g	Magnitude of the gradient of the trend (taken as being in the range 0.2 to 0.5)
$g_n(x_i)$	Intensity Function
High	Maximum of the scaled range (equals 1)
Input (i)	Input value for processing element i
k	Parameter determining the shift position (k = 0 before the shift position; k=1 at the shift position and thereafter)
L	Sampling Length (m)
Low	Minimum of the scaled range (equals 0)
m	Number of Spatial Locations in Intensity Curve
Max	Maximum allowed value for inputs (equals 110)
Min	Minimum allowed value for inputs (equals 50)
n	Number of Sampling
o	Offset of the scaling
O_k	Neural Network Output
Ra	Average Surface Roughness (m)
r	Normally distributed random number (between -3 and 3)
RMS Error	Root Mean Square Error

s	Magnitude of the shift (taken as being in the range 7.5 to 20)
Sj	Weighted sums of the unit j
S_B	Bias Uncertainty
S_P	Precision Uncertainty
SEE	Standard Estimate of Error
sf	Scaling factor
t	Discrete time at which the pattern is sampled (taken as being within the range 0 to 59)
T	Period of a cycle (taken as being in the range 4 to 12 sampling intervals)
T_i	Threshold Function
t_j	Target Value of Unit j
TR1, TR2	Threshold values for weight pruning
ΔTR	Threshold Increment for weight pruning
$t_{m,n}$	t-statistics with m and n degrees of freedom
U	Total Uncertainty
Uninput (i)	Scaled input values for processing element I
$W_{i,j}$	Weights between Input/Hidden and Hidden/Output layers
$W_{1,max}$	Maximum value of the weight between Input/Hidden Layer
$W_{2,max}$	Maximum Value of the Weight between Hidden/Output Layer
x	Distance (m)
$y(t)$	Sample value at time t
y_i	Measured Response
y_{eq}	Calculated Response
Z_i	Maximum peak to valley height (m)
β	Gaussian function constant
δ_j	Error for unit j in the Hidden Layer
η	Learning Rate
μ	Mean value of the process variable being monitored (taken as 80)
σ	Standard deviation of the process (taken as 5)
τ	Sample Standard Deviation

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 1

INTRODUCTION, LITERATURE SURVEY AND SCOPE OF THE WORK

1.1 INTRODUCTION

In industrial manufacturing, a continuous improvement in quality requires to monitor various process and product conditions. The surface finish of manufactured parts is a key attribute in determining the quality. The surface roughness affects several functional attributes of manufactured components. These include fatigue life, contact surface friction, contact stress distribution, etc. Consequently, desired surface roughness values are, in general, specified for a specific part and pertinent processes are selected in order to achieve the specified conditions. Most industrial processes involve uncertainty or time variant characteristics due to variable properties such as tool wear, vibration, material defects and etc. Therefore, the measurement of surface roughness needs to be performed to ensure the desired quality of a part.

Surface is defined as the boundary between one material and another In the applications, with which we are concerned, one material is invariably air and other a solid material such as metal. The texture produced by machining covers the whole surface so ideally assessment is made over an area.

Fortunately most surfaces produced by the common manufacturing processes have a texture which will differ only in detail from that obtained from a parallel trace made at another location a short distance away. Even a surface that appears to have an irregular pattern tends to have profiles, which are similar, wherever they are traced. Although experience has shown that a trace along a single track is usually adequate to obtain a typical profile, it is sometimes necessary to assess the texture over an area. Some instruments, differing from the stylus instrument in principle, can do this, but even with a stylus instrument it is possible to take several closely-spaced parallel traces to build up 3-dimensional picture of the surface.

The mechanical surface roughness measurement system, in general, results in a measurement limit and, when the high resolution for the mirror finish surfaces are concerned, the use of the mechanical method has disadvantages and gives limited accuracy [1.1].

These limitations and disadvantages may be listed as follows:

1- On surfaces with deep valleys, the stylus (measuring probe), because of its tip size, may not be able to penetrate fully to the bottom. The larger the tip radius the less will be the measured excursion of the stylus and therefore the roughness values, both computed or recorded on the graph, will be smaller.

2- When a spherical stylus passes over a sharp peak, the point of contact moves across the stylus, from one side to other. This causes the stylus to follow a path which is more rounded than the peak, although the stylus is raised to its full height when it is directly on the crest, the true peak height may not be measured with any accuracy.

3- Whenever a re-entrant feature is encountered, the stylus tip will lose contact with the profile and will, therefore obliterate this feature on the graph. Surfaces with re-entrant asperities exist chiefly in cast iron, sintered and porous materials, and this is one of the reasons why surface roughness (Ra) measurement on these materials may be misleading.

4- Because of the mechanical inertia of the pen and friction between the pen and the chart paper, the frequency response of the recorder is usually the factor which limits the upper overall stylus-to-graph frequency response or (bandwidth) of the system. This in turn determines the maximum permissible traverse speed.

5- The surface of the workpiece may be damaged during the mechanical contact between the stylus and the surface. This is very important in the case of mirror finish surfaces. Therefore, new and alternative methods for surface roughness measurement, which do not damage the surface need to be investigated.

Due to the limitations in the applicability of the mechanical measurement system the use of non-destructive surface roughness measurement may be fruitful in manufacturing industries. In general, two types of measurement techniques are availbale in this regard. These include optical and ultrasonic measurement techniques. The ultrasonic technique relies on the sensing of reflected ultrasonic waves.

However, the accuracy of this method is relatively poor as compared to optical method, and the measurement is limited to the surfaces having roughness of higher than 0.8 μm [1.2]. Therefore, this technique was excluded in the present study. On the other hand, the application of optical methods enables the measurement of surface roughness with considerable accuracy [1.3]. In general, three basic optical techniques can be used to monitor and measure the surface roughness of metallic surfaces. These include; interferometry, surface image analysis and light scattering techniques.

The underlying principle of an interferometer is that when two coherent beams of light waves from different paths intersect, fringe patterns are formed due to the interaction of these waves. The intensity of these fringe patterns depends on the path or phase difference between the waves. Consequently, the fringe pattern intensity and phase difference define the roughness of the surface. This method offers considerable advantages for the inspection of super-finished surfaces. However, the equipment involved is expensive and the measurement is very sensitive to vibration.

In the surface image analysis technique, a grey level analysis of microscope images is considered. This method usually employs a white light sources, which preclude any diffraction effect. Therefore, the measurement range is limited to a large scale roughness or waviness, i.e. a high grade roughness.

The light scattering technique relies on the light reflection from the metallic surfaces; in this case, when light is incident on a rough surface it scatters into various directions with each of different intensities. The reflected light travels back in definite directions as though coming from a surface placed in some new position. The reflecting surface then acts as an aperture and limits the reflected beam [1.4]. Such a reflection is said to be regular or specular. However, when a surface is not well polished, every irregularity of the surface reflects the light in a different direction and the light will not return as a definite beam. This type of reflection is considered to be irregular or diffused, therefore, the light is scattered. The various properties of the reflected light can be measured from the scattering pattern. These include light intensity, angular distribution, and degree of polarisation. Consequently, some of these properties measured may be related to the characteristics of a surface texture of that particular surface.

In the light of the above arguments, it is the light scattering technique, which is considered in the present study. Moreover, the patterns measured from the diffused reflections are classified by using a neural network. The specular reflection technique is employed to corelate the broadness of the reflected beam intensity profile, detected by the probe, and the surface roughness value (Ra) obtained from the stylus measurement. The study is extended to include the self-pruning process to simplify and optimise the back-propagation network used in the present study.

1.2 LITERATURE SURVEY

The literature survey presented in this section consists of two parts which cover the optical surface roughness measurement and classification of resulting surface patterns through a neural network. Consequently, the literature survey is given under the relevant sub-headings.

1.2.1 OPTICAL METHOD FOR THE SURFACE ROUGHNESS MEASUREMENT

A lack of progress in mechanical surface texture measurement for in-process or on-line quality control and condition monitoring has prompted reserachers to develop new surface texture measurement techniques and/or to improve the mechanical measuring system. Independence of surface roughness parameters in the assessment of the surface texture was introduced by Kothandaraman et al [1.5]. They developed the correlation between the various surface roughness parameters and discussed the relative merits in the assessments of surface finish. They investigated the surface topography of C45 steel after surface grinding and established the correlation between the surface roughness and the various parameters.

A non-contact optical technique for rapidly measuring surface roughness using a laser beam was studied by Kurita and Sato [1.6]. The reflected light beam intensity was measured by a plasma-coupled device (PCD). The reflected beam intensity distribution was approximated by a Gaussian function and the broadness of the intensity distribution curve was evaluated with a Gaussian curve parameter. They showed that the surface roughness value (Ra) within the range of 0.1 to 0.5 μm could be measured with considerable accuracy.

Stylus instrument for roughness and profile measurement of ultra-fine surfaces was developed by Garratt and Nettleton [1.7]. They introduced a stylus profiling instrument and nano-step performance was discussed in respect of noise, straightness and profile. They developed a practical instrument for a nano-step surface roughness measurement. However, the repeatability of the measurement was not indicated.

Domanski and Wolinski [1.8] introduced two electro-fiber-optic systems to measure the surface roughness. The measurement included the intensity and polarisation based methods, which required the use of either multi-mode or single-mode polarisation preserved optical fibers. Some configurations, including different constructions of a fiber optic head and the specific requirements for optical fibers were discussed. However, they did not fully explaine the limitations of the measurement.

The evaluation of limits to the performance of the surface roughness meter developed for remote sensing was carried out by Johnson et al [1.9]. The instrument developed comprised of rugged, portable, cost effective field component and a personal computer-based image analysis system for data processing. The results demonstrated very accurate estimates of root mean square height. The standard approach of estimating the correlation length from the auto-correlation function was subjected to greater errors.

Stone and Shafer [1.10] investigated possible determination of surface roughness from the reflected step edges. They introduced a technique based on the Beckmann-Spizzichino [] model of reflection for simulation of the images, which ordered the sample roughness. The height distributions of the surfaces were studied by stylus profilometry and were found to be non-Gaussian.

An optical non-contact profilometer was introduced by Dawei [1.11] for super-smooth surfaces and applied a common-mode rejection technique. The laser amplitude noise and environmental disturbances were eliminated through a technique developed by the authors. However, the practical applicability and limitations were left obscure.

Surface roughness monitoring using computer vision was investigated by Sodhi and Tiliouine [1.12]. A comparative experimental surface roughness measurement method using the speckle pattern caused by a laser beam on a rough surface was introduced. The online use of the method introduced for tracking the roughness of a workpiece being processed on a surface grinder was successfully applied.

El-Baradie [1.13] investigated the computer aided analyses of a surface roughness model for turning 140 BHN steel at a very high cutting speed of 500 m/min. He discussed the effect of cutting parameters on the resulting surface roughness contours. He then determined the optimum cutting conditions to achieve a required surface roughness.

The surface roughness detection by tool image processing was studied by Galante et al [1.14]. They introduced a technique for the on line control of the surface roughness in turning operation. The model developed enabled an estimate of the value of the effective roughness of the workpiece to be made from one related to the ideal profile. They indicated that the model was not significantly influenced by the feed employed.

Alauddin et al [1.15] studied factors affecting the surface due to an end milling process. They developed a model predicting the surface roughness of the milled workpieces. The parameters employed in the mathematical model included cutting speed, feed rate and axial depth of cut. They constructed the surface contours using a computer, and used the contours for determining the optimum cutting conditions for a required surface roughness.

A metal surface roughness was investigated by Peiponen and Tsuboi [1.16]. They introduced dispersion of the roughness distribution as a new parameter when describing the surface roughness. They investigated experimentally the surface reflectance of various metals and indicated that the laser based reflection system could be used to monitor the surface roughness in industry.

Bengtessen [1.17] developed a laser profilometer for surface roughness detection. The incident laser beam was directed normal to the surface. The device was calibrated using a relocation table. He monitored the roughness height within an error limit of 3 μm. He indicated that the profilometer might be used on curved surfaces and for on-line wear monitoring.

Theory and development of user friendly software for the measurement of surface roughness and geometrical parameters was developed by Parsons and Tabenkin [1.18]. They discussed the development of a touch screen and related software functions using a dedicated keyboard which represented a very limited number of operating options at any one time. The program developed presented only specific questions or options relevant to the measurement.

Westra and Thomson [1.19] investigated the effect of tip shape on surface roughness measurements from atomic force microscopy images of thin films. The investigation was based on the distortion of images by the finite radius of the tip in atomic force microscopy when imaging of thin films. They developed the relation between the horizontal and vertical roughness, and power spectral density. They indicated that the vertical roughness was the least affected by the tip induced distortion.

Correlation of light scattering measurements and visual ranking of optical surfaces was carried out by Guenther et al [1.20]. A surface roughness criterion was presented in relation to differential interference contrast microscope. They provided the scatter measurements and equipment, and choices for the assessment of optical surface roughness.

Mhyshkin et al [1.21] studied the quantitative analysis of surface topography using a scanning electron microscopy (SEM). They introduced a supplement to the existing microscopy through a image processing technique. A method of evaluating topography using a scanning microscopy and a personal computer was described. However, the method introduced was expensive and required an SEM facility.

Interferometric methods of speckle optics in the laser diagnostics of the workpiece surfaces was investigated by Ryabukho et al [1.22]. The peculiarities of partially developed speckle fields formed by focused laser beam diffraction on rough surfaces were considered. The methods of profile and surface roughness parameter measurements were discussed and their characteristics were analysed.

The light scattering from flat periodic rough surfaces for surface roughness estimation was carried out by Fan and Huynh [1.23]. A special scattering geometry was adopted in which the plane of incidence was parallel to the lay and grazing angles were small. The Beckman's scattering model was applied to calculate the scattering intensity and numerical solutions were obtained for sunisoidal rough surface profiles. The results were used to determine the influence of the surface roughness amplitude and surface spatial wavelength on the scattering patterns. They indicated that for any given scattering order, the scattered intensity oscillated as the angle of incidence increased. The degree of oscillation was found to be directly related to the surface roughness amplitude since the higher roughness amplitude resulted in a higher oscillation frequency. However, the study was limited to well defined geometric surfaces such as surface of sine and cosine structures.

1.2.2 NEURAL NETWORKS PERTINENT TO PATTERN CLASSIFICATION AND NETWORK SIMPLIFICATION

A neural network is used in this study to classify the surface profiles obtained from the measurement. Therefore, the relevant literature on neural network simplification and pattern classification is introduced here briefly.

A neural network can be roughly characterised by three classes of features: their architecture or connectivity, their performance or operating rules, and their learning rules. Artificial neural networks developed over the past decade exhibit a range of features along these three dimensions, allowing them to be roughly grouped according to three architectural features, number of layers, feed-forward only or feedback connectivity and unidirectional versus bi-directional connections. Although many subdivisions are possible and, in the limit, each neural network could be placed as a single member of its own category, the above dimensions provide fairly good covering of the existing realm of the techniques used in artificial neural networks as described in the literature. Novel networks and variants exhibit distinct computational behaviour, and much remains to be learned about the value of the various attributes of such networks in terms of their contribution to particular critical operation features, such as learning speed, generalisation, capacity and error rate.

The structures developed makes the neural network non-algorithmic massively parallel computing devices. When constructing an artificial neural network, a difficulty arises in choosing a network of the right size for a problem. If it is possible to solve the problem with a network of a given size, it can also be solved with a larger net which imbeds the smaller network with (hopefully) all the superfluous connections or synapses having a zero strength. The connections are key to the ability of neural network to exhibit learning and memory capabilities. In a network with superfluous connections the learning algorithm will typically produce a different structure, with non-vanishing synaptic weights spreading over all the network. For this reason it is very important to choose an optimum size of network for a particular task. A larger network may provide improved learning rate, but tends to memorise the training patterns, which in turn results in poor generalisation capability. The cost of computation, measured by the number of arithmetic operations, grows almost linearly with the number of synaptic connections. Hence the smaller the net the more efficient it is in both forward computations and learning, but when the network is too small it may not be efficient in solving the problem. For these reasons, it is desirable to start with a large network then minimise its size considering its learning rate and generalisation capacity [1.24].

The aim of the minimisation of the network is to obtain an optimal neural network without loss of performance. When a neural network can provide an adequately high performance after deleting some of its processing elements or connections then that neural network can be thought of as being unnecessarily large. Neural network minimisation is a process where the non-contributing processing elements or connections within a network are deleted, ideally to leave a minimum configuration with positively contributing elements. The network should still be able to provide a satisfactory performance, but as a general rule this is not always expected [1.25].

Determination of the minimum set of a neural network is important for two distinct reasons. First, explaining the decisions of a neural network is impractical with highly superfluous processing elements and connections. Explanation is vital because neural networks will not be used to make decisions with important consequences without being required to demonstrate the basis for the decisions. Second, for a serial machine, response time is proportional to the number of connections, and learning time is proportional to its square [1.26]. The weight elimination in a back-propagation network was discussed by Santini [1.27]. He introduced a cost function to prune the inactive neurons. Won and Pimmel [1.28] introduced connection-pruning algorithm for network simplification. However, the method used underestimated the effect of early pruned weights on the dynamic response of the network, since the pruning was introduced before the start of the training iteration. Moreover, it was shown that the initial weights had significant effect on the early stages of the back-propagation learning process [1.29]. A pruning procedure for back propagation trained networks was investigated by Karnin [1.30]. He introduced shadow arrays, which keep track of the incremental changes to synaptic weights during a single pass of back-propagation learning. However, the threshold for weight pruning was not stated clearly. The study on the network size minimisation was conducted by Kamruzzaman et al [1.31]. They modified the activation function in the hidden layer and introduced encoding of target vectors. The method introduced left obscure the simplification stages.

On the other hand, a considerable amount of research work has been carried out for pattern recognition and system identification. Kayata and Kawato [1.32] developed a control structure where neural networks were employed to model system dynamics. Yamamura et al [1.33] formulated several neural network control schemes where neural networks were used as dynamic or inverse dynamic identifiers.

Ananth et.al. [1.34] presented an optimal pruning algorithm for a neural network which was developed and called Neural Tree Network (NTN). The NTN was grown by a constructive learning algorithm that decreased the classification error on the training data recursively. They were presented to show improved generalisation using the pruning algorithm on a speaker independent vowel recognition task..

Youngjik et. al. [1.35] proposed an efficient method which yielded to an appropriate number of hidden nodes by iteratively removing unnecessary hidden nodes. The criterion for selecting the removable nodes derived from the analysis of information transformation of individual neurons and the network as a whole. They applied their method to some problems including hand-written digit recognition, and this resulted in a significant reduction of the number of hidden nodes.

Abraham et. al. [1.36] studied the suitability of Artificial Neural Networks in solving pattern classification problems that required massively parallel computation. A neural network simulation was developed for pattern recognition tasks. Their work was later extended to further study simple process control problems. A diagnostic problem was developed and neural network approach was introduced to solve the pattern classification problems. The issues related to this form of diagnostic problem solving were discussed relative to the more common use of an expert system based design and implementation.

Hutton [1.37] worked on Neural Network (NN) approaches to pattern classification problems both complement and compete with statistical approaches. Each approach had unique strengths that could be exploited in the design and evaluation of classifier systems. Three points were made in this article. First, classical (statistical) techniques could be used to evaluate the performance of NN classifiers. Second, NN classifiers often outperformed classical techniques. Third, NN classifiers might have advantages. He suggested that methods which are frequently used in statistics, but not in NN approaches, should be adopted for the latter as well.

Patriker [1.38] studied an Artificial Neural Network (ANN) architecture termed a dual network which was proposed for pattern classification problems. A dual network was described as a network of densely connected simple processing elements and it presented a structured way to implement polynomial classifiers. A supervised learning algorithm was developed for these networks and their ability to solve complex pattern classification problems was verified through experimental studies.

Murphy [1.39] presented an algorithmic approach to design feedforward neural networks for pattern classification. The technique computed a single hidden layer of nodes by adding one node at a time until the desired classification had been achieved. At each iteration, a node that maximised an information was selected from a collection of candidates. The methodology was heuristic in nature, intended to solve the problem of constructing a neural network with a minimum number of nodes. He presented two strategies for computing a collection of candidates and obtained some experimental results.

Fine and Turman [1.40] investigated the trade-offs among network complexity, training set size, and statistical performance of feedforward neural networks. Nets were labelled as given functions and being classified as given input points. The use of error probability as a measure of network performance was considered.

1.3 SCOPE OF THE WORK

The commonly used instrument for surface roughness measurement is the mechanical stylus-profilometer. The measurement is based on a simple principle; in this case, a stylus having a small tip radius, around 1 μm, is dragged across the surface at a constant speed. Its vertical movement is detected by displacement transducer, and is amplified to create a time varying signal, which is approximately proportional to the amplitude of the surface roughness profile. In the past years, various improvements have been made to enhance the resolution and data processing capability. However, at present, there is a lack of suitable texture measurement sensors for in-process or on-line quality control and condition monitoring. These applications require fast measurement speeds and the ability to operate in a hostile environment such as dirt, cutting fluid, etc. Due to the demands for increased reliability and quality, research into new surface roughness measurements has accelerated in recent years. In addition, to overcome the disadvantages of the stylus contact measurements, various innovations in the area of non-contract roughness sensing have been introduced. These include the capacitance-based, the ultrasonic and the optical techniques providing the most favourable ones mainly because of their fast measuring speed. Although great improvements have been achieved in the first two techniques in recent years, only the optical technique has experienced rapid progress and holds the greatest promise for surface roughness measurement. This is because the optical techniques potential for providing high resolution by virtue of relatively short illumination wavelengths. Moreover, new AI techniques such as neural networks enhances the performance of existing techniques, and also, enables a classification of the resulting surface roughness pattern.

In the light of above arguments, the present study was conducted to measure the surface roughness and to classify the resulting surface roughness profiles. A computer controlled fiber-optic surface roughness measurement system was developed while neural network was introduced to classify the resulting surface roughness patterns. The measurement of the surface texture was based on the reflection of incident laser beam from the surface of the steel substrate. Diffusive and specular reflections are considered; in this case, the measurement relies on the diffusive reflection for rough surfaces ($Ra \geq 2\mu m$) and specular reflection is considered for relatively smooth surface ($0.1 \leq 2\mu m$). In both cases, measurements were carried out to establish the reference intensity levels. When $Ra \geq 2\mu m$, the surface roughness of the workpiece can be determined from the intensity of the reflected beam. On the other hand, a relatively smooth surface modifies the intensity distribution of the incident beam at reflection. Therefore, the amount of distortion of the incident beam profile after the reflection can be related to the surface roughness of the workpiece.

An experimental set up has been designed and built to measure and monitor the surface roughness. During the measurements, the workpiece is placed on a holder, which enables 2-dimensional plane motion of the workpiece. The axial motion parallel to the workpiece surface is precisely controlled using a micrometer with graduation of 0.0254 mm, the micrometer is driven by a stepper motor, which is linked to a computer. The He-Ne laser beam is focused onto the workpiece surface and an optical probe receives the reflected beam. Two probes are utilised in this regard: one for the rough surfaces ($Ra \geq 2\mu m$) and the other for relatively smooth surfaces ($0.1 \leq Ra \leq 2\mu m$). The first probe consists of a fiber-optic cable and fast response photodiode, while the second probe is made of a bundle of fiber-optic cables and photo detector arrays. To monitor the variation in the power intensity of the incident He-Ne laser beam, a reference fiber-optic probe was employed; in this case, about 8% of the incident He-Ne laser beam was sampled using a beam spliter. The reflected beam power intensity was normalised using the sampled beam. Therefore, variation in the fiber-optic probe(s) output due to intensity fluctuation of the incident beam is eliminated. Software was developed, which provides the real time plot of the probe response and enables to monitor the position of the workpiece during the measurement. The surface roughness value and relative estimate error were calculated accordingly.

A neural network employing a back-propagation algorithm was developed to classify the resulting surface roughness patterns, i.e., the fiber-optic probe output was fed to the neural network, which enabled a classification of the patterns according to the control chart patterns introduced previously. In order to simplify the structure of the neural network developed a new concept of weight pruning mechanism based on automatic thresholding was introduced. This provided an improved learning and rapid testing of the resulting patterns. The study was extended to include the uncertainty analysis employing a 95% confidence level.

CHAPTER 2

SURFACE CHARACTERISTICS AND GENERATION OF PATTERNS ASSOCIATED TO SURFACE ROUGHNESS

2.1 INTRODUCTION

The characteristics of a surface can be very complex and composed of roughness, waviness and form, which exist in combination. In addition, there is an inherent microstructure due to the crystalline, or molecular structure of the material, i.e. very few natural surfaces are molecularly smooth. Consequently, in the present chapter, the texture and defining parameters, the formulation of the surface roughness and the control chart parameters governing the resulting surface roughness are introduced.

2.2 SURFACE TEXTURE

The surface texture produced by a manufacturing process is very characteristic of the particular process employed, i.e. the surfaces produced by grinding tends to have an irregularly spaced but directional texture while the texture on surfaces produced by cutting processes tend to be both uniformly spaced and directional. Moreover, materials are manufactured and finished by innumerable different processes and to quote a single dividing value between texture parameters for all surfaces would be arbitrary and have little practical significance [2.1]. Consequently, surface texture may be classified according to their characteristics. These include; roughness, waviness and form as shown in figure (2.1). The irregularities, which are inherent in the production process, and left by the actual machining agent (such as the cutting tool, spark, etc), define the roughness. However, waviness is a component of the surface texture upon which roughness is superimposed. It may result from such factors as machine or work deflections, vibrations, chatter, various causes of strain in the material, and extraneous influences. On the other hand, it is almost impossible to specify uniquely where waviness and the irregularities become part of the general form of the surface. Consequently, the form may be defined as the general shape of the surface, neglecting variations due to roughness and waviness.

The distinctions in surface texture are qualitative and therefore difficult to express numerically. However, roughness is produced by the method of manufacture, resulting from the process rather than from the machine. A number of causes can contribute to roughness. These include the mark left by the tool or grit, which will be a periodic nature for some processes, the finer structure due to tearing of the material during machining, the debris of a built-up edge and small irregularities in the shape of tool tip. On the other hand, the waviness can be attributed to the characteristics of an individual machine error, which are often caused by insufficient rigidity of the setup used to support the workpiece, allowing it to bend or deflect under the machining forces.

The surface texture characteristics are always in combination, since most surfaces are the result of the combined effect of roughness, waviness and form are different, and it is fruitful to be able to measure them separately.

2.3 SURFACE ROUGHNESS ASSESSMENT

The surface roughness consists of peaks and valleys. Consequently, all the methods of surface roughness assessment need to take account of both peak-to-valley parameters. However, in practice, the peaks and valleys may have some different functional importance. In some applications the peaks have a major effect, but in others the valleys are predominantly of more practical significance [2.2].

The machining pattern on many surfaces has distinctly directional characteristics, which is called the lay as shown in figure (2.2). These surfaces produce different profiles and results in different measured values according to the direction taken into account during the measurement. A measuring probe, such as a stylus, making a trace at right angles to the lay produces the peaks more closely spaced than one made at another inclination [2.3]. In an extreme case a trace made parallel to the lay may reveal no peaks at all. A trace only slightly inclined to the perpendicular may have very little effect on the measured surface roughness value (Ra), however, as the trace becomes more inclined the peak spacing increases and may fall outside the cut-off length over which Ra value is determined.

Lay direction is important in surface texture characterisation and measurement, and the lay needs to be defined correctly. However, on surfaces having multi-directional lay, it is usual to make measurements in several different directions, and to accept the maximum value as the roughness. Some surfaces have roughness and waviness lays at right angles to each other. This is not the same as cross lay because the relative heights and spacing of the two lays differ considerably. Therefore, it is desirable to specify the functional importance of the lay before the measurement.

On the other hand, the surface roughness-measuring device needs to trace accurately the outline of every peak and valley it encounters, retaining contact with each flank from top to bottom. Therefore, it is expected that all the information, that may be obtained from a line trace across the surface, should be contained in the profile signal. Owing to the limitations imposed by the measuring equipment some of the surface, which may not be traced by the equipment used. It is also possible that the spiky appearance of the peaks and valleys on the signal may lead much misunderstanding when assessing the surface visually from the signal. However, the relationship between the surface characteristics and the signal may be formed using the international surface roughness standards [2.4].

The surface texture is very complex and no one parameter has been found to quantify all this various characteristics. In most applications, however, it is only a few of characteristics that have practical importance. Therefore, the surface texture measurement is carried out through quantifying these characteristics. Although the surface texture measurement is largely standardised, each country defines parameters to suit the methods of assessments in use in that country [2.5]. However, the surface roughness value (Ra) is the most internationally agreed parameter used to predict the profile amplitude. On the other hand, the peak curvature is an important parameter for measuring wear on account of its sensitivity to changes of the type being monitored. The surface texture parameters fall into three groups depending on the characteristics of the profile that they quantify. These include:

- Amplitude Parameters: These are determined peak or valley heights, or both, irrespective of horizontal spacing, e.g. peak count.

- Hybrid Parameters: These are determined by amplitude and spacing in combination, e.g. average wavelength.

- Spacing parameters: They are determined solely by the spacing of irregularities along the surface, e.g. peak count.

2.4 SURFACE ROUGHNESS ANALYSIS

Three characteristic lengths which can be associated with the numerical assessment of the surface roughness are indicated in figure (2.3). These include:

i) Sampling length: This is the length of surface over which a single assessment of a measurement is made.

ii) Evaluation Length: The length of surface over which measurements are made.

iii) Traverse Length: The total length of surface traversed by the measuring equipment in making a measurement.

It should be noted that for short surfaces, the measurement should be confined to one sampling length; in this case, the sampling length becomes the same.

An important requirement when quantifying the surface roughness is the provision of some datum within the profile to which measurements can be related. In practice, the reference adopted for parameters is the centre line, i.e. the mean line, which is a straight line that passes centrally through the peaks and valleys. Within the sampling length, the sum of areas enclosed by the profile above the centre line equals the sum of these below.

Another important parameter is the meter cut-off, which is the instrumental equivalent of the metrology sampling length. The distinction between the sampling length and the cut-off length is that the sampling length is a physical length of the surface while cut-off is the means whereby the resulting waveform is made to stimulate the effect of restricting the assessment to the sampling length.

Derivation of Roughness Average (Ra):

The derivation of Ra can be illustrated graphically as shown in figure (2.4). The portions of the profile below the centre line within the sampling length are inverted and placed above the centre line; Ra is, then, mean height of the resulting profile. Mathematically, Ra is the arithmetic average value of the departure of the profile from the centre line throughout the sampling length.

In determining the Ra, the following points are considered:

- The Ra value over one sampling length represents the average roughness, therefore, the effect of the single spurious, non-typical peak or valley will be averaged out and have only a small influence on the value.

- The Ra value is meaningless unless the cut-off is quoted, or can be assumed. The Ra can only be repeatable if assessed with the same cut-off.

- Measurement should be made perpendicular to the direction of the lay.

- Ra gives no information as to the shape of the irregularities or profile.

- No distinction is made between peaks and valleys.

Determination of surface roughness:

The derivation of average surface roughness (Ra) due to an optical method can be obtained by employing the international surface texture measuring technique [2.5]. In this case, the portions of the profile below the centre line within a sampling length L are inverted and placed above the centre line. Ra is then considered as the mean height of the resulting profile, i.e. Ra is the arithmetic average value of the departure of the profile from the centre line throughout the sampling length (figure (2.4)). The Ra value is [2.5]:

$$Ra = \sum_{n=0}^{n} \frac{1}{L} \int_{x=0}^{x=Le} Z_i dx$$

where L, n and Z_i are the sampling length, number of sampling and maximum peak to valley height within the sampling length respectively.

It is sometimes desirable to specify the maximum roughness height, rather than the mean height that Ra gives. This parameter is defined as Rauhtiefe (Rt), which is the German for roughness depth. Since the international standards for the surface roughness is considered, it is, therefore, the mean roughness value (Ra) which is considered in the present study.

2.3 GENERATION OF PATTERNS FOR SURFACE ROUGHNESS PROFILES

The surface roughness profiles may be presented in terms of control chart patterns, Figure (2.5). Moreover, control charts provide information about the process capability and also provides information of diagnostic value to the network. It is possible to produce a series of numbers resembling the specific surface roughness patterns.

A typical control chart is a graphical display of a quality characteristic that has been measured from a sample versus the sample length or time. The chart contains a centre line that represents the average value of the quality characteristics corresponding to the in-control state (that is in between the upper and lower limits). In order to assess any type of pattern, the control charts defining the pattern may be used in this regard, i.e. the comparison between the pattern and control chart resembling the same pattern gives information about that particular pattern.

Each control chart pattern is a time series consisting of several points [2.6]. These points are the most recent average measurements of the quantity characteristic being monitored. Six control chart patterns are formed to resemble the surface roughness profiles. These include:

- Normal Pattern
- Increasing Trend
- Decreasing Trend
- Upward Shift

- Downward Shift
- Cyclic Pattern

When formulating the control chart patterns, some useful assumptions are made, which are given below:

- In a normal pattern, the parameter variations conform to a normal distribution.
- The mean and standard deviation of patterns are known.
- The maximum allowed deviation from the pattern mean for normal pattern is known.
- The minimum averaged slopes for the trend patterns are known.
- The maximum least-mean-square linear regression error for trend, cyclic, and trend patterns are known.

