Hidden Markov models for tool wear monitoring in turning operations

Gideon van den Berg

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Hidden Markov models for tool wear monitoring in turning operations

by

Gideon van den Berg

A dissertation submitted in partial fulfillment
of the requirements for the degree

Master of Engineering

in the

Department of Mechanical and Aeronautical Engineering
Faculty of Engineering, the Built Environment and Information Technology

University of Pretoria
Pretoria

July 2004

Synopsis

Author	Gideon van den Berg
Supervisor	Prof P.S. Heyns
Department	Mechanical and Aeronautical Engineering
Degree	M.Eng

The classification of the condition of a machining tool has been the focus of research for more than a decade. Research is currently aimed at online methods that can process multiple features from more than one sensor signal. The most popular technique so far has been neural networks.

A new technique, very popular in speech recognition namely, hidden Markov models has recently been proposed for studies in classification of faults in mechanical systems. Hidden Markov models have excellent ability to capture spatial as well as temporal characteristics of signals, which is harder to do with neural networks.

This study applies the techniques of hidden Markov models to turning operations from strain signals recorded on a tool holder during cutting. Two classes of tool condition, "sharp" and "worn" are appointed in the data. A hidden Markov model is trained for each class and classification is done.

From unseen data the "sharp"-model achieved a 95.5% correct classification and the "worn"-model achieved a 94.5% correct classification. This is compared to a maximum likelihood classifier that achieved a "sharp" classification of 96.8% correct and a "worn" classification of 72.7% correct.

Dimensional reduction was done on the feature space extracted from the data in order that it may be used by the hidden Markov model. This technique shows how multiple features from more than one sensor signal can be used by a hidden Markov model for robust recognition.

KEYWORDS: dimensional reduction, hidden Markov model, HMM, principal component analysis, PCA, strain signals, turning, tool wear, tool condition monitoring.

Sinopsis

Outeur	Gideon van den Berg
Promotor	Prof P.S. Heyns
Departement	Meganiese en Lugvaartkundige Ingenieurswese
Graad	M.Ing

Klassifikasie van die werkstoestand van snygereedskap in die vervaardiningsindustrie is al vir meer as 'n dekade die fokus van navorsing. Huidige navorsing konsentreer op prosesse wat die seineienskappe van meervoudige sensors aanlyn kan verwerk. Kunsmatige neurale netwerke is op die oomblik die mees populêre tegniek wat hiervoor gebruik word.

Baie onlangs is 'n tegniek wat algemeen vir outomatiese spraakherkenning gebruik word genaamd, verskuilde Markov modelle, voorgestel vir klassifikasie van foute in megnaniese stelsels. Verskuilde Markov modelle se vermoë om die temporale en ruimtelike kenmerke van seine vas te vat maak hulle baie geskik vir die taak.

In hierdie studie word tegnieke van verskuilde Markov modelle toegepas op vervormingsseine vanaf 'n beitelhouer tydens 'n snyproses op 'n draaibank. Twee toestande naamlik, "skerp" en "stomp" is aangewys vanuit die data. 'n Verskuilde Markov model is opgelei vir elk van die twee toestande.

Die modelle is getoets met data wat nie vir die opleiding gebruik is nie. Die "skerp" model het 'n korrekte klassifikasie van 95.5% behaal terwyl die "stomp" model 'n korrekte klassifikasie van 94.5% behaal het. Hierdie resultate is vergelyk met die van 'n maksimum waarskynlikheid klassifiseerder. Dié tegniek het 'n korrekte klassifikasie van 96.8% behaal op "skerp" beitels en 72.7% op "stomp" beitels.

'n Tegniek van dimensionele reduksie is gebruik om die dimensionaliteit van die seineienskappe te verminder, sodat dit deur die verskuilde Markov model gebruik kon word. Hierdie tegniek toon aan hoe seineienskappe van verskillende sensors deur 'n verskuilde Markov model gebruik kan word vir 'n kragtige klassifikasietegniek.

SLEUTELWOORDE: beitelslytasie, dimensionele reduksie, draaiproses, PCA, toestandmonitering, verskuilde Markov model, vervormingsseine, HMM

Acknowledgements

I would like to thank the following people:

- Professor Stephan Heyns for his belief in me and the lively guidance for the project.
- Dr Cornie Scheffer for the solid foundation he left for the project as well as the suggestion to apply hidden Markov models.
- Frans Windell, Jan Brand and At du Preez for technical expertise with strain gauges, preparation of the workpieces, etc.
- AIDC for financial support of this project.

Gratitude toward my friends, especially: Reghard, Dewald, Servaas, Flubber, Mariechen and Dirk, for forming and inspiring me. And Carl for the help with LaTeX 2_ε.

Also I would like to thank my parents for love and support and an almost blind belief in me.

Finally in humble submission, utmost gratitude to my Saviour and Lord, Jesus Christ for the talents and abilities with which He has graced me.
James 1:17 (Afrikaans version)

> "Elke goeie gawe en elke volmaakte geskenk kom van Bo. Dit kom van die Vader wat die hemelligte geskep het, maar wat self nie soos hulle verander of verduister nie."

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of symbols

Acronyms

TCM	Tool Condition Monitoring
HMM	Hidden Markov Model
AR	Auto Regressive
ARMA	Auto Regressive Moving Average
RMS	Root Mean Square
NN	Neural Network
KBES	Knowledge Based Expert System
ANNBFIS	Artificial Neural Network Based Fuzzy Inference System
ANN	Artificial Neural Network
RF	Radio Frequency
SOM	Self-Organising Map
BMU	Best Matching Unit
DWT	Digital Wavelet Transform
SSE	Sum of Squares of Error
EM	Expectation Modification
PDF	Probability Density Function
AI	Artificial Intelligence
DHMM	Discrete hidden Markov model

Mathematical symbols

A	State transition probability matrix
a_{ij}	State transition probability
B	State probability distribution matrix
b_{ik}	$i-th$ state $k-th$ symbol emission probability
π	Initial state distribution vector
λ	HMM model
O	Observation sequence vector
o_i	Observation
α	forward probability
T	Time vector
t_i	Time instant
N	Integer denoting number
P()	Probability
σ	Standard deviation
x	sample
S	Skewness (Third statistical moment)
K	Kurtosis (Fourth statistical moment)
CF	Crest factor
E	Shannon Entropy
D	Dynamism
Ψ	Energy contained in a frequency band
fl	Lower frequency band
fh	Higher frequency band
S_x	One sided power spectral density
M	Dimensionally reduced feature set
P	Transformation vector
f	Frequency
f_s	Sampling frequency

CHAPTER 1

Introduction

1.1 Background

Economic forces drive the need for high availability of machining equipment, and demands high quality of machined parts. Tool Condition Monitoring (TCM) is a means to this end. Through the optimised use of cutting tools and process monitoring, TCM supports this trend of economic forces. A spin off of TCM is of course the potential for substantial cost savings in terms of less scrapped parts and more efficient use of expensive machine tools. The lack, on the other hand of a proper TCM may include excessive power take-off, inaccurate tolerances, serrations and an uneven workpiece surface finish. This may eventually lead to machine tool and/or machine peripheral damage, according to Dimla (2000). Research into these systems has been continuing for some time now and as sensor and computing technology have advanced, their presence is starting to be felt in industry. To quote Byrne et al. (1995):

> *Despite the huge amount of research, not many of these strategies for TCM have found their way into commercial products. This is mainly due to the following:*
>
> - *The nature of machining processes, which can be complex and chaotic.*
> - *Non-linear relationships between tool wear and process parameters.*
> - *Changes in sensor signals due to tool wear are very small in some cases.*
> - *An adequate sensor that can satisfy all the requirements for TCM does not exist yet.*
> - *A number of different tool wear modes exist which cannot always be monitored with the same strategy.*

The above quote is given from an international perspective. In a small country like South Africa this lack of TCM is even more apparent. According to Scheffer (2003), South Africa does not have any commercial tool condition monitoring systems currently installed anywhere in the country. The reason for this is that manufacturers consider the currently available systems still too unreliable and/or too expensive. The need for cheap and efficient TCM systems is clear. Scheffer et al. (2003) developed such a system and have proved it to be both effective and cheap. As research progresses, more and more techniques become available for process modelling and monitoring. This forces us to review our current methods and explore new options that are made available as time progresses. This dissertation will do just that, and explore TCM using hidden Markov models.

The methods used in this dissertation belong to the continuous, indirect methods of TCM. There exist a number of philosophies on how TCM should be done. The first two schools of thought are continuous and intermittent TCM. The former advocates that monitoring should be done continuous, while the latter encourages monitoring at intervals (e.g surface finish of every 10th component manufactured). The next level of separation is that of the type of monitoring scheme used, direct or indirect. Direct monitoring is concerned with volumetric loss at the tool tip. This may be done by electrical sensing methods or visually. Indirect methods seek patterns in sensor data from the process, e.g. torque on spindle increasing when a cutting tool becomes blunted. A taxonomy is presented in figure 1.1, which should give a brief overview of some methods in TCM. Most research in TCM have gone into continuous systems and only the continuous branch is expanded in the figure.

Various authors (Byrne et al. (1995); Scheffer and Heyns (2001) and Leem and Dornfeld (1996)) state that the establishment of a TCM system can be divided into a number of stages:

1. Sensor selection and deployment
2. Generation of a set of features indicative of tool condition
3. Classification of the collected and processed information to determine the amount of tool wear.

In the next three subsections, these stages will be elaborated on.

1.1.1 Sensor selection and deployment

Byrne et al. (1995) has described the requirements for a tool condition monitoring system. This is listed in table 1.1 on the selection and deployment of sensors for TCM.

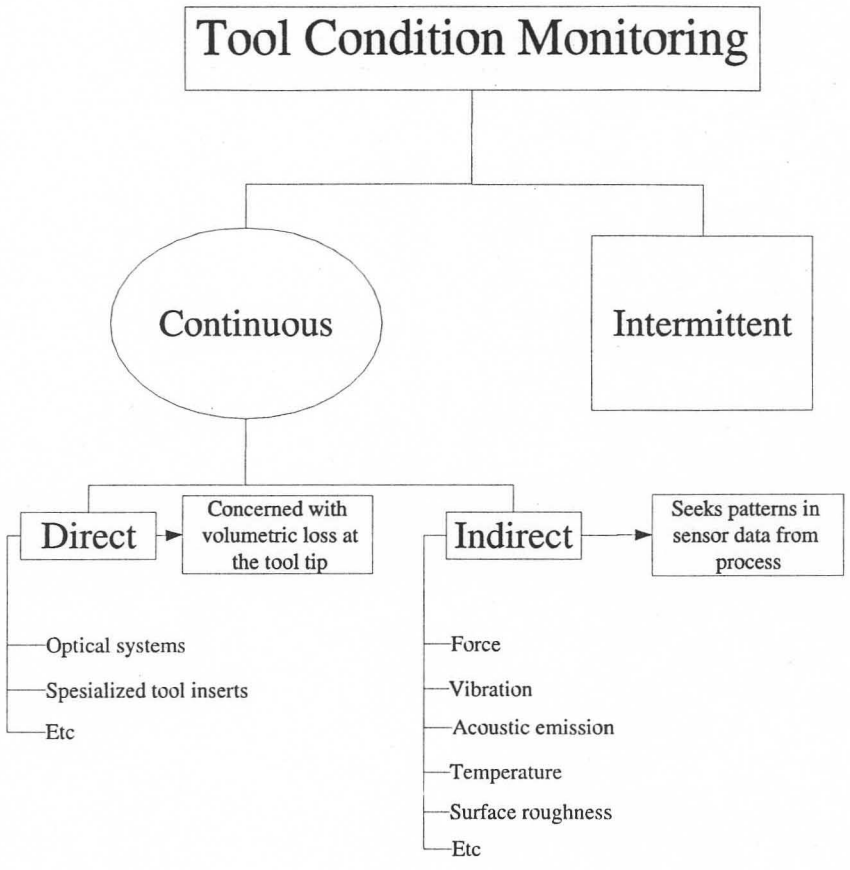

Figure 1.1: A taxonomy of continuous tool condition monitoring systems

Table 1.1: Requirements of a TCMS

Requirements
Measurement as close to the machining point as possible.
No reduction in the static and dynamic stiffness on the machine tool.
No restriction of working space and cutting parameters.
Wear and maintenance free, easily replaceable and cost-effective.
Resistant to dirt, chips and mechanical, electromagnetic and thermal influences.
Function independent of tool and workpiece.
Adequate metrological characteristics.
Reliable signal transmission, e.g. form rotating to fixed machine components.