The mathematics for generating the control chart patterns is given as follows[2.7]:

(1) Normal patterns:

$$y(t) = \mu + r(t) * \sigma \qquad ....(2.1)$$

(2) Increasing and/or decreasing trends:

$$y(t) = \mu + r(t) * \sigma \pm g\,t \qquad ....(2.2)$$

(3) Upward and/or downward shifts:

$$y(t) = \mu + r(t) * \sigma \pm k\,s \qquad ....(2.3)$$

(4) Cyclic patterns:

$$y(t) = \mu + r(t) * \sigma + a * \mathrm{Sin}\,(2 * \pi * t / T) \qquad ....(2.4)$$

where:

μ : Mean value of the process variable being monitored (taken as 80)

σ : Standard deviation of the process (taken as 5)

a: Amplitude of cyclic variation (taken as 51 or less)

g: Magnitude of the gradient of the trend (taken as being in the range 0.2 to 0.5)

k: Parameter determining the shift position (k = 0 before the shift position; k=1 at the shift position and thereafter)

r: Normally distributed random number (between -3 and 3)

s: Magnitude of the shift (taken as being in the range 7.5 to 20)

t: Discrete time at which the pattern is sampled (taken as being within the range 0 to 99)

T : Period of a cycle (taken as being in the range 4 to 12 sampling intervals)

y(t) : Sample value at time t

Patterns in the training set are scaled between 0 and 1 according to maximum and minimum input values. Scaling procedure may be given as:

s: Scaling factor

o: Offset of the scaling

Input (i) : Input value for processing element i

Unitinp (i) : Scaled input values for processing element i

High : Maximum of the scaled range (equals 1)

Low : Minimum of the scaled range (equals 0)

Max : Maximum allowed value for inputs (equals 110)

Min : Minimum allowed value for inputs (equals 50)

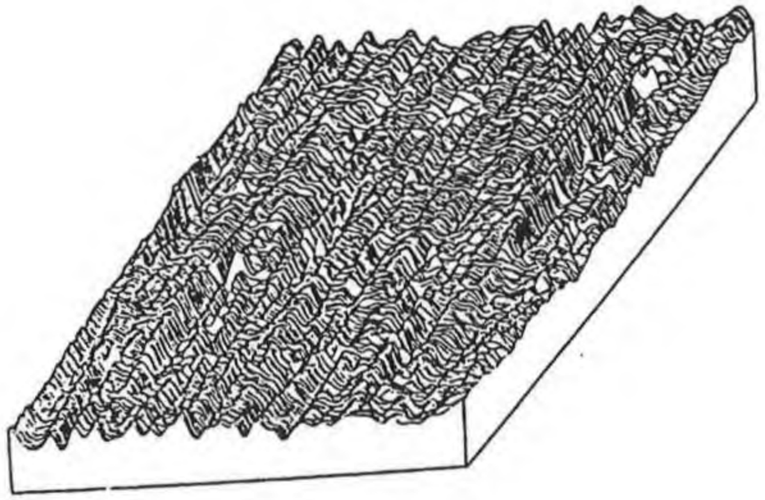
Scaling procedure is as follow:

s = (High - Low)/(Max - Min)

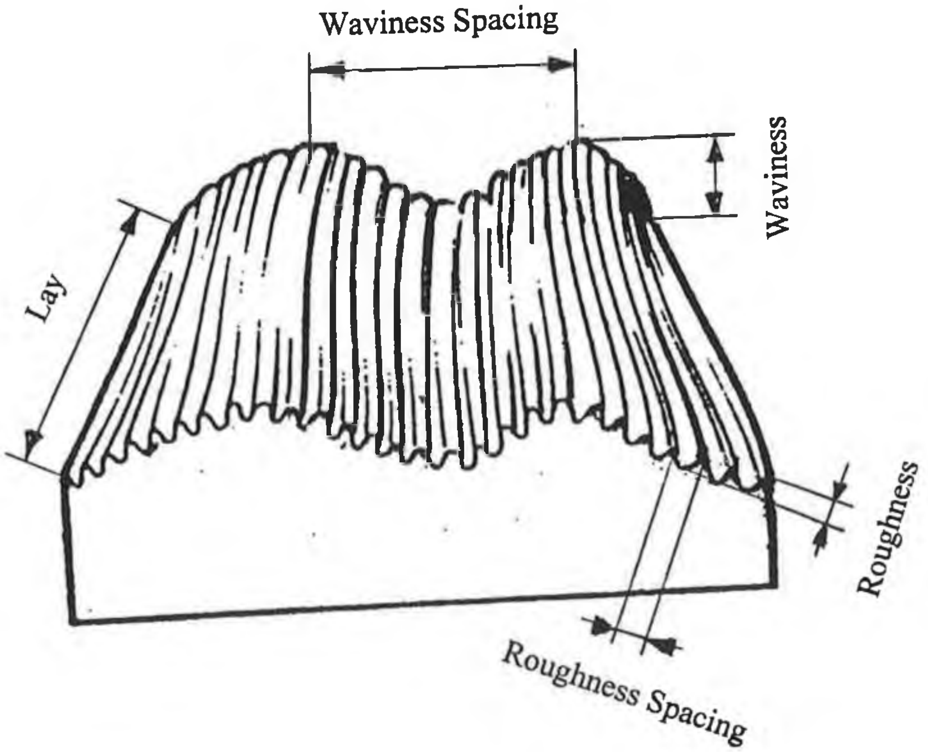
o = (Max *Low) - (Min*High)/(Max - Min)

Unitinp(i) = s*Input(i) + o

In the present study two different structures of data are introduced. In the first structure, the coefficients and random variables are varied. In the second structure the coefficients of equations 2.1, 2.2, 2.3 and 2.4 are kept constant while random variables are changed. The resulting data show different appearance. As a result, (6 x 100) data sets resembling the control chart patterns were developed. The control chart patterns developed were used to train the neural network constructed for classifying the resulting surface roughness profiles. Consequently, (6 x 100 x 20) input data strings were developed to train the network and a further (6 x 100 x 20) data strings were developed to test the resulting network. A view of typical chart patterns is shown in figure (2.5). In the training and testing processes, both data structures were employed independently.



General view of the surface.



Close view of the surface.

Figure 2.1 – 3-dimensional view of surfaces.

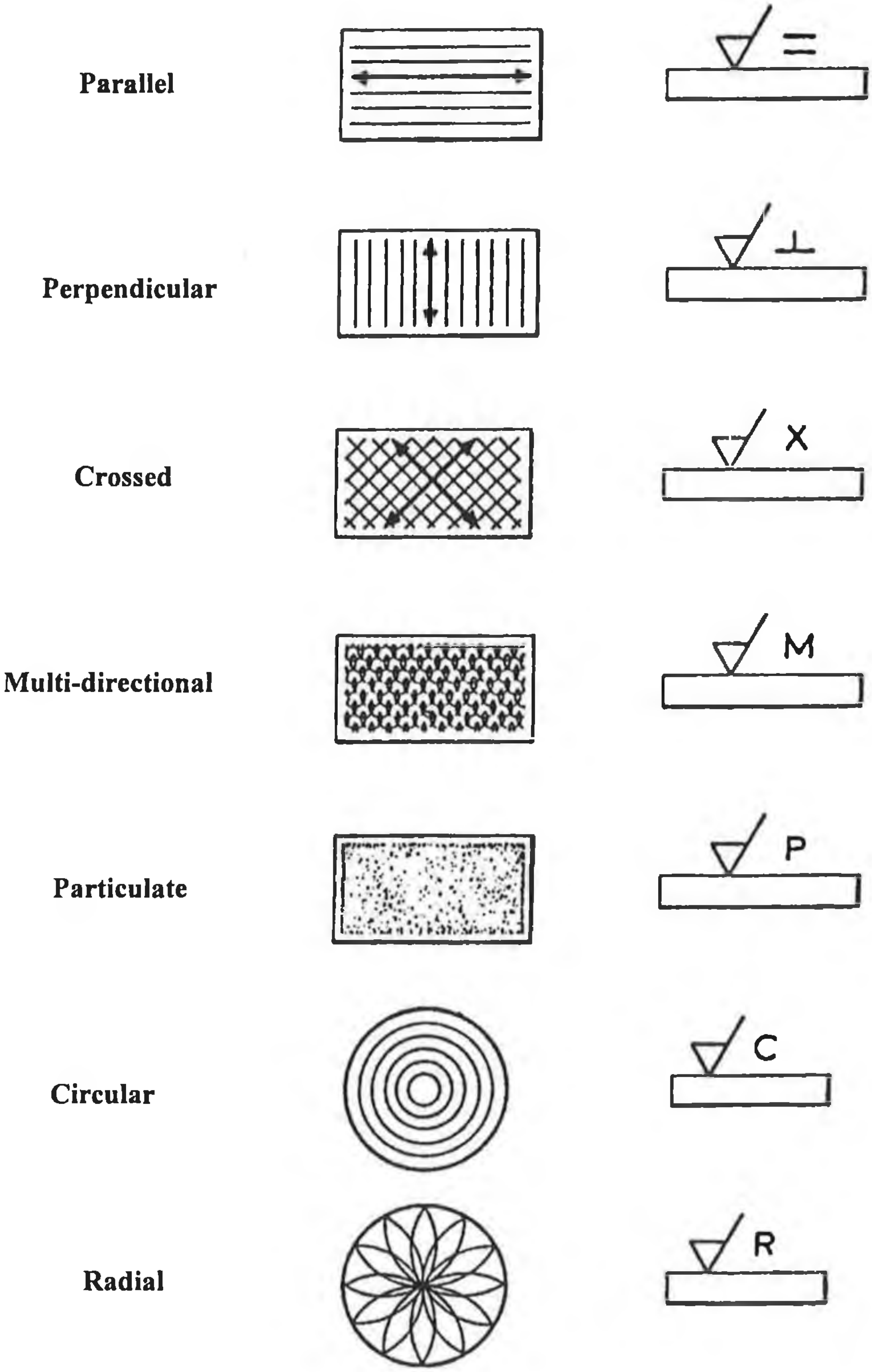


Figure 2.2 – Lays and corresponding symbols.

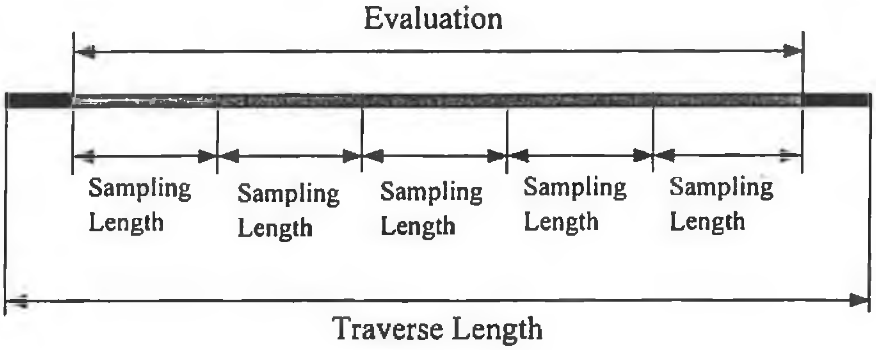


Figure 2.3 - Schematic view of characteristic length.

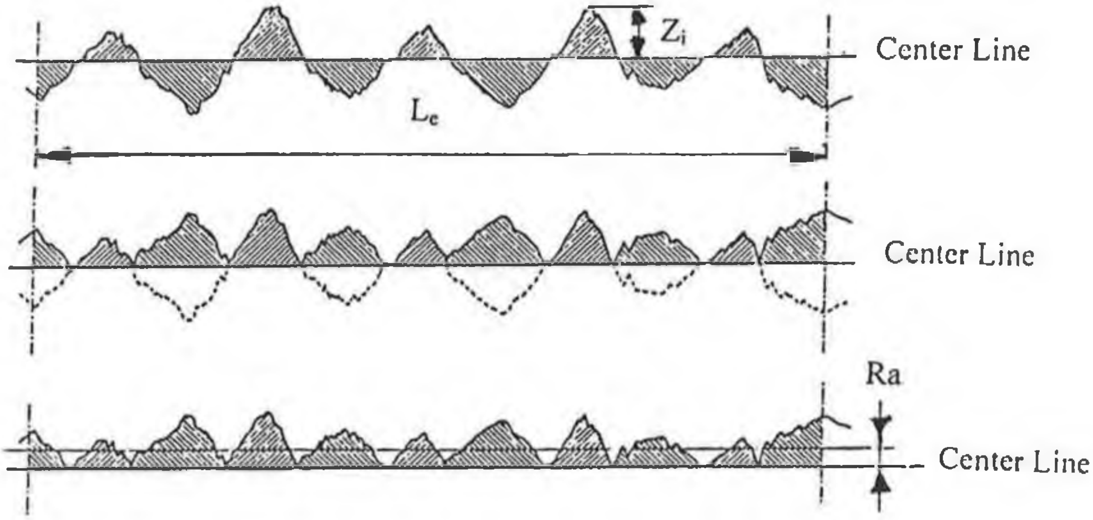
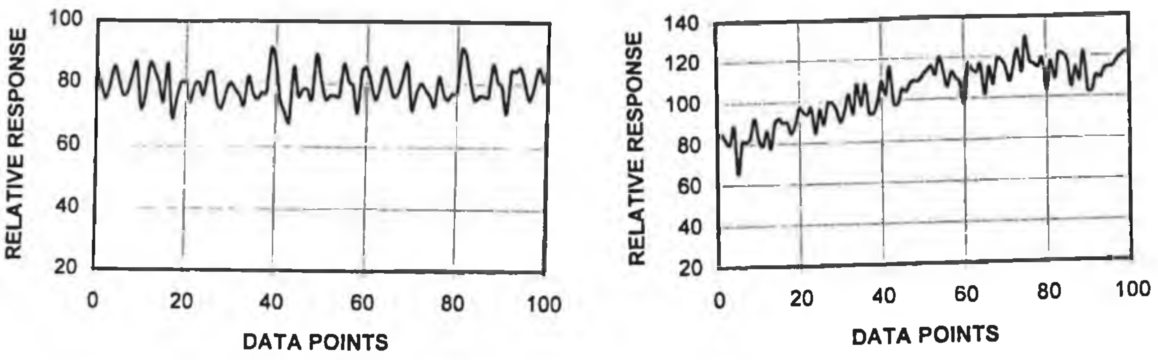
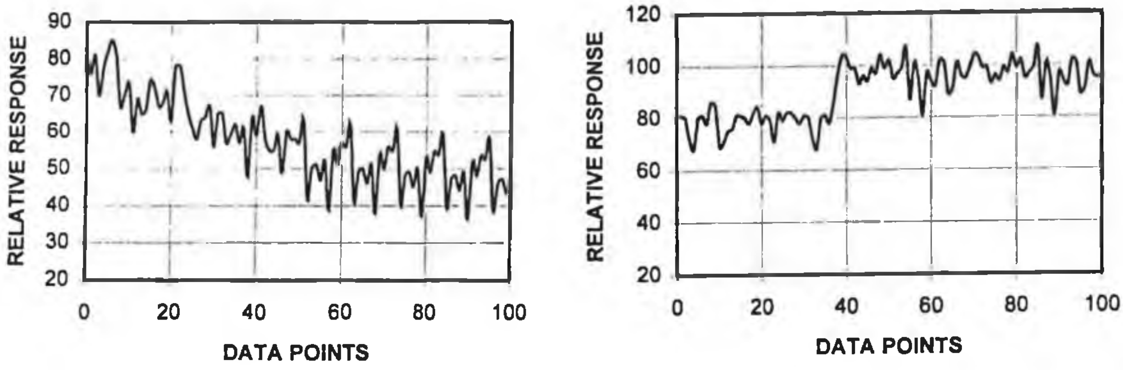


Figure 2.4 – Profile used to determine the surface roughness value (Ra).



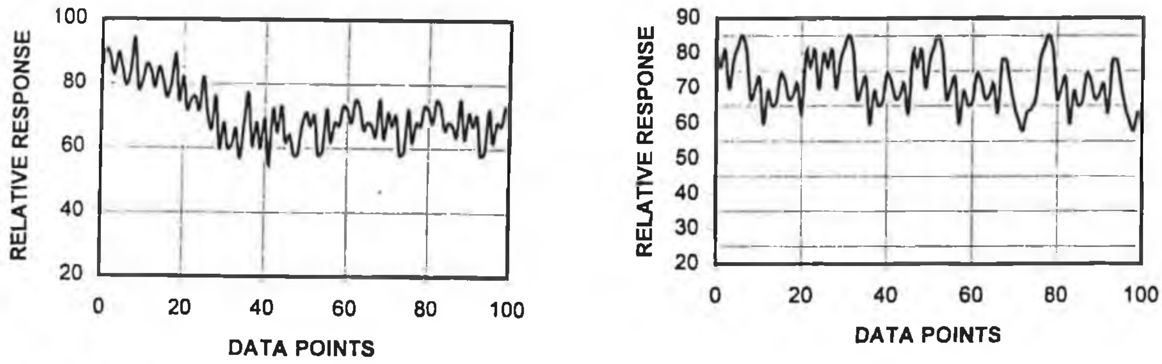
Normal Pattern

Increasing Trend



Decreasing Trend

Upward shift



Downward shift

Cyclic Pattern

Figure 2.5 - Control chart patterns.

CHAPTER 3

EXPERIMENTAL WORK

3.1 INTRODUCTION

The optical method used in the present study relies on the reflection of an incident light from the metallic surfaces. The equipment involved consists of a laser beam source, a beam splitter for the incident beam sampling, a sample holder, an optical probe for the detection of the reflected beam, and a data acquisition system. The experiments covered two different ranges of surface roughness value (Ra measurements); therefore, two optical probes were developed. The present chapter describes the experimental apparatus and relevant measurement techniques and details are presented on each under the appropriate sub-headings.

3.2 EXPERIMENTAL APPARATUS

A photograph showing the experimental apparatus and its layout is presented in figures (3.1) and (3.2) respectively. The experimental apparatus consist of mechanical, electro-optic and electronic parts, which are similar to those used in a previous study [3.1]. The mechanical parts are designed to hold the workpiece, the laser source and the optical probe, and to provide lateral motion of the workpiece movement. The mechanical parts include 10 units. These are given as follow:

Reference probe and a laser holder: This was designed to support the laser source (including prismatic lens) and the reference detector probe (figure (3.3.a, b, c)).

Attached Arc (movable): The arc is fastened by screws and moves about the main arc carrying the micrometer and fiber-optic probe (figure (3.4)).

Sample Holder: This unit is made of steel and is designed to hold the sample as it is being scanned. A micrometer precisely governs its motion and a stepper motor produces lateral displacement.

Main Arc (stationary): It has an external radius of curvature of 112 mm and a total angle of 120°. It is used to adjust both the incident laser beam angle and fiber-optic probe at appropriate positions (figure (3.5)).

Main Plate: It supports the whole design and it has dimensions of $280*150*20 \text{ mm}^3$ (length * width * thickness). It rests on four adjustable screws to insure the stability of the mechanical parts.

Threaded Screws: Two of them are installed to allow the movement of laser source and the fiber-optic probe.

Micrometer: Two identical micrometers of precision range of 0.001" are installed. They are provided with a non-rotating spindle, a ratchet stop, and a lock nut.

Threaded Fasteners: Four of them are mounted to hold the workpiece on the workpiece holder.

Washers: Two pieces are used to prevent any tilt of the workpiece.

Adjusting Screws: Four screws are fixed at the bottom of the main plate to secure the whole design from vibration and instability.

Test Specimen: Five samples of stainless steel were prepared. They were polished at different grades. The samples were 70*50*4 mm^3 (length * width * thickness).

Stepper Motor: 1.8° /step stepper motor type is used [3.2]. It has a holding torque of 460 mNm and a weight of 600 g. The motor is driven by a computer and it allows the workpiece to be scanned by a laser beam until the required displacement is achieved. The stepper motor with 4-phase construction results in higher working torque and stepping rates than those available from the permanent magnet types. It maintains very high resolution due to the small step angle. Applying the correct electrical pulse results in a 1.8° step angle rotation of the spindle (i.e. 200 steps / revolution). When correctly loaded and driven these motors can produce discrete output steps. The number of steps and speed of rotation are determined by the number of pulses and frequency of the input signal respectively.

Stepper Motor Drive Board: A uni-polar drive board (figure (3.6)) is connected to the motor which is capable of driving any 4-phase stepper motor up to 2A, 30 V D.C/ phase [3.2]. The board can be used as a stand-alone unit in conjunction with external logic. The board has a number of features; among these are:

- 4-phase uni-polar motor drive up to 2 A / 30 V D.C per phase.
- Full step and half step drive modes (for higher resolution, greater performance stability and faster stepper rates).
- On-board 12 V, 50 mA D.C output for external use.
- Drive board and the motor can share the same D.C power supply.

Amplifying Unit: Two amplifying units are used; one for the detector receiving the reference laser beam and the other for the detector collecting the reflected laser beam from the specimen surface. Each photodiode is connected to a integrated circuit instrumentation amplifier chip of an INA101 HP type, the chip is highly accurate, multistage integrated, which is designed for signal conditioning where high performance is required. The amplifying circuit is connected to A/D converter, which is interfaced with the computer by a standard PIO card. The circuit diagram of the amplifier unit is shown in figure (3.7).

Photodiode: Two photodiodes are used; one for the reflected laser beam and the other for the reference laser beam. The photodiodes are very sensitive to a visible light and the maximum forward current is 100 mA. The maximum dissipation of heat is 200 mW at room temperature (25 °C). They have a capacitance of 12 μF and a response time of 250 ns.

Laser Source: A high quality 3 mw He-Ne laser diode module consisting of a laser diode, with a wavelength of 670 nm is used. The adjustable lens provides a highly collimated beam making it ideal for precession applications such as alignment, edge detection and targeting. The spot size can be easily adjusted using the focusing tool provided. The laser can be amplitude modulated using digital or analogue signals over a frequency range of 100 kHz to 50 MHz.

Analog-to-Digital (A/D) Converter: The A/D conversion is a quantising process whereby an analog signal is represented by equivalent binary states; this is opposite to the D/A conversion process. Analog-to-Digital converters can be classified into two general groups based on the conversion technique. One technique involves comparing a given analog signal with the internally generated equivalent signals. The second technique involves changing an analog signal into time or frequency and comparing these parameters against known values. The successive approximation and the flash types are faster, but generally less accurate then the integrator and the voltage-to-frequency type converters. Furthermore, the flash type is expensive and difficult to design for high accuracy. The A/D converter of 427 E is used in the present study.

I /O Card: A lab-Tender I/O card is used [3.2]. The block diagram of the card is shown in figure (3.8). The lab-Tender card combines all the functions necessary for most data manipulations on one multifunction board. It can be programmed using a Modified Basic Code. The I/O card possesses the following; 5 independent 16 bit timer/counters, 16 lines available for external use, event counter speeds up to 5 MHz, alarm comparators on 2 counters and one shot or continues frequency outputs.

Electro-Optic Probes: Two types of electro-optic probe were designed. The first probe (probe type 1) is used to monitor the surface roughness of the workpieces for Ra ≥ 2 μm while the other probe (probe type 2) is utilised for surfaces having the roughness of 0.1 ≤ Ra ≤ 2 μm. The probes consist of fiber-optic cables and a detector system.

Probe Type 1: The fiber-optic probe consists of fiber- optic cable and photodiode. The BPX 65 photodiode [3.2] with fast response and high detectivity is used. The fiber-optic cable has a core diameter of 0.5 mm and gives attenuation of the order of 1000 dB/km [3.3]. The selection of the core diameter of the fibre cable is based on the experimentation in which the maximum probe response is obtained. The length of cable used is 1 m. 3 mW output power He-Ne laser emitting 6348 nm wavelength with 0.8 mm beam diameter is used. A lens is employed to focus the He-Ne laser beam on to the workpiece surface.

Probe type 2: This probe sees the spatial distribution of the reflected beam intensity across the reflected beam size. The probe consists of a fiber optic bundle and a linear photodetector array of fast response (4ns response time) and high detectivity (800 mV/μW/cm^2) [3.2]. The fiber optic bundle consists of 20 closely spaced multi-mode fiber cables of 150 μm in diameter. The bundle of 20 multi-mode fibre cables, each of 150 μm diameter, is selected due to the fact that this arrangement only enables to see the spatial distribution of the reflected beam intensity curve. The characteristics of the fiber cable are given in [3.3].

Stylus Equipment: The Bendix proficoder (RCL-9) linear profile system (stylus) is used to record the surface roughness and measure the average surface roughness of the workpiece. The system incorporates a meter display and a chart recorder, which show the magnified surface profile along a path traced by a stylus. The system records the total profile, roughness or waviness only, or averaged roughness of part of surfaces. This provides the basis for the development of the mathematical relationships between the Gaussian curve parameter and the average surface roughness value.

3.3 MEASUREMENT

The measurement of the surface roughness by an electro-optic system must be made using two separate techniques. This is due to the measurement limitations for the surface roughness value (Ra). In this case, when $Ra \geq 2\ \mu m$ diffusive reflection can be considered while specular reflection suitable for $0.1 \leq Ra \leq 2\mu m$ and the methods used in the present study are introduced under the relevant sub-heading.

3.3.1 BACKGROUND

The light from a source in an isotropic homogeneous medium continues to travel in straight rays and expanding wavefronts until it meets the surface of some other medium. At this surface a number of effects can occur, depending on the nature of the two media and of the surface between them. In all cases, some of the light is sent back or reflected. The reminder passes into the new medium, where some will be changed into some other form of energy or absorbed and some will continue travelling on through the medium being transmitted.

When a surface is highly polished nearly all the reflected light travels back in definite directions, as though coming from a source placed at some new position. The polished surface then acts as an aperture and limits the reflected beam. An observer in this reflected beam sees an image of the original source of the light without being aware of the reflecting surface. Such a reflection is said to be regular or specular (figure 3.9.a). When a surface is not highly polished, every irregularity of the surface will reflect light in a different direction and the light will not return as a definite beam. The surface itself will be visible from all directions, because the light is spreading out in all directions from each point on the surface. This type of reflection is said to be irregular or diffuse (figure 3.9.b) and the light is diffused or scattered, although actual reflection at each minute portion of the surface will be regular.

The amount of light reflected at a surface will vary greatly with different media. For a specular reflection, a lot depends on the quality of polish. Mirrors made of steel do not reflect as much light as those made of silver. The polished surfaces are not perfectly smooth, because of the molecular structure of materials. However, once the surface irregularities are less than the wavelength of the incident light, the reflection becomes more and more specular.

A mathematical formulation for the light scattering from rough surfaces was introduced by Beckmann [3.4]. However, this formulation is limited in applications due to the assumptions made in the analysis. These assumptions include:

i) The light beam is a coherent plane wave

ii) The scattering field of interest is far from the scattered surface (a far field or Founhofer region)

iii) The surface has continuous, gently rolling slope which prevents shadowing and multiple scattering

iv) The surface is perfectly conducting.

Therefore, because of these restrictions the mathematical analysis presented by Beckmann is not considered in the present. However, it is given in [3.4].

3.3.2 MEASUREMENT METHODS

First Method (Ra ≥ 2 μm):

The measurement of the surface profile is based on the reflection of an incident laser beam from the surface of the workpiece. The time variation of the reflected beam intensity is a function of the surface roughness of the workpiece, i.e. the measurement of the diffusive reflections.

The geometric configuration of the fiber-optic probe and its position relative to the surface being measured is very important. In the experimental tests conducted for this study the incident laser light was initially set at a 45° angle to the workpiece surface, and the fiber-cable carrying a reflected beam was also set at 45° angle to the surface. Consequently, both cables were positioned perpendicular to each other. However, this probe setting did not always produce sufficient response to be detected by the photodiode, therefore, the intensity of He-Ne laser output power was regulated using a beam attenuator. To study the effect of probe position on the measurement accuracy, the probe setting was altered. It was found that the 45° probe setting gave the maximum probe response. The experimental set up and the probe setting is shown in figure (3.2).

In order to minimise the effect of the oscillation that occurs in the incident He-Ne laser beam, a reference beam was sampled from the incident beam using a beam splitter. The ratio of the intensity of the reflected beam to the reference beam intensity was considered. This minimizes the fluctuations of the reflected beam due to any oscillation in the incident beam.

The distance between the fiber-cable end and the workpiece surface was varied to obtain maximum probe response. Once the probe response was maximised, it was set to this length. The software developed for this particular task provides real-time plots on the computer screen of the probe response and sample holder position during the measurement. With this setup complete surface profile obtained while Ra value and neural network test results classifying the resultant pattern are being documented. The experiment was repeated for the sets of stainless steel surfaces ground under different conditions.

To validate the optical measurement, a mechanical measurement of surface roughness was carried out using stylus equipment. This provided the basis for comparison of both measurement results.

Second Method ($0.1\ \mu m \leq Ra \leq 2\ \mu m$):

The principle of the measurement technique used for this surface condition relies on an evaluation of the broadness of the reflected beam intensity curve. A laser beam reflected from the surface of a workpiece will broaden as the surface roughness increases. However, the broadness of the reflected beam may be evaluated by using the standard deviation of a Gaussian function. Therefore, a relationship between the Gaussian curve parameter (Gaussian standard deviation) and the average surface roughness value may be developed.

A He-Ne laser of 3 mW power was used to illuminate the workpiece surface with the beam making an incident angle of 45° to the surface. The He-Ne laser beam is expanded to obtain a beam 6mm in diameter before focusing onto the workpiece. A fiber optic probe was employed to collect the reflected beam from the surface. The probe sees the spatial distribution of the reflected beam intensity profile. A high precision adjustable holder is used to mount the workpiece. A stepper motor with 1.8° step angle and 12 Nm holding torque is used to activate the workpiece holder. The laser beam source and the fiber optic probe are mounted on an adjustable stand, which provides for ease of alignment.

In an experiment, the He-Ne laser beam scans the surface and a fiber optic probe collects the reflected beam from the workpiece surface. At each point on the workpiece surface, the reflected beam intensity profile may be approximated by a Gaussian function. Since the data are obtained at every 1 ms, a considerable number of reflected beam profiles associated with the different points at the workpiece surface are collected (figure (3.10)). It is also known that for a given surface, there is a unique average surface roughness value [3.4]. Therefore, the data collected from each reflected beam profile, associated with each point at the surface, should be combined to produce a profile representing the average surface roughness value for that particular surface. Consequently, the reflected beam intensity profile is processed using two different methods. These methods comprise: (i) averaging the data points corresponding to each point in the intensity curve before applying a curve-fitting technique to obtain a Gaussian function constant (β) for a resulting profile, and (ii) introducing the curve-fit to obtain the constants of the Gaussian function for each intensity profile associated with each point at the workpiece surface, and then averaging these constants over the number of reflected beam intensity profiles. The surface roughness profiles and average surface roughness value (Ra) are measured initially using a Bendix proficoder (stylus) instrument. The Gaussian curve parameter is introduced when correlating the reflected beam intensity profiles with the average surface roughness values obtained from the stylus instrument. Consequently, a regression analysis is introduced to develop a mathematical relation between the Gaussian curve parameter and the average surface roughness value. Stainless steel workpieces with different surface profiles were used as workpieces. The surface of each workpiece was ground and then polished to obtain the average surface roughness values in the range 0.1 - 2 μm.

A computer based data acquisition system is used to control the equipment and store and process the data received from the fiber optic probe at every 1 ms. A flow chart of the computer program developed is given in Appendix 1. In addition, the incident He-Ne beam is sampled to closely monitor any variation in the laser beam power intensity.

Gaussian curve parameter:

A Gaussian function, as shown in figure (3.11), can be written as [3.5]:

$$f(x) = A \exp[-\beta (x - r)^2] \qquad (3.1)$$

where A, β are the constants and r is the main axis of the Gaussian function. The broadness of this function can be defined as (B^2) [3.5]:

$$B^2 = \frac{\int_{-\infty}^{\infty} x^2 f(x)dx}{\int_{-\infty}^{\infty} f(x)dx} \qquad \text{or} \qquad B = \frac{1}{\sqrt{2\beta}} \qquad (3.2)$$

B is also defined as the standard deviation of the Gaussian function. However, the Gaussian function can be normalized by dividing the function by its peak (maximum) value; in this case the constant (A), in equation. (3.1) attains a value of 1 at the peak value of the function.

3.4 DATA STORING AND ANALYSIS

The resolution of the reflected light intensity depends very much upon the profile of the reflecting surface, i.e. higher average surface roughness values correspond to a broadening of the reflected beam intensity profile. In the present case, the data from the fiber optic probe are continuously received and are stored in the computer. The set of data associated with a point and the surface consisted of 20 analog inputs at any instant of in time during the measurement, and each analog input was associated with a spatial location in the reflected beam intensity. The data are sampled every 1 ms. It should be noted that a sampling period less than 1 ms requires considerably greater computer memory capacity than used in this study. A typical data structure associated with each point on the surface is shown in Figure (3.10). For a known surface, there exists a unique average surface roughness value (Ra). Consequently, the reflected beam profile associated with each point at the workpiece surface needs to be combined to give a unique profile representing that particular surface. However, the reflected beam intensity distribution, as seen in Figure (3.10), can be approximated by a Gaussian function. Therefore, the standard deviation of the Guassian function (B) for the beam profile, associated with the each point along the surface, should be determined first. This can be achieved using a standard regression analysis (curve fitting). To combine the Gaussian profiles of the reflected beam corresponding to each point at the surface, two techniques may be used. In the first technique, the intensity values corresponding to spatial points in the intensity curves are averaged, i.e.:

$$g_n(x_i) = \frac{1}{m}\sum_{i=1}^{m} f_n(x_i) \qquad (3.3)$$

where n is the number of reflected intensity profiles corresponding to each point at the surface, m is the number of spatial locations in the intensity curve. The resulting intensity profile ($g_n(x_i)$) can be used to determine the standard deviation of the Gaussian curve parameter (B) as given in equation(3.2). This technique is called the First Technique throughout the text.

In the second technique, the B values of each reflected beam profile corresponding to respective points at the surface are determined and the average value of B over the several intensity profiles is calculated. The average B value (B_{av}) representing the reflected intensity profile of the surface is:

$$B_{av} = \frac{1}{n}\sum_{i=1}^{n} B_i$$

where n is the number of reflected beam intensity profiles.

3.5 CURVE FITTING AND ESTIMATION OF STANDARD ERROR ANALYSIS

One way of achieving an estimation of error is the use of a regression analysis to fit the data into a curve, which can be represented by a mathematical formula. A more general statement about the uncertainty that should be associated with the value of the response determined from curve fitting is that the uncertainty should reflect the uncertainty associated with data points used in the regression [3.5]. The uncertainty band around each data point represents the range in which the true value lies with 95% confidence. Therefore, a comparison of the curve fitting and the data point uncertainty limits enables an appropriate uncertainty to be associated with a value calculated from the curve fitting. However, for simplicity the standard error estimation (SEE) is used to define the limits of the curve fitting in the present study. Therefore, the SEE is calculated for the curve fitting, and the ± 2 SEE band is considered around the regression line. It should be noted that this band encloses the data point. The formula governing the standard estimation of error is [3.5]:

$$SEE = [\frac{1}{n-2}\sum_{i=1}^{n}(y_i - y_{eq})^2]^{1/2} \qquad (3.4)$$

where n is the data size, y_i is the measured response and y_{eq} is the response calculated from the curve fitting.

Uncertainty analysis:

To determine the uncertainty involved in the experiment, the bias and precision uncertainties should be considered [3.6]. The total uncertainty may be written as:

$$U = \sqrt{(S_B)^2 + (S_P)^2} \qquad (3.5)$$

where S_B and S_P are the bias and precision uncertainties respectively. The precision uncertainty can be determined using the t-statistics, in this case, the precision uncertainty is [3.7]:

$$S_P = t_{m,n} \frac{\tau}{\sqrt{q}}$$

where t is the t-statistics, τ is the sample standard deviation, q is the sample size and m, n are the degrees of freedom. In the present experiment, a 95% confidence level is considered and the sample size is 18, consequently the value of t is of the order of 2. The source of measured error may be classified as equipment and data acquisition system errors [3.8]. Therefore, the source of error may include oscillation in laser output power and variation in workpiece speed as well as errors from the fiber-optic probe and data acquisition system. The laser output power oscillation is monitored via sampling of the incident beam. The workpiece speed is determined from the stepper motor rotations. The calculation procedure for the bias and precision uncertainty analysis is not given here, but refer to [3.9]. The calculations for $(S_B)^2$ and $(S_P)^2$ rely on the following formulas:

$$(S_B)^2 = \sum_{i=1}^{n} \left(\frac{\sigma_B}{T_B}\right)^2 \qquad \text{and} \qquad (S_P)^2 = \sum_{i=1}^{n} \left(\frac{\sigma_P}{T_P}\right)^2$$

where n is the number of variations, T_B and T_P are the measured value of the source of variation under consideration, σ_B and σ_P are the bias and precision limits, and $(S_B)^2$ and $(S_P)^2$ are the squares of total bias and precision errors, respectively. The uncertainty analysis results are given in table 3.1. The total uncertainty determined from the present experiment using table 3.1 is about 0.05.

Source of Variation	**Bias Limit (σ_B)**	**Precision limit (σ_P)**
Laser output power	± 0.01 mV	± 0.01 mV
Laser power meter	± 0.1 mV	± 0.5 mV
Fiber optic probe	± 1 dB/km	± 2 dB/km
Workpiece Holder	± 0.01 mm	± 0.01 mm
Stepper motor step angle accuracy	± 0.02	± 0.03
Data acquisition system	± 0.01	± 0.02
	$(S_B)^2 = 6.06E-4$	$(S_P)^2 = 2.45E-3$

Table 3.1 – Uncertainty analysis results for the equipment used in the experiment.

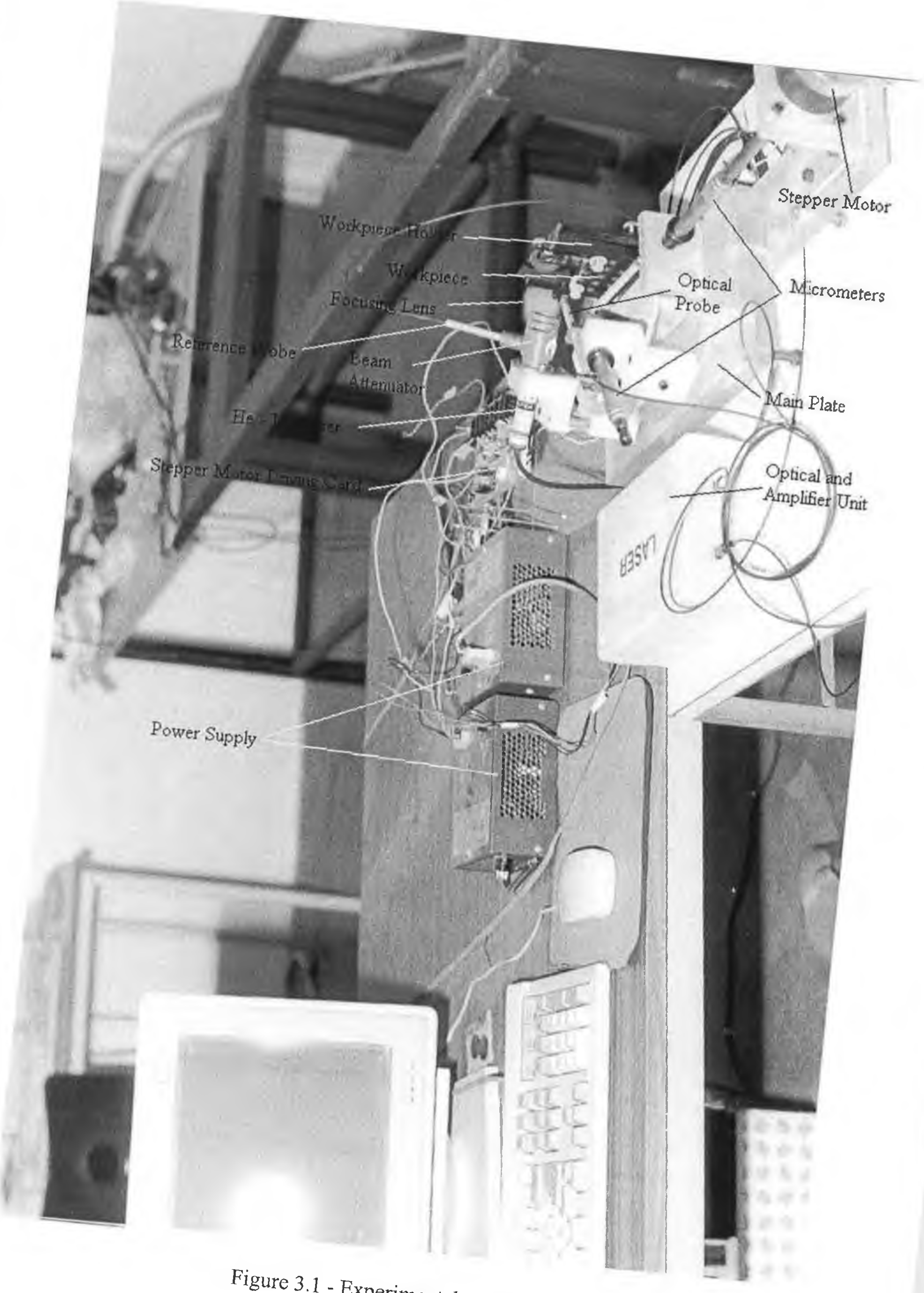


Figure 3.1 - Experimental apparatus

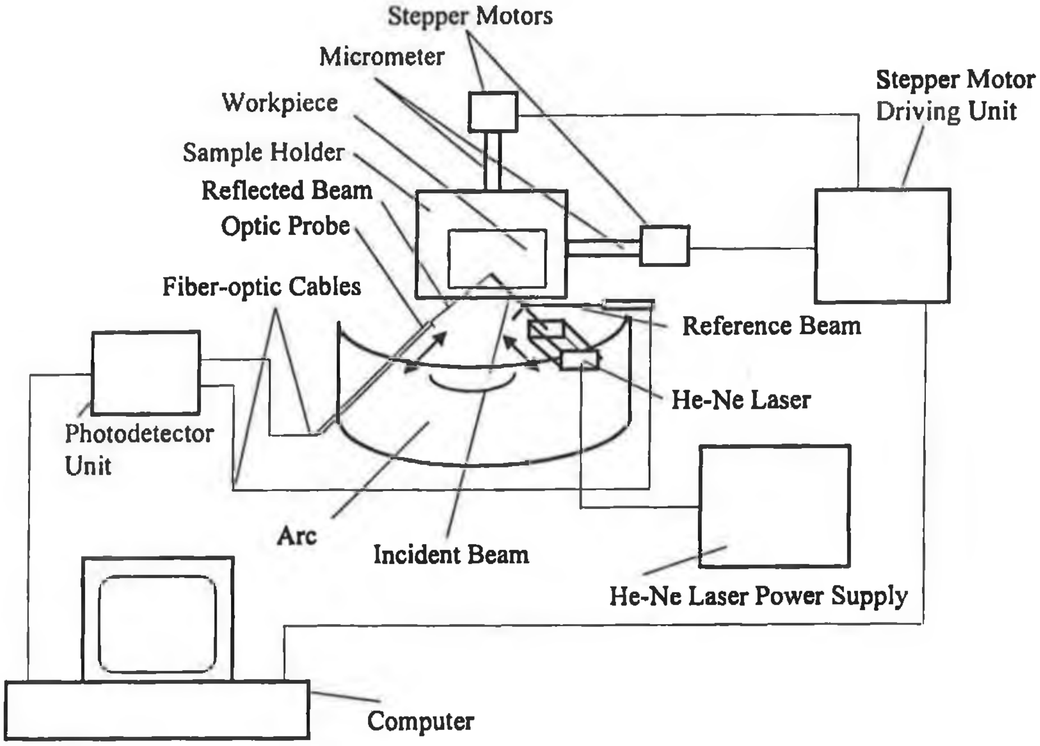


Figure 3.2 - Experimental apparatus.

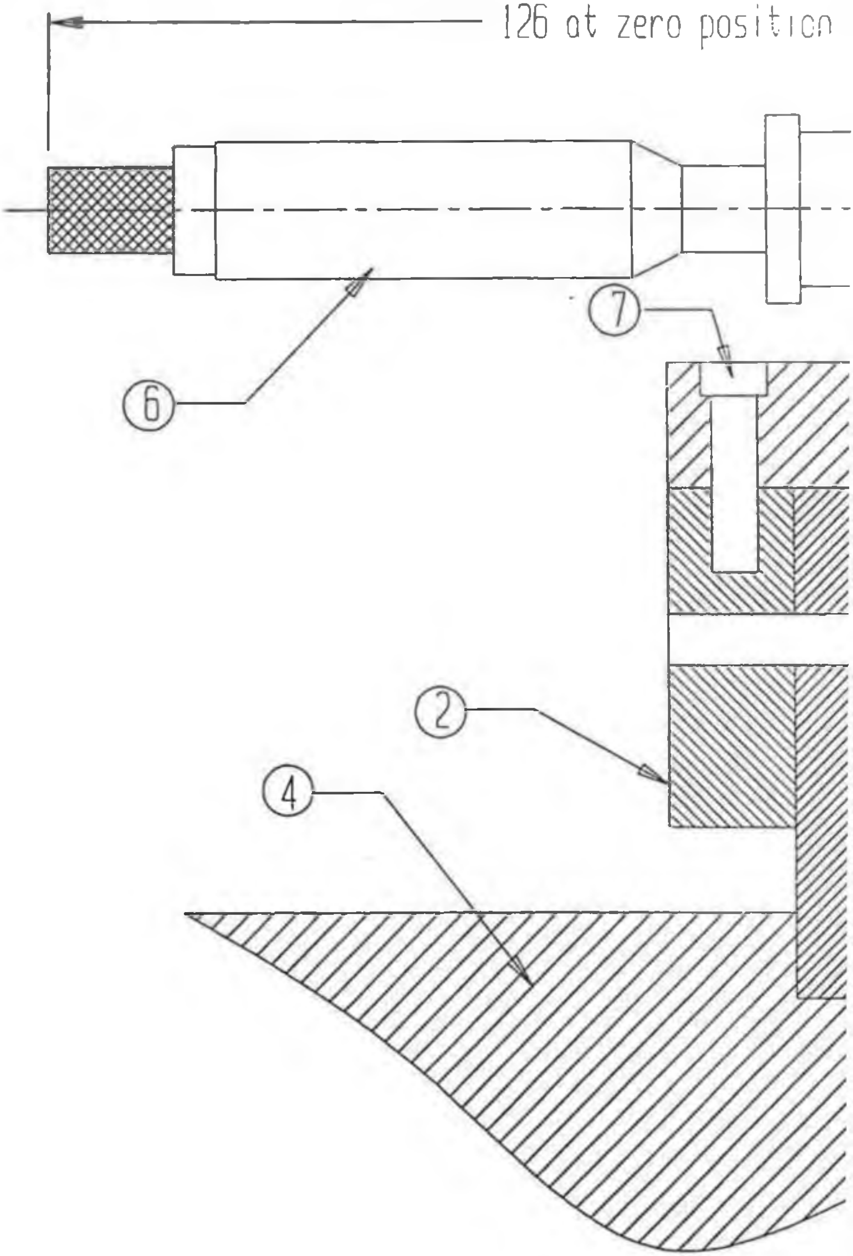
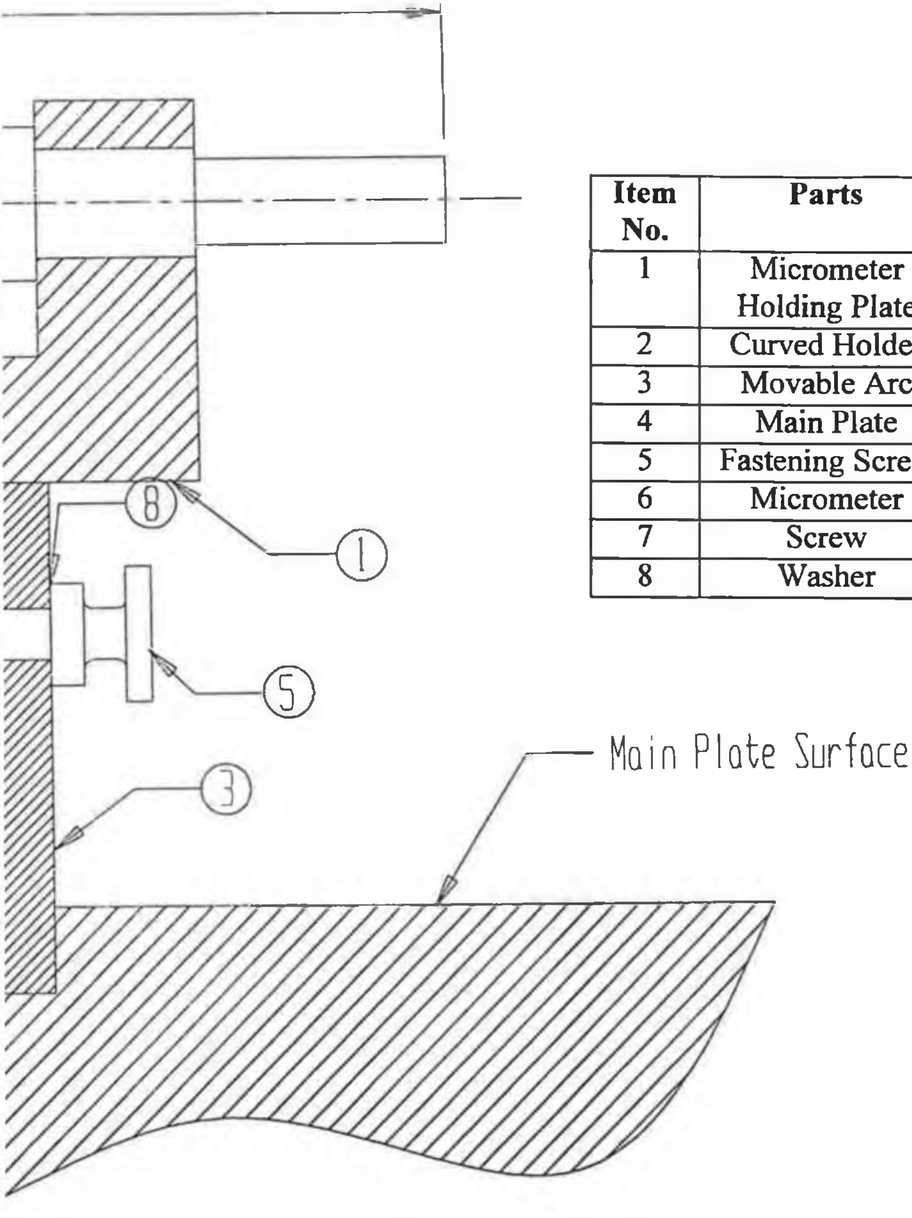


Figure 3.3.a – He-Ne laser and reference probe holder.



Item No.	Parts
1	Micrometer Holding Plate
2	Curved Holder
3	Movable Arc
4	Main Plate
5	Fastening Screw
6	Micrometer
7	Screw
8	Washer

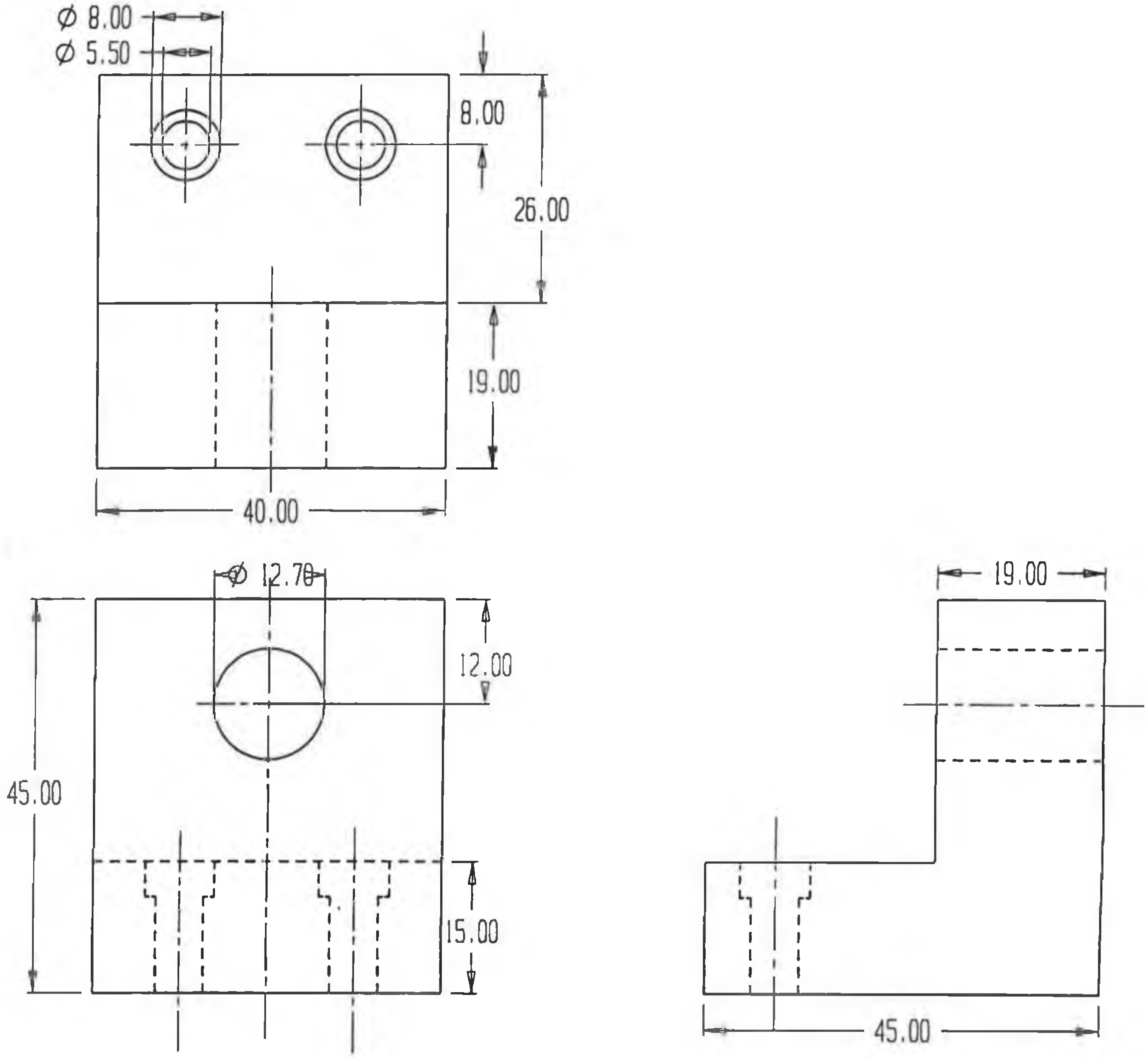


Figure 3.3.b – View of micrometer holding plate (item # 1).

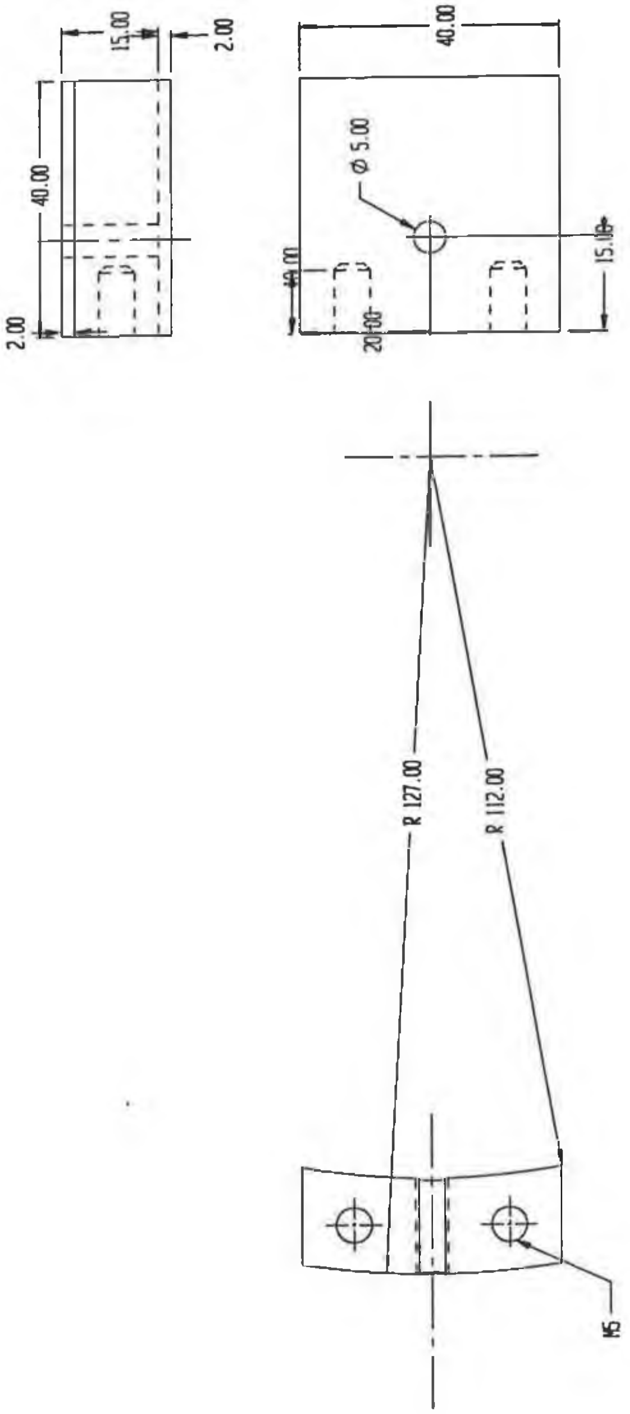


Figure 3.3.c – View of curved holder (item # 2).

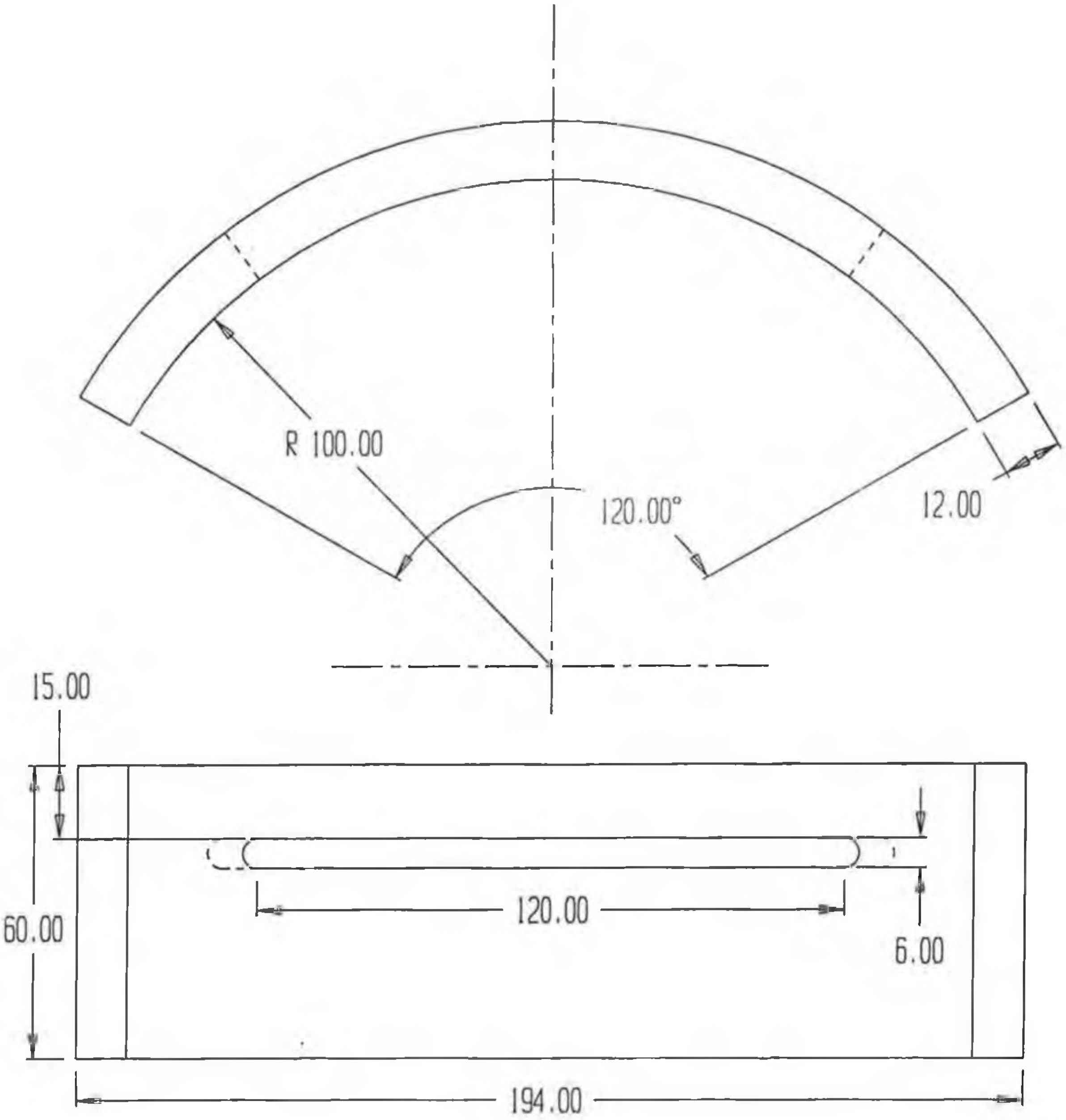


Figure 3.4 – Movable arc.

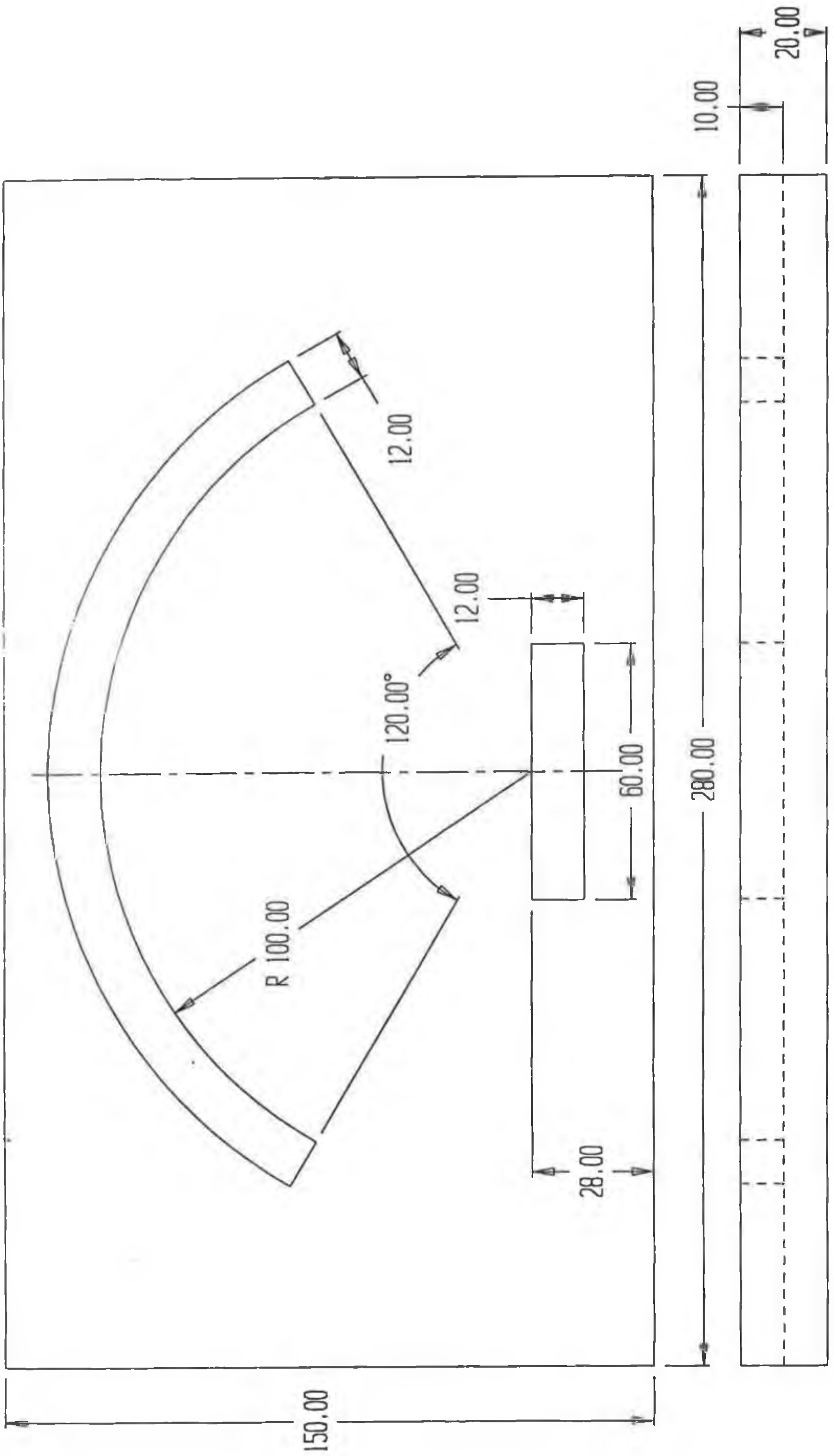


Figure 3.5 – Main arc.

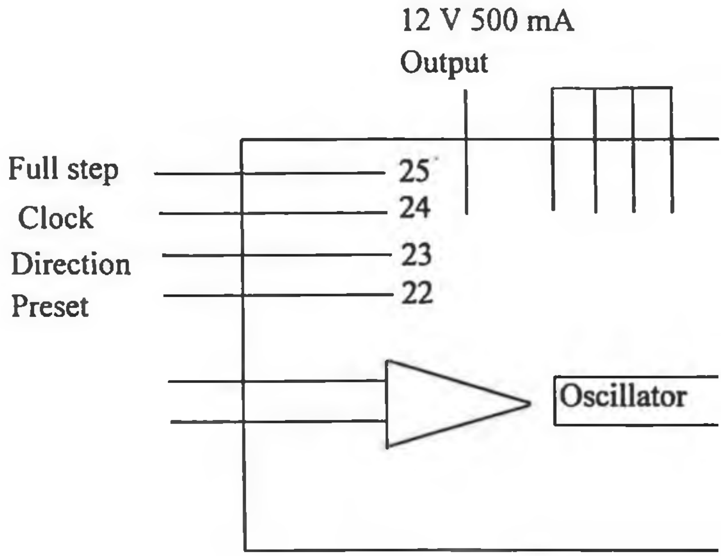
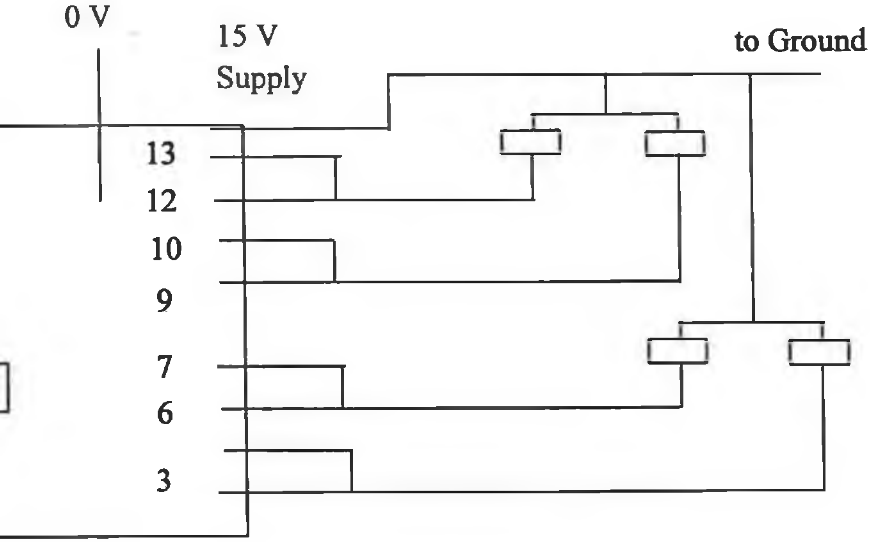


Figure 3.6 Stepper motor driving port.


0 V
15 V
Supply
to Ground
13
12
10
9
7
6
3

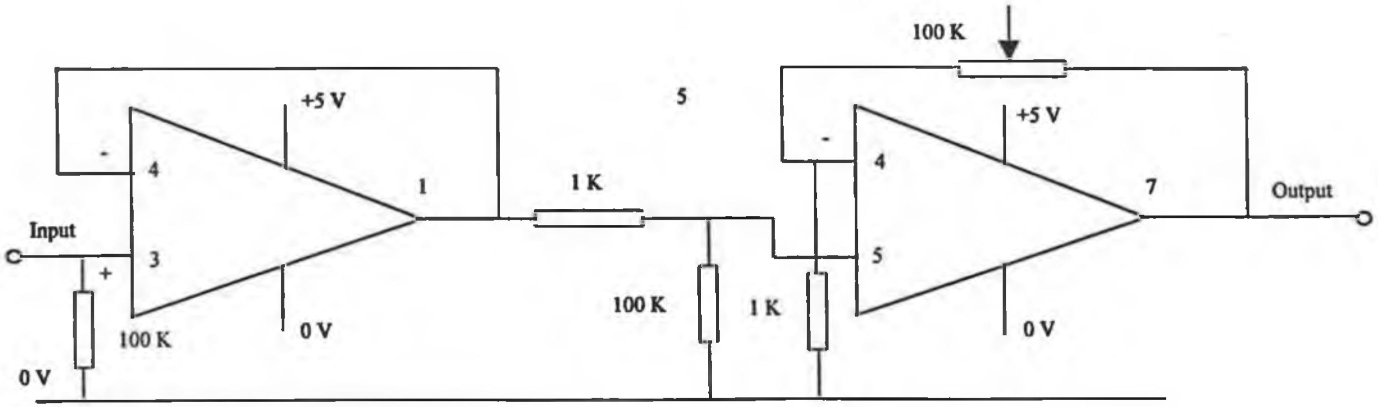


Figure 3.7 - Circuit diagram of one channel of LM 324 Instrument Amplifier.

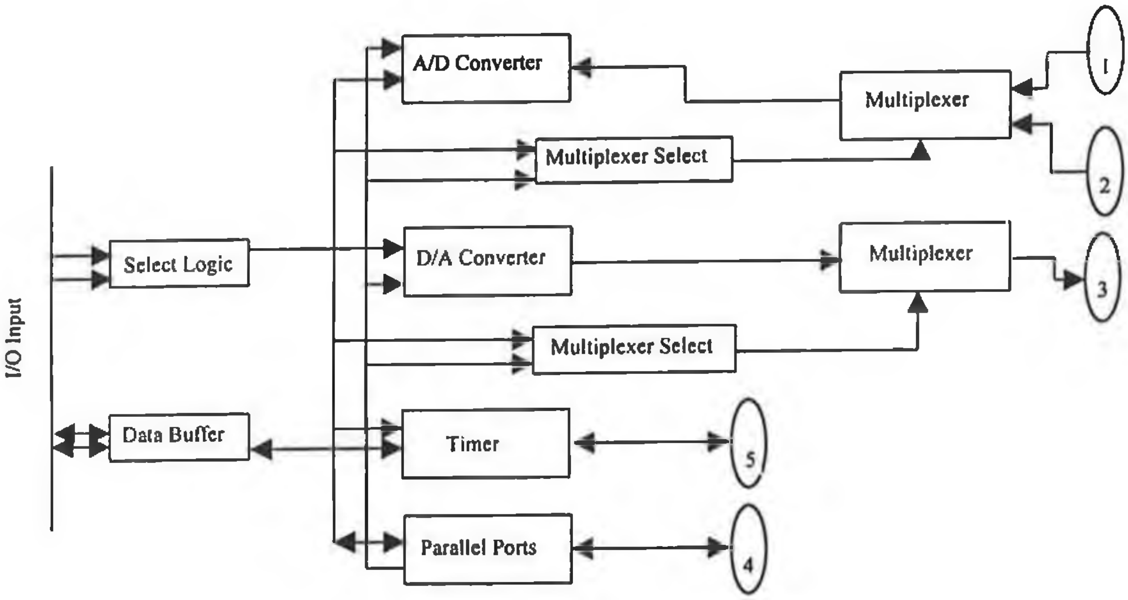
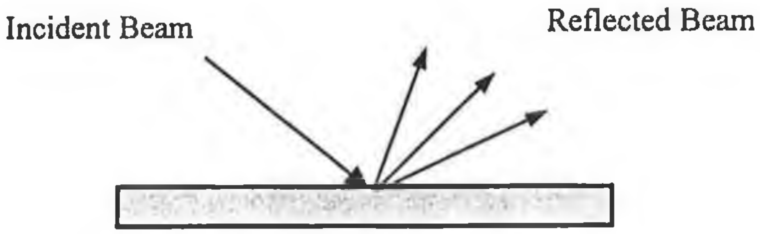
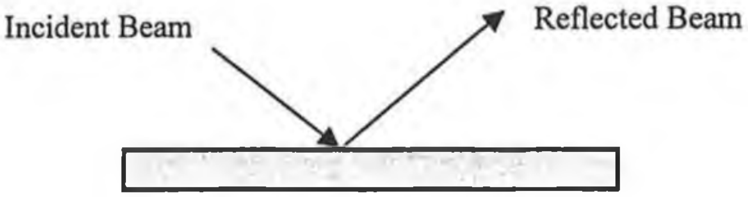


Figure 3.8 - Lab. Tender block diagram.