Table 1.1 puts in full view what is ideally expected of a sensor and how it should be employed. A real sensor system will always end up as a trade-off between performance and cost.

1.1.2 Generation of features sensitive to tool wear

This subsection deals with the generation of suitable features that are indicative of tool wear and will be continued in the next chapter on the theory of feature extraction and selection. Features are also referred to as monitoring indices. This requires that one keeps in mind the disadvantages of using certain sensors on a TCM-system (e.g. the reduction in stiffness of the tool holder when using a dynamometer). Some common monitoring indices are listed in table 1.2

Table 1.2: Common features for TCM

Common Features
Mean
Variance
Root mean squares (RMS)
Skewness
Kurtosis
Crest factor
Power in a specific frequency band
Auto Regressive (AR) and Auto Regressive Moving average (ARMA) coefficients
Wavelet packet energy

Sensor fusion is also applied in order to get the most from the measured data. During sensor fusion the signals from different sensors are combined. According to Dimla (2000), sensor fusion serves the following purposes:

- Enhances the richness of the underlying wear-level information contained in each signal.

- Increases the reliability of the monitoring process as loss of sensitivity in one signal could be offset by that from another.

1.1.3 Classification of signals to establish tool wear

In this stage a signal model classifies the features from the collected data to obtain a useful conclusion about the tool life. This is a decision making technique. There are a variety of methods, which include trending, threshold and force ratios methods. One such is by Choudhury and Kishore (2000) who used different force ratios. Artificial intelligence (AI) approaches, however are arguably the most popular method currently used for signal classification. Park and Kim (1998) provide an introduction and a review of the use of AI. Broadly, the methods of classification can be put into two categories:

1. Weighting methods. Which include:
    - Neural networks (NN). This seems to be the most popular method because of its robustness to noise and its ability to handle more than one simultaneous input and to extract underlying information. An excellent review of online and indirect tool wear monitoring methods with artificial neural networks was done by Sick (2002).
    - Fuzzy logic. Fuzzy systems have the advantage of being able to directly encode structured knowledge. A number of fine articles can easily be found on the web. One such is by Li and Elbestawi (1996).
2. Decomposition methods. These are:
    - Signal understanding. Signal understanding is a technique based on the blackboard system, which was an artificial intelligence technique created in the 1980's. Du (1999) gives an application of signal understanding in tool condition monitoring.
    - Decision trees
    - Knowledge-based expert system (KBES)

These decision making techniques are often combined, so as to ensure a more robust output from the decision making algorithm. An excellent example of this is the work done by Balazinski et al. (2002). In this work, a fuzzy inference system and a backpropagation network were compared with an Artificial Neural Network Based Fuzzy Inference System (ANNBFIS). The neuro-fuzzy system was found to be quite adequate for wear prediction because of its short training time.

The three stages highlighted above are also given in a schematic form in figure 1.2. This is the generic form of a TCM setup.

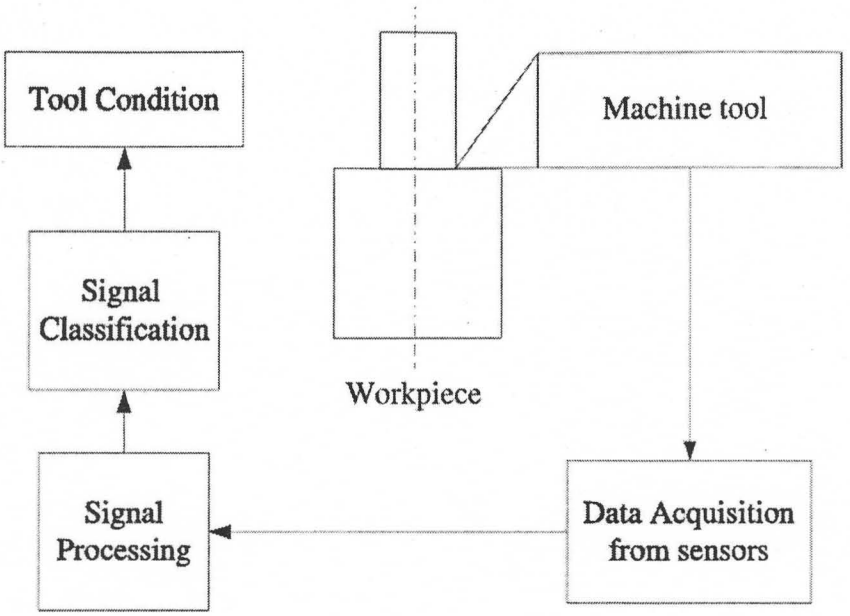

Figure 1.2: A generic TCM system setup

1.2 Complexity

An important issue which must be addressed is that of complexity. Does a TCM system really have to be so complicated, could similar results not be achieved by a simpler system? The answer to this is negative. There are numerous reasons for this; the non-linear nature of the machining process and the information lost in sensing and the signal processing corrupts the monitored indices. Measured signals are also not only correlated with tool wear, but also with machining conditions. On the other hand, monitoring tool indices that are only related to tool wear (methods of direct monitoring) are very expensive. Another reason is that the definition of tool condition is typically vague. There are also a number of different tool wear patterns, each with its own characteristics. Condition indices are also usually very small changes in processes with very wide dynamic ranges, which make them very hard to track. This, as Byrne et al. (1995) has stated, is why TCM has not yet properly found its way into commercial systems.

1.3 Some trends in tool condition monitoring

According to Sick (2002) the most popular method for tool condition monitoring in turning operations are methods that use neural networks. One reason for this is because of the emphasis that has been placed on online, indirect methods for TCM. Another reason is that usually during monitoring, several process parameters have to be measured and evaluated. Neural networks provide a very natural way to do this.

Sick (2002) also states that the most popular sensor signals are cutting force signals and the second most popular are vibration signals. Most of these sensor signals are from the cutting force and almost as much from the feed force. The reason why the monitoring of cutting forces is effective is because it provides a direct link to the cutting tool and workpiece interaction.

Also clear from the literature survey from Sick (2002), is the fact that most research into online and indirect tool wear monitoring has gone into systems that only classify wear. The author states that systems that use only two states may be sufficient to establish a practical tool monitoring strategy.

Currently (as previously mentioned) in TCM, neural networks that use a feature set generated from fused signals from the process have become the "state of the art." Because the understanding of cutting processes and neural networks has increased, the way in which neural networks are applied has moved toward the continuous wear estimation. Practical systems have recently achieved by Scheffer et al. (2003) and Balazinski et al. (2002). Scheffer et al. (2003) used a neural network configuration proposed by Ghasempoor et al. (1999) and reviewed by Sick (2002). Balazinski et al. (2002) used a neuro-fuzzy system and compared it to plain neural and plain fuzzy techniques.

The reason why practical continuous wear estimation has only recently been achieved is provided by Leem and Dornfeld (1996). The authors have also identified that the main problem with on-line systems is the problem of feature selection and suggest an unsupervised method. Indeed the greatest problem with most classification systems or methods are that they are sensitive to cutting conditions. Scheffer et al. (2003) also suggest that research into TCM using NNs be focused on this area.

Silva et al. (1998) has shown that there exists a zone of influence where NNs are insensitive to a change in cutting parameters. The network recognition then performs adequately for system conditions for which it was not trained. This zone is small but usable according to the authors. The authors experimented using Adaptive Resonance Theory (ART) and Self Organising Maps (SOM) network paradigms.

In answer to this problem, methods for various force ratios have been proposed by Choudhury and Kishore (2000), Novak and Wiklund (1996), Lee et al. (1998). Empirical formulae were created by Choudhury and Kishore (2000) and Novak and Wiklund (1996) for the prediction of tool life. Lee et al. (1998) continues from the force ratios to train an 1-step-ahead ANN predictor to forecast tool wear.

1.4 Document overview

The document is divided into the following sections.

1. Introduction. An overview on TCM and the associated problems.

2. Literature. Trends in TCM as well as presentation of work done on measuring equipment, specifically tool holders. Work done on hidden Markov models as mechanical fault identification is also shown. Based on this the scope of the research is defined.

3. Theory. Basic knowledge on hidden Markov models is supplied as well as the process of feature selection and extraction.

4. Experiment. This elaborates on the detail of the equipment used in the experiment as well as the operating parameters.

5. Results. The results of the applications of the theory on the experimental data are presented.

6. Conclusion. The results are discussed and suggestions are made with regard to future directions which the research might take.

CHAPTER 2

Literature Study

An overview of work with relevance and/or similarity to this project is presented in this chapter. Reviews will be categorised into:

The tool holder and the integration of sensors into it.

Hidden Markov models and Condition monitoring. Finally the scope of the present research is presented.

2.1 A sensor integrated tool holder

Looking at table 1.1, it is not hard to deduce that in turning operations, a sensible location for sensors would be on a tool holder of some sort. This is why a lot of work has been done in this area. From this work force and vibrations on the machining equipment have been noted to be the most sensitive carriers of tool wear information. The literature is explored with regard to measurement techniques on and around the tool holder. Work in the literature can be subdivided into two parts, authors who have developed tool holders with:

- sensing capability only, as well as
- sensing and actuating capability

This distinction is important since these tool holders were clearly designed with different goals in mind. Sensor/actuator tools are usually developed for active vibration control applications. Tools with only sensing ability are designed for monitoring. Both can however have very useful roles in TCM.

A sensor integrated tool holder that uses strain gauges to measure cutting forces has been developed by Santochi et al. (1996). The latest version of the tool incorporates

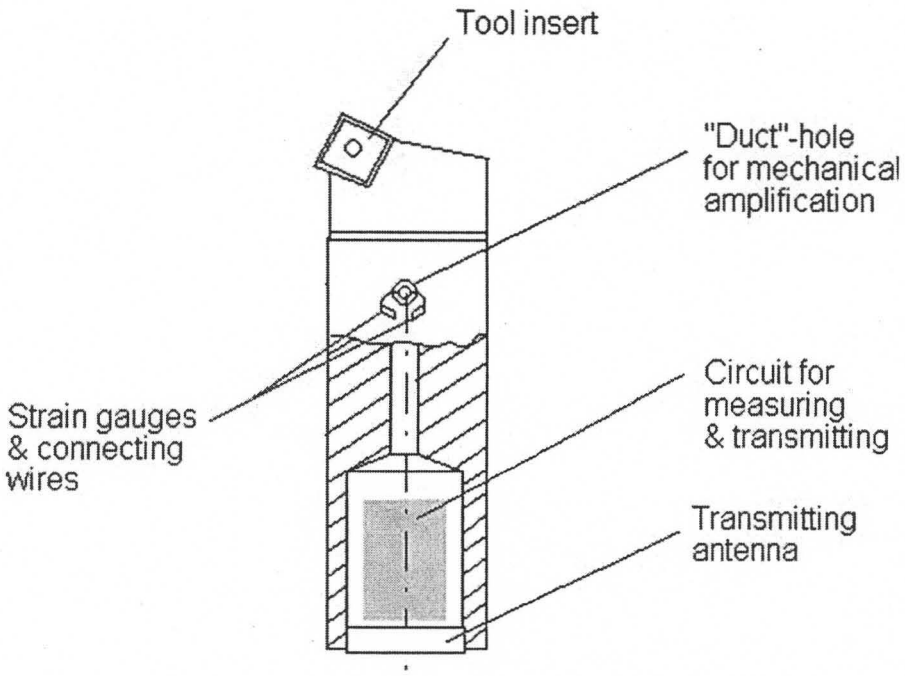

Figure 2.1: The tool holder by Santochi et al. (1996) uses strain gauges to measure cutting force.

a very clever technique of mechanical amplification, whereby the stress concentration caused by a hole in the tool is used to uncouple some of the measured forces. The hole is actually a "duct" for wiring of the strain gauges to the inside of the tool holder where the electronics are housed. The sensing tool uses a transmitter (RF) to transmit the strain to a computer. This tool is only capable of sensing and has no actuating capability. The tool is shown schematically in figure 2.1. It is not mentioned whether these holes in the structure significantly reduce the stiffness of the tool holder. A smart cutting tool for in-line boring was produced by Min et al. (2002). Feed force is measured using a piezoelectric actuator. This piezoelectric element gives the tool the ability of actuation. The actuator is used to compensate for the increased compliance of a long boring bar without support. A capacitance proximity sensor is used as an observer for controlling of the actuator. The actuator can unfortunately measure only force in one direction because of the flexure hinge mechanism (see figure 2.2).

A project which has been going on for a number of years and which is now evolving into a commercial product, is a chatter control system for turning and boring applications by Lägo et al. (2002). (see figure 2.3). The tool holder is capable of sensing as well as the active control of machine tool vibration. Because of patent rights very little is revealed of the inner working of the tool in the article. It is however mentioned that it uses piezo-ceramic actuators that were developed for the tool holder. The tool holder has a significant advantage in that the actuators are embedded in the shank of the tool

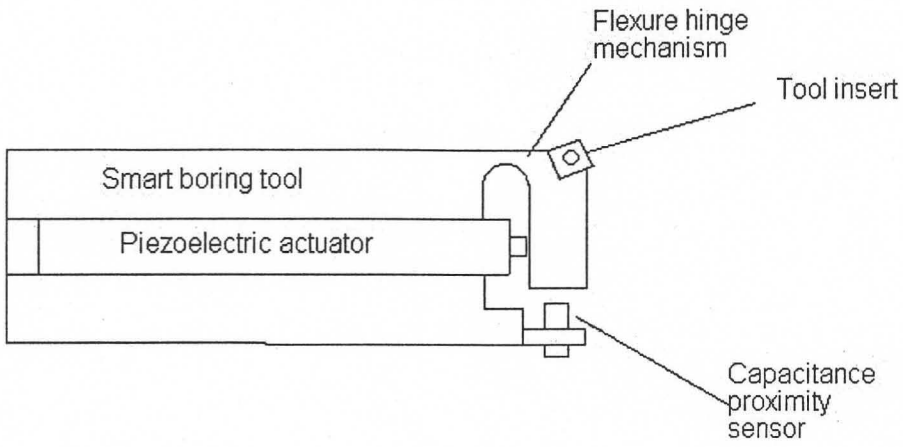

Figure 2.2: The smart tool produced by Min et al. (2002).

holder. This means that no alteration of the tool turret on the lathe is needed. It is not mentioned whether the tool can measure the forces in more than one direction. Håkansson et al. (2001) verified that the vibration pattern of a boring bar is usually dominated by the natural frequencies of the bar.

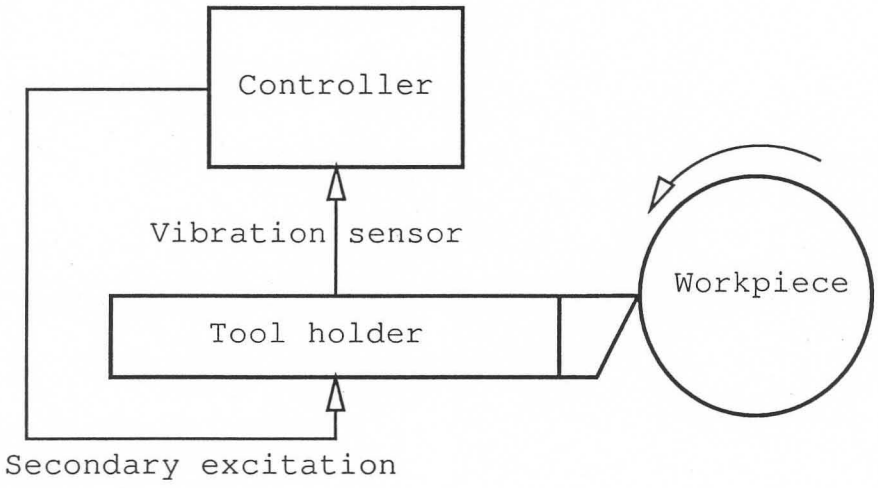

Figure 2.3: This is almost the generic setup for sensor/actuator tool holders. This is also the setup that Lägo et al. (2002), used.

Li and Ulsoy (1999) developed a method of high-precision vibration measurement of a beam using strain gauges. This method is based on the fact that the vibration displacement can be expressed in terms of an infinite number of vibration modes. Vibration modes can also be related to the measured strains through the strain-displacement relationship. By placing multiple sets of strain gauges on a beam, multiple modes could be taken into account to achieve high-precision measurement.

Scheffer (2003) suggested that the following issues should be addressed when one is constructing a sensor/tool holder with strain gauges. Keeping this in mind will allow for

the easy upgrading of the system into a commercial product.

1. Optimise the number, size and position of the sensors which are to be used on the tool.
2. If strain gauges are to be used, investigate the possibility of an on-board strain gauge amplifier on the tool holder. Mechanical amplification via stress concentrations should also be kept in mind.
3. Develop mechanical protection for sensors.
4. Investigate the industrial implementation of wireless data transfer.
5. Attempt constructing a sensor integrated tool for larger tool holders or holders that carry more than one tool.
6. Facilitate Internet monitoring capabilities.

A very well documented review of sensor signals for tool-wear monitoring in metal cutting was done by Dimla (2000). The author provides insights into different phenomena encountered with different monitoring techniques.

2.2 Hidden Markov models and condition monitoring

The first question that should be answered is why HMMs should be used for condition monitoring. According to Blimes (2002) most "state of the art" automatic speech recognition systems today are based on/use HMMs. HMMs provide a method for very robust classification of signals that are non-stationary. If it is considered that HMMs can correctly classify spoken words from time domain data from speakers with different voices, one can immediately see that this is indeed a very robust technique. In pattern recognition problems (such as TCM) there is always some randomness or incompleteness that is inherent to the sources. Byrne et al. (1995) places the signals from machining operations in this category by classifying them as typically chaotic and non-linear. Atlas et al. (2000) states that these signals require advanced classification procedures for monitoring and prognostication tasks.

The chaotic and non-linear nature of cutting signals implies that in the time domain these signals will be non-stationary. According to Rabiner (1989) the rich mathematical structure makes it possible for HMMs to easily handle non-stationary, chaotic data. This has also been confirmed by Bunks et al. (2000) who also agree that HMMs are well suited to handle quasi-stationary signals. Kwon and Kim (1999) state that NNs cannot provide proper solutions for temporal variations in data that are to be classified. The authors

also state that: *"The notion that artificial neural networks can solve every problem in automated reasoning, or even all pattern-recognition problems, is probably unrealistic."*

In the light of these facts it is clear that HMMs are very well suited for the purposes of tool wear classification although it has not been widely applied.

HMMs have only been used by a very small group of researchers and some of their work is reviewed in this section. For this section to be as lucid as possible one technique from the literature needs to be explained beforehand. This technique is called "scoring" and is used for classification.

2.2.1 Scoring of the forward probabilities

Assume a system needs to be monitored and that in this system, there are certain conditions which the user wants to be able to classify (e.g. immanent bearing failure; shaft unbalance; tool breakage; unacceptable tool vibration). Signals that carry information about the system condition can then be recorded and relevant signal features can be extracted. (Vector quantisation can then be done if needed). One HMM can then be trained for each system condition to be classified. Once the HMMs have been trained it is possible to calculate the forward probabilities of each HMM for a new signal that needs to be classified. The forward probabilities are a measure of how close the signal is to the training signals of a particular HMM. A signal with a high correlation to the training data of a HMM will produce a high forward probability and vice versa. The signal is then classified into the category of the HMM with the highest forward probability. Figure 2.4 shows such a classification system that uses "scoring". There is another classification technique that can be used with HMMs. This is called "alignment", this technique will not be discussed here as it is not implemented in this study.

2.2.2 Relevant literature

Ertunc et al. (2001) investigated two methods of using HMMs to establish the condition of a drilling tool. The first method is the bar graph monitoring of the HMM output probabilities. The second method is the multiple model method, whereby three different models were trained on data from a drilling process. Each model represented a different tool condition (i.e. one model was trained on sharp tool data, while another on workable and yet another on worn tool data). The recognition procedure used then was scoring. The data signals were typically force and torque data. The authors concluded that thrust was a better indicator of tool wear than torque for their particular experimental setup. It was also concluded that this technique was suitable for other machining operations as long as there are readily available data signals that are sensitive to tool wear.