Diffusive Reflection



Specular Reflection

Figure 3.9 - A schematic view of the diffusive and specular reflection from the surface.

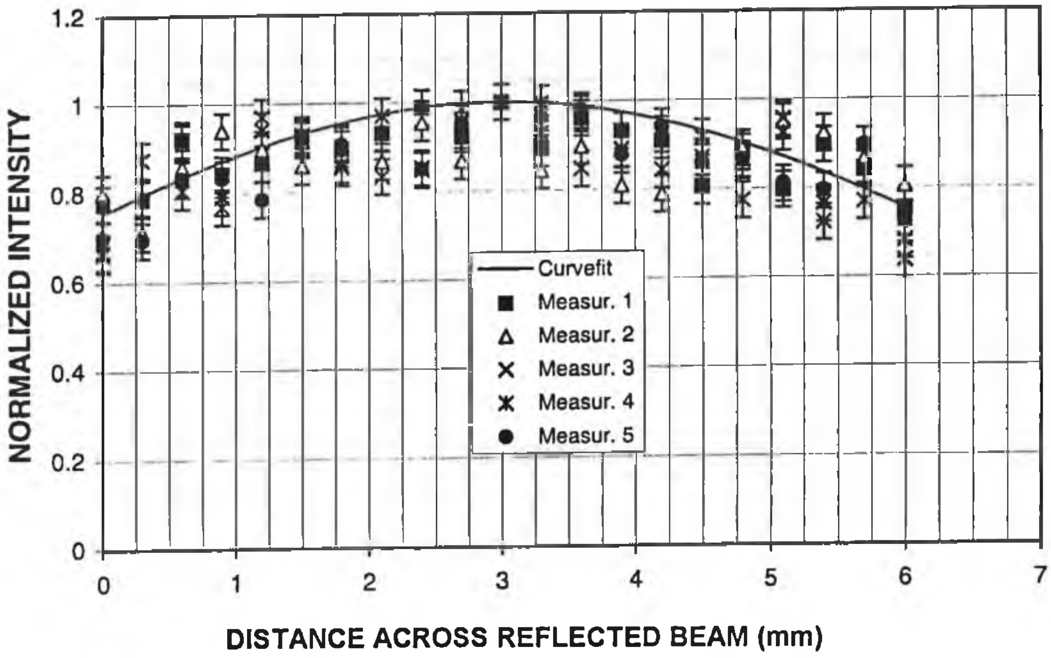


Figure 3.10 - The second probe output (normalised to maximum intensity) at five locations on the same workpiece surface and a Gaussian curve-fitting (Ra = 0.39 μm).

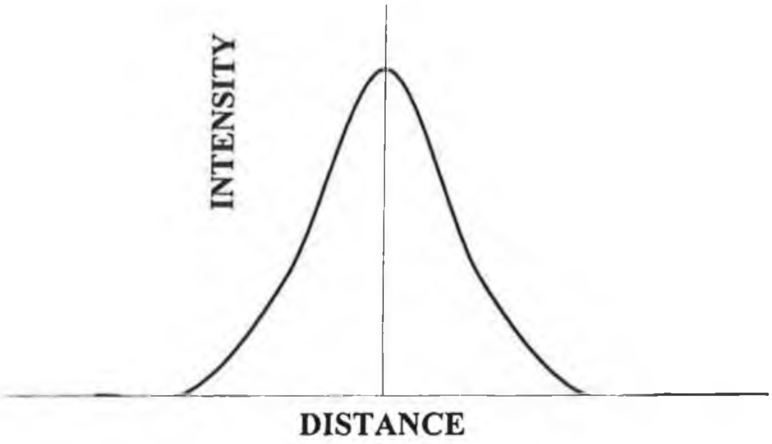


Figure 3.11 - A typical Gaussian function curve.

CHAPTER 4

NEURAL NETWORK : SIMPLIFICATION AND APPLICATON TO SURFACE PATTERN CLASSIFICATION

4.1. INTRODUCTION

Neural networks can be roughly characterized by three classes of features: their architecture or connectivity, their performance or operating rules, and their learning rules. Artificial neural networks developed over the past decade exhibit a range of features in these three areas, allowing them to be roughly grouped according to three architectural features namely; number of layers, feed-forward only or feedback connectivity and unidirectional versus bi-directional connections. Although many subdivisions are possible and, in the limit, each neural network could be considered as a single member of its own category. Novel networks and variants exhibit distinct computational behaviour, and much remains to be learned about the value of the various attributes of such networks in terms of their contribution to particular critical operational features, such as learning speed, generalization, capacity and error rate. The present chapter includes; the terminology related to neural networks, the rules governing the back-propagation learning mechanism, basic operations, network simplification through self running weight deleting process for pattern classification and architectures of the neural networks pertinent to surface roughness measurement.

4.2 HISTORICAL BACKGROUND OF NEURAL NETS

In 1943, Warren McCulloch and Walter Pitts [4.1], published a watershed paper entitled "A Logical Calculus of Ideas Imminent in Nervous Activity". This paper provided the inspiration, which helped to launch three diverse fields. One of the areas the paper influenced was an early digital computer. Neumann [4.2] saw the paper as a blue print for "electronic brains". Minsky [4.3], became enthralled with the idea of macroscopic intelligence, the birth place of expert systems. Rosenblatt, a compatriot of Minsky [4.4], became intrigued with computations of the eye. This interest led to his investigation of perception. In 1956, artificial intelligence pioneers Minsky, McCathy, Rochester, and Shannon organized the first conference on artificial intelligence. In this conference, number of interesting papers were presented. Using several hundred simulated neurons and interconnections, it might become possible to construct a system to explore how such a network would respond to environmental stimuli.

In 1962, Rosenblatt [4.5] published the first major research project in neural computing: the development of an element called a "perceptron". Rosenblatt's perceptron sparked a great amount of research interests in neural computing. The perceptron is a pattern classification system that possesses a great deal of plasticity, meaning that it could be retrained to a point of no errors after some cells had been destroyed. In addition, the perceptron is capable of making limited generalizations and could properly categorize patterns despite noise in the input. Because it is a developmental devise, the perceptron has also certain limitations. The perceptron does not have state structures and is inappropriate for description in terms of automata theory. Another major limitation, emphasized by Minsky and Papert, was the inability of the perceptron to represent the basic Exclusive OR (XOR) function. This is a result of the linear nature of the perceptron. One of the most significant changes made in the neural computing of Rosenblatt in the 60's has been complex class category [4.5]. This is typically achieved by using a non-linear transfer function, but can also arise from normalization and competition.

Rosenblatt's novel perceptron spurred research in neural computing systems until 1969, when Minsky and Papert [4.6] published their critic of neural computing and perceptron. In 1960, Widrow [4.7], developed un-adaptive linear element, called "Adeline" (Adaptive Linear Neuron) based on neuron like elements. The Adeline and a two-layer variant, the "Madeline" (Multiple Adeline) were used for a variety of applications. Including speech recognition, character recognition, weather prediction and adaptive control. Later, the Adeline was modified to produce a continuous rather than discrete output. Widrow [4.7] used the adaptive linear element algorithm to develop adaptive filters, that eliminate echoes on phone lines this was the first time a neural computing system was applied to a measure real world problem.

In the mid 1960's Minsky and Papert [4.6], began work on in-depth critique of the perceptron. The main result of the conclusion that the perceptron and neural computing were basically not "interesting" subjects to study - drastically decreased the amount of funds and therefore research that was going on in neural computer at the time was almost stop. A handful of researchers continued their work in neural computing despite Minsky and Papert's work [4.6]. One such researcher was Anderson [4.8], and his work focused on the development of a linear model, called linear associator, based on models of memory storage, retrieval and recognition. The linear associator is a parallel, distributed associative model based on the Hebbian principle that connections between neural like elements are strengthened every time they are activated. Anderson also devised an extensional linear associator, called the brain-state-in-a-box (BSB) model. Anderson describes this model: "if we start with an activity pattern inside the box, it receives positive feedback on certain components which have the effect of forcing it outward. When its elements start to limit (when it hits the wall of the box), it moves into a corner of the box where it remains for eternity. The box resides in the stage space one axis for each neuron of network and represents the saturation of each state."

Kohonen [4.9] carried out fundamental work in adaptive learning and associative memories in early 1970's. He was responsible for the description and analysis of a large class of local adaptation rules: rules in which weights are modified in a manner only dependent on the previous weight value and post and pre-synaptic values. The memories which have as special cases, auto-correlation and brain-state-in-a-box. Such memories are a formalization of the Hebbian principle that neural pathways are reinforced each time they are used.

Another researcher who continued work in neural computing in addition to Minsky and Papert was Stephen Grossberg [4.10]. Grossberg was especially interested in using neurological data to build neural computing models. He spent much of his time devising neurally inspired mechanisms in perception and memory. He translated an associative rule for a synaptic model through explicit equations. Most of his formal rules and postulates were derived from physiological studies. Grossberg has contributed a wealth of research to the design and construction of neural models. In his systems, the basic units can take only real activation values between a minimum and maximum value. One class of network he studied falls under the title of "Adaptive Resonance Theory". This involves "one shot" learning opposed to long training sessions required by some other networks. Learning and recall depend only on local information; that is, the modification of synaptic strength depends only on pre-synaptic and post-synaptic values. Knowledge already in the network is not corrupted by showing the network or noisy input data. The network self-organizes, that is, no external teacher is required to provide desired response patterns. Another important contribution Grossberg made was the proof of the Cohen-Grossberg theorem, which concerns stability recall of a general class of networks. This stability theorem is formulated in terms of an energy or Lyapunov function and shows that the response to any external stimulus in one of these networks converges to equilibrium.

In 1984, John Hopfield [4.11], presented a paper to the National Academy of Sciences. This was the first paper on neural computing presented to this distinguished body since the 1960's. In fact, the response mechanism and energy function of his network are special case of the general class of networks considered by Grossberg. Hopfield's model [4.11] represented neural operation as a thresholding operation and illustrated memory as information from many storage sites for a given input. His enthusiasm and clarity of presentation was well received and prompted researchers to become interested again in the fascinating field of neural computing. The increasing attention the field was receiving also opened up many sources of funding necessary for the pursuit of neural computing research.

There are now many research groups researching neural networks. Each group has a different emphasis and motivation with neuro-scientists and mathematicians, physicist, computer scientists relevant to the subject. One of the major research groups of recent years has been the Parallel Distributed Processing (PDP) group started by Rumelhart, McClelland and Hinton in 1982 [4.12]. Rumelhart is one of the individuals credited with the development of the back propagation technique, which is probably the most popular network for current applications of neural computing. Hinton together with Seznowski [4.13] developed the Boltzman machine, which modifies a Hopefield network in two significant ways. First, a stochastic update rule is used during recall. This enables the system to escape from local minima in the energy surface. Secondly, a learning rule is proposed which constructs distributed representations of stimuli shown to the network. Both learning and recall use a search method called stimulated annealing in analogy to the metallurgical term annealing in which low energy states of a metal are achieved by raising the metal to very high temperature and then gradually cooling it.

Levine [4.14] has designed a network called the “bi-directional Associative Memory” (BAM) network. This is a Grossberg like network in which recall is achieved by oscillation back and both between the two layers of the networks until an equilibrium state is guaranteed by means of an energy function which is shown to be reduced at each iteration of the recall process.

4.3 TERMINOLOGY

Neuro-biologists interested in how a mammalian brain works are faced with trying to understand the function of networks composed of many billions or millions of complex neurons making, in total billions or trillions of equally complex connections. Traditional theoretical approaches to this problem have emphasized the construction of abstract "top down" models in which the detailed anatomical and physiological structure of the nervous system itself is taken into consideration only to a relatively minor degree. At the same time much experimental neurobiology appears to be based on the implicit assumption that, if enough structural details are available, as understanding of brain function will naturally follow.

A neural net is a collection of simple processing elements, which are called neurons, each having limited computing capacity. Each neuron receives a number of simple scalars as input, and generates a simple scalar as output. The output of a neuron may be used as the input to many other neurons. This passing of scalars from neuron to neuron forms the only method of interaction between the neurons. Each neuron receives many inputs (in the brain this may be as many as 10,000) but produces only one output, which may in turn be used as input to many neurons. In the brain, the output of neurons passes through a synapse before being used as input to another neuron. These synapses mediate the output of the neuron (the axonic tree) and the structure collecting the input to a neuron (the dendritic tree). There are many different types of synapse in brains. One can consider the synapse to "belong" to the neuron whose dendritic tree is on. Thus, a processing element consist of: axon hillock - which does computation, axonic tree - which collects the inputs, synapses - which are the junction points between neurons. Figure 4.1 shows the neuron, axon hillock, axonic tree, dendritic tree and synapses.

The system's adaptivity is accomplished by altering the way in which the synapse mediates the transfer between processing elements. Thus learning rules are actual rules for the modification of synapses. Many different researchers used different nets and the terminology used by the researchers is almost the same.

The Synapse:

Synapses take the output of a unit and modulate it, thus providing the input to another unit. Generally, this modulation is simply multiplication by a constant. Writing Y_i for the output of unit i, and X_j for the input to unit j, the effect of such synapse can be written

$$X_j = W_{i,j} . Y_i$$

$W_{i,j}$ is frequently referred to as the weight from unit i to unit j. If $W_{i,j}$ is positive the synapse is said to be excitatory, and if it is negative, the synapse is said to be inhibitory.

The Dendritic Tree:

The dendritic tree of a neuron gathers together the inputs through its synapses. In synthetic neural units (e.g., processing elements), it is usually referred to as an activation function, since it computes the activation of the unit from its input. In some neural elements

$$A_i = \sum_i X_i$$

where A_i is the activation of the unit i.

On other occasions, this is not sophisticated enough; for example, one might like the previous activation, or one might want the inputs to the unit to interact in a more complex fashion.

The Axon Hillock:

So far the unit has done a linear mixing of its inputs. The axon hillock (as it is usually called in real neurons) or output function (as it usually referred to in synthetic neural units) computes the output from the activation level collected by the activation function. A variety of different functions have been used by different workers ranging from the pure identity function (so that the whole processing element is simply a linear mixer) to the threshold function.

$Y_i = 1$ if $A_i > T_i$

$Y_i = 0$ if $A_i < T_i$

where Y_i is the output of neurons, T_i is the threshold function, and A_i is the activation of unit i. The different processing elements produce result in nets with different properties; for example, threshold units can only produce 0 or 1 as output, whereas units using a logistic function can produce outputs between 0 and 1.

Axonic Tree:

The axonic tree provides the transport of the output of a unit from the axon hillock to the synapses. In this simple model, we shall consider that it simply passes Y_i on. A delay may be associated with this axonic transport.

4.4 BACK-PROPAGATION ALGORITHM

Back propagation is the most widely used of the neural network paradigms and has been applied successfully in application studies in a broad range of areas [4.15]. Back propagation networks are usually layered, with each layer fully connected to the layers below and above. When the network is given an input, the updating of activation values propagates forward from the input layer of the processing units, through each internal layer to the output layer of the processing units. The output units then provide the network's response. When the network corrects its internal parameters, the correction mechanism starts with the output units and back propagates through each internal layer to the input. The process repeats until the weights of the network reaches a final state where the RMS error becomes a minimum.

The back-propagation algorithm involves a forward-propagating step followed by back-propagation. Both the forward and the back propagation steps are carried out for each pattern presentation during training. In each successive layer, every processing unit sums its inputs and then applies, in general, a sigmoid function to compute its output. The output layer of units then produces the output of the network. The mathematical formulation of the back-propagation algorithm is not given here, but refer to [4.16] for full details.

As mentioned above the back propagation algorithm consists of two propagations, namely forward and backward propagations, which will be presented under relevant sub-headings.

4.4.1 FORWARD PROPAGATION

The forward propagation step is initiated when an input pattern is presented to the network. Each input unit corresponds to an entry in the input pattern vector, and each unit takes on the value of this entry. After the activation levels for the first layer of units is set, the remaining layers perform a forward-propagation step, which determines the activation levels of the other layers of the units.

Figure (4.2) illustrates the schematics of the forward-propagation step. Incoming connections to unit *j* are at the left and originate at units in the layer below. Output values from these units arrive at unit *j* and are summed by:

$S_j = \Sigma A_i W_{ji}$

where W_{ji} is the weight from unit i to unit j (unit i is one layer below unit j) and S_j is the weighted sum. After the incoming sum S_j is computed, a function f is used to compute $f(S_j)$. The function f, a sigmoid curve, is illustrated in figure (4.3.a).

The sigmoid curve is relatively flat at both ends, and has a rapid rise in the middle. There is transition from 0 to 1 that takes place when x is approximately 0 ($-3<x<3$). The sigmoid function performs a sort of "soft" threshold that is rounded (and differentiable) compared to step function (figures (4.3.b)).

The equation of the sigmoid function is:

$$f(x) = 1/(1+e^{-x})$$

Since the operand is the weighted sum of unit j (S_j), we have

$$f(S_j) = 1/(1+e^{-S_j}) = 1/(1+e^{-\Sigma a_i W_{ji}})$$

After the sigmoid function is computed on S_j, the resulting values become the activation level of unit j. This value, the output of unit j, is sent along the output interconnections. The same output value is sent along all of the output interconnections. The input layer of units is a special case. These units do not perform the weighted sum on their inputs because each input unit simply assumes the corresponding value taken from input vector.

We consider the input layer to be a layer of the network even though it does not perform the weighted sum and sigmoid calculations.

Some back-propagation networks employ a bias unit as a part of every layer, except the output layer. This unit has a constant activation value of 1. Each bias unit is connected to all units in the next higher layer, and its weights to them are adjusted during the back-error propagation. The bias units provide a constant term in the weighted sum of the units in the next layer. The result is sometimes an improvement on the convergence properties of the network.

The bias unit also provides a "threshold" effect on each unit it targets. It contributes a constant term in the summation S_j, which is the operand in the sigmoid function. This is equivalent to translating the sigmoid curve in figure (4.3.c) to the left or to the right.

4.4.2 BACKWARD PROPAGATION

Figure (4.4) illustrates the backward propagation step. Here the δ values are calculated for all the processing units while weight changes are calculated for all interconnections. The calculations begin at the output layer and progress backward through the network to the input layer.

The error correction steps take place after a pattern is presented at the input layer and the forward-propagation step is completed. Each processing unit in the output layer produces a single real number of its output, which is compared to the target output specified in the training set. The basic dynamics of the back-propagation algorithm is shown in figure (4.4). After forward propagation, the target pattern is compared to the output pattern. The error values, denoted by δ, are calculated for the output layer. In the next step, the weights are adjusted. Further, δ values are calculated for the hidden layer. After this calculation, its incoming weights are adjusted. Based on the difference in error value calculated for each unit in the output layer, then, the weights are adjusted for all of the interconnections that go into the output layer. Later, the weights are adjusted for all interconnections that go into the hidden layer. The process is continued until the last layer of weights has been adjusted. δ_j value for the output layer is:

$$\delta_j = (t_j - A_j) f'(S_j)$$

where t_j is target value for unit j, a_j is the output value for unit j, $f'(x)$ is the derivative of the sigmoid function f and S_j is the weighted sum of input to unit j. The quantity $(t_j - A_j)$ reflects the amount of error.

The error value of δ_j for a unit in a hidden layer can be calculated as:

$$\delta_j = [\sum_k \delta W_{kj}] f'(S)_j$$

In this case, a weighted sum is taken of the δ values of all units that receive from unit j.

The adjustment of the connection weights is carried out using δ values of the processing units. Each connection weight is adjusted by taking into account the δ value of the unit that receives input from that interconnection. The connection weight adjustment is carried out as [4.17]:

$$\Delta W_{i,j} = \eta \delta_j A_i$$

Where, η is the learning rate, which has a value in between 0.25 – 0.75 [4.17].

The size of the weight adjustment is proportional to δj, the error value of the target unit. Thus a larger error value for unit j results in larger adjustment to its incoming weights.

4.5 A SELF-PRUNING THRESHOLD APPROACH FOR NETWORK PRUNING

The pruning of superfluous connections may be achieved through two different techniques, which are introduced for automatic weight deleting process in the present chapter. In the first technique, the weight ratios, which are obtained by dividing the values of weight by the maximum weight in the layer, are compared with the threshold value. Once the weight deleting process starts, new values of the weight ratios are determined for every training iterations and these ratios are then compared to threshold value, therefore, ratios less than the threshold value are deleted. In this case, the number of weights deleted changes during each training iteration. This provides dynamic weight deleting process. In the second technique, The weight ratios less than the threshold value are disregarded and weights corresponding to these ratios are deleted. Once the weights are deleted, these weights are kept as deleted weights (set to zero), throughout the learning process.

In order to develop a new model for automatic thresholding in weight deleting process, variation of RMS error with number of training is taken into account by the author. In this case, once the RMS error drops to desired level, weight deleting process starts. However, to avoid meeting the local minimum trap of the RMS error, derivative of RMS error (DRMS) is also considered such that weight deleting process starts when both RMS error and DRMS error drop simultaneously to the desired values. The present chapter is extended to include two different data structures. This provides the ability to investigate the effect of data structure on the network generalization capacity. In the first data structure, the coefficients and random variables, as defined in chapter 3, are varied. In the second data structure, the coefficients of control chart equations are left constant while random variables are changed. In this case resulting data become different in nature, therefore, data corresponding to the same pattern also show a different appearance. Consequently, (6 x 100 x 20) input data were developed to train the network and (6 x 100 x 20) data are developed to test the resulting network. In training and testing processes, both data structures are employed independently.

4.5.1 MATHEMATICAL ANALYSIS

The mathematical analysis carried out by the author for the weight pruning process is given below.

The output of the network (O_k) may be written as:

$$O_k = \frac{1}{(1 + e^{-S_{2,k}})} \qquad (4.1)$$

or

$$\frac{\partial O_k}{\partial t} = \frac{e^{-S_{2,k}}}{(1 + e^{-S_{2,k}})^2} \frac{\partial S_{2,k}}{\partial t} \qquad (4.2)$$

and

$$S_{2,k} = \sum_j A_j W_{2,j,k}$$

where, where j and k are the number of neurons in the hidden and output layers respectively, a_j is the neuron activation in the hidden layer and $W_{2,j,k}$ is the weight matrix between the hidden and output layers.

Therefore:

$$\frac{\partial S_{2,k}}{\partial t} = \sum \frac{\partial W_{2,j,k}}{\partial t} A_j + \sum \frac{\partial A_j}{\partial t} W_{2,j,k} \qquad (4.3)$$

but,

$$A_j = \frac{1}{(1+e^{-S_{1,j}})}$$

and

$$S_{1,k} = \sum_i b_i W_{1,i,j}$$

Where i is the number of neurons in the input layer, b_i is the input vector and $W_{1,i,j}$ is the weight matrix between the input and hidden layers.

$$\frac{\partial A_j}{\partial t} = \frac{e^{-S_{1,j}}}{(1+e^{-S_{1,j}})^2}\frac{\partial S_{1,j}}{\partial t} \qquad (4.4)$$

When equation 4.4 is substituted in equation 4.3, it yields;

$$\frac{\partial S_{2,k}}{\partial t} = \sum\frac{1}{(1+e^{-S_{1,j}})}\frac{\partial W_{2,j,k}}{\partial t} + \sum\frac{e^{-S_{1,j}}}{(1+e^{-S_{1,j}})^2}\frac{\partial S_{1,j}}{\partial t}W_{2,j,k} \qquad (4.5$$

or

$$\frac{\partial S_{2,k}}{\partial t} = \frac{1}{(1+e^{-S_{1,j}})}[\sum\frac{\partial W_{2,j,k}}{\partial t} + \sum\frac{e^{-S_{1,j}}}{(1+e^{-S_{1,j}})}\frac{\partial S_{1,j}}{\partial t}W_{2,j,k}] \qquad (4.6$$

When equation 4.6 is inserted into equation 4.2, it yields;

$$\frac{\partial O_k}{\partial t} = e^{-S_{1,j}} O_k^{\,2} A_j [\sum \frac{\partial W_{2,j,k}}{\partial t} + \sum e^{-S_{1,j}} \frac{\partial S_{1,j}}{\partial t} W_{2,j,k}] \qquad (4.7)$$

When learning is completed $\frac{\partial O_k}{\partial t}$ approaches to 0 or iteration time approaches to ∞. In accordance with equation (4.7), the following should be satisfied:

$$\sum \frac{\partial S_{1,j}}{\partial t} \rightarrow 0 \qquad \text{and} \qquad \sum \frac{\partial W_{2,j,k}}{\partial t} \rightarrow 0$$

However, RMS error of the iteration process can be determined from;

$$RMSerror = \sqrt{[\frac{1}{n} \sum_{k}^{n} [O_{Target}(k) - O(k)]^2]}$$

Where O_{target} is the target output vector. When $\frac{\partial O_k}{\partial t}$ approaches to 0, therefore, $\frac{\partial (RMSerror)}{\partial t}$ should approach to 0, since O_{target} is constant. Consequently;

$$\lim_{t \rightarrow \infty} \frac{\partial (RMSerror)}{\partial t} \cong 0$$

It is evident from the above expressions that in a weight pruning process, the mechanism which minimizes the RMS error value needs to be considered. In this case the threshold value, based on the selected maximum weight ratio, minimizing the RMS error in a weight pruning process may be introduced in between input/hidden and hidden/output layers. Consequently, the threshold values resulting in the minimum RMS error may be selected as the threshold value for the weight pruning mechanism. The details of the pruning mechanism are given in the following section.

4.5.2 NETWORK TOPOLOGY USED IN THE OPTICAL PATTERN CLASSIFICATION

A neural network using a back-propagation algorithm was developed in the present study. The topology of the network includes input, hidden and output layers, i.e. a three layered network. To develop input vectors for the neural network, the data file obtained from the optical measurement for each workpiece surface was divided in 10 sub-data files such that each sub-data file contains 100 data points, i.e. the number of neurons in the input layer is fixed as 100. In the second step, the type of output used should be decided. Consequently, a six digit binary vector was used as output relevant to the number of workpiece surfaces, i.e. six processing elements take part in the output layer of the back-propagation network. The process of determining the number of hidden layer elements was carried out on a trail-and-error basis. If the number of processing elements is high in the hidden layer, the network will lose its ability to generalize, but there is a small number of processing elements in the hidden layer, the network will not be able to learn. As a consequence of this, it was found that 53 processing elements in the hidden layer gave an improved performance. Each connection weight was initialized to a small random number. A data file was developed mathematically to train the network. The network was, then, trained and tested to recognize the input vectors under noisy or uncertain conditions.

A computer program for a back-propagation neural network was developed to recognize the input vectors relevant to each surface and interpret the output of the network. Consequently, the network developed had 100-53-6 neurons structure in input-hidden-output layers (figure (4.5)).

The performance of the network can be defined as:

$$\text{Performance} = \frac{\text{Number of Patterns Correctly Classified}}{\text{Total Number of Tests for the Pattern}} \text{x}100$$

Therefore, all the data obtained from the optical method was tested and a performance analysis was obtained.

4.5.3 PRUNING MECHANISM

The aim of this analysis is to establish a model, which classifies the control chart patterns while allowing automatic thresholding in a weight deleting process. Two different techniques are introduced here to accomplish the automatic weight deleting process for network simplification. In the first technique, new weight ratios are determined for every training iteration once the weight deleting process starts i.e. new weight ratios are performed dividing the weights by their maximum value. In this case, the number of weights deleted changes during each training iteration. This provides a dynamic weight deleting process. This technique is referred to as deleting process-I- (dynamic deleting) throughout the text. In the second technique, once the weights are deleted, these weights are kept as deleted weights (set to zero) throughout the training process. This technique is referred to as deleting process-2- (static deleting) throughout the text.

To start the weight deleting process, the weight ratios are computed by dividing the values of the weights in each layer to value of a maximum weight corresponding to that layer, i.e.:

$$TRW1 = \frac{W_{1,i,j}}{W_{1,\max}} \qquad 0 \le TRW1 \le 1$$

where $W_{1,i,j}$ refers to the weights in between the input/hidden layers, $W_{1,\max}$ is the maximum value of the weights in between these layers, and TRW1 is the weight ratios in between input/hidden layer.

However, for the weights in between hidden/output layers, the weight ratios (TRW2) may be written as:

$$TRW2 = \frac{W_{2,i,j}}{W_{2,\max}} \qquad 0 \le TRW2 \le 1$$

where $W_{2,i,j}$ is the weights in between the input/hidden layers and $W_{2,\max}$ is the maximum value of the weights in between these layers.

The computed weight ratios are then compared to the threshold value assigned ($0 < TR_{1,2} < 1$). Therefore, weight ratios smaller than this threshold value are disregarded and corresponding weights are deleted (set to zero) during the training process, i.e. static weight deleting process (Delete-2 process), or change in these deleted weight occur according to new computed weight ratios during the training process, i.e. dynamic weight deleting process (Delete-1 process). To make the weight deleting process self-pruning, the assigned threshold value is divided by 10 obtaining a threshold increment:

$$\Delta TR= TR_{1,2}/10$$

where $TR_{1,2}$ is an assigned threshold value. The network is initially trained without employing the weight deleting process up to the level where both the RMS and DRMS error values meet the desired values. Once this condition is reached the weight deleting process is initiated, i.e. a triggering condition to start the weight deleting process. In this case, the number of training to meet this condition is called as triggering training number (TTRN) and the resulting weights in the network are saved to utilize in a later stage of the weight deleting process. Initially, the network is trained without any weight deleting to twice the triggering train number (iteration) (TTRN) and the resulting RMS and DRMS errors are recorded. Later, the network is then initialized to obtain the triggering condition (TTRN condition) at which both the RMS and DRMS errors meet the desired levels. Following this, the threshold value is set to equal ΔTR while training the network to twice the triggering training number which corresponds to twice TTRN. After this training period, the resulting RMS and DRMS errors are recorded. Similarly, the network is then initialized to obtain triggering condition (TTRN condition). As a next step, the threshold value is increased gradually by ΔTR:

$(TR_{1,2})_{new} = (TR_{1,2})_{old} + \Delta TR$

The network is then trained employing a new TR value up to the training number that equals to twice of TTRN and resulting RMS and DRMS errors are recorded. In a similar way, the above procedure is repeated for the other thresholds which increase gradually by an amount of ΔTR, and resulting RMS and DRMS errors are recorded. Consequently, the RMS and DRMS errors of the network corresponding to each of the threshold values (ΔTR, 2ΔTR, 3ΔTR,....9ΔTR, 10ΔTR) and RMS error value corresponding to no deleting process are compared. The threshold value resulting in the smallest RMS error than those results from No-Deleting process is selected as the best threshold value. If the threshold value selected is equal to the threshold value assigned then it is increased by an amount equal to ΔTR and new weight deleting process starts from the triggering training number. On the other hand, if the threshold value resulted is equal to ΔTR of the threshold value assigned then the new threshold value is assigned as less than ΔTR and new weight deleting process starts from the triggering training number with this new assigned threshold value.

To determine the desired levels of the RMS and DRMS errors initiating the weight deleting process, a number of tests are carried out. When the desired levels of RMS and DRMS errors are assigned as less than 10^{-2} and 10^{-3} respectively, early starting of weight deleting process occurs. Consequently, poor network performance is obtained. Alternatively, when RMS and DRMS errors desired levels are assigned as greater than 10^{-5} and 10^{-6} respectively, in this case, weight deleting process starts late and this gives rise to performances being very close to those obtained for No-deleting process. Therefore, in the present study, RMS and DRMS errors desired levels are assigned in the range 10^{-2} - 10^{-5} and 10^{-3} - 10^{-6} respectively (Table 4.I). A flow chart of the computer program developed for the weight pruning mechanism is given in appendix 2.

Source of Variation	Triggering RMS	Triggering DRMS
Level 1	1.0E-02	1.0E-03
Level 2	1.0E-03	1.0E-04
Level 3	1.0E-04	1.0E-05
Level 4	1.0E-05	1.0E-06

Table 4.1 - Triggering error root mean square (RMS) and derivative of error root mean square (DRMS).

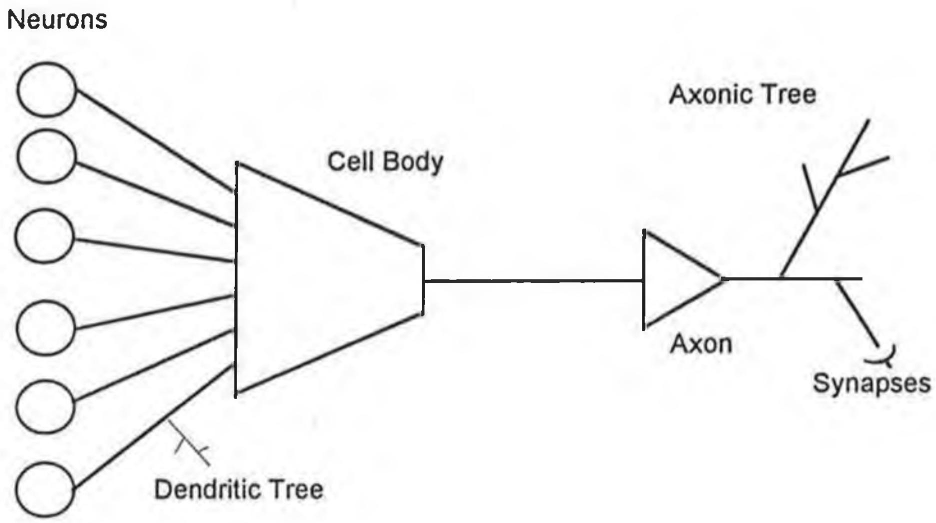


Figure 4.1 - A schematic view of neuron connections.