Tool wear monitoring on milling processes using hidden Markov models have been done by Atlas et al. (2000). The evolution of vibration signals for the real-time transient

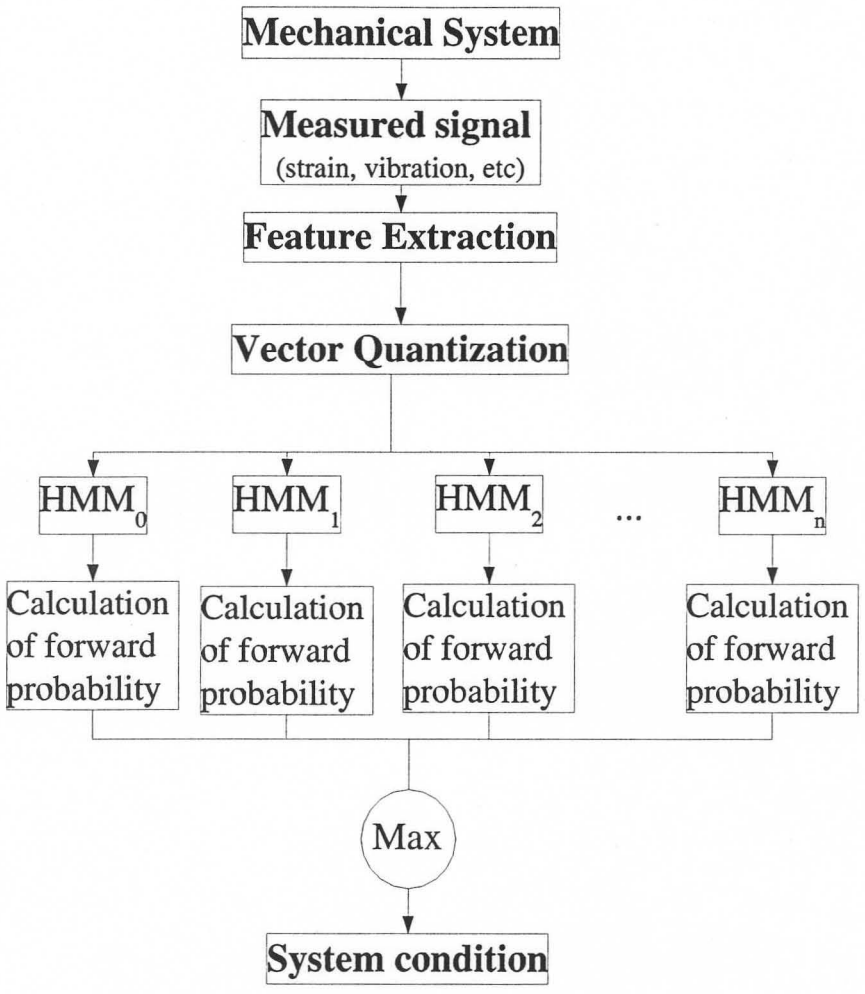

Figure 2.4: Hidden Markov model based fault diagnosis system based on scoring

classification of tool wear is done using HMMs. The authors stated that tool wear, although inherently continuous, can be represented at a quantised level with a HMM with a left-to-right state topology. The authors used an alignment method for the recognition procedure. Alignment is done with the Viterbi algorithm or a variant thereof. The Viterbi algorithm predicts what the most likely state sequence would be if a particular HMM were to produce a specific signal.

An end milling process was used where vibration was measured with accelerometers. The accelerometers were mounted on the spindle housing of a CNC machining centre in a climb-cutting process for machining notches in hard metal. The data was segmented into passes. One pass was defined as the period from when a tool touches the metal until it cuts air after it leaves the workpieces. Three time scales were investigated:

1. The progression of the tool from sharp to worn.
2. The dynamics of the tool in:
    - entering the workpiece
    - bulk machining
    - leaving the workpiece.
3. Very short potentially meaningful transients.

It was found that the HMMs trained with the very short transients did not generalise very well and that the models would have to be retrained each time that the tool was changed. For the intermediate time scale the HMMs achieved excellent classification in assigning binary ("worn" and "not-worn") wear labels based upon simple RMS energy and energy derivative features.

Another interesting application of hidden Markov models was done by Bunks et al. (2000), where HMMs were implemented for Condition-Based Maintenance (CBM). The objective was to:

1. Collect vibrational characteristics which correspond to physical changes (which indicate abnormal operation) in a machine.
2. Determine the statistics of this vibration data for various defects, either by modelling or by experiment.

The authors applied this to vibration data from a Westland helicopter gearbox. The measurements were taken with 8 accelerometers placed on the casing of the helicopter gearbox. Measurements were taken in a laboratory environment. Data consisted of 68 distinct operation conditions. These were obtained from 8 different seeded defects and 9 different torque levels. From the test it was concluded that the data is not stationary as

a function of operating torque levels. This is therefore an ideal place to apply HMMs. An HMM with 68 states was created for the classification problem. The state process models were 8-dimensional (because of the 8 accelerometers used in the experiments) Gaussian distributions. The distributions were estimated using the first 10000 samples from each of the operating condition runs. When trained the HMMs achieved very good classification of the data. The recognition procedure used was alignment. The classification of the HMMs is shown to be quite robust. Hidden Markov models are also shown to a provide natural framework for diagnostics and prognostics.

In more recent article on HMMs and mechanical systems, Lee et al. (2003) produced a rig on which rotor faults could be created. The authors concentrated on oil whirl and unbalance. Time signals were sampled and the autospectrum was used to train the HMMs. Both continuous and discrete HMM types were studied. The HMMs were then trained with a small set of data. The trained models were then scored with unknown data in order to classify the signal. It was found that the continuous HMMs give better results from scoring but the discrete HMMs give more robust and hence more consistent classification. The authors mentions nothing of the topology of the models that they used.

Kwon and Kim (1999) produced a high level fault detection for nuclear power plants using hidden Markov models. The authors advocate HMMs for their ability to model temporal as well as spatial information. Rapid accident identification in nuclear power plants is very crucial in order for authorities to select appropriate actions to mitigate the consequences of the accident. Signals from 22 different sensors are combined into a 1-dimensional signal using a self organising map (SOM). This is a technique for vector quantisation, where the input signals are shown to a fully-trained SOM and the Best matching Unit (BMU) is returned as an output. The best matching unit is simply the neuron that best matches the input signal. This sequence of BMUs produced is then used to train the HMMs. Several HMMs are trained, 1 for each accident type. The authors show that this technique correctly identifies the accident types.

In stamping processes Ge et al. (2003) have produced a hidden Markov model based fault diagnosis system. The system diagnoses 6 different operating conditions encountered during a stamping process. The system uses a strain signal from the press. An autoregressive (AR) model is fitted onto the signal from the press. The signals were detrended before this was done. The sum of squares of error (SSE) was used as the signal feature from which to train the HMMs for the diagnosis. The authors found that for their application, a 6-state HMM fitted onto the SSE of an AR-model of 8th order showed the best results. Classification results for the six different operating conditions ranged between 100% and 70%. The authors do not mention the state topology, but it is suspected to be ergodic. It was also found the HMM trained on the AR models showed only a marginal improvement over HMM trained directly on the signals.

Wang et al. (2002) used a 3-state ergodic HMM with discrete outputs to create a system for TCM. Accelerometers were mounted on the cutting tool holder to measure the vibration in the feed force direction. The HMM was trained on the coefficients of a Discrete Wavelet Transform (DWT) for 5 different scales which were derived form the acceleration data. The average energy of each scale was calculated and inserted into a feature vector. Vector quantisation was then done on this "scale-energy" vector and a codebook of size 10 was created. The features are then normalised to make them independent of the signal magnitudes. The codebook size is also equivalent to the number of distinct observation symbols in the HMM. Any input feature can then be represented by simply calculating the index of the pattern in the codebook that best matches it. This is done by using the Euclidean distance. A continuous cutting signal was then detrended and segmented into non-overlapping parts which were then quantised using the above procedure of DWTs. The HMMs were then trained and tested on the observation lengths of 5 observations. The HMM achieved a 97% correct classification of the testing data. The classification was a worn/sharp decision test.

2.3 Scope of the research

The use of HMMs is a very new technique in condition monitoring of mechanical systems. The models seem to have great potential in this field but research is only in the beginning stage and their actual worth will only be discovered as more study is done in this field. It is therefore proposed in this study to firstly create a wear classification system.

2.3.1 Summary of research goal

The aim of this research was to apply the techniques of HMMs to create a tool wear classification system for turning operations. This system should be able to distinguish between two classes of tool condition using signal features that are common in NN research for TCM. Attention will be paid to the following considerations.

- The ability to be to do a sharp/worn classification on sensor signals to bring the system in line with what has already been done.
- For the issue of operating system compatibility as well as continuity of the research, it was decided to use an open source software toolbox that runs on MATLAB to train and infer the HMMs. No custom algorithms were therefore used for the inferring of the HMMs. As it is the case with Wang et al. (2002) a HMM with a discrete output will be used.
- The same technique namely, the "scoring" of the forward probabilities of the HMMs

will be used as the method of classification. This is similar to many word recognition systems Rabiner (1989).

- Because HMMs have mostly been used for speech recognition, there has not been very much development in the use of multi-dimensional signals (speech signals are recorded with a single sensor). This project is therefore set aside by the fact that is uses dimensional reduction on multiple features. A framework is therefore created which can be used if more than one sensor is used and where multiple-features from the time and frequency domains will be used. A scheme for dimensional reduction will be used to fuse the features into a single dimension. This technique differs from the work of Bunks et al. (2000) in the fact that it does not use N-dimensional state process distributions.

2.3.2 Measuring of forces

Dimla (2000) states that the feed and radial forces are influenced more by tool condition that the cutting force itself. The feed force is also known to be rather insensitive to most cutting parameters. Therefore:

- the feed force will be monitored with strain gauges, and
- because of the constraint to measure as close as possible to the cutting process, as few as possible number of sensors will be applied to the tool holder. A single sensor consisting of two strain gauges in a rosette will be used for the data acquisition system. The use of a single sensor will make the system very minimalistic which is ideal for a first study.

With regard to the measurement techniques and sensors, this project is in line with current measurement techniques. It should however be stated that in the case where force measurements are made, that the norm is to use a dynamometer

2.3.3 More on features for HMMs

The nature of speech signals are so that HMM can be trained directly on the time domain data. This is not exactly the case with signals that are correlated to machine condition. The raw signals are usually not directly helpful for the classification of machine condition. An exception from this is the work done by Bunks et al. (2000) where classes could be identified more clearly. Usually for TCM where the changes in machine dynamics are more subtle, features which are sensitive to tool wear need to be extracted from the data. The signal features that have been derived by researchers of NN techniques have not been investigated in HMM research. These features have been tried and tested and are known

to work well. There is therefore a need to investigate (or at least demonstrate) the use of features from NN research in the application of HMMs. This study will do this by investigating some popular time and frequency domain features. Because DHMMs are used it implies that the signals have to be discretized. This will be done with a custom nearest neighbour algorithm which is very similar to a histogram algorithm.

CHAPTER 3

Theory

This section will provide an overview of aspects of the theory of the models and signal processing techniques used in this dissertation. This will include an explanation of, and an introduction to:

Hidden Markov models and how they are used for recognition

Signal processing, feature extraction and selection

3.1 Hidden Markov models

Physical processes generally produce observable outputs that can be represented by signal models. These models allow us to learn a great deal about the process, without having the actual signal source around. There are several choices for a user when it comes to the types of signal model that can be used to characterise the properties of the signal of interest. According to Rabiner (1989), signal models can broadly be divided into two groups:

- Deterministic models which exploit the known properties of the signal (e.g. the signal is a sine wave or a sum of exponentials).
- Statistical models where one tries to characterise only the statistical properties of the process. (e.g. Gauss processes, Poisson processes, Markov processes).

Under the statistical model it is assumed that the process can be described by a parametric random process for which the parameters can be estimated by means of a well defined formulation.