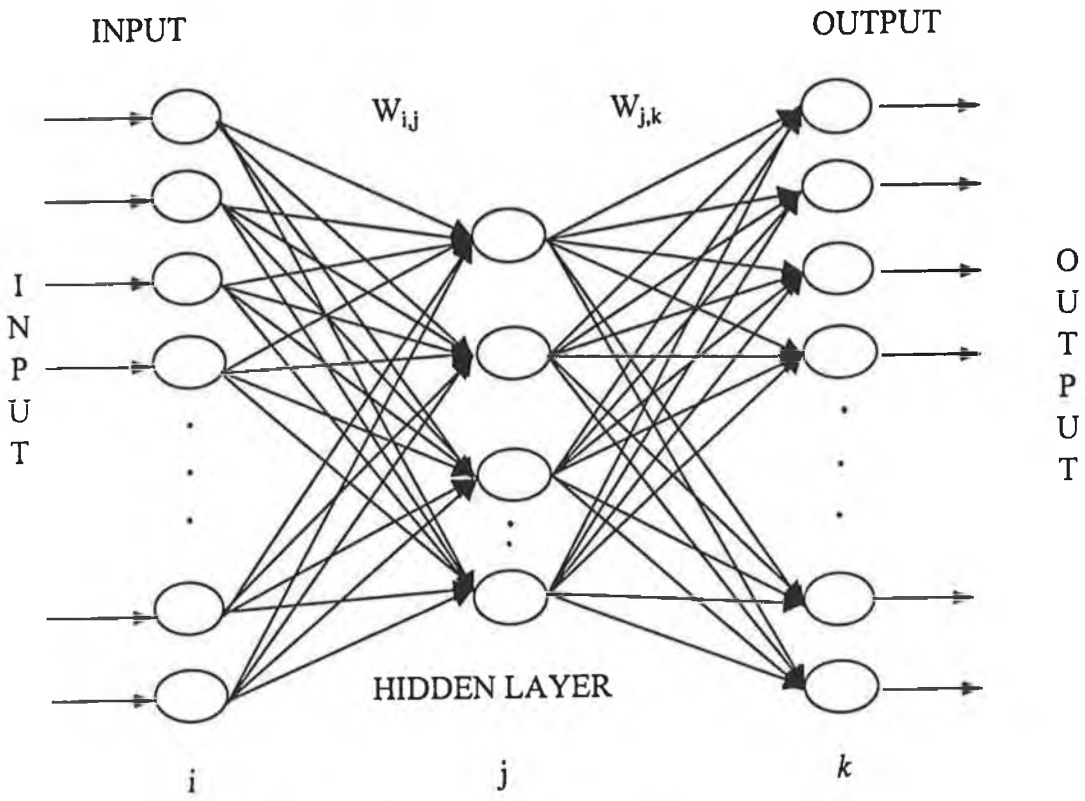


Figure 4.2 - Forward propagation step.

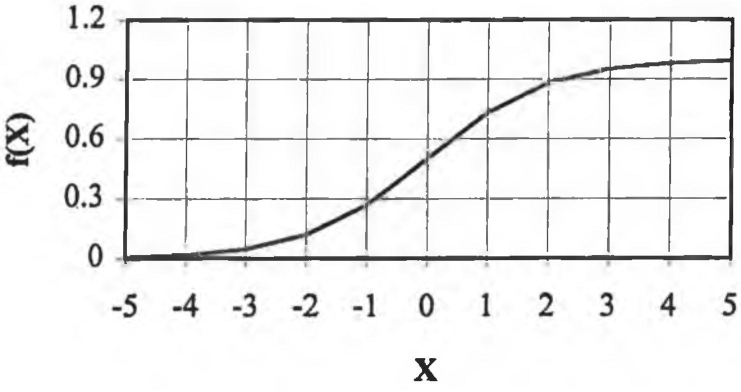


Fig. 4.3.a - Sigmoid function.

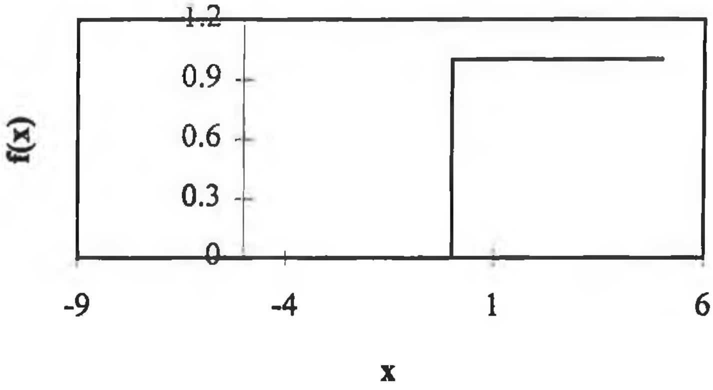


Figure 4.3.b - Step function.

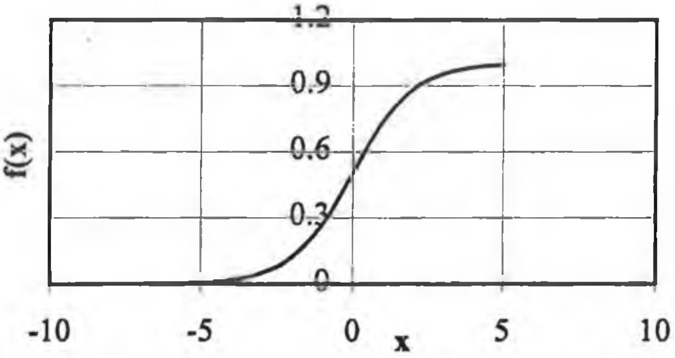


Figure 4.3.c - Translating sigmoid function

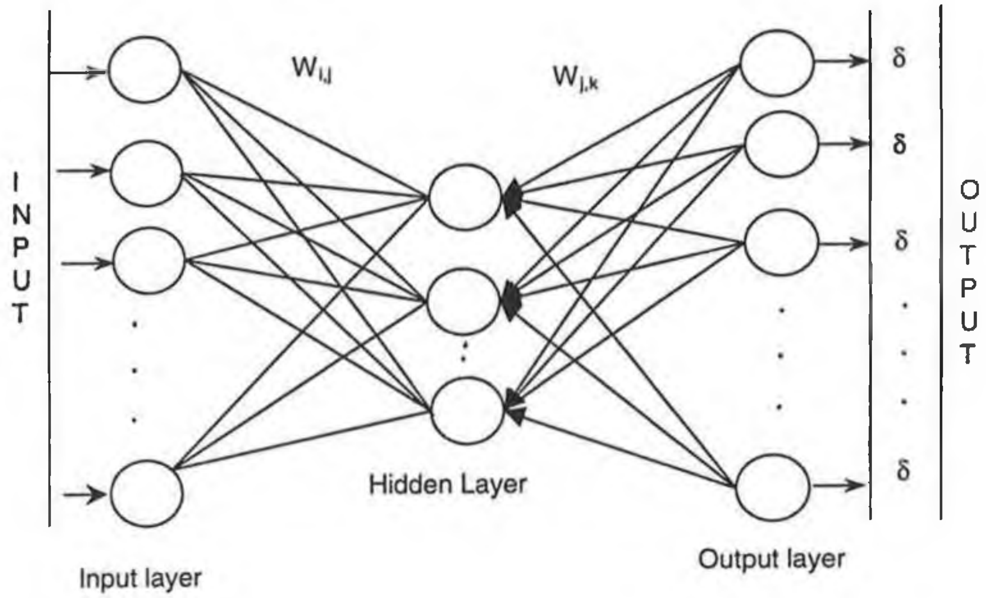


Figure 4.4 - Back-propagation step

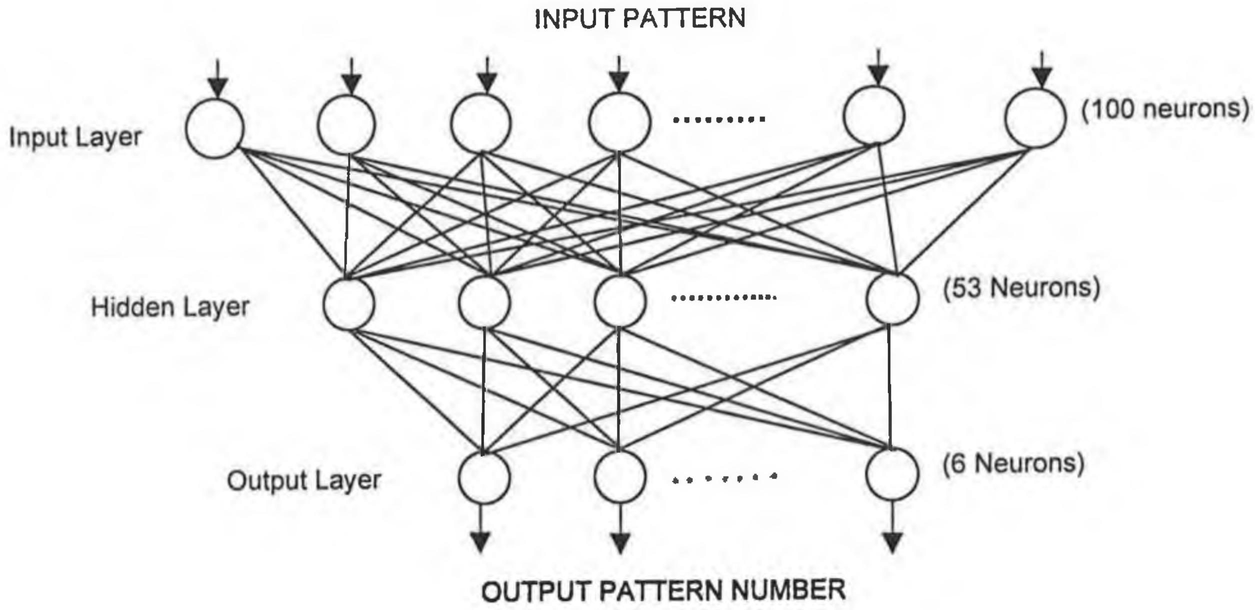


Figure 4.5 - Schematic view of back-propagation network used in training and testing the surface patterns.

CHAPTER 5

RESULTS AND DISCUSSIONS

The results and discussions relevant to the present work will be presented under a number of sub-headings. These include; results and discussion of network simplification through weight pruning, results and discussion of surface roughness measurement for Ra ≥ 2 μm, and results and discussion of surface roughness measurement for $0.1 \leq$ μm Ra $2 \leq$ μm.

Results and Discussion of Network Simplification through Weight Pruning

In order to relate the results obtained from the automatic weight deleting process with the variables employed, the following definitions are introduced:

Percentage of Delete Start:

This can be defined as the ratio of number of training iterations when the weight deleting starts to the total number of training iterations.

$$\text{Percentage of Delete Start} = \frac{\text{Number of Training Iteration when Deleting Starts}}{\text{Total Number of Training Iterations}} x100$$

Triggering RMS Error Level:

It is the value of the RMS error where the first weight deleting starts.

Triggering DRMS Error Value:

The value of the change of RMS error in consecutive training iteration, i.e.

$$\text{DRMS Error} = (\text{RMS})_{N+1} - (\text{RMS})_{N}$$

where N is the training iteration at which weight deleting starts.

DELTA RMS Error Value:

It is defined as the difference in RMS Error values corresponding to deleting and no deleting processes at any instant of training time, i.e.:

$$\text{DELTA RMS Error} = (\text{RMS})_{\text{No-Deleting}} - (\text{RMS})_{\text{Deleting}}$$

Percentage of Weight Deleted:

It is the ratio of total number of weights deleted in each layer to the total number of weights in the corresponding layer. Therefore:

$$\text{Percentage of Weight Deleted} = \frac{(\text{Number of Weights Deleted})_{\text{Layer}(1,2)}}{(\text{Total Number of Weights})_{\text{Layer}(1,2)}} \text{x}100$$

In figure (5.1.a), the variation of selected thresholds with triggering RMS error is shown for both Deleting-1 and Deleting-2 processes. Fluctuation in the selected threshold occurs with triggering RMS errors due to the randomness involved in the learning process.. The general trend is that the selected threshold increases as the triggering RMS error value decreases. This indicates that the late start of the deleting process corresponds to a low triggering RMS error value causes the network to select a high threshold value. Variation of selected threshold values corresponding to second data kind with triggering RMS error is show in figure (5.1.b). Similar discussion can be made for the variation of selected thresholds as made for figure (5.1.a). However, when comparing selected threshold values corresponding to Delete-1 and Delete-2 processes, it can be observed that both thresholds have the similar values. However, in the case of first data kind, they differ significantly and Delete-2 process results in relatively high threshold values (Figure (5.1.b)). This may be due to that in Delete-1 process weights are deleted dynamically, i.e. (each number of iteration) number of weights deleted varies with learning iteration, consequently, network selects relatively low threshold values which minimizes the RMS error.

In figure (5.2), variation of percentage of delete start with triggering RMS error is shown. Percentage of delete start drops rapidly with increasing triggering RMS error value. This may indicate that automatic weight deleting starts earlier at higher triggering RMS error value. The percentage of delete start is higher in first data type than that corresponding to second data type. This may again be related with data structures.

In figure (5.3), variation of percentage of weight deleted with triggering RMS error value is shown for two deleting processes in which first data kind was used when training the network. The general trend of the curves shows that decreasing triggering RMS error results in increased percentage of weight deleted in the first and second layers. However, percentage of weight deleted in the second layer is higher than that corresponds to first layer. When comparing Figures (5.3.a) and (5.3.b), the percentage of weight deleted for both layers is higher in Delete-2 process than that obtained from Delete-1 process. This may be due to the fact that in Delete-2 process, once the weight-deleting threshold is obtained, it is kept constant throughout the training process. Therefore, some contributing weights to learning may be omitted in the late training stages.

In Figure (5.4.a), percentage of weight deleted corresponding to second data kind with triggering RMS error value is shown for two layers (input and hidden layers). Similar discussion may be made as it is made for the first data type. Low triggering RMS error results in high percentage of weight deleted. When comparing Figures (5.4.a) and (5.4.b), it is clear that percentage of weight deleted in Delete-1 process is higher than that corresponding to Delete-2 process. The percentage of weight deleted in the first layer is relatively higher than that occurs in the second layer for Delete-2 process, however, this is opposite for Delete-1 process. This may again be due to the effect of mechanism of weight deleting process.

In figure (5.5), variation of RMS error corresponding to first data kind with number of training iteration is shown for No-Delete, Delete-1 and Delete-2 processes. It should be noted that the overlapping of the curves corresponding to different triggering RMS errors occurs in figures (5.5.b) and (5.5.c). In general, oscillation in RMS error occurs after about 2400 training iteration and amplitude of this oscillation is higher for training iteration in the range 12000 - 30000. This oscillation may be due to the structure of input data. In this case, data is developed both varying the random variables and coefficients in equations 2.1, 2.2, 2.3, and 2.4.). However, when the training iteration increases the amplitude of this oscillation reduces. When examining the figures (5.5.a), (5.5.b) and (5.5.c)). Delete 1 and Delete-2 processes have similar effects on the resulting RMS error.

To accomplish the differences between No-Delete and Delete-1, and No-Delete and Delete-2 processes (resulting Delta RMS error values) figures, (5.6) and (5.7) are plotted. In figure (5.6) variation of Delta RMS error corresponding to first data type with number of training iterations is shown for Delete-1 process. The negative values of RMS error representing those RMS errors corresponding to any of deleting process is lower than that corresponds to no deleting process. Delta RMS error shows the fluctuations (figure (5.6.a), (5.6.b), (5.6.c) and (5.6.d)). When this fluctuation is negative; in this case the network learns faster than No-Delete process. This indicates that for a certain type of data generated, the network learns at a faster rate thereby resulting in a considerable drop in RMS error. Since varying the random variables together with parameters govern the data, RMS error increases when the network input pattern data differs than the previous data. In addition, early triggering of deleting process results in network to learn at a fast rate for some patterns and at a slow rate to others when compare to No-Deleting process. In this case, oscillation occurs in Delta RMS error. In the case of Delete-2 process, Delta RMS error oscillates with training iteration. In this case, similar discussions can be applicable as it is made for figure (5.6). When comparing figures (5.6) and (5.7), it may be seen that Delete-1 process results in lower Delta RMS error (higher in negative magnitude) than that corresponding to Delete-2 process. This suggests that the late start of deleting process is fruitful in learning mechanism and Delete-1 process makes the network able to learn at a faster rate than the case corresponds to No deleting process. Consequently, when examining figures (5.6.a), (5.6.b), (5.6.c) and (5.6.d), it is evident that RMS error corresponding to triggering RMS error level 4 (figure (5.6.d)) is lower than that corresponding to other triggering RMS error levels. This shows that network trained employing Delete-1 process and having triggering RMS error level 4 learns faster than the network corresponding to No-Deleting process.

Figure (5.8) shows the variation of RMS error due to second data kind with number of training iteration for Delete-1, Delete-2 and No-Deleting processes. When comparing figures (5.5) and (5.8), the oscillation in RMS error is minimal in the case of second data kind. This may be due to the structure of data, i.e. it is only the random variables which are changed while keeping the coefficients constant in equations 2.1, 2.2, 2.3, and 2.4.

In figure (5.9), variation of network performance corresponding to first data type with pattern number is shown. Network performance corresponding to Delete-1 process gives very close results to that obtained from No Deleting process. The performance of Delete-1 process increases for the case of pattern number 5. On the other hand, it decreases for pattern number 4. This variation may be due to the effect of data structure on the learning mechanism. Similarly, figure (5.10) shows the variation of performance, corresponding to second data kind, with pattern number for Delete-1, Delete-2, and No-Deleting processes. It is evident that some small variation occurs in performance and this variation is large for pattern numbers 2 and 5, i.e. performance corresponding to third triggering RMS error level (TRG. RMS 3) results in relatively lower performance than No-Deleting process. When comparing figures (5.9) and (5.10), both deleting processes (Delete-1 and Delete-2) give close results to that obtained from No-deleting process. It should be noted that due to the most of the points have the same values for No-Deleting and TRG. RMS 4, the corresponding curves are overlapping. However, Delete-1 process shows slightly better performance than Delete-2 process. This may be due to one or all of the following facts: i) - In Delete-1 process, weight-deleting process occurs dynamically. As a result, number of weights deleted at each training iteration changes, i.e. the number of non-contributing weights reduces gradually after each training iteration. This provides network to settle dynamically with progressing training iteration. ii) - Keeping the weights deleted as the same throughout the training process may constraint the learning mechanism, i.e. weights deleted initially may become important in the later stages of learning mechanism.

It is found by the author that the dynamic self-pruning process improves the performance of the network and minimizes the network structure.

Results and Discussion of Surface Roughness Measurement for Ra ≥ 2 µm

To calibrate the probe response relative to the workpiece surface the probe position was varied from the surface of the workpiece. The initial probe position was set 3 mm away from the workpiece surface, and the selection of this position was based on the probe response. It was found that a distance closer to the surface and less than the initial probe position did not effect the probe response. A gold plated mirror was used as the reflecting workpiece surface.

The variation of the probe output (intensity ratio of reflected beam to reference beam) with distance from the workpiece surface is shown in figure (5.11.a) when the sample holder surface is parallel to the workpiece surface. The probe output is almost constant up to the distance about 12 mm away from the workpiece surface. However, as the distance increases beyond 12 mm away from the gold plated surface, the intensity ratio decreases. When examining the case at which the attenuator is installed in the path of incident beam, the probe response does not alter considerably from the non-attenuator case. This indicates that the effect of the beam attenuation is negligible on the probe response noting that only 8% of the incident beam are attenuated. On the other hand, some fluctuations are evident in the probe response, especially, when the probe is close to the workpiece surface. This may be due to the noise in the measurement system as well as the environmental noise.

The probe response, when the workpiece is tilted at -5°, is shown in figure (5.11.b). The similar behavior of the probe response occurs, as it is the case in figure (5.11.a). However, the probe response decays after 11 mm from the workpiece surface. As it is depicted from figure (5.11.b), probe response drops further for –5° at shorter distance than that corresponding to no-tilt case. This may be due to the occurrence of divergence of the reflected beam after tilting the sample with an angle of -5°.

When the sample is tilted 5°, the probe response remains constant up to a distance of 9 mm from the workpiece surface. (figure (5.11.c)). Once the probe position is changed to a distance greater than 9 mm from the surface, sudden drops in the probe response is resulted. The removal of attenuator has no effect on the behavior of the probe response at a distance closer to the surface. As the distance increases, early decay of probe response is resulted when the attenuator is installed. This may be due to the different attenuation behavior of the cone of the reflected beam, since the cone angle of the reflected beam changes with tilting.

Figure (5.12) shows the surface profiles obtained for 100 data points from the optical method for different workpiece surfaces while figure (5.13) and (5.14) show the detailed surface profiles obtained from both optical and mechanical (stylus) measurements for a particular sample surface. It should be noted that the spatial resolution of the data points in the case of optical method is considerably high which is 600 data points/cm of workpiece surface. When comparing figures (5.13) and (5.14), the variation in surface profiles is more apparent in the case of optical measurement. Consequently, it is expected that the optical method provides more precision measurement than that obtained from the stylus measurement.

Table 5.1 gives the surface roughness (Ra) values obtained from both types of measurements. Ra behaves almost the same for both measurements, except its values. It should be noted that Ra values measured by the optical method are dimensionless, since the intensity ratio (reflected beam intensity/reference beam intensity) is taken into account. Nevertheless, the qualitative agreement in both sets of results is evident.

Figure (5.15) shows the performance of the neural network when tested using the mathematically developed test patterns. The network gives high performance of learning for all the patterns developed mathematically. On the other hand, when testing the actual surface profiles, the network classifies the patterns with high performance. However, the network gives mixed performance for all the patterns (figure (5.16)). In this case, the resulting surface profile for the workpieces 1, 2 and 3 may appear similar to the patterns of normal, cycle and upward shift. The surface profiles show more towards to the normal pattern, since the network performance for normal pattern is higher than those corresponding to other patterns. In addition, the performance of the network is higher in the case of normal pattern when testing the surface profiles of workpieces 1, 2, and 3. Consequently, surface profiles corresponding to different workpieces differ in appearance and this may be easily classified by the neural network developed.

Results and discussion of Surface Roughness measurement for $0.1\ \mu m \leq Ra \leq 2\ \mu m$.

The results obtained from the present study are presented according to the workpiece number. Due to the lengthy and similar arguments, the reflected beam intensity profiles obtained for locations at the surface of the workpiece no.1 are discussed in detail. However, only the resulting intensity profiles associated with the first and second methods are discussed for all the workpiece surfaces.

Figure (5.17) shows the surface profiles obtained from the Bendix proficoder instrument for six workpiece surfaces. The appearance of regular pattern is due to the partially polishing process. However, figure (5.18) shows the normalized intensity distribution of the reflected beam obtained for four locations at the surface of the workpiece no.1. The data points scatter slightly in the region close to the peak intensity. This may occur because of the irregular structure of the surface profile of the workpiece and the error related to measurement system. However, care is taken during the experiment and this case is tested several times. In the experiment similar scattering is observed at every trial. When introducing the curve fitting technique, it is observed that the data best fit in a Gaussian curve provided that small scattering occurs around the peak intensity, i.e. the reflected beam is of a Gaussian form with small distortions. When examining figures (5.18.a) and (5.18.b), the data points in Gaussian curves, associated with different locations at the surface, vary slightly. It should be noted that each curve in these figures corresponds to a point at the surface of the workpiece no.1 and they are sampled from the data recorded at 1 ms periods. The β values obtained after curve fitting differs for each intensity profile. Consequently, variation in β values can be attributed to surface roughness of the workpiece.

Figure (5.19), shows the curve fitting obtained from both methods for workpiece no. 1. The β values obtained from both methods are almost the same with slight differences. A small scattering of data points in the region close to the peak intensity is apparent. However, similar observations can be made for the other workpieces as shown in figures (5.20), (5.21), (5.22), (5.23), and (5.24). It should be noted that all the data points and curve fitting based on the averaged β value are shown in figure (5.20.b - 5.24.b) while averaged data and a resulting curve fitting is shown in figure (5.20.a - 5.24.a). When comparing these figures, it may be seen that the degree of scattering increases as the β value decreases. This occurs especially for workpieces no. 4, 5, and 6. In this case, the surface roughness profile changes considerably along the workpiece surface. In addition, as the surface roughness increases, the broadness of the reflected intensity increases, therefore, the detected size of the reflected beam reduces, i.e. only the central region of the reflected beam is seen by the fiber optic probe. In this case, the curve fitting may not be accurate due to the degree of data scattering in this region.

Figure (5.25) shows the β and standard estimation of error (SEE) with workpiece number provided that each workpiece number represents the different surface roughness value. When comparing figures (5.25.a) and (5.25.b), it may be observed that the first method, in which the data are averaged then the curve fitting is introduced to find β value, results in higher SEE values than that corresponding to the second method. It should be noted that β values are obtained using a curve fitting technique for each intensity distribution curve associated at each point at the surface. The β values are averaged, and then a new curve fitting is introduced using the averaged β value, i.e. the curve resembling all the intensity profiles is obtained. This is also evident from table 5.2 in which SEE values for β and workpiece number are given. However, SEE increases slightly as the workpiece number increases. This is due to the data being scattered because of the relatively irregular surfaces. Consequently, the use of the second method is more accurate than the first method, since it represents the data in a better order. The maximum error involved in the second method is of the order of 10% for relatively rough surfaces and less than 5% for relatively smooth surfaces.

Figure (5.26) shows the variation of average surface roughness value (Ra) with the standard deviation of Gaussian function (B). The B values used in the curve result from the second method where the SEE values are minimum. However, the B with error curve includes the B and the SEE values together. Consequently, the influence of SEE on the average surface roughness can also be observed. The Ra varies almost linearly with B values. Moreover, the influence of SEE on the Ra is minimal when the Ra value is small. Therefore, the equation related to Ra with B is in a linear form, which is:

$$Ra = 0.08369B + 0.022774 \qquad (5.1)$$

However, error related to equation (5.1) increases as Ra increases. This is due to the scattering of the data points across the reflected beam intensity profile. This suggests that as the surface roughness increases further, the degree of diffusive reflection increases, which in turn distorts the reflected beam intensity profile considerably. Therefore, the applicability of the present measurement is limited to the reflected beam intensity profile. On the other hand, it is apparent from table 3.1 that the uncertainty (U, where $U=\sqrt{(S_B)^2+(S_P)^2}$) involved in the present study is of the order of 5% , which is acceptably small. Consequently, the present measurement gives good results for the surfaces of low Ra values, since for high Ra values the reflected beam profile is distorted considerably due to the degree of diffusive reflection involved.

Workpiece Number	**1**	**2**	**3**	**4**	**5**
Optical Measurement	0.10149	0.086421	0.066143	0.072682	0.069563
Stylus Measurement (μm)	7.61	6.22	5.1	5.64	5.66

Table 5.1. Ra values obtained from optical and stylus measurements for different workpiece surfaces (the values obtained from the optical measurement are normalised).

	First Scheme		Second Scheme	
	Averaged β	**SEE**	**β**	**SEE**
Workpiece 1	0.54	0.035204977	0.5	0.045391188
Workpiece 2	0.2245	0.052521387	0.222	0.055366704
Workpiece 3	0.125898	0.062542036	0.125	0.06888684
Workpiece 4	0.055899	0.116070617	0.0555	0.074341871
Workpiece 5	0.0252	0.1423	0.025	0.074401493
Workpiece 6	0.03125	0.1546	0.031	0.084329148

Table 5.2 - SEE and β values with different workpiece surfaces.

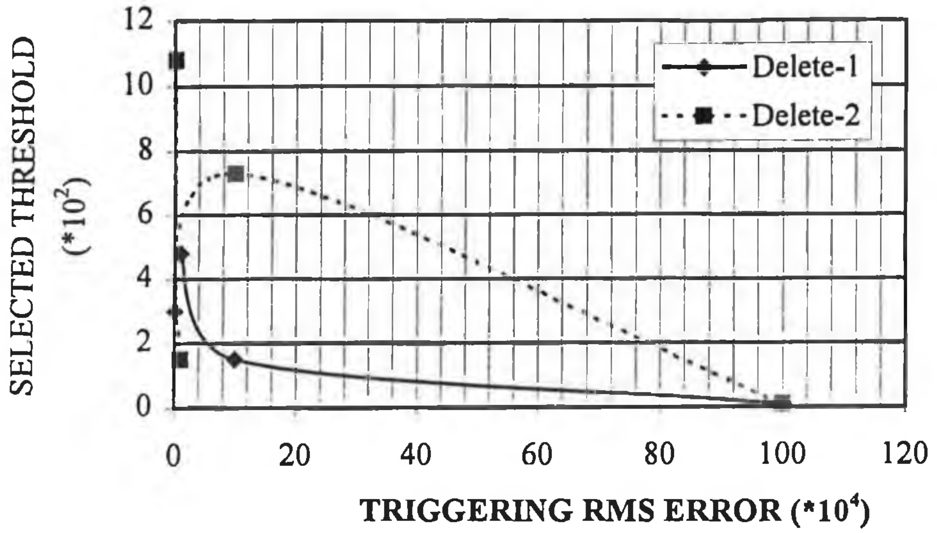


Figure 5.1.a – The neural network results for the first data type.

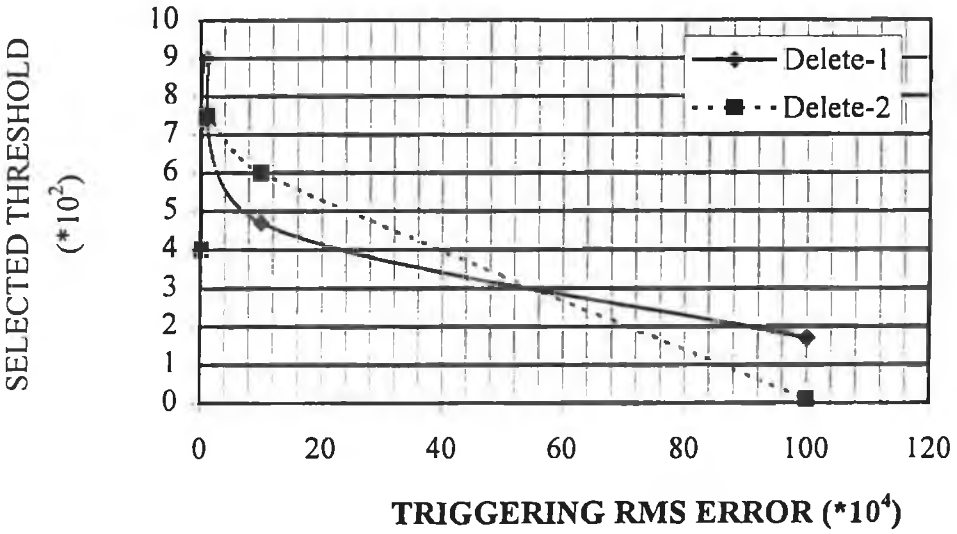


Figure 5.1.b – The neural network results for the second data type.

Figure 5.1 - Selected threshold values with triggering RMS error.

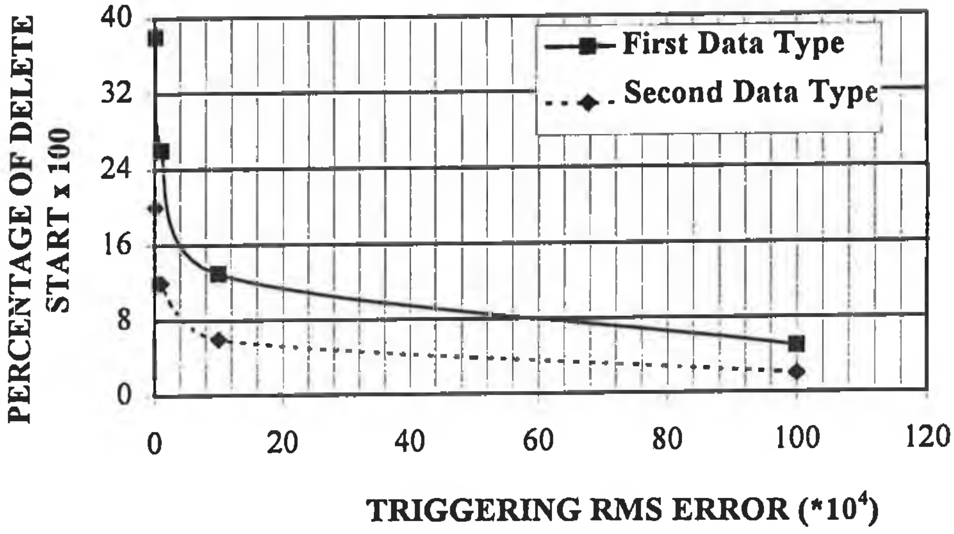


Figure 5.2 - Percentage of delete start with triggering RMS error.

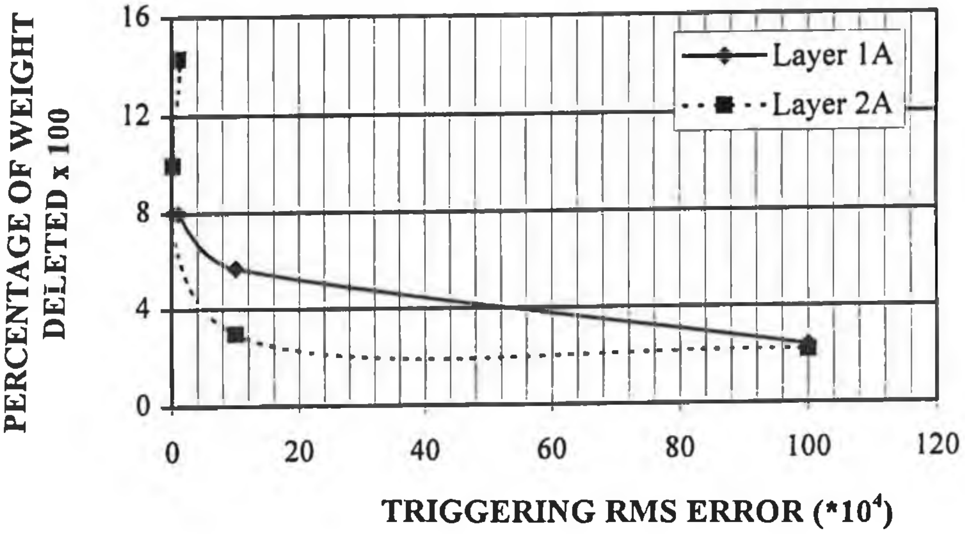


Figure 5.3.a - Delete-1 process

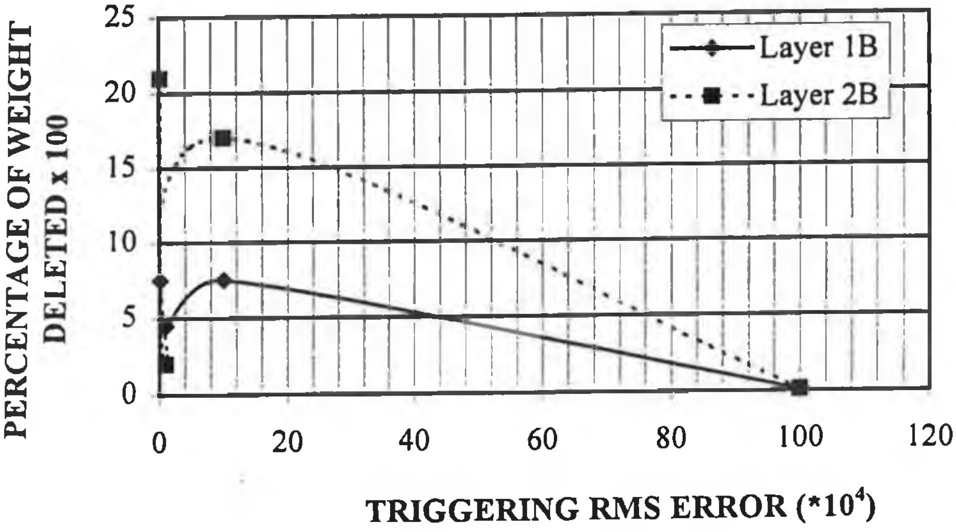


Figure 5.3.b - Delete-2 process

Figure 5.3 - Percentage of weight deleted corresponding to first data kind with RMS error.

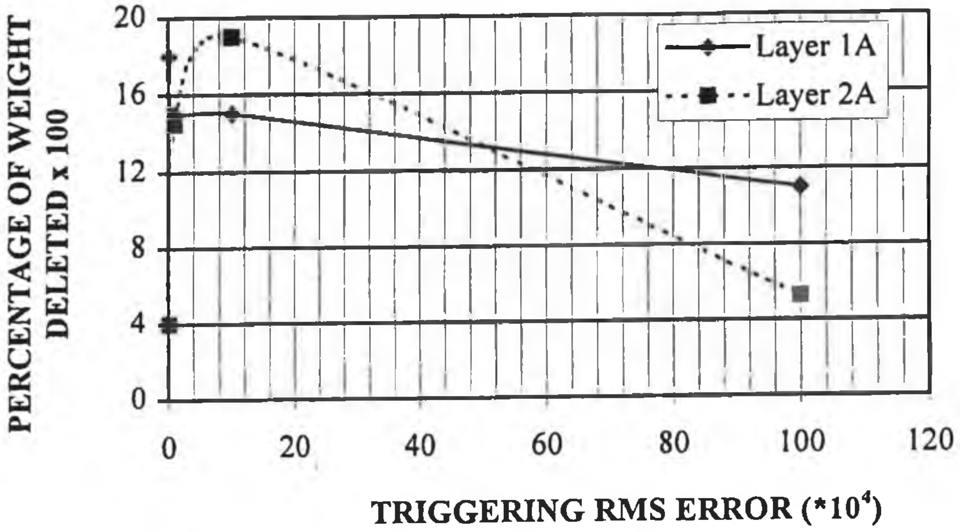


Figure 5.4.a - Delete-1 process.

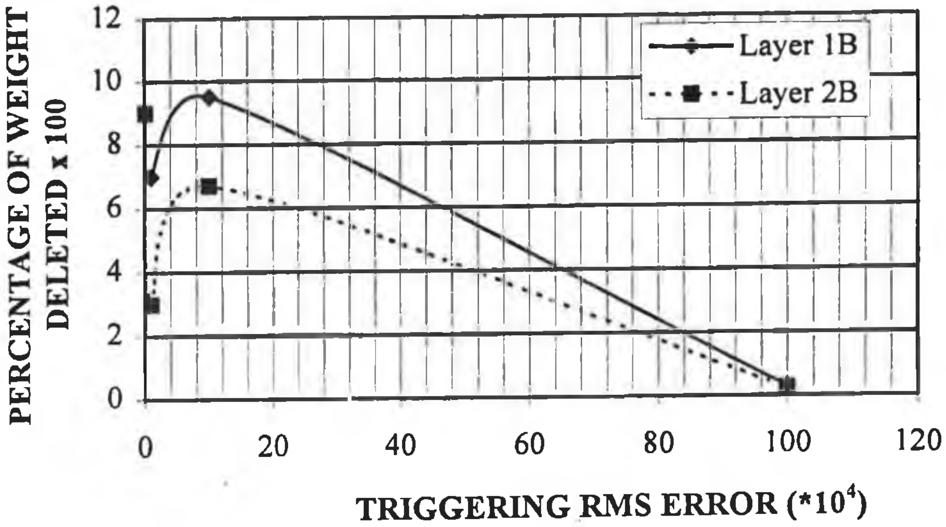


Figure 5.4.b - Delete-2 process.

Figure 5.4 - Percentage of weight deleted corresponding to second data kind with RMS error.

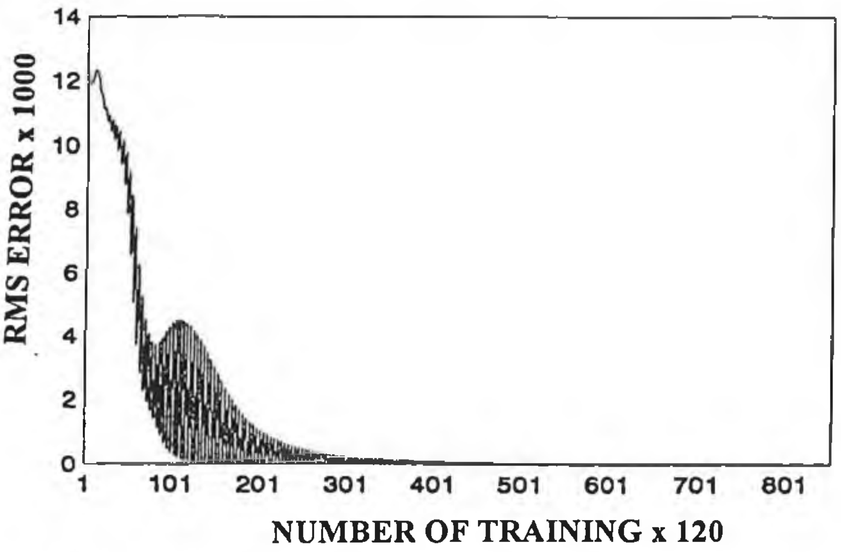


Figure 5.5.a - No-Deleting Process

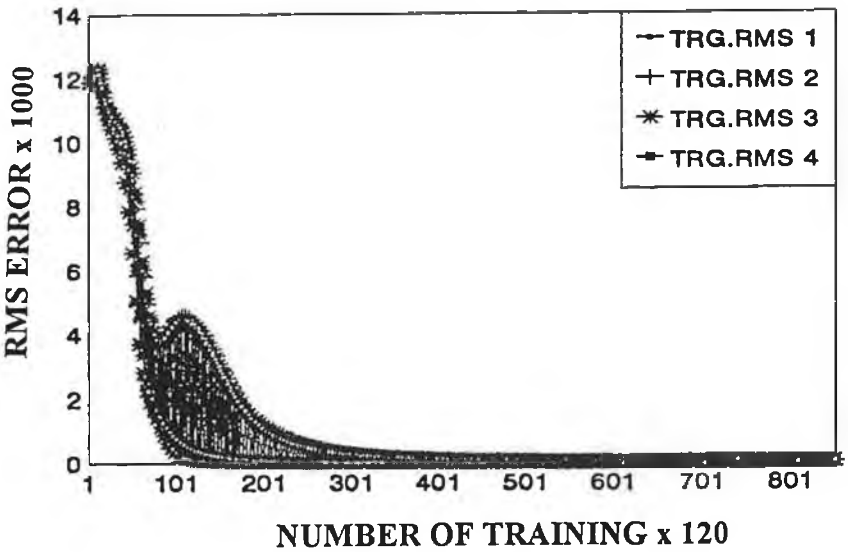


Figure 5.5.b - Delete-1 Process

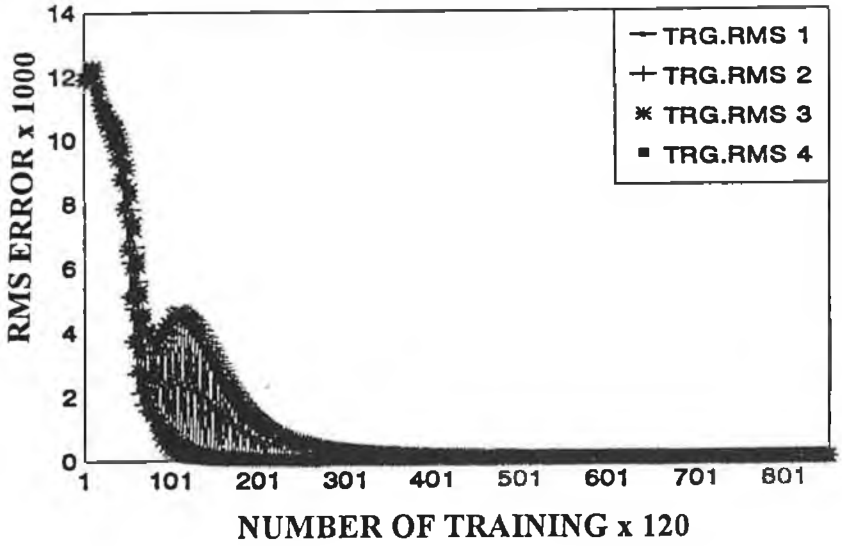


Figure 5.5.c – Delete-2 Process

Figure 5.5 – RMS error with number of training corresponding to No-Delete, Delete-1 and Delete-2 processes.

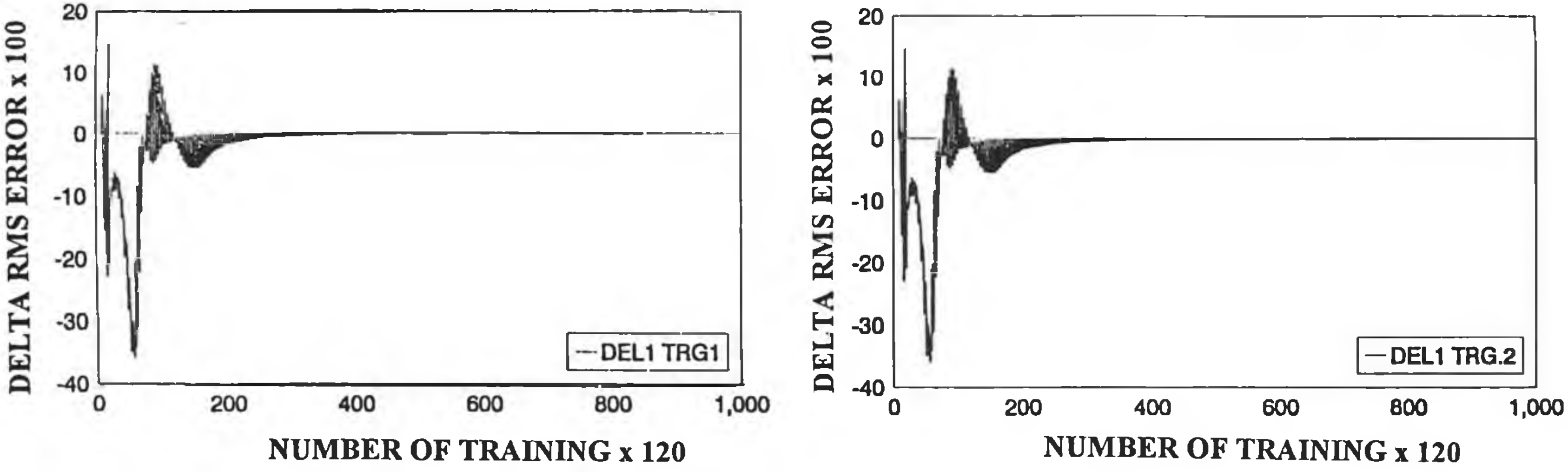


Figure 5.6.a - Triggering RMS error level 1.

Figure 5.6.b - Triggering RMS error level 2.

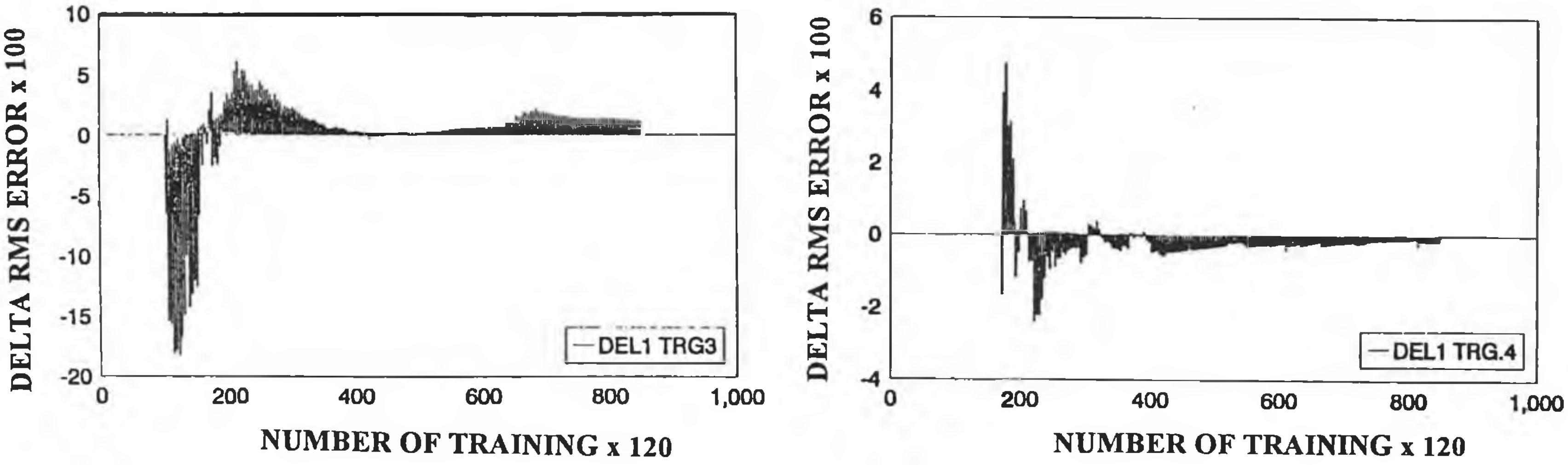


Figure 5.6.c - Triggering RMS error level 3

Figure 5.6.d - Triggering RMS error level 4.

Figure 5.6 – Delta RMS error with number of training corresponding to Delete –1 process.

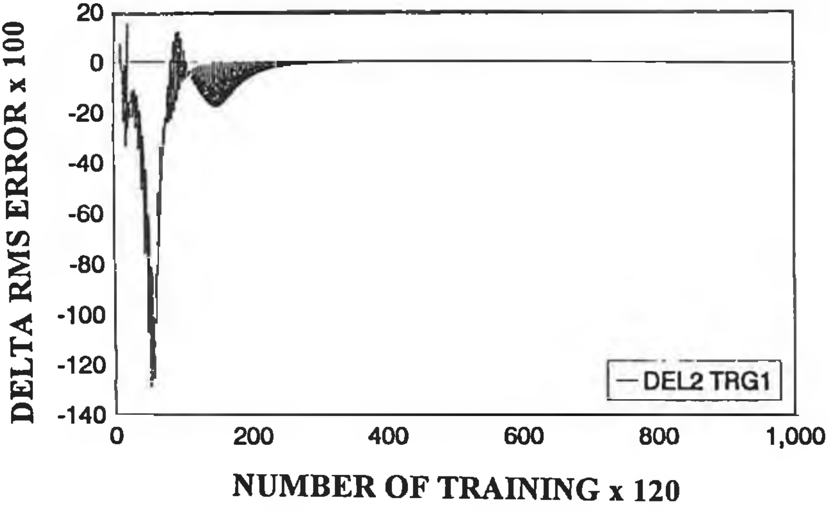


Figure 5.7.a - Triggering RMS error level 1.

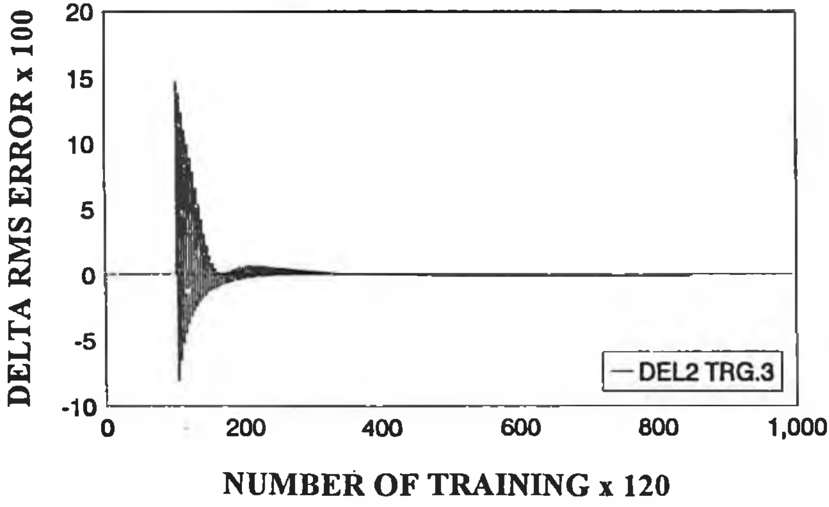


Figure 5.7.c - Triggering RMS error level 3

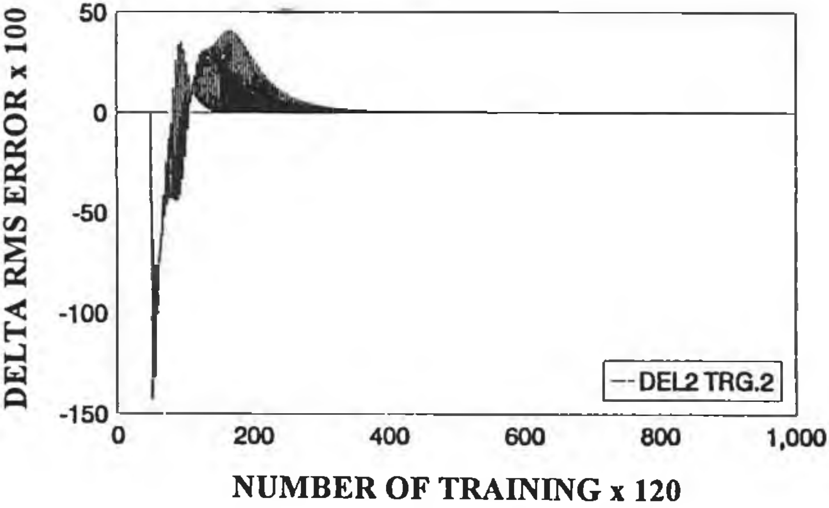


Figure 5.7.b - Triggering RMS error level 2.

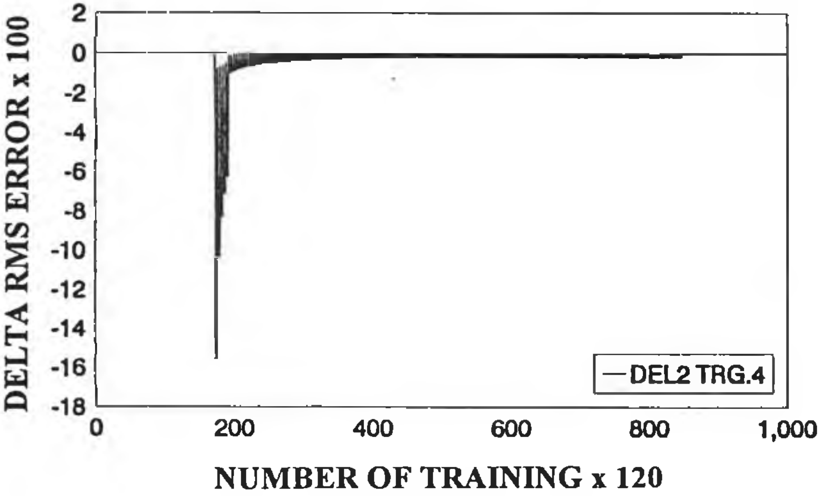


Figure 5.7.d - Triggering RMS error level 4.

Figure 5.7 – Delta RMS error with number of training corresponding to Delete –2 process.

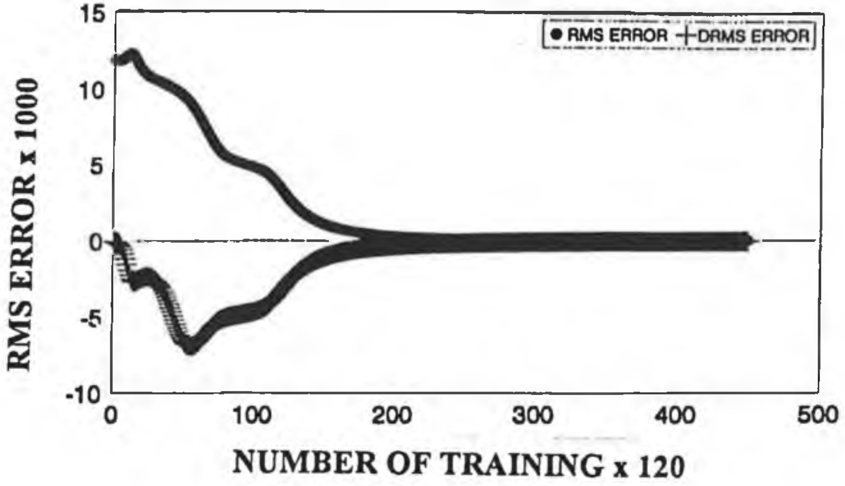


Figure 5.8.a - No Delete process

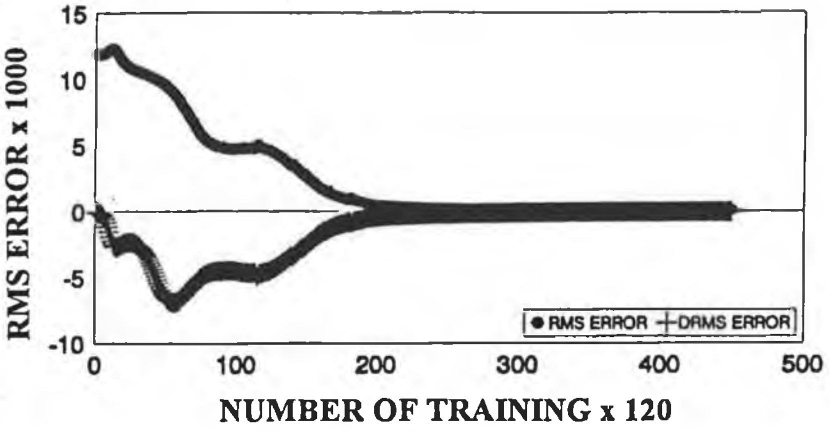


Figure 5.8.b - Delete-1 process

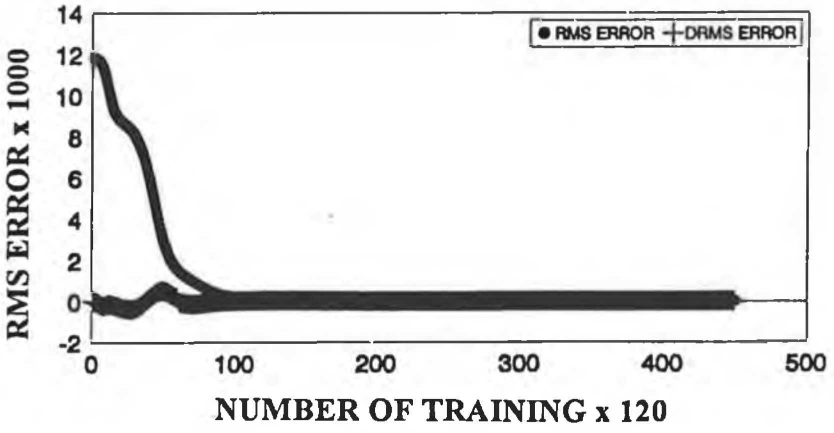


Figure 5.8.c - Delete-2 process.

Figure 5.8 – RMS error with number of training corresponding to second data kind.

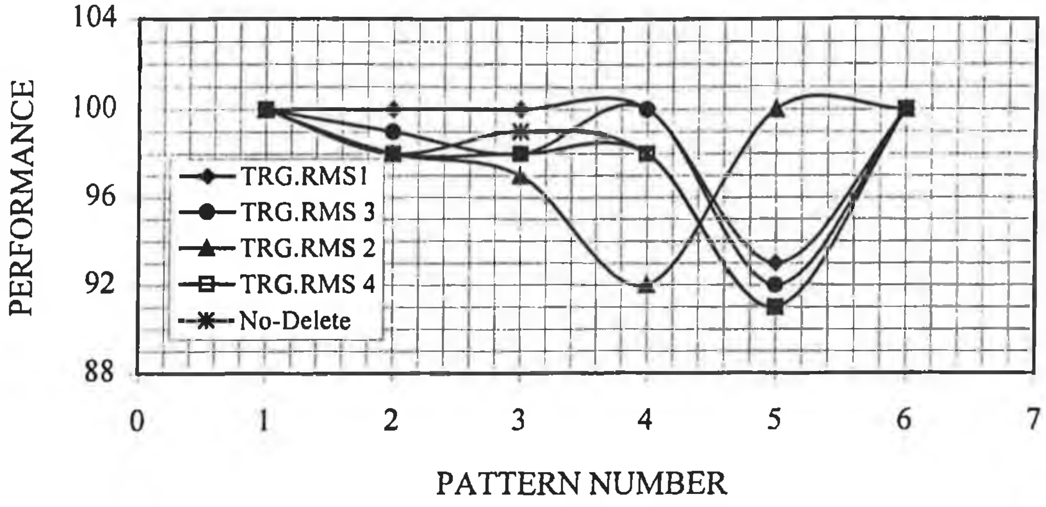


Figure 5.9.a - No-Delete and Delete-1 process

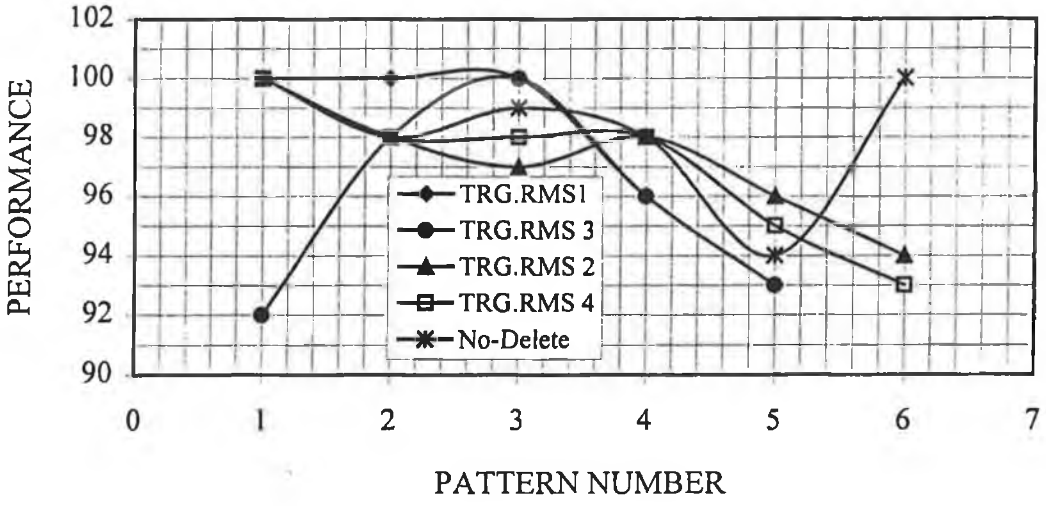


Figure 5.9.b - No-Delete and Delete-2 process

Figure 5.9 - Performance corresponding to first data kind and number of training.

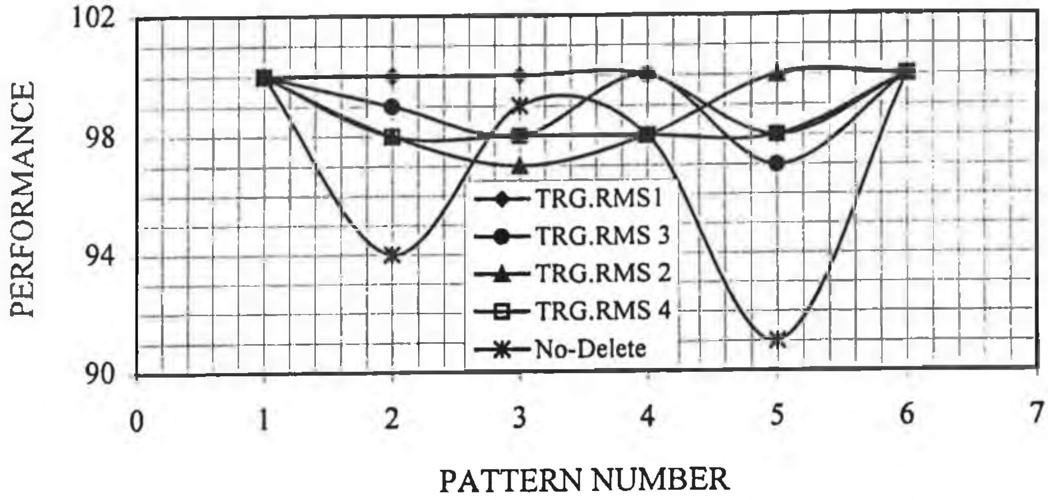


Figure 5.10.a - No-Delete and Delete-1 process

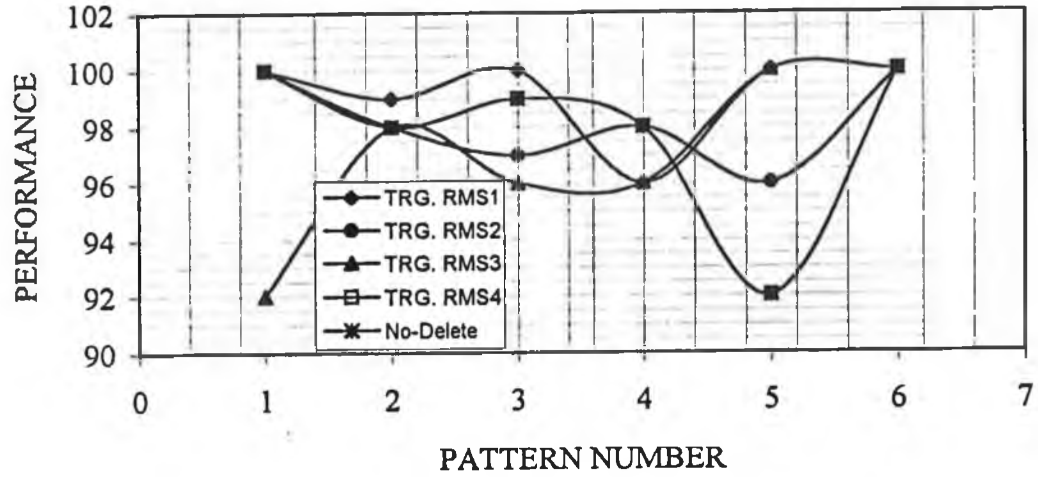


Figure 5.10.b - No-Delete and Delete-2 process

Figure 5.10 - Performance corresponding to second data kind and number of training.

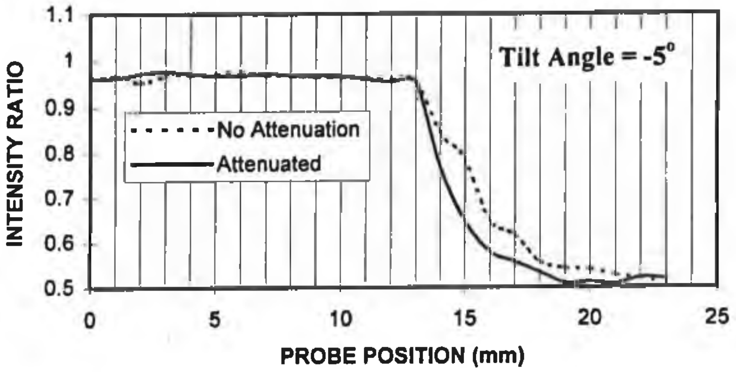


Figure 5.11.a

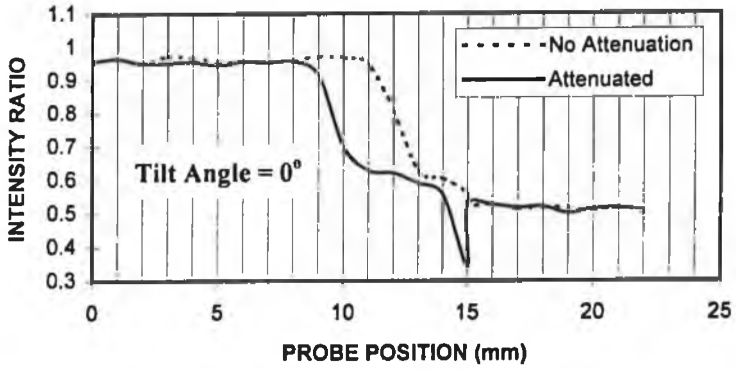


Figure 5.11.b

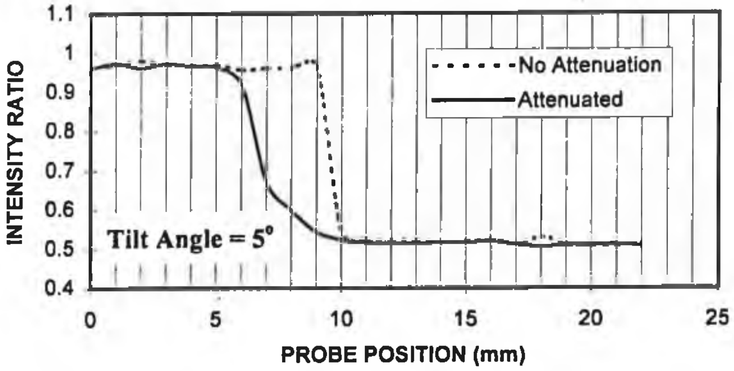
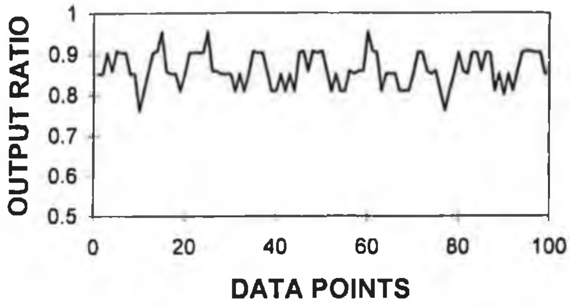
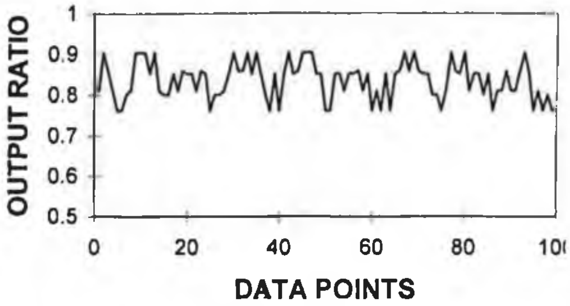


Figure 5.11.c

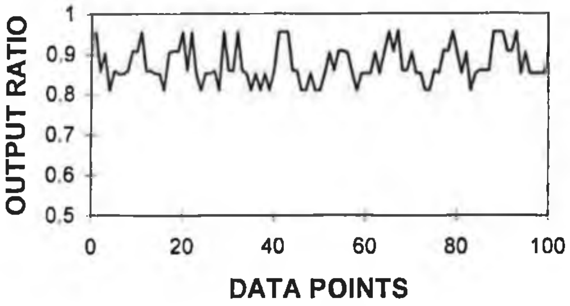
Figure 5.11 – Ratio of probe response (intensity ratio due to reflected and incident He-Ne beam) with probe position from the surface for different tilt angle of the workpiece surface and when incident beam is slightly attenuated.