These signals can further be divided into discrete and continuous signals. Statistical models can also be stationary (statistical properties are time invariant) or non-stationary

(statistical properties vary with time). Hidden Markov models (or Markov source in older literature) falls into the category of non-stationary statistical models.

3.1.1 Defining the HMM

A hidden Markov model can be defined as finite state machine that functions in discrete time. Each state in the HMM contains the definition of some stochastic process (i.e. a Probability Density Function (PDF) or an AR-model). At each time step the HMM emits an observation from one of its states. A signal/observation sequence may then be produced by taking a random walk (defined by a Markov process) "within" the states. This random walk is dependant on the transition probabilities. To clarify this consider figure 3.1 which shows a network diagram of a 3-state HMM. The lines connecting the states (numbered $1-3$) represent state transition probabilities. The state transition probabilities are the probabilities that the HMM, currently in state i will transit to state j for the next time step. An HMM can also stay in its current state for the next time step. This is shown as little "loopbacks" on the figure. The HMM is therefore a doubly stochastic process in the fact that it is a random process for which the variables are determined by a random Markov process. The HMM is also in actual fact, a statistical signal generator although it is not used as a signal source. It is rather used as a vehicle for probabilistic inference. This will be explained later on in this chapter.

The reason for its name is that, during training the state sequence cannot be observed from the training sequences. The state sequence is therefore "hidden", hence the name. The training goal is therefore to infer the state sequence and to determine the state process parameters from the training sequences. It should be noted here that training sequences are the same as the observation sequences. Once the state transition probabilities and the state process parameters are determined the model can be used for classification. The technique for classification used in this study is called "scoring" and is described on page 13.

The emissions from the states can be of a continuous or a discrete nature. Discrete emissions are usually symbols while continuous emissions may be a real valued numbers within a certain range.

Definitions

The HMM used for this project will have discrete emissions and discrete states. This is a very specific subclass of HMMs and the interested reader should consult Elliott et al. (1995) for a more advanced and general description of HMMs [1] [2]. The notation used

[1] Some additional theory on HMMs can be found in appendix A

[2] These definitions are from Narada Warakagoda´s website at http://jedlik.phy.bme.hu/ ˜gerjanos/HMM/node3.htm.

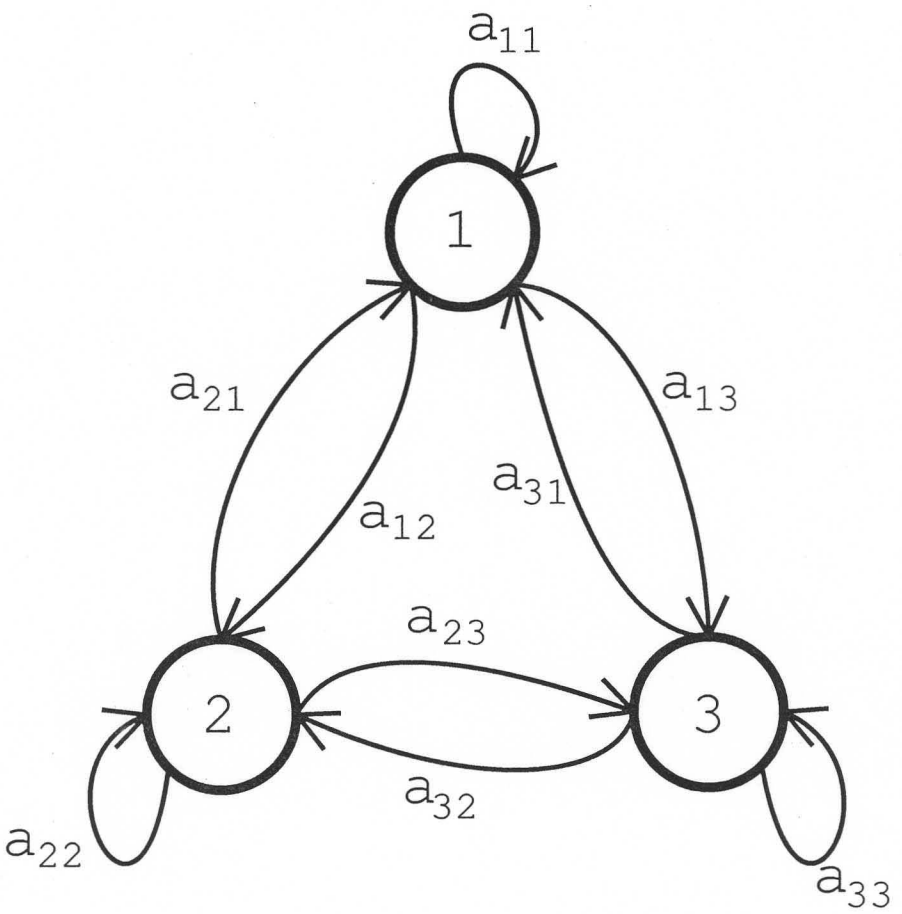

Figure 3.1: A directed state-transition graph of an ergodic 3-state HMM

throughout this text will be that of Rabiner (1989)[3]

An HMM is completely defined by the following parameters:

- State transition matrix, A. This defines the probability that the model, currently in state i will transit to state j for the next time step. This will be written as a_{ij}. The reader is again referred to figure 3.1. A will thus always be square and have the form:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots \\ a_{21} & a_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{3.1}$$

  The number of states that the model can then assume, N is equal to the number of columns in A. A is also subjected to the normal stochastic constraints namely:

$$a_{ij} \geq 0 \qquad \text{with } 0 \leq i, j \leq N$$

  and

$$\sum_{j=1}^{N} a_{ij} = 1 \qquad \text{for all } i$$

- Probability distribution for each state. This probability distribution will be denoted with B and is defined as follows: b_{ik} is the probability that the model, currently at state i will emit the k-th symbol in the defined alphabet of discrete emissions. As was previously mentioned the HMM will have discrete emissions defined within an alphabet with total number of M symbols. For a HMM with i states and M symbols, B will then have the form:

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{i1} & b_{i2} & \dots & b_{iM} \end{pmatrix} \tag{3.2}$$

  As with A, B is also subjected to the normal stochastic constraints:

$$b_{ik} \geq 0 \qquad \text{with } 0 \leq i, k \leq M$$

  and

$$\sum_{k=1}^{M} b_{ik} = 1 \qquad \text{for all } i$$

- Initial state probability distribution named π. π is the probability distribution that describes the likelihood that a HMM will start in state i. The normal stochastic constraints apply.

[3]This is also a very good starting place for readers who are new to the subject of HMMs.

Once A, B and π are defined, one has a complete HMM. To shorten the notation a specific model will be denoted as λ. $\lambda = f(A, B, \pi)$, viz. Given λ, the HMM can be used to generate a sequence of observations,

$$O = \{o_1, o_2, o_3, \ldots, o_T\}$$

O is a vector that contains the emitted observations from time, $t = 1$ to time $t = T$, with T being the length of the sequence in discrete-time. When an HMM is to be trained, the signals need to be segmented into these observation sequences.

3.1.2 The three problems of HMMs

Once one has defined a HMM, λ, there are certain things that one usually wants to be able to do with it. In HMM literature one will read of the three problems of HMMs, which describe what the HMM will be used for. These are discussed in Rabiner (1989) in the form of 3 problems. These are:

1. Given a HMM model, λ and an observation sequence, $O = \{o_1, o_2, o_3, \ldots, o_n\}$, how is the probability, $P(O|\lambda)$ efficiently calculated? ($P(O|\lambda)$ is the probability that the HMM, λ produces the emission sequence, O.)

2. Given a HMM model, λ and an observation sequence, $O = \{o_1, o_2, o_3, \ldots, o_n\}$, how is the state sequence, that in some way optimally describes the observation sequence, chosen?

3. How can the model parameters A,B and π be chosen so as to maximise $P(O|\lambda)$?

The solution to problem 1 is used in this dissertation to score HMMs. Consider the scenario where one has different competing models that describe an observation set. The solution to problem 1 can then be used to select the model with the highest probability of producing the observation set in question.

The solution to problem 2, called the Viterbi algorithm is not used in this dissertation and thus falls outside of the scope of discussion. The reader is referred to Rabiner (1989) and Bengio (1999) for an in-depth description of this procedure.

Problem 3 does not have a known analytical solution to choose the model that maximises $P(O|\lambda)$. This makes it the most difficult problem of the HMMs.

The Forward Procedure

It was mentioned previously that classification can be done with HMMs with a technique called scoring. This is a procedure where the probability is calculated that a given HMM, say λ_1, will emit a certain sequence. In order to do this one needs to calculate

the emission probabilities for the given sequence, for each possible state sequence. This quickly becomes intractable. Fortunately there exists an efficient recursive algorithm to do this. This algorithm is called the forward procedure and is discussed in Rabiner (1989). The result of the Forward procedure is called the forward probability and is denoted with an α.

For an HMM with N states and a sequence length of T the procedure works as follows:

1. Initialisation:

$$\alpha_1(i) = \pi_i b_i(O_1) \tag{3.3}$$

2. Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad \text{and} \quad 1 \leq j \leq N \tag{3.4}$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{3.5}$$

There is no analytical solution to show what number of states will produce the best HMM for a specific application. It can however be said that a model with more states may perform better. This is because the amount of states in the HMM is directly related to its ability to model signal non-stationarities. More states unfortunately require more training data which may be difficult to come by.

Another problem encountered with the calculation of probabilities using HMMs is that of underflow. The numbers tend to be extremely small, well under machine precision for most computers. For this reason the probabilities are scaled and use is made of logarithmic probabilities. As the name implies, the logarithm of the probabilities are calculated and used in the algorithms. The properties of the probabilities are now slightly different. Whereas in the normal case where probabilities lie between 0 and 1, logarithmic probabilities lie between $-\infty$ and 0. With HMMs it is usually not strange to work with probabilities in the range of -100, which is a very small number indeed!

Training the hidden Markov model

This is the most difficult problem of the HMM. According to Rabiner (1989) there is no known way to analytically solve for the model parameters that maximises the probability of the observation sequence. The most common technique usually employed is the Baum-Welch method which, locally maximises λ for $P(O|\lambda)$. This method is equivalent the Expectation-Modification (EM) algorithm, which is a maximum likelihood approach.

There also exists some gradient based methods but usually the EM algorithm is preferred for its fast convergence properties. The EM technique also guaranties a finite improvement on each iteration. Conditions can be formulated so that gradient based method can be applied to the HMM and this is presented by Rabiner (1989). Kwon and Kim (1999) have devised a method that uses the EM algorithm together with a genetic algorithm to train the HMM and to a select a state topology. Good results are achieved but training is slow.

As with neural networks, HMMs also have an architecture that needs to decided on, eg. the number of states and the state topology. Faced with this problem Bicego et al. (2003) presented a strategy to sequentially prune the number of states in an HMM.

It is important to know what is being done when one trains an HMM. Training implies that the parameters that define the HMM are updated. As mentioned previously these parameters are:

- the state transition matrix, A
- the emission probability density function for each state B
- the initial state distribution, π

To do this the α-parameter is once again used. Three other similar variables are also introduced in order to make training possible. It is because of these three other quantities that the training algorithm will not be shown here. A thorough description can be found in Rabiner (1989). Alternatively there is also a shorter version in appendix A.

3.2 Signal processing

In order for any intelligent system to be applied to the data, the data first needed to be processed into a different form that would be usable by the system. There are some similarities between speech data and vibration data and the signals processing techniques used on them. Bunks et al. (2000) compares speech data to acceleration data from a helicopter gearbox. This can unfortunately not be used directly because machining data is fundamentally different from acceleration data form gearboxes. An altered version will be presented.

Data from machining processes and speech are both quasi-stationary. The speech data however stays stationary over intervals of approximately $10ms$, according to Bunks et al. (2000). Cutting processes may be of one of two types. Interrupted cutting, in which the cutting tool is in contact with the workpiece for only a fraction of each revolution, produce signals which are stationary for intervals of milliseconds. Continuous cutting, where the tool is in contact with the workpiece for the whole period of each revolution, on the other hand produce cutting signals which are stationary for longer periods of time.

Another difference is that speech data is recorded in relatively "quiet" environments. Vibration data from working environments may, in the very worst cases, have a signal to noise ratio, orders of magnitude lower than that of speech data.

Another is difficulty is that changes in tool condition produce only slight changes in the response of the tool holder which are recorded. This necessitates the use of signal features which compress the information content of the signal. This is also why, in this study, features for NN studies will be investigated for the use of HMM applications. Therefore owing to the different nature of the data from speech signals, the raw signal was not used. The data had to undergo a number of preprocessing steps. These were:

- *Segmentation* of the raw signals into intervals for which the features are calculated.
- *Detrending*, which removes the most dominant linear trend from the data. This is usually done for FFT analysis. After this the observation sequences have a mean of zero.
- *Feature extraction* whereby the salient features of the signals are extracted.
- *Feature Selection*, is applied so that only the features with the most information with respect to tool wear is used.
- *Feature space reduction* which condenses the selected features into the final product which was a 1-dimensional feature vector.
- *Discretisation and Construction* of observation sequences. The signals are firstly discretized into a number of levels then consecutive samples from the feature space are constructed into rows of observation sequences of a specific length.

3.2.1 Feature extraction

In order to learn most about tool wear, certain features are extracted from the data. Each feature has a characteristic behaviour that can be followed over time to reveal information about the health of the tool. It is in this way that features will be used in this study.

The extraction of features also compresses the data into a form, which can be handled with much more ease and efficiency. This is important for real-time implementation, which is the longterm goal for any project that hopes to see an industrial application.

Two types of features were investigated, time domain and frequency domain. These two will be discussed in the sections.

Features in the time domain

Features in time are usually figures that one would normally find in most statistical analyses. As a tool wears, in the case of flank wear, the wear land increases. The interaction

surface of the workpiece and tool is then changed. This also alters the interaction of frictional forces between the two elements in the system. The result of this are changes in the dynamic characteristics of the system.

The features of the time domain are usually of a statistical nature. These features are also very fast to calculate which makes them very attractive for on-line applications. The interested reader may also review the implications of some of the statistical parameters used in a text such as Miller and Miller (1999).

The features that were investigated were:

- *Variance*, which is the second statistical moment of the data. Because of detrending the mean of the data is 0 which makes the variance of the data equal to the square of the RMS of the data. RMS is an indicator of energy content of a signal. As tool wear progresses, more energy is needed to drag the tool insert through the workpiece, it follows to reason that the RMS (or variance in this case) should increase. The variance is calculated using:

$$\sigma^2 = \frac{1}{T}\int_0^T x(t)^2 dt \tag{3.6}$$

  In equation 3.6 σ is the standard deviation. The variance is by definition the square of this. T is the time interval for which the integral is calculated. $x(t)$ is the signal for which the variance is calculated.

- *Skewness* is the third statistical moment and describes the distribution of the data in terms of symmetry or lack thereof, hence the term skewness. The skewness is calculated using:

$$S = \frac{1}{\sigma^3 T}\int_0^T x(t)^3 dt \tag{3.7}$$

- *Kurtosis* is the fourth statistical moment and is very popular in bearing condition monitoring. The kurtosis is a measure of the relative peakedness of the distribution, this is similar to the variance. The kurtosis is also a measure of how close the distribution is to the Gaussian distribution. It thus carries valuable information for condition monitoring. The kurtosis is calculated using:

$$K = \frac{1}{\sigma^4 T}\int_0^T x(t)^4 dt \tag{3.8}$$

- *Crest factor* is another feature which is widely used in bearing condition monitoring and is a measure of the impulsiveness of a vibration signal. A truly random signal has a crest factor generally less than 3. The crest factor is calculated using:

$$CF = \frac{X_{max}}{X_{rms}} \tag{3.9}$$

- *Entropy* is a measure of the uncertainty or disorder of of a given signal. One can intuitively see that a signal with a higher energy content, as in the case of a worn tool, will display more disorder. The entropy measure used was Shannon entropy which is often used in wavelet analysis. Shannon entropy is calculated using:

$$E = -\sum_{i=1}^{N-1} x_i{}^2 log(x_i{}^2) \tag{3.10}$$

  In 3.10, x_i is the value of x at time $t = i$. N is the number samples the feature is calculated for. This is the same as the time interval for the statistical features.

- *Dynamism* is a measure of the rate of change of a quantity. This feature also captures dynamic behaviour of a signal in a similar way to the crest factor. Dynamism was used for speech and music segmentation by Ajmera et al. (2003). Dynamism is calculated with:

$$D = \sum_{i=1}^{N} \Big[x_i - x_{i+1} \Big]^2 \tag{3.11}$$

Features in the frequency domain

Of the more salient features are usually those in the frequency domain. These features are directly connected to changes in the dynamic behaviour These are calculated from the one-sided power spectral density (PSD) using:

$$\Psi = \int_{fl}^{fh} S_x(f) df \tag{3.12}$$

In eq. 3.12 S_x is the one-sided PSD function and fl and fh are the frequency band for which this number is calculated. Ψ can increase, or decrease with increasing tool wear. The case where Ψ increases is where the cutting process changes from smooth cutting to a breakaway process. This causes an increase in vibration amplitudes. The case where Ψ decreases is where the dynamics of the process is altered so much by the change in the contact interaction caused by tool wear, that a shift in the peak occurs. When the peak starts to move out of the frequency band, the spectral energy decreases.

Håkansson et al. (2001) showed that the frequency bands that are most likely to show an increase, are those around the natural frequencies of the tool holder. Other characteristics of the cutting process, such as the chip forming frequency may also be monitored for signs of tool wear. On the whole there is not a method that can be used to predict which frequency bands are most likely to be useful in TCM. Allen and Shi (2001) suggested monitoring two frequency bands. A lower and a higher. The higher band then captures the natural frequencies of the system. Scheffer (2003), Lim (1993) and Jiang et al. (1987) have each derived their own frequency bands which were useful for their

work. These band are also process specific and also dependant on cutting parameters.

In this study frequency bands are also derived. This was done by hand and was therefore more of an art than a science. The approach that was used to find these frequency bands was firstly a summing algorithm. This algorithm summed the PSD functions of the tool during its lifetime. Peaks on this summed PSD show the regions where the most energy in the system is at. The search for relevant peaks were then focused on these areas. Two types of peaks in the energy spectrum my be found in these high energy regions:

1. peaks that are insensitive to tool wear and subsequently do not significantly increase or decrease during the life of the tool.

2. peaks that grow with tool wear.

It is because of this that the selection of frequency bands has not yet been automated.

3.2.2 Feature selection

Having extracted a number of features from the data, one usually wishes to reduce the number of features. This is because not all the features are sensitive to tool wear. To do this Scheffer (2001) proposed that the correlation coefficient be used for this selection process.

It is assumed that the progression of tool wear over time can be approximated by a straight line with a arbitrary gradient. This was chosen to be 40°. The correlation coefficient for each feature and the theoretical tool wear is then calculated. The correlation coefficient is a measure that describes to what degree certain values of one signal occurs with certain other values of another signal. The correlation coefficient is calculated using:

$$\mathrm{corr}(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2\right]^{1/2}} \tag{3.13}$$

X and Y are the two signals which are to be compared. $\overline{x}$ and $\overline{y}$ denote the mean values of the variables.

A value close to 1 is indicative that high values of one signal occurs with high values of the other signal. In the case where the correlation coefficient is close to -1, large values of one signal coincides with small values of the other signal.

Once the correlation coefficients have been calculated the highest ones can be chosen as the ones that carry the most information on tool wear. Correlation coefficients in the negative range are also very valuable because it guaranties the independence of features on each other. A combination of both was thus used for the recognition system.

3.2.3 Feature space reduction

From the theory of HMMs it has been implied that this technique uses 1-dimensional arrays for training and recognition. The theory of HMMs may be extended to use multidimensional arrays, but it was decided to use an existing HMM toolbox, feature space reduction is necessitated.

Dimensional reduction is a common technique in pattern recognition. These techniques reduce the dimensionality of the data for easier handling. According to Fugate et al. (2000) it is futile to expect good estimates from the tails of multidimensional data unless there is a very large amount of independent data available. This is what is referred to as "the curse of dimensionality.". The curse of dimensionality is simply that the amount of data required for training increases exponentially if the dimensionality is increased.

A simple and well known method namely, principal component decomposition was applied to the data. All the data is then projected onto the first principal component to reduce the dimensionality from N dimensions to 1 dimension. This was chosen conveniently in order to use the HMM toolbox directly on the application. If needed another set of HMM could be created. These HMM would then use the some of the other principal components. The output of the HMM committees could then be combined to form a more robust recognition. This study will use only one principal component to establish the technique.

The principal component analysis (PCA) is a standard function in the statistics toolbox for MATLAB that uses a singular value decomposition to calculate the principal components of a data matrix. The principal components can also be calculated as the eigenvectors of the covariance matrix of the feature space. The eigenvector with the highest corresponding eigenvalue will then be the unit vector of the first principal component. When the feature space is projected onto this vector it becomes the first principal component. The eigenvalues are then a measure of the total variance explained by each principal component.

3.2.4 Discretisation and construction

To accommodate the DHMM the dimensionally reduced feature is discretized into a number of levels. This is done with respect to the maximum and minimum values of the samples used for training. All the values that fall in-between these two values are rounded to the "nearest" level. These levels are similar to the bins in a histogram.

The observation sequences are constructed from the discretized feature vector. This is simply done by segmenting the feature vector into lengths of N consecutive samples. This number N is a parameter that determines how much temporal information is contained in the sequence. The strength of HMM recognition lie in these observation sequences.

CHAPTER 4

Experimental setup

In this chapter the experimental procedure will be explained and the equipment setup will be shown. The technique for wear measurement will also be explained.

4.1 The procedure

In order to take a tool insert through its natural life cycle a cutting experiment was set up. Cutting parameters were selected and were kept constant as much as possible. The experiment consisted of a number of cylindrical workpieces which were cut repetitively on a lathe with the depth of the cut being kept constant. A cool-down time of approximately 2 minutes was allowed for between each cut. The tool inserts were removed during certain intervals to measure tool wear.

4.2 The setup

The experiments were conducted on a Graziana Tortona SAG14 lathe. This was a "manual" lathe meaning that the machine is not of the CNC type. Operator experience therefore plays a role in the consistency of the data and should be kept in mind when the results are shown.