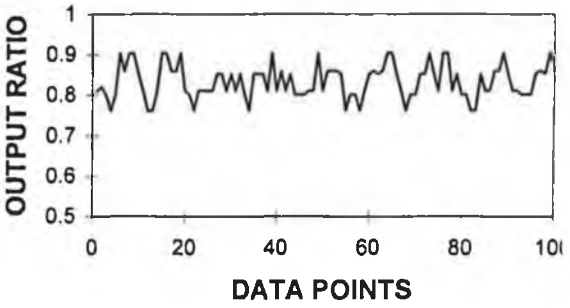
WORKPIECE 1 SURFACE (Ra = 7.61 μm)



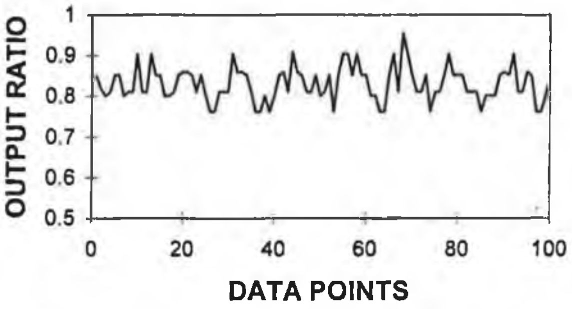
WORKPIECE 2 SURFACE (Ra = 6.22 μm)



WORKPIECE 3 SURFACE (Ra = 5.1μm)



WORKPIECE 4 SURFACE (Ra = 5.64 μm)



WORKPIECE 5 SURFACE (Ra = 5.66 μm)

Figure 5.12 - The first probe output (the ratio of incident and reflected laser beam intensities) for different workpiece surfaces.

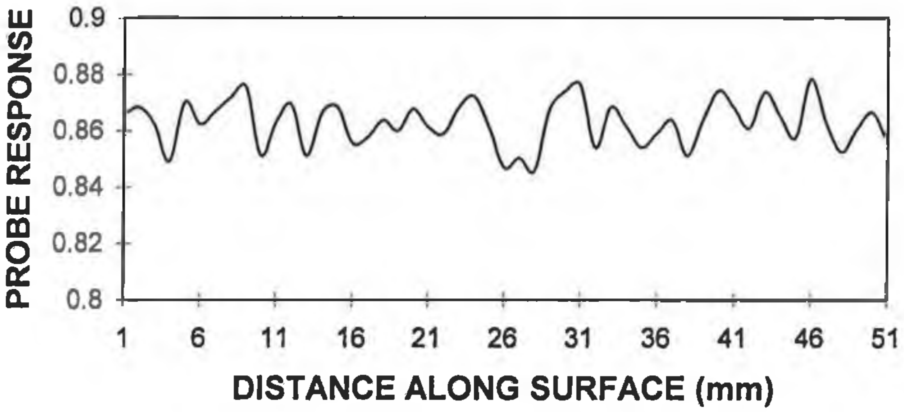


Figure 5.13 - First probe response with distance along workpiece surface 1.

(Ra = 7.61 μm).

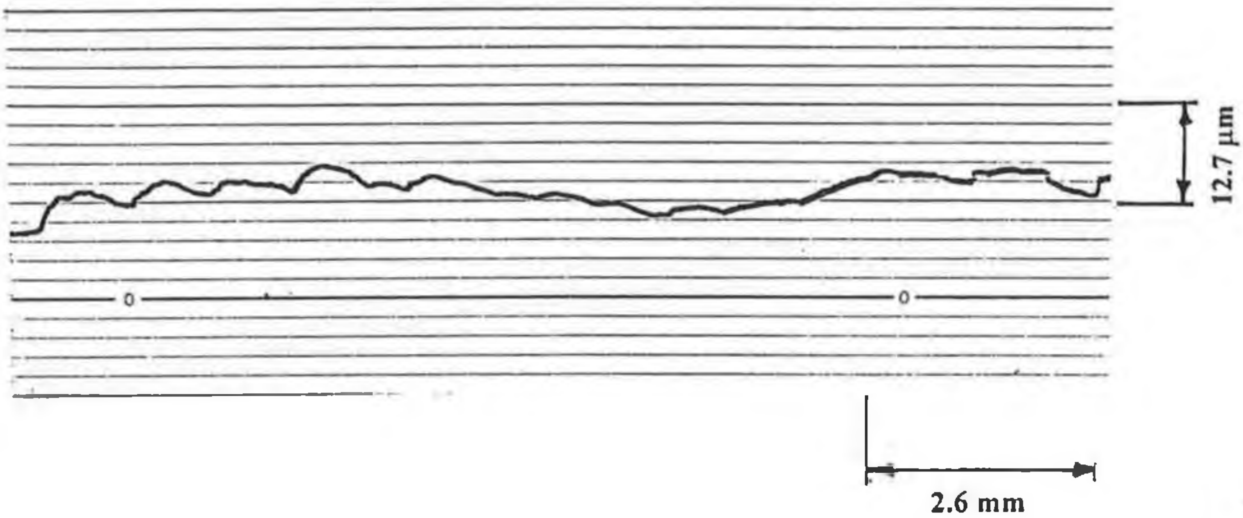


Figure 5.14 – Surface profile obtained from stylus instrument for workpiece no. 1.

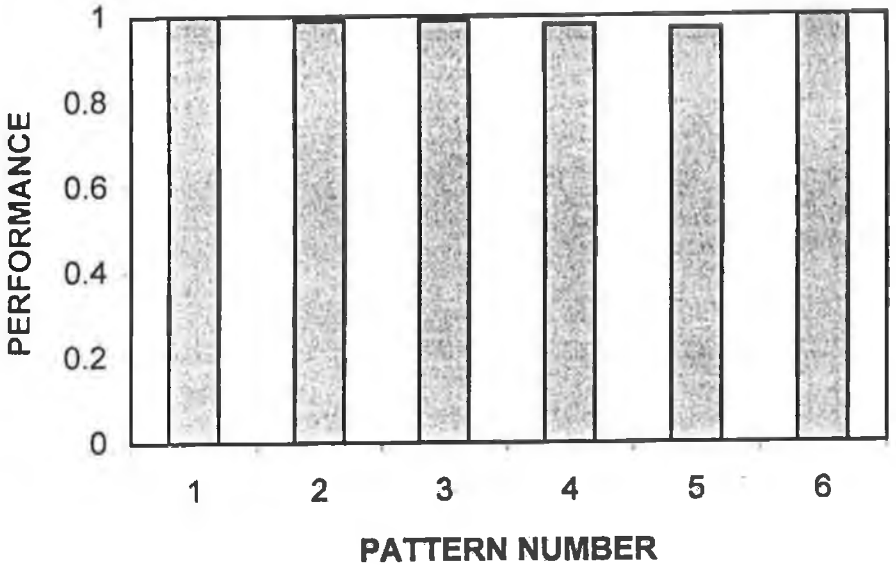
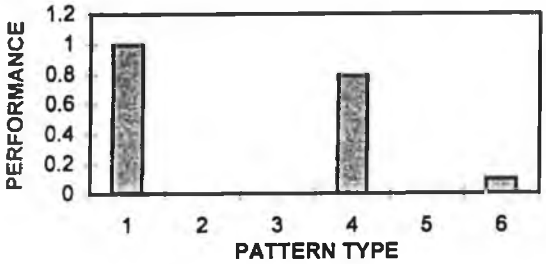
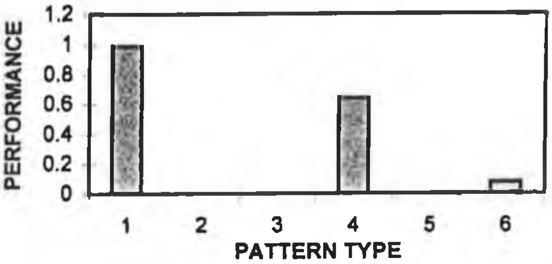


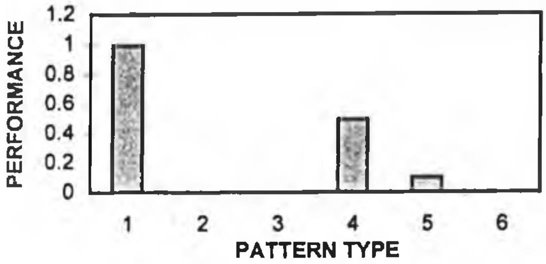
Figure 5.15 - Performance of the network output when testing the mathematically developed charts.



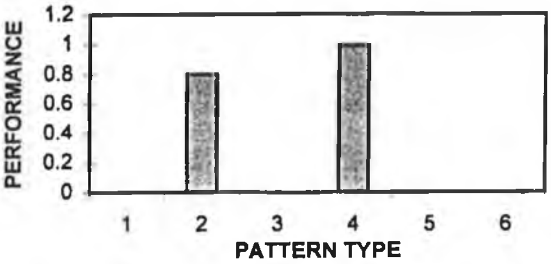
WORKPIECE 1 SURFACE (Ra = 7.61 μm)



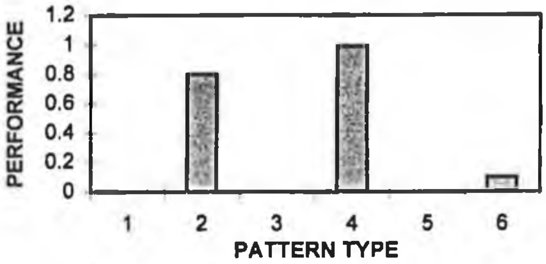
WORKPIECE 2 SURFACE (Ra = 6.22 μm)



WORKPIECE 3 SURFACE (Ra = 5.1μm)



WORKPIECE 4 SURFACE (Ra = 5.64 μm)



WORKPIECE 5 SURFACE (Ra = 5.66 μm)

Figure 5.16 - Neural network output for all workpiece surfaces. Patterns 1-6 are normal, increasing trend, decreasing trend, upward shift, downward shift and cycle.

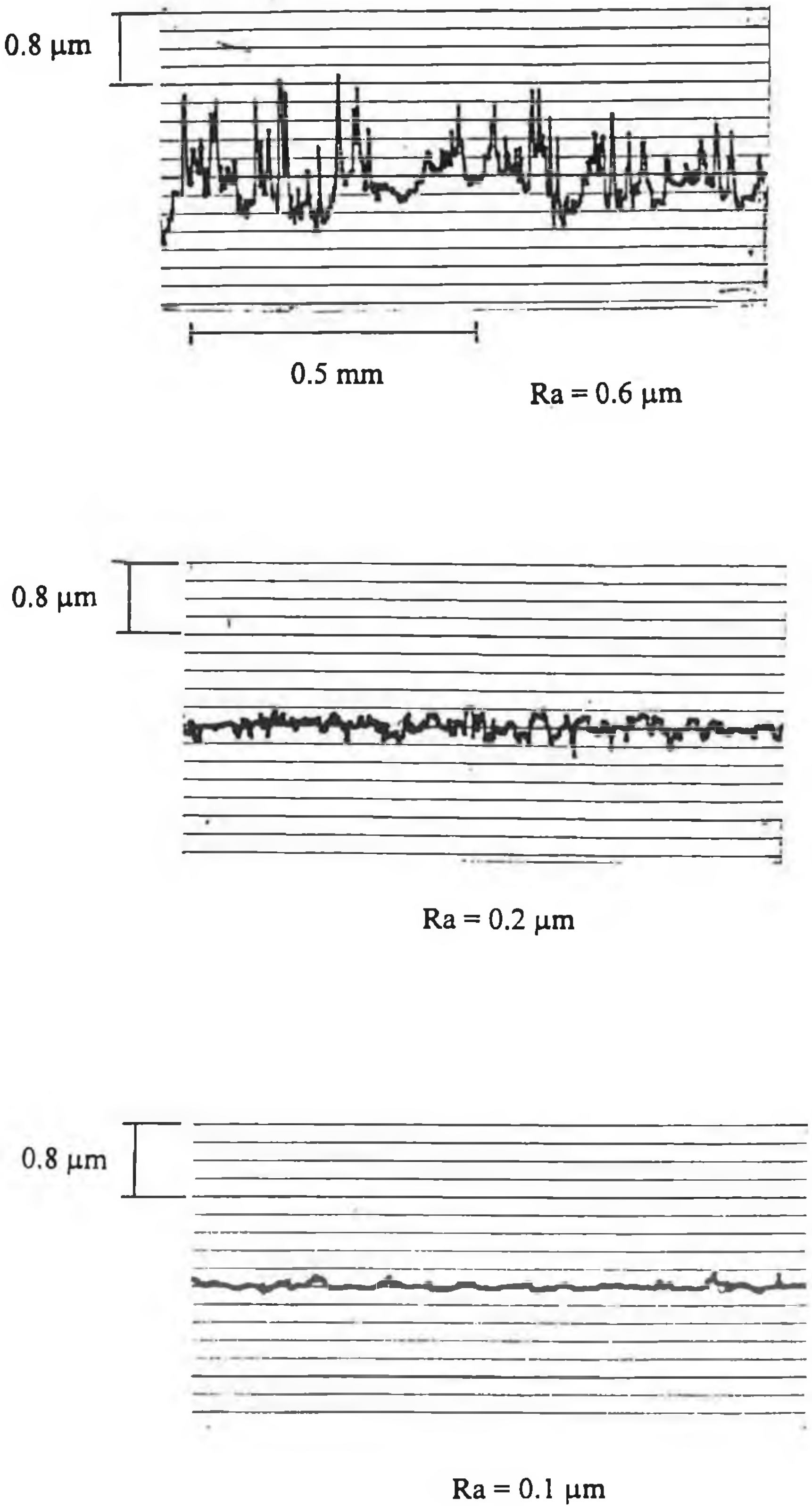


Figure 5.17 – Surface profiles obtained from stylus instrument for various roughness grades.

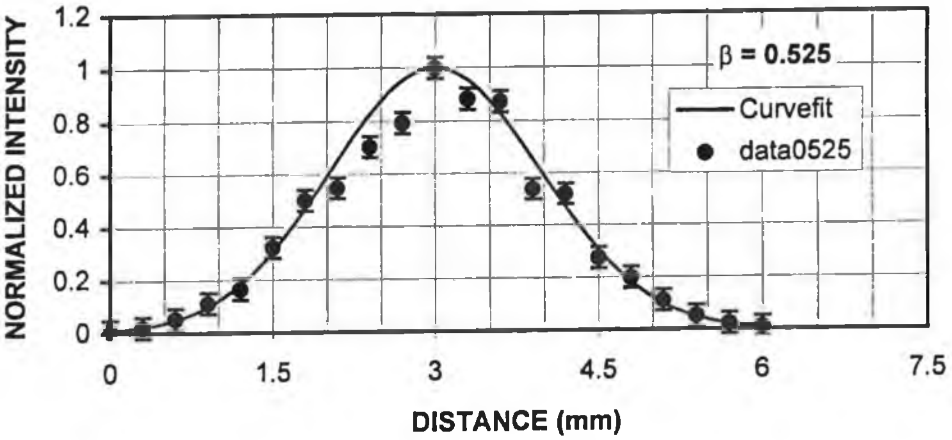


Figure 5.18.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (β is a Gaussian function constant).

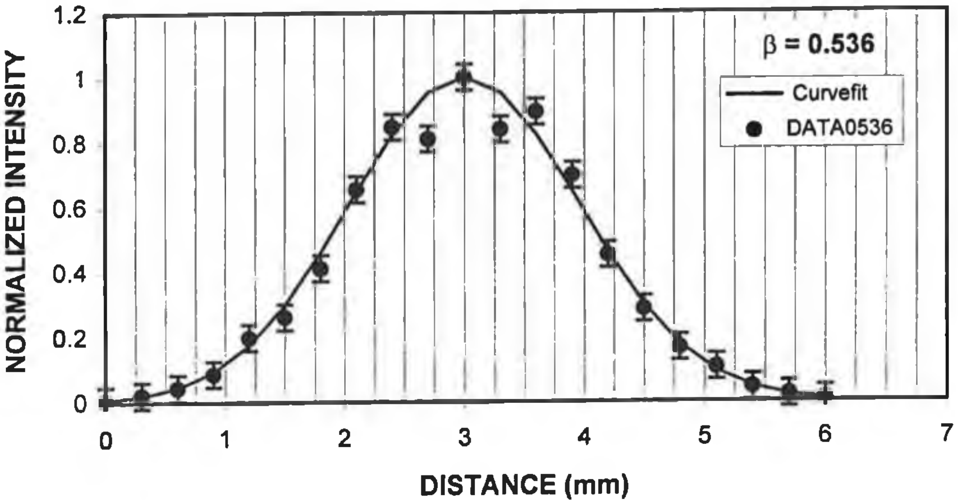


Figure 5.18.b – Normalised intensity with distance across the reflected beam spot as seen by probe (β is a Gaussian function constant).

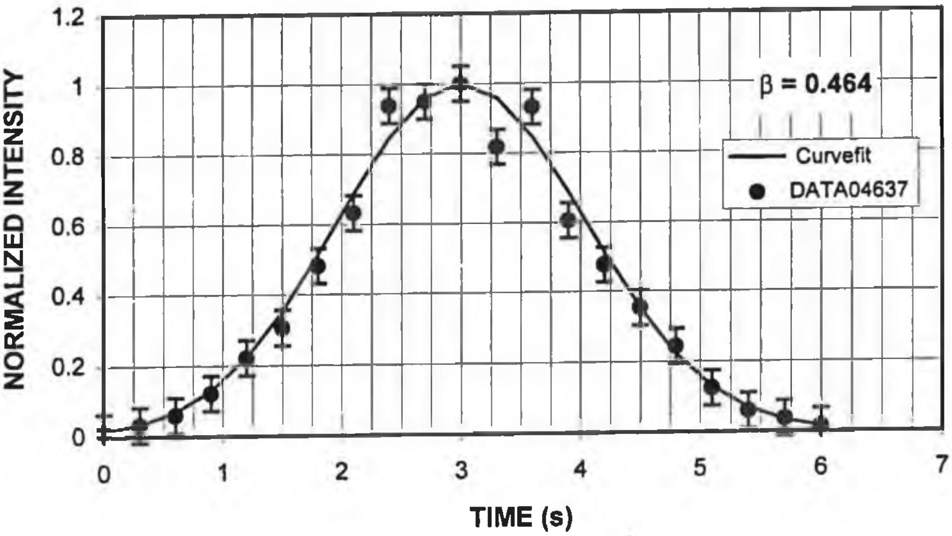


Figure 5.18.c – – Normalised intensity with distance across the reflected beam spot as seen by the probe (β is a Gaussian function constant).

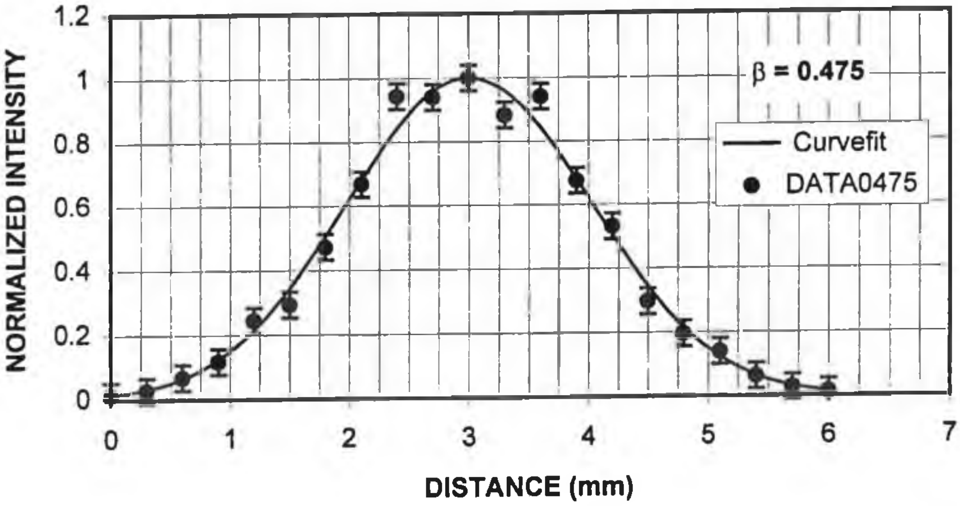


Figure 5.18.d – – Normalised intensity with distance across the reflected beam spot as seen by the probe (β is a Gaussian function constant).

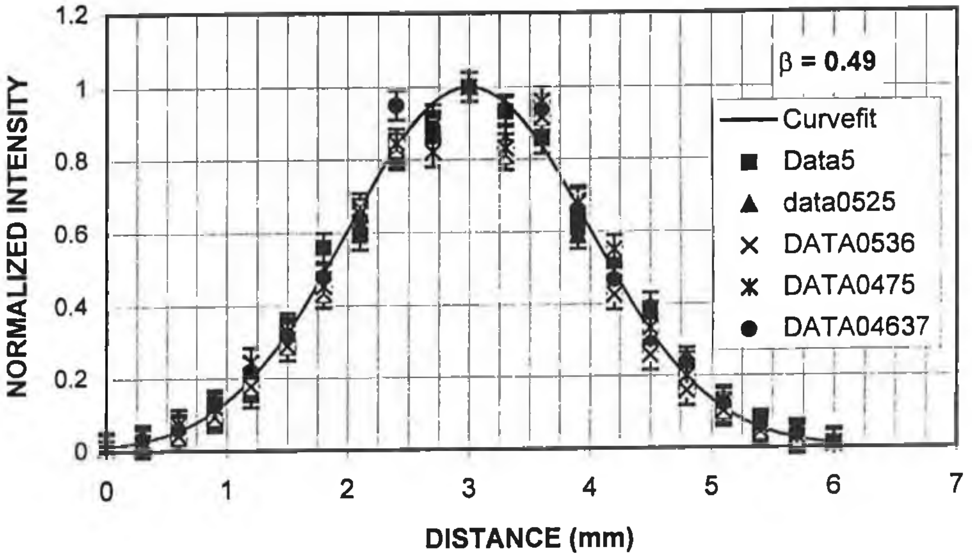


Figure 5.19.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (first method).

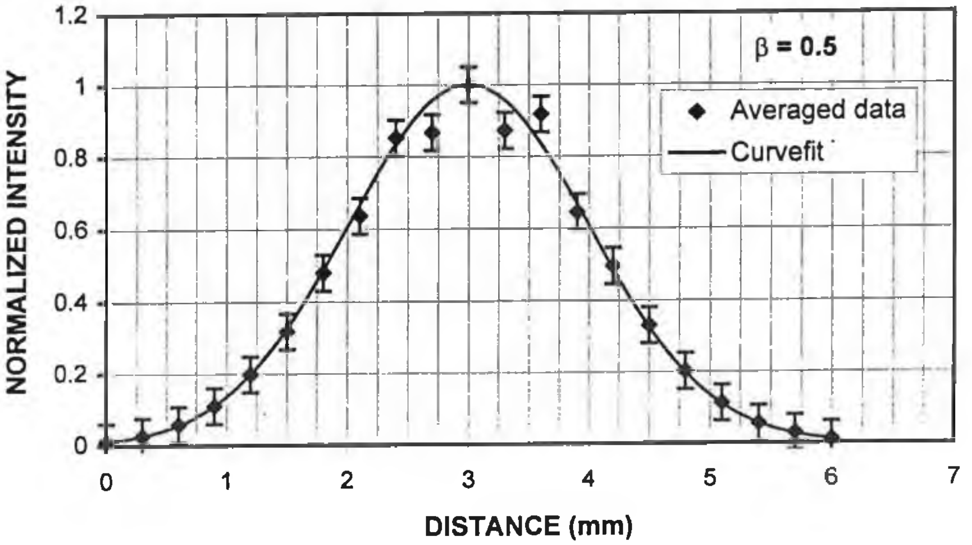


Figure 5.19.b – Normalised intensity with distance across the reflected beam spot as seen by the probe (second method).

Figure 5.19 - First and second method results for the workpiece no.1.

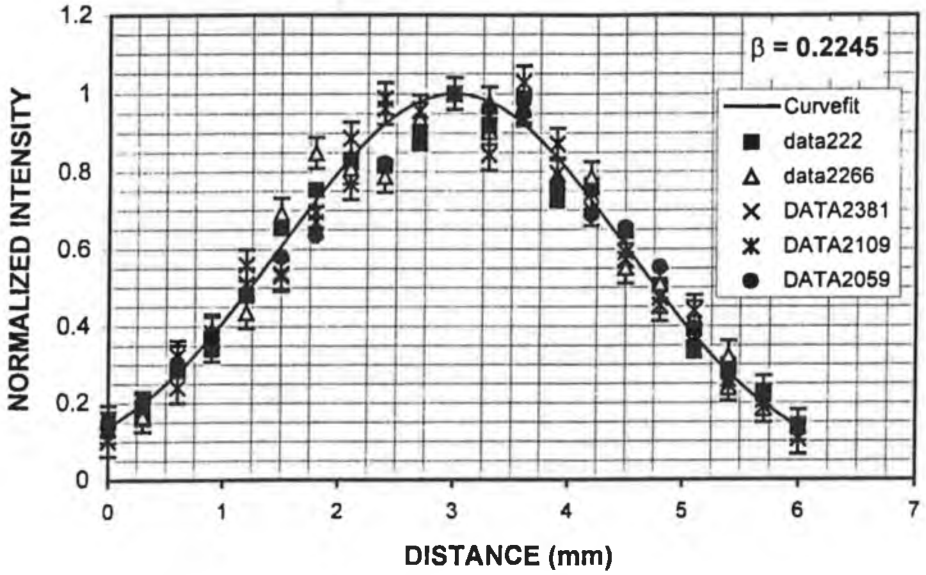


Figure 5.20.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (first method).

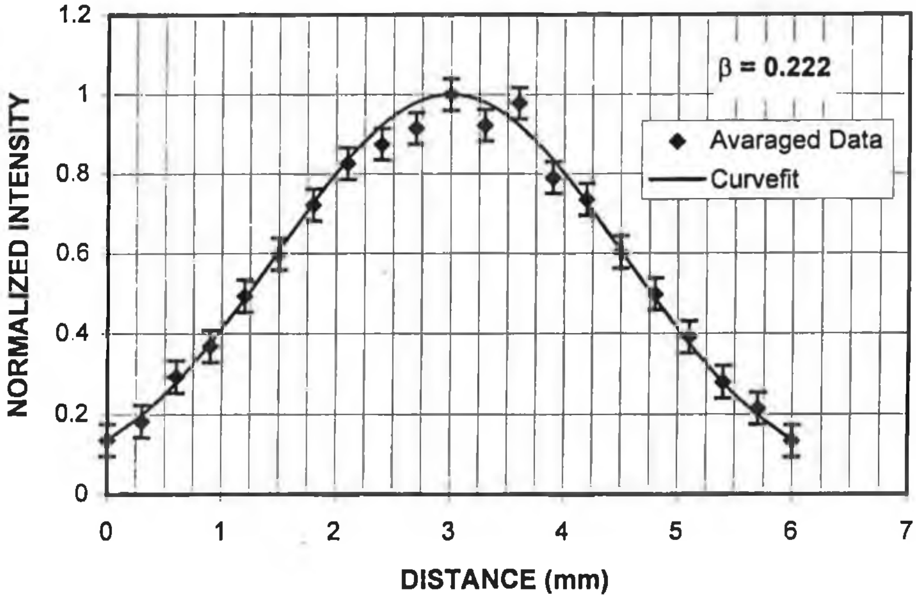


Figure 5.20.b – Normalised intensity with distance across the reflected beam spot as seen by the probe (second method).

Figure 5.20 - First and second method results for the workpiece no.2.

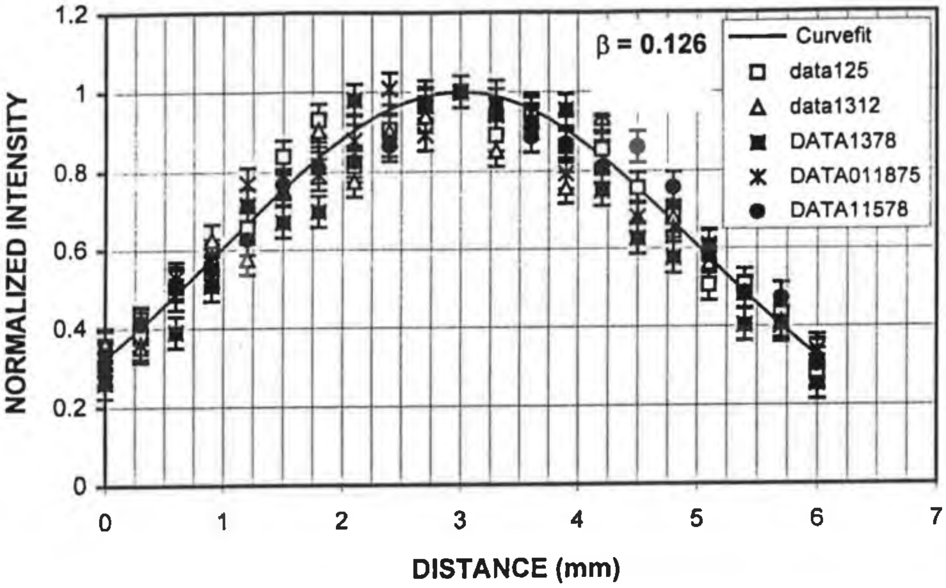


Figure 5.21.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (first method).

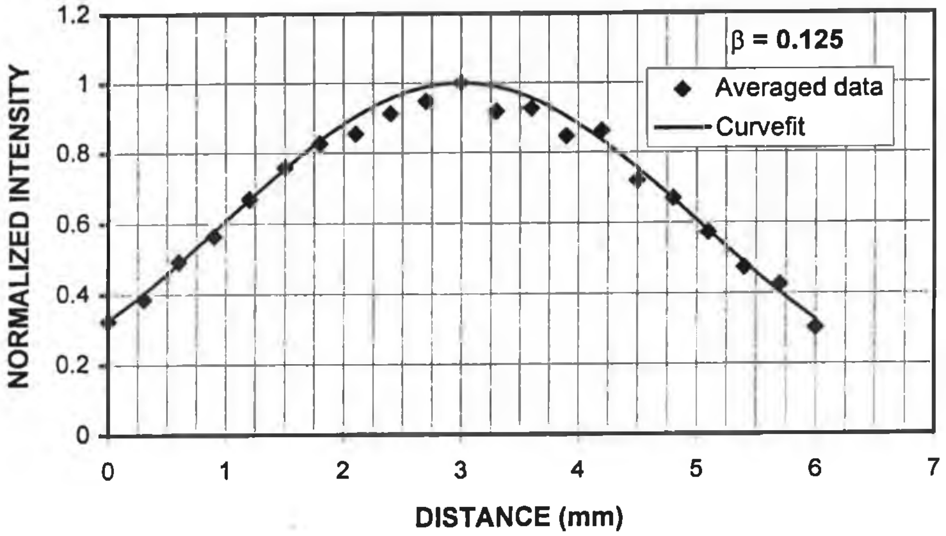


Figure 5.21.b – Normalised intensity with distance across the reflected beam spot as seen by the probe (second method).

Figure 5.21 - First and second method results for the workpiece no.3.

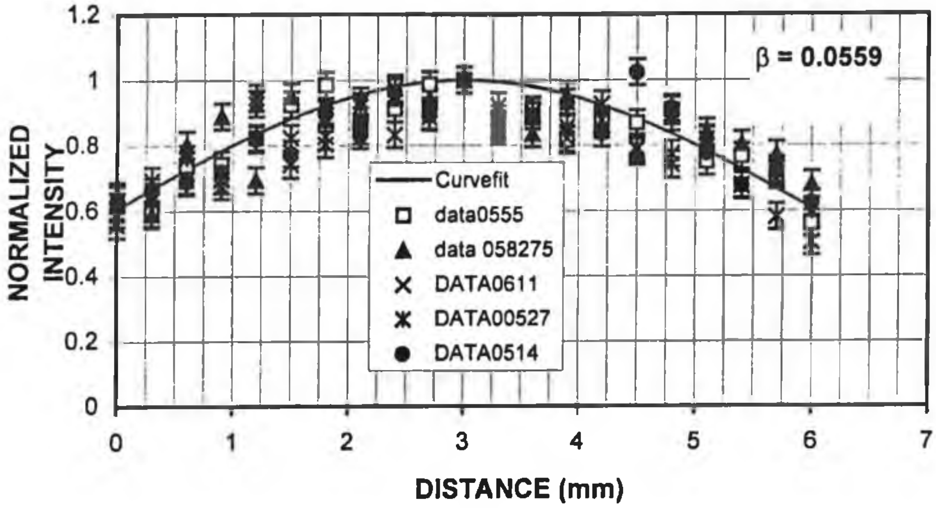


Figure 5.22.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (first method).

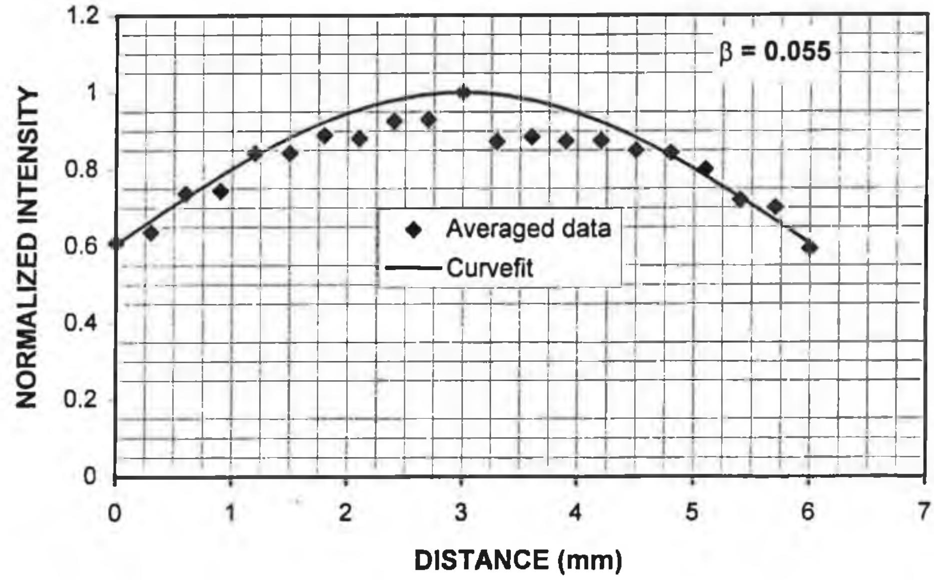


Figure 5.22.b – Normalised intensity with distance across the reflected beam spot as seen by the probe. (second method).

Figure 5.22 - First and second method results for the workpiece no.4.