The type of cut is a very important consideration. During interrupted cutting, the shock impulses excite all the natural frequencies of the system. These natural frequencies are very strong indicators of tool wear. In the case of a continuous cut, the excitation of natural frequencies are not as prominent. This fact also complicates the matter of recognition for a TCM system and also influences the quality of the data.

Cutting was done using a boring bar. Boring bars are used to machine on the inside of a component. A boring bar usually has a more slender shape than a normal tool holder.

For this project the boring bar was not used for boring but for normal cutting. Boring bars are less rigid that normal tool holders for lathes, but it can still be used for normal cutting operations. (The availability of the bar also made it a very good choice.)

The boring bar was instrumented with strain gauges on one side. Figure 4.1 shows a schematic of the front end of the tool holder. This figure shows the approximate location of the strain gauge rosette. HBM 1.5/120XY91 strain gauges were used for the experiments. These strain gauges measure $1,5mm$ by $1,5mm$ with the complete padding and packaging measuring $3mm$ by $3mm$. Each strain gauge rosette contained two strain gauges orientated in perpendicular directions. The two strain gauges were connected in a half bridge configuration into strain gauge amplifiers. A Clip AE 101 strain gauge amplifier was used for this. The fact that the signals of two strain gauges measuring in two perpendicular directions were combined implies that sensor fusion was implemented. (see also figure 4.6)

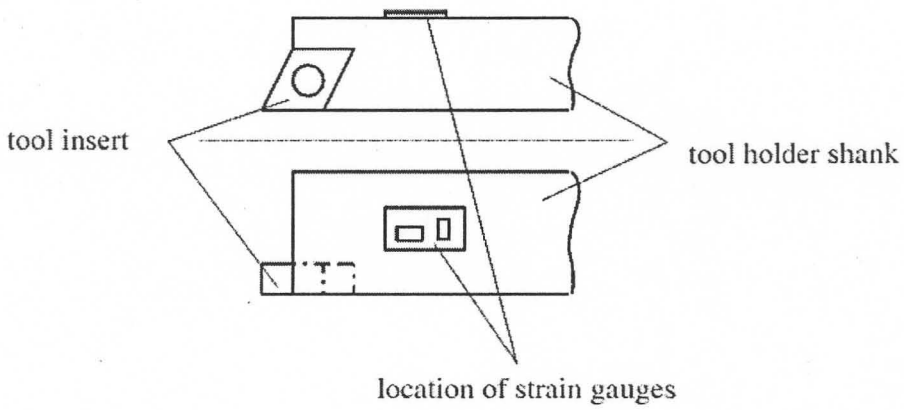

Figure 4.1: The approximate location of the strain gauges.

Low-pass filtering was done on the signal to prevent aliasing. The filter design was that of a $4th$ order Chebyshev type which had a roll off of $-3dB$ at $4350Hz$. The filter was built in-house and designed with the FilterLab Low Pass program[1]. The filtered signal was then captured on a personal computer in a MATLAB environment using the Data Acquisition Toolbox (DAQ toolbox). The DAQ toolbox allows for the easy creation of rather elaborate monitoring systems.

To automate the data processing as much as possible the data had to have a uniform structure. This means that all the recorded signals had to have the same length. In order to achieve this a triggering mechanism is needed. This trigger mechanism was created so that recording was started when the signal crossed a certain threshold. Recording was then allowed for a set amount of time before it was stopped. This recording time as set to be longer that than the cutting time of the each experiment run.

The experiment went as follows. The tool was set at the correct depth for the cut,

[1]Available from Microchip Corporation at http://www.microchip.com

but just "outside" the workpiece so that when the lathe is started and the autofeed is engaged, the tool will enter the workpiece within approximately 1 second. When the tool touches the workpiece, the recording is triggered and continued until the set time is exhausted. The signal is then stored on the computer. This is shown schematically in figure 4.2.

The analog to digital conversion was done with a National Instruments PCI-6024E analog to digital card. A schematic of the data acquisition system can be seen in figure 4.3. The signals were sampled at a rate of $f_s = 20kHz$.

In appendix D some photos are shown of the equipment.

4.2.1 Machining parameters

The purpose of the experiment was to monitor the progression of wear for a tool during a normal life cycle. Parameters were chosen so that they were to fall for the "medium" range for most tool inserts. As with some design situations some of the parameters were chosen arbitrarily for a first iteration. The machining parameters that were decided on are shown in table 4.1.

Table 4.1: The machining parameters for the experiment.

Machining Parameter	value	unit
feed	2	mm/rev
depth of cut	0.5	mm
cutting speed	120	m/min

By the nature of a cutting process, there is a fundamental problem with keeping experimental conditions constant. Each time a cut is made the workpiece looses $0,5mm$ from the circumference. This changes the circumferential velocity for the next cut. If experiments were to be kept 100% constant, cutting could only have been done on workpieces of exactly the right circumference. A lot of workpieces would therefore be needed for the experiments. This would have been a very expensive experiment. To counteract this problem it was decided to introduce a tolerance band of about 8% around the cutting speed. This means that for a certain rotational speed on the lathe, a certain amount of cuts could be made that all lie within the cutting speed band.

Tool wear is renowned for its dependence and sensitivity to machining conditions. It is therefore desirable to have cutting conditions that are not always exactly the same so that a classification algorithm that proves its effectiveness on those signals will also have proved its robustness a priori.

Typical shavings can be seen in figure 4.4. This shows a long continuous chip of about $1m$ length. The chip has a bright blue metallic colour, which is indicative of martensite

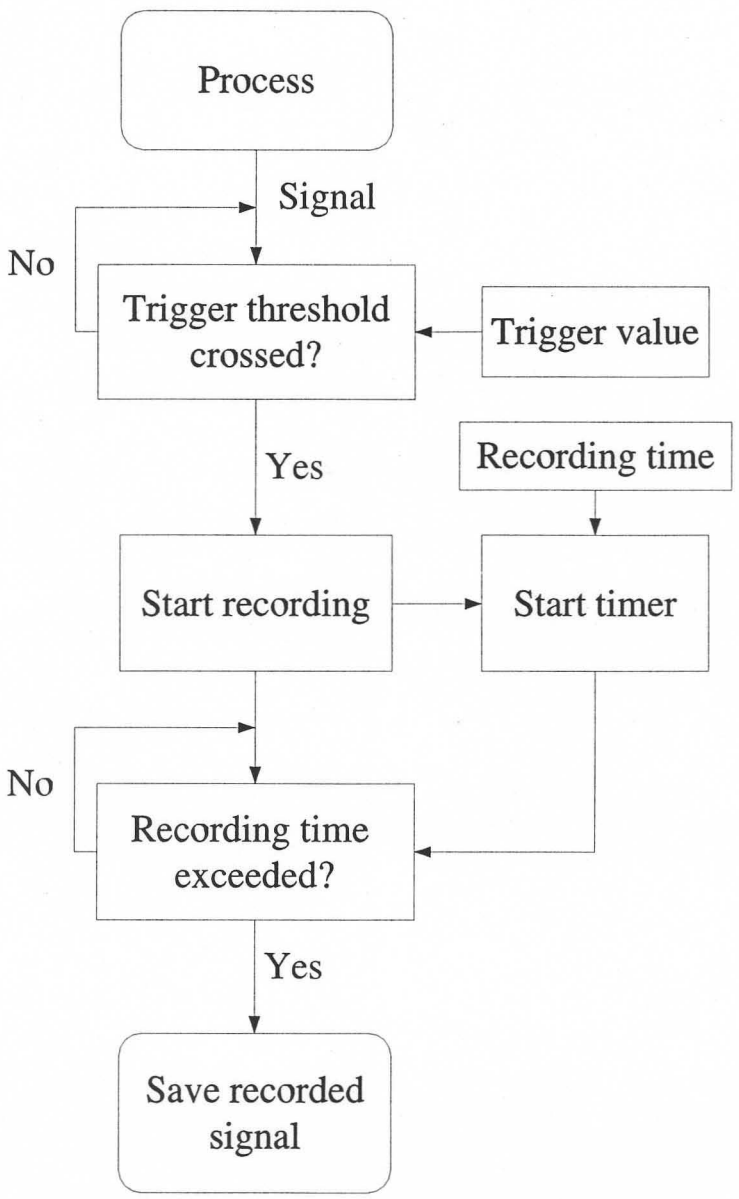

Figure 4.2: A schematic of the data acquisition program.

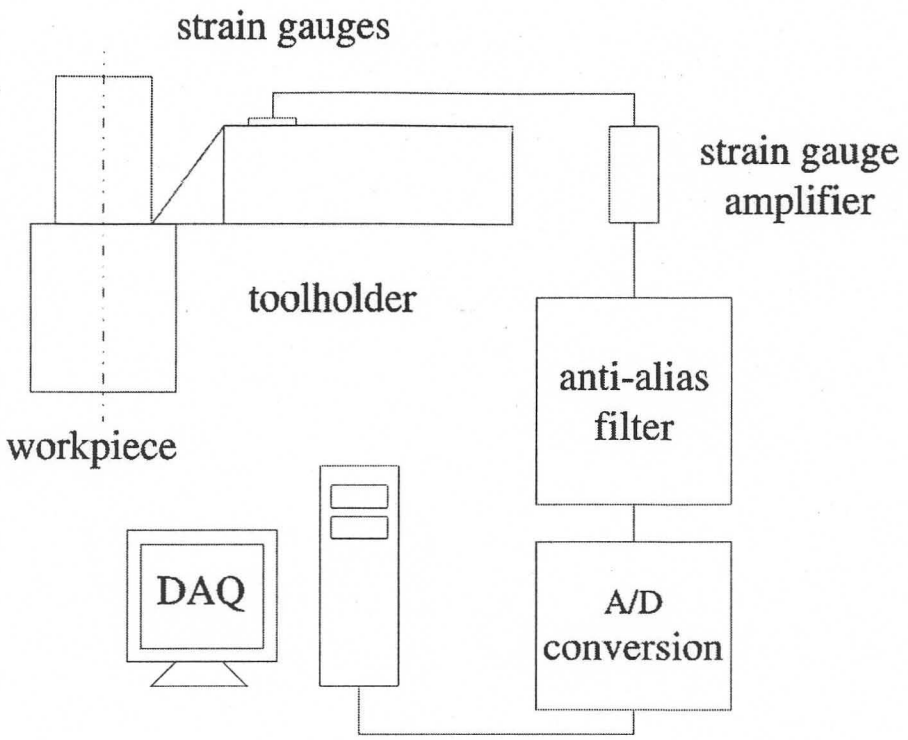

Figure 4.3: The schematic overview of the data acquisition system used for the experiments.

and subsequently a very high cutting temperature. This may also be because of the high carbon content of the workpiece material. This continuous chip is actually undesirable in real life situations according to (Cho et al., 1999). Continuous chips usually have an adverse affect and surface roughness and may cause tangling problems around the tool.

It is interesting to note that if the depth of cut was increased to $1mm$, the cutting chips would break up into small curls with lengths of about $25mm$. The colour of these chips were also the bright metallic blue. If the depth of the cut was however slightly decreased, the chips continued to be long but the colour turned silvery and bright. This silver colour is an indication that the tool is being utilised to its capacity.