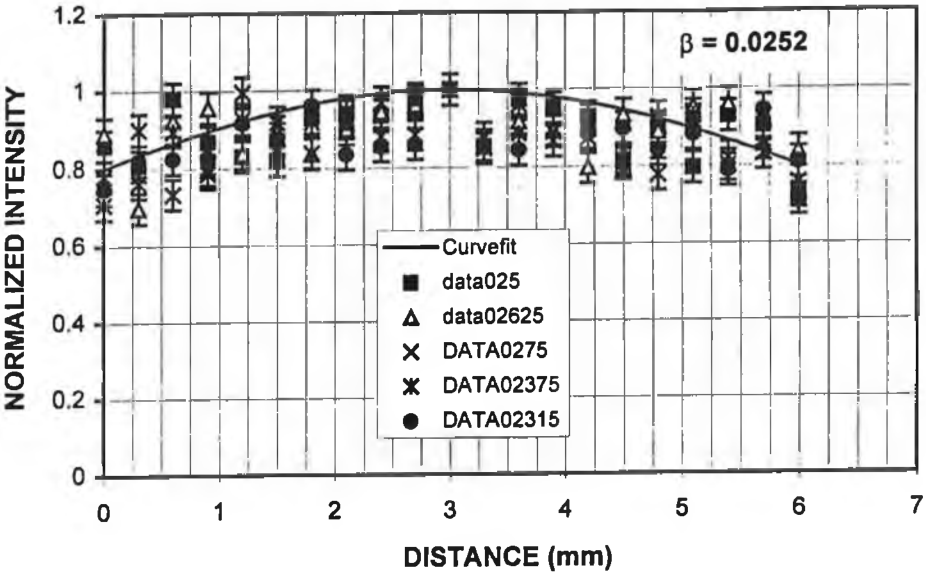


Figure 5.23.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (first method).

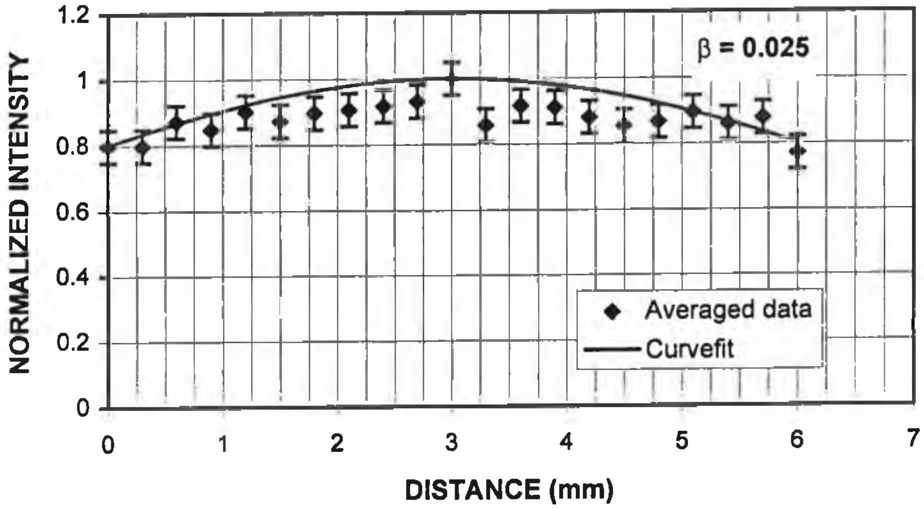


Figure 5.23.b – Normalised intensity with distance across the reflected beam spot as seen by the probe (second method).

Figure 5.23 - First and second method results for the workpiece no.5.

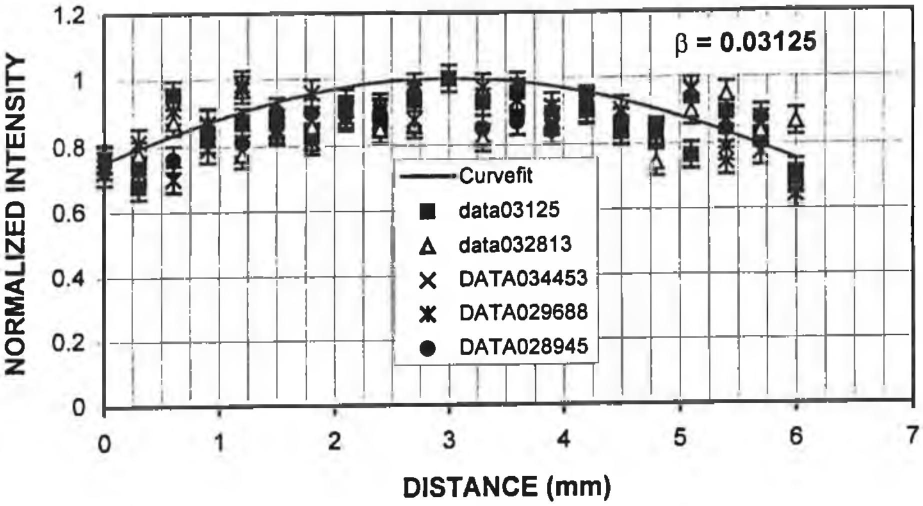


Figure 5.24.a – Normalised intensity with distance across the reflected beam spot as seen by the probe (first method).

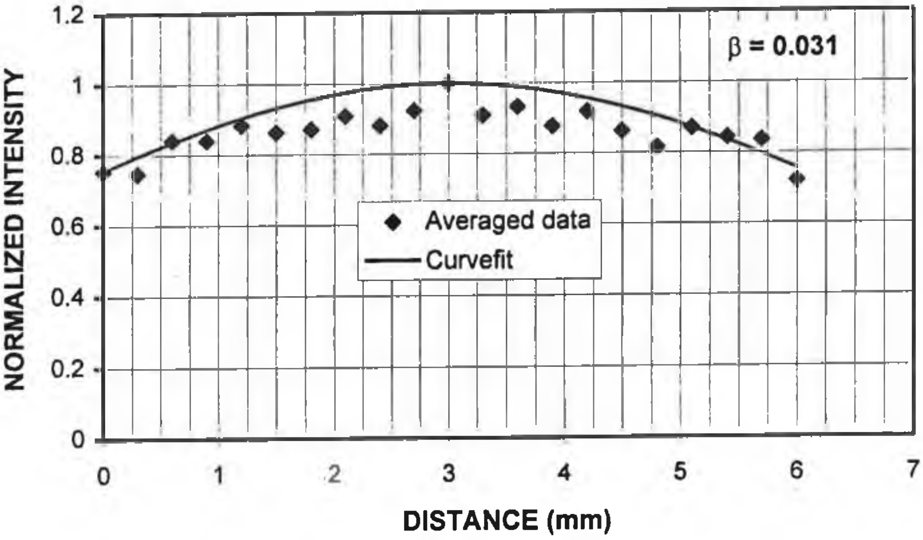


Figure 5.24.b Normalised intensity with distance across the reflected beam spot as seen by the probe (second method).

Figure 5.24 - First and second method results for the workpiece no.6.

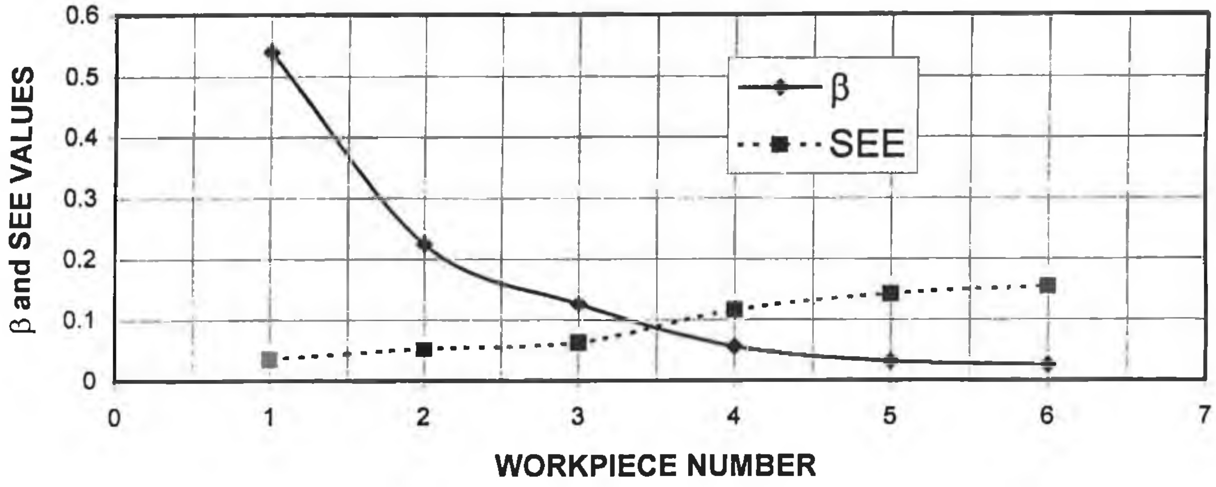


Figure 5.25.a -β and SEE values obtained from the first method.

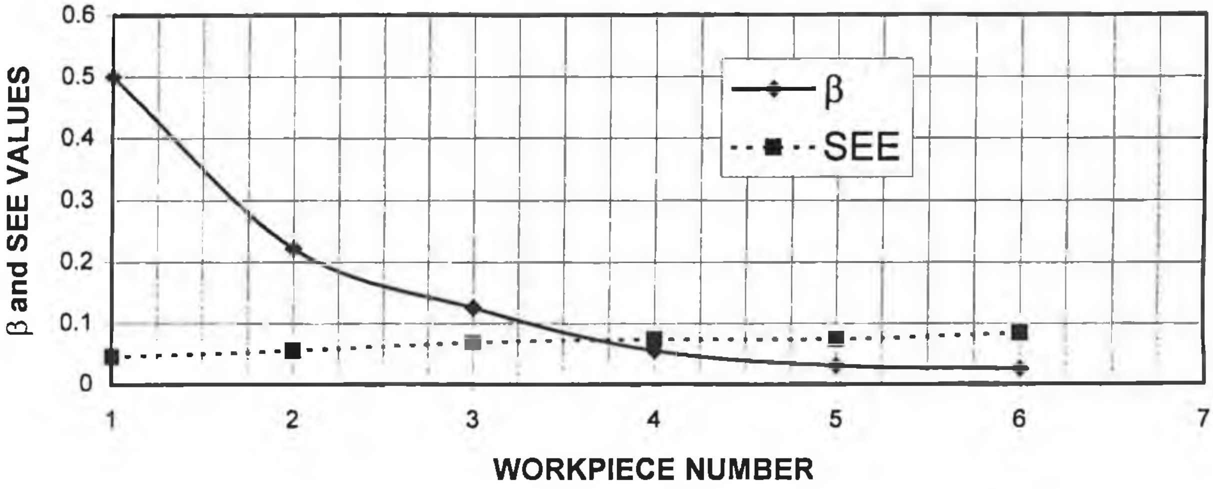


Figure 5.25.b -β and SEE values obtained from the second method.

Figure 5.25 -β and SEE values with workpiece numbers.

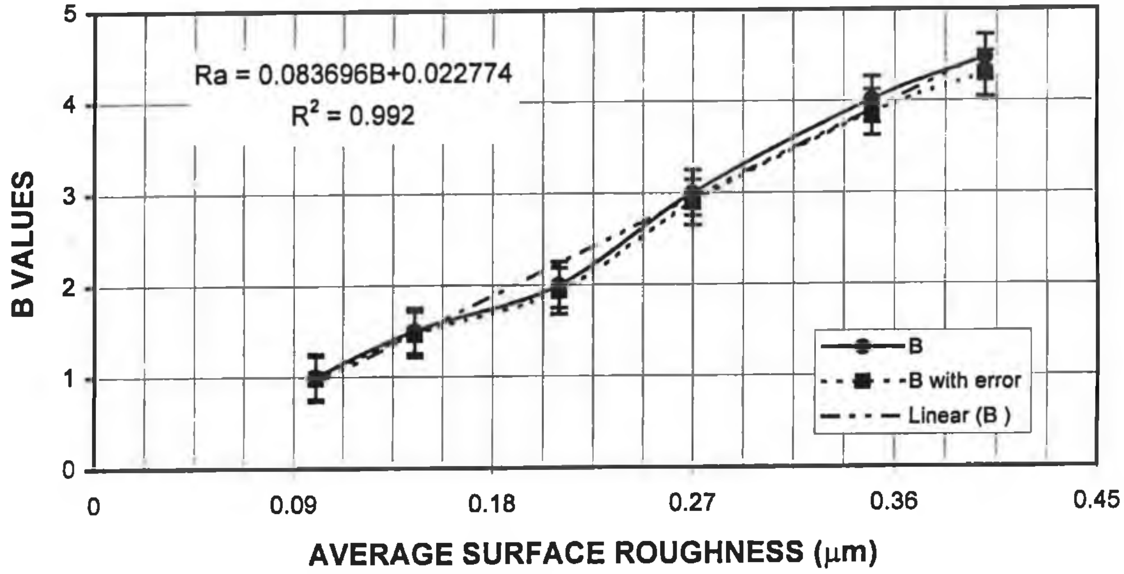


Figure 5.26 - Standard deviation of Gaussian function (B) with average surface roughnes values. (B with error, curvefit and the resulting equation are included).

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

The conclusions derived from the present study covers the neural network pruning and surface roughness measurement based on diffusive (Ra ≥ 2 μm) and specular ($0.1 \leq$ Ra$\leq$ 2 μm) reflections. Therefore, the conclusions will be presented in accordance with the relevant study.

In the neural network pruning, two data types, two deleting methods and four triggering RMS levels are considered for automatic thresholding process. Consequently, the conclusions derived from the network pruning may include the following.

The threshold selected by the network decreases with increasing triggering RMS error. Delete-1 process results in smaller thresholds selected than that occurs in Delete-2 process. This may be explained in terms of Delete-1 process in which the weight deleting takes place dynamically. The effect of data structure on the selected threshold is significant; i.e. the second data kind results in increased threshold values selected.

The percentage of weight deleted in the first and second layers varies for the two delete processes. The percentage of weight deleted in the first layer is lower than that of the second layer. Delete-2 process results in higher percentage of weight deleted than that occurs in Delete-1 process. In this case, some contributing weights to learning in late stages are disregarded in Delete-2 process. These effects the percentage of weight deleted such that this effect increases in the case of second data kind.

When comparing Delete-1 and No-Deleting processes, drop in RMS error occurs with training time. Network learns faster in the case of Delete-1 process than that corresponding to both Delete-2 and No-Deleting processes. Late start of deleting process is fruitful in learning mechanism. The dynamic deleting process (Delete-1) enables the network to learn faster than No-Deleting process. The effect of data structure appears as oscillation in the resulting RMS error. The cause of this oscillation is due to that in the first data kind, random variables and coefficients in equations 2.1, 2.2, 2.3, and 2.4 are varied simultaneously, however, when generating the second data kind, only random variables are changed. Consequently, when training the network with first data kind, RMS error oscillates as the input data for patterns differ from the previous data.

Delete-2 process produces relatively lower performances than Delete-1 and No-Deleting processes. This may be due to the fact that keeping the deleted weights constant throughout the training iterations may constrain the learning mechanism in late learning stages, i.e. initially deleted weights may become important in the late stages of the learning mechanism. On the other hand, Delete-1 process shows a slightly improved performances comparable to the No-Deleting process. This is especially true for pattern number 5. This may be due to the fact that in the Delete-1 process, weight deleting takes place dynamically and number of weights deleted at each training iteration changes, which in turn provides gradual reduction in number of unimportant weights during the training operation, i.e. network settles dynamically with progressing train iterations. It is clear that network performance is maximum in Delete-1 process and triggering level 4.

Although the use of the second data kind may cause a sudden decrease in the RMS error, which in turn provides high learning rate, and improved performance, but in this case generalisation capacity of the network drops considerably.

The conclusions derived from the surface roughness measurement based on diffusive and specular reflections may include the following:

The probe position relative to the workpiece surface becomes critical as the distance between the probe and the workpiece increases. However, beam attenuator installed in the path of incident laser beam does not have a significant effect on the probe response providing that attenuation of the order of 8% occurs.

Tilting of the workpiece surface effects the probe response such that an early decay in the probe response occurs when the distance between the probe and the workpiece surface increases.

The optical method results agree well with the stylus measurement results qualitatively noting that the spatial resolution of the data obtained from the optical method is considerably high.

The back-propagation network classifies the surface profiles, in accordance with the patterns developed mathematically, with considerably accuracy. Some surfaces have mix pattern profiles, however, a normal pattern is the dominant pattern evident from the network output.

The surface roughness and averaged surface roughness values measured using a Bendix proficoder instrument (stylus) show that the workpiece surfaces possess almost similar surface patterns with different Ra values.

The optical method based on specular diffusion samples the reflected beam intensity profile at every 1 ms period. Consequently, the intensity profiles seen by the fiber optic probe vary at every point on the workpiece surface. The beam intensity profile characteristics can be related to the Ra values of the surface. Moreover, the incident beam intensity profile is Gaussian in nature, therefore, the reflected beam intensity profile is also Gaussian with some distortions. The standard deviation of the Gaussian function may play a key role when developing the relationship between Ra and reflected beam intensity profile. On the other hand, for a known surface there exists a unique Ra value. Hence, the reflected beam intensity profiles associated at every point at the surface needs to be combined to give an intensity profile, which represents this particular surface. Two methods were used to obtain the representative intensity profiles. In addition, standard estimation of error and uncertainty analysis were carried out to determine the limitations and errors associated with the measurement.

In the case of the specular reflection measurements, a linear relationship between Ra and B values is developed provided that small values of B result in small values of Ra. The SEE reduces as the Ra value decreases. This may be due to the distortion of the reflected beam reducing as the surface roughness decreases. In this case, less scattering of the data in the reflected beam intensity curve occurs. In addition, increase in B results in broadening of the reflected beam intensity curve. Here, the data collected by the optical probe covers only the region close to the peak intensity. Therefore scattering of data, due to high degree of diffusive reflection, results in high values of SEE. Moreover, the degree of scattering increases as Ra increases; in this case, the measured results involve a high degree of error. Therefore, the present experiment gives good results for low Ra values, i.e. the error involved is less than 5%. In addition, the measurement is limited to a certain range of Ra values, i.e. as the Ra reduces beyond 0.1 µm, the measurement error increases considerably.

6.2 FUTURE WORK

The future work of the present study may be listed as follows:

1 - The optical method for surface roughness measurement may be generalised, for the measurement of all Ra values, introducing two probes at different angles for seeing the reflected beam intensity profiles. In this case, a set of detector arrays detecting the total reflected beam size may be used as an optical probe. Therefore, a number of input ports for I/O card needs to be expanded and the computer program should be rewritten for this purpose.

2 - The experimental apparatus developed at present may be redesigned to accommodate cylindrical objects: in this case, the rotational motion of the object enables the optical probe to scan the workpiece surface.

3 - The neural network simplification can be extended to include pruning of the processing elements in the hidden layer. In this case, the neural network topology will be optimised for a given task. Consequently, the computational time and cost will be minimised.

REFERENCES

[1.1] Whitehouse D.J., " Comparison between stylus and optical methods for measuring surfaces", Annals of CRIP Vol.37/2, pp.649-653, 1988.

[1.2] Shin Y.C., Oh S.J. and Coker S.A., "Surface roughness measurement by ultrasonic sensing for in-process monitoring", ASME Transaction, J. Engineering for Industry, Vol.117, pp.439-447, 1995.

[1.3] Shiraishi M., "A consideration of surface roughness by an optical method", Transaction of ASME, J. Engineering for Industry, Vol.109, pp. 100-105, 1989.

[1.4] Stover J.C., "Optical Scattering: Measurement and Analyses", Mc Graw-Hill, New-York, 1990.

[1.5] Kothardaraman X.G., Arunachalem V.P. and Shankar B., "Independence of surface roughness parameters in the assessment of surface texture", Modelling, simulation and control B : Mechanical and Thermal Engineering, Material and Resources, Chemistry, Vol.30, No.1, pp.1-4, 1990.

[1.6] Kurita M. and Sato M., "A Technique for rapidly measuring surface roughness using a laser", JSME, Int.J., Series 1, Vol.35, No.3, pp.335-339, 1992.

[1.7] Garratt J.D. and Nettleton D.J., "A Stylus instrument for roughness and profile measurement of ultra fine surfaces", Int. J. Mach. Tools Manufact., Vol.32, No.1/2, pp.233-238, 1992.

[1.8] Domanski A.W. and Wolinski T.R., "Surface roughness measurement with optical fibers", IEEE Transactions on Instrumentation and Measurement, Vol.41, No.6, pp.1057-1061, 1992.

[1.9] Johnson F., Brisco B. and Brown B.J., "Evaluation of limits to the performance of the surface roughness meter", Canadian Journal of Remote Sensing, Vol.19, No.2., pp.140-145, 1993.

[1.10] Stone R.A. and Shafer S.A., "Determination of surface roughness from reflected step edges', J. Optical Society of America A: Optics and Image Science, and Vision, Vol.11, No.11, pp.2969-2980, 1994.

[1.11] Dawei T., "In-process sensor for surface profile measurement applying a common-mode rejection technique", Optics and Laser Technology, Vol.27, No., pp.351-353, 1995.

[1.12] Sadhi M.S. and Tiliune K., "Surface roughness monitoring using computer vision ", Int. J. Machine Tools and Manufacture, Design, Research and Appl.,Vol.36, No.7, pp. 817-829, 1996.

[1.13] El-Baradia M.A, “Computer aided analyses of a surface roughness model for turning”, J. Material Processing Technology, Vol.2, No.2, pp.207-216, 1991.

[1.14] Galante G., Piacentini M., Ruisi V.F., “Surface roughness detection by tool image processing “, Wear, Vol.148, pp.211-220, 1991.

[1.15] Alauddin M., El Baradie M.A. and Hashmi M.S.J., “Computer-aided analysis of a surface roughness model for milling.”, J. Materials Processing Technology, Vol.55, No.2, pp.123-127, 1995.

[1.16] Peiponen K. E. and Tsuboi T., “Metal Surfaces Roughness and Optical Reflectance”, Optics and Laser Technology, Vol.22, No.2, pp.127-130, 1990.

[1.17] Bengtssen A., “On three-dimensional measurement of surface roughness.”, Ph.D.Thesis, Published by Chalmers Tekniska Hogskola, Goteborg, Sweden, 1991.

[1.18] Parsons F.G. and Tabenkin A.N., “Theory and development of user-friendly software for the measurement of surface roughness and geometrical parameters.”, pp.255-262, 1962.

[1.19] Westra K.L. and Thomson D.J., “Effect of tip shape on surface roughness measurements from atomic force microscopy images of thin films.”, J. Vacuum Science and Technology B: Microelectronics Processing and Phenomena, Vol.13, No.2, pp.344-349, 1995.

[1.20] Guenther K.H., Mc Candless J.A. and Orazio F. D. Jr., “Correlation of light-sacttering measurements and visual ranking of optical surfaces”, Applied Optics, Vol.32, No.19, pp.3425-3432, 1993.

[1.21] Mhyshkin N.K., Grigoriev A.Y. and Kholodilov O.V., “Quantative analysis of surface topography using scanning electrom microscopy”, Vol.153, pp.119-133, 1992.

[1.22] Ryabukho V.P., Tuchin V.V. and Ulyanov S.S., “Interferential methods of speckle optics in laser diagnostics o surface”, LAMILADIS’91 Intl. Workshop: Laser Microthechnology and Laser Diagnostics of Surfaces, Chernowsky, Ukraine, Apr. 1-19, 1991: SPIE Proceedings, Vol.1723, pp.143-151, 1991.

[1.23] Fan Y.Y. and Huynh V.M., “Investigation of light scattering from flat periodic rough surfaces for surface roughness estimation”, Precision Engineering, Vol.16, No.3, pp.205-211, 1994.

[1.24] Hancock P.J.B.,”Pruning neural nets by genetic algorithm”, Artificial Neural Networks 2, by Aleksander and J. Taylor, Elsevier Science Publishers, B.V., pp.991-994, 1992.

[1.25] Mirzai, A.R. ,Higgins A. and Tsaptsinos, D. “Techniques for the Minimization of Multi-Layer Perceptrons.” Engng. Artif. Intell. Vol.6, No.3, pp. 265-277,1993.

[1.26] Simon, E.W., and Carter, J., "Removing and Adding Network Connections with Recursive Error Minimization (REM) Equations." SPIE, Vol.1294. Applications of Artificial Neural Networks ,pp.601-606, 1990.

[1.27] Santini S.,"The bearable lightness of being reducing the number of weights in backpropagation networks", Artificial Neural Networks 2, by Aleksander and J. Taylor, Elsevier Science Publishers, B.V., pp.139-142, 1992.

[1.28] Won Y. and Pimmel R.,"A comparison of connection pruning algorithms with backpropagation training", Proceedings of Artificial Neural Networks in Engineering, (ANNIE'91), StLouis, Missouri, pp. 112-118, Nov. 10-13, 1991.

[1.29] Saratchandran P.,"Dynamic programming approach to optimal weight selection in multilayer neural networks", IEEE Transactions on Neural Networks, Vol.11, No.2, pp.239-243, 1990.

[1.30] Karnin, E.D., "A Simple Procedure for Pruning Back-Propagation Trained Networks." IEEE Transactions on N eural Networks, Vol.l, No.2, pp.239-243. June, 1990.

[1.31] Kamruzzaman J., Yukio Kumagai and Hikita H.,"Study on minimal net size, convergence and generalization ability of heterogeneous backpropagation network"Artificial Neural Networks 2, by Aleksander and J. Taylor, Elsevier Science Publishers, B.V., pp.203-207, 1992.

[1.32] Katayama M. and Kawato M., "Parallel-hierarchical neural network model for motor control of a musculo-skeletal system", Systems and Computers in Japan, Vol.22, No.6, pp.95-105, 1991.

[1.33] Yamamura A.A. , Slerides A., Ji C, Psaltis D., "Neural network control of a two-link manipulator", Proceedings of 29th IEEE Conference on Decision and Control Part 6, Honolulu, Dec. 5-7, 1990.

[1.34] Ananth S. and Richard J. M.,"Optimal pruning of neural tree networks for improved generalization", Proceedings of International Joint Conference on Neural Networks, Seattle - WA, Vol. II, pp.229-224, 1991.

[1.35] Youngjik L., Sang-Hoon O. and Mymung W.K.,"The effect of initial weights on premature saturation in back-propagation learning", Proceedings of International Joint Conference on Neural Networks, Seattle - WA, Vol. I, pp.765-770, 1991.

[1.36] Abraham A., Niccolai M.J. and Gharpure P.K., "A neural network simulation in ADA with application to process control problems", Proceeding of 21st Annual Pittsburgh Conference, Vol.21, part 2, pp.553-538, Pittsburgh, USA, May 3-4, 1990.

[1.37] Hutton L.V., "Using statistics to assess the performance of neural network classifiers", John Hopkins APL Technical Digest (Applied Physics Lab.), Vol.13, N0.2, pp.291-299, 1992.

[1.38] Patrikar A., "Dual networks and their pattern classification properties", IEEE Comput. Soc. Conf. Computer Vision Pattern Recognition, Publ. by IEEE Service Center, Piscataway, NJ, USA, pp.686-687, 1991.

[1.39] Murphy, O.J., "An information theoretic design and training algorithm for neural networks", IEEE Transactions on Circuits and Systems, Vol.38, No.12, pp.1542-1547, 1991.

[1.40] Fine, T.L., Turmon, M.J., "Sample size requirements of feedforward neural network pattern classifiers", Proceedings of the 1993 IEEE International Symposium on Information Theory, Publ. by. IEEE Service Center, Piscataway, NJ, USA, pp.432, 1993.

[2.1] Shaw M. C., "Metal cutting principles", Clarendan Press, Oxford, 1984.

[2.2] ANSI/ASME B4 6.1, "Surface texture (surface roughness, waviness and lay)", The American Society of Mechanical Engineers, New York, 1985.

[2.3] Peters J., Vanherck P. and Sostrodinoto M., "Assessment of surface topology analyses techniques", CIRP Annals, Vol.28, No. 2, pp.539-554, 1979.

[2.4] R468, "Surface roughness ", International Standards, 1960.

[2.5] Bowker R.R.,"Exploring surface texture", Rank Taylor Hobson Ltd., London, 1980.

[2.6] Montgomery J. W., "Introduction to statistical quality control", John Wiley and Sons, New York, 1985.

[2.7] Pham D.T. and Oztemel E.,"Control chart pattern recognition using nueral network", J. Systems Eng. Vol.2, pp.256-262, 1992.

[3.1] LabTender Users Manuel, Scientific Solutions Inc., New York, 1981.

[3.2] R.S. Compenents, Product Cataloque, R.S. Component ltd., Box 253, Birmingham B8 1BQ, UK.

[3.3] Gorur A., Ciftlikli A., Yilbas B.S., and Sahin A.Z.,"Measurement of variation in optical properties of fiber cables - produced by HESFIBEL- subjected to mechanical bending", SPIE Proceedings, Vol. 2101, pp. 774-781, International Symposium on Measurement Technology and Intelligent Instruments, Wuhan China, Oct.29- Nov.5, 1993.

[3.4] Beckmann P., "Scattering o light by rough surfaces ", Progress in Optics, E. Wolf Ed., Vol. VI, pp. 55-69, 1967.

[3.5] Coleman H.W. and Steele W.G., "Experimentation and uncertainty analyses for engineers", J. Wiley Sons, New York, 1989.

[3.6] Beckwith T.G., Marangani R.D. and Lienhard V.J.H., "Mechanical measurements", Addision-Wesley Publishing Comp., New York, 1993.

[3.7] Kline S.J., "The purposes of uncertainty analysis", J. Fluids Eng., Vol.107, pp. 153-178, 1985.

[3.8] Lassahn G.D., "Thermal-hydraulic uncertainty analysis and measurement techniques", Transaction of American Nuclear Society, Vol. 44, pp. 610-611, 1983.

[3.9] Smith R.E. and Wehofer S., "From measurement uncertainty to measurement communications, credibility and cost control in population ground test facilities", Trans. ASME, J. Fluids Eng., Vol. 107, pp.165-172, 1985.

[4.1] - McCulloch. W.S., & Pitts, W.. "A logical calculus of the ideas immanent in nervous activity", Bulletin of Mathematical Biophysics, 5, 115-133, 1943.

[4.2] - Neumann, J.Von, "Probabilistic logic and the synthesis of reliable organisms from unreliable components", The Hixon Symposium, New York, John Wiley & Sons, 1951.

[4.3] - Minsky, M., "Neural nets and the brain-model problem", Ph.D. Dissertation. Princeton University, 1954.

[4.4] - Rosenblatt. F. ,"Two theorems of statistical seperability in the perceptron", In Mechanization of thought processes: Proceedings of a symposium held at the National Physical Laboratory, November 1958. Vol. 1 (pp. 421-456) London: HM Stationery Office, 1959.

[4.5] - Rosenblatt, F., "Principles of neurodynamics", New York: Spartan, 1962.

[4.6] - Minsky, M., & Papert, S.,"Perceptrons", Cambridge, MA: MIT Press, 1969.

[4.7] - Widrow, G., & Hoff, M.E., "Adaptive switching circuits", Institute of Radio Engineers, Western Electronic Show and convention, Convention Record, Part 4, 96-104, 1960.

[4.8] - Anderson, J.A., "Two models for memory organization using interacting traces" Mathematical Biosciences, 8, 137-160, 1970.

[4.9] - Kohonen T.,"Associate memory – a system- theoretical approach", 1st Ed., Springer, New York, 1977.

[4.10] - Grossberg, S. "Adaptive pattern classification and universal recording, Part 1. Parallel development and coding of neural feature detectors", Biological Cybernetics, 23, 121-134, 1976.

[4.11] - Hopfield, J.J., "Neural networks and physical system with energent collective computational abilities", *Process of the National Academy of Sciences, Vol. 81*, pp 3088-3092, 1984.

[4.12] - Reumalhart, D.E., & McClelland, J.L., "An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model", Psychological Review, 89, 60-94, 1982.

[4.13] - Hinton, G.E., & Sejnowski, T.J.,"Analysing co-operative computation", Proceedings of the Fifth Annual Conference of the Cognitive Science Society, 1983.

[4.14] - Levine D.S.,"Introduction to neural and cognitive modeling", Lawrence Erlbaum Associates Publishers, New Jersey, 1991.

PAPERS PUBLISHED AND PRESENTED BASED ON THIS RESEARCH

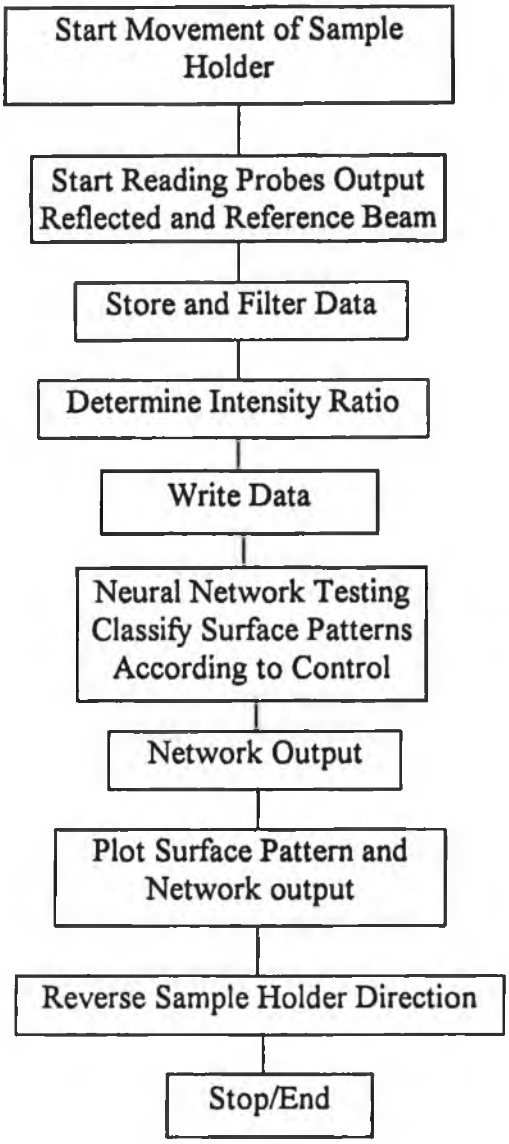
1 - Yilbas Z. and Hashmi M.S.J.,"An optical method and neural network for surface roughness measurement", Optics and Lasers in Engineering (in press), 1998.

2 – Yilbas Z. and Hashmi M.S.J., "Simulation of weight pruning process in backpropagation neural network for pattern classification: a self-prunning threshold approach", Computer Methods in Applied Mechanics and Engineering, (in press), 1998.

3 – Yilbas Z. and Hashmi M.S.J., "Surface roughness measurement using an optical system", Journal of Materials Processing Technology, (in press), 1998.

4 – Yilbas Z. and Hashmi M.S.J., "Simulation of weight pruning process in neural network – a threshold approach", submitted to International Conference on Advances in Materials and Processing Technologies AMPT'98, Kuala Lumpur, 24-28 August 1998.

APPENDIX 1 – Flow chart of the program developed for probe reading.



APPENDIX 2 – Flow chart of the program used in network training , testing and pruning.