As was previously mentioned, because the machine is operator driven, the quality of the data is dependant on the experience of the operator. The depth of each cut was measured directly after each cut with normal vernier callipers. This is shown in figure 4.5 in the form of a histogram of the depth of cut during the experiment. There is quite a large variance around the mean. This large variance has the same effect on tool wear as the cutting speed tolerance band.

The last machining parameter to take into account is that of cooling fluid. On recommendations from technical personnel it was decided to cut the material dry. No cooling fluid was used since dimensional stability was not an important issue. Also, the cooldown period and the continuous chip that transports the heat away from the workpiece, provided a steady experimental temperature.

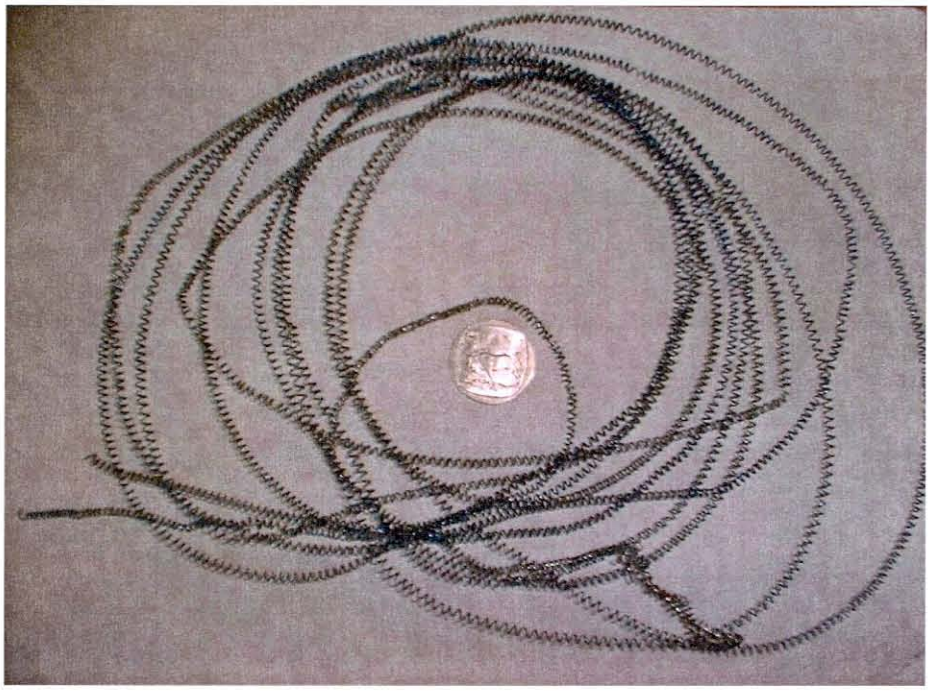

Figure 4.4: A typical shaving from a cut.

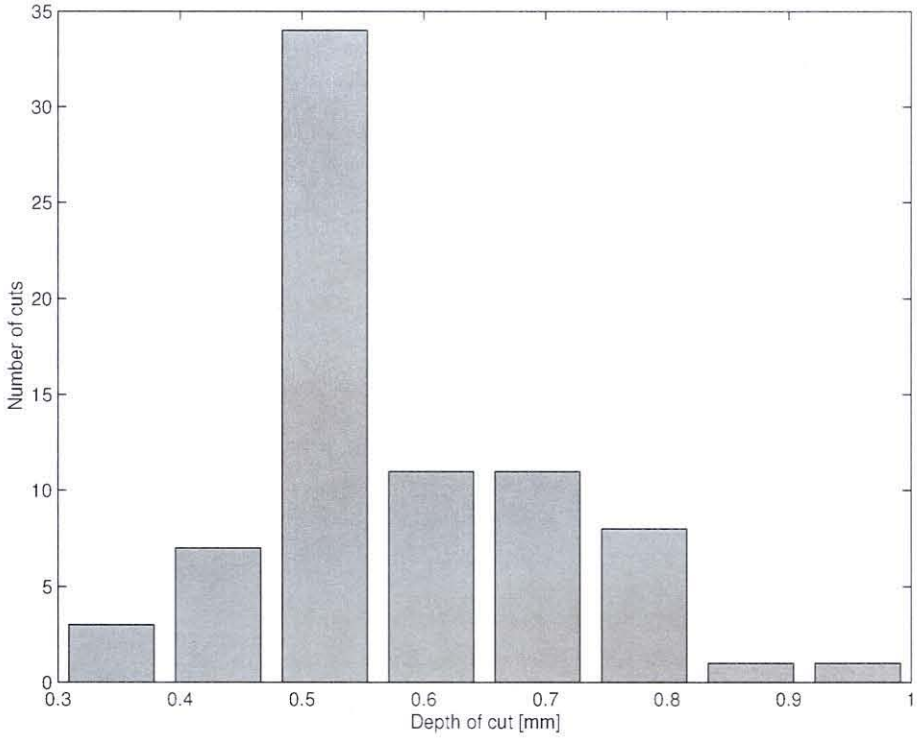

Figure 4.5: A histogram for the depth of cut.

4.2.2 The tool holder

Previously it was mentioned that a boring bar was used. This was a Mitsubishi S160 SCI PR09[2]. The boring bar was machined on the side and top so that there would be space for the stain gauges. The stain gauges were covered with an epoxy mixture to protect them from the machining environment. It will be assumed that the epoxy does not change the modal properties of the bar in any significant way.

In the setup the bar was given an overhang of $30mm$. This is almost the minimum amount that the geometry of the bar and the epoxy coating allows for.

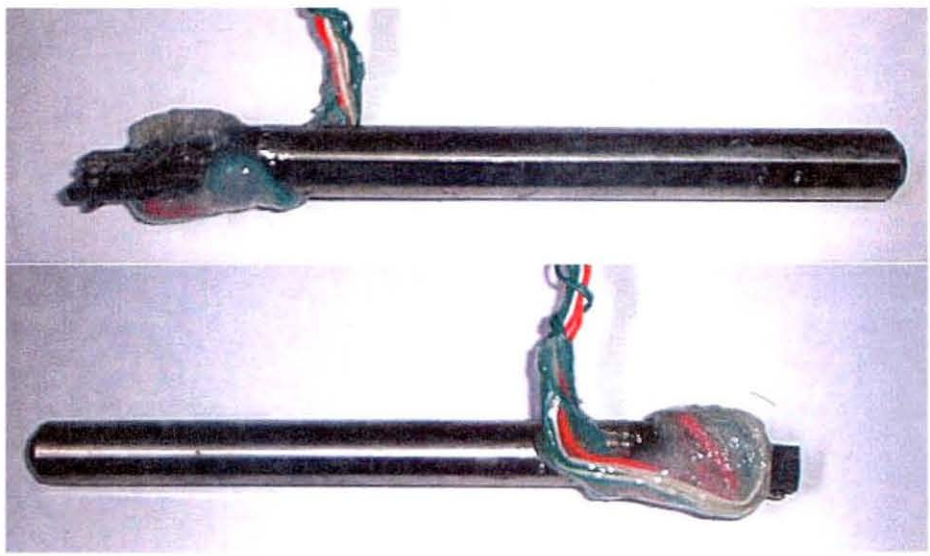

Figure 4.6: The boring bar was instrumented with strain gauges on one side.

4.2.3 The insert and measurement of tool wear

With the machining parameters listed on page 34, Mitsubishi suggests a medium finishing tool insert. It was decided from references from the Mitsubishi website to select a *US7020 MV* tool insert.

In order to measure tool wear during the life of the tool it was necessary to remove the tool for inspection during certain time increments. This was adjusted as experience was gained with the tool inserts. In the end, wear was measured after every $10 - 15$ minutes of cutting time.

The measurement of tool wear was done on an optical microscope. Since the machining parameters were chosen to be in the "mid range", efforts under the microscope were focused on finding traces of flank wear. Flank wear and nose wear are the most common forms of wear to be found on tools.

[2]Information on this boring bar can be downloaded in pdf format from: http://wwww.mitsubishicarbide.com

It was found from the measurements under the microscope that nose wear is the dominant wear mode for the machining parameters and workpiece combination. Nose wear is commonly found at low cutting speed and the mechanism of wear is caused by abrasion wear on the cutting tool's major edges. Tool sharpness is caused by plastic or elastic deformation of the cutting edge. A built-up edge may also be formed at low cutting speeds. In figure 4.7 a plan view the worn nose of the tool is shown. In contrast with this, figure 4.8 shows a new tool insert (this is however at a lower magnification).

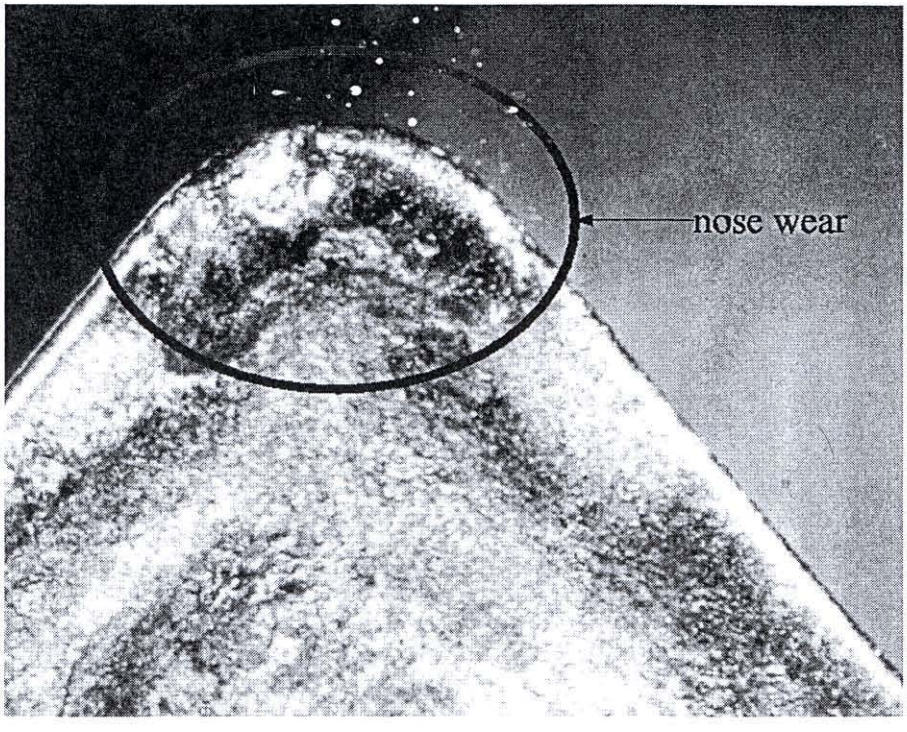

Figure 4.7: The nose of an insert under a microscope. Nose wear is shown on this photo.

In a similar situation to the dilemma of the changing cutting speed for each experiment, was that of the removal of the tool insert. The removal and re-insertion of the insert changes the dynamic characteristics of boring bar and insert system. This change is caused by the difference in clamping conditions at each "iteration" of the wear measurements. This again is once again not necessarily a bad thing, since it offers a chance to, at least in a qualitative manner, to prove the robustness of the recognition system which is to be implemented.

4.2.4 Machining material

The experiments were conducted on EN 19 alloy steel. This is a tough steel which is mostly used for shafts and gears. Because of its high carbon content this steel is ideal for hardening. This can sometimes present a problem for machining. If the material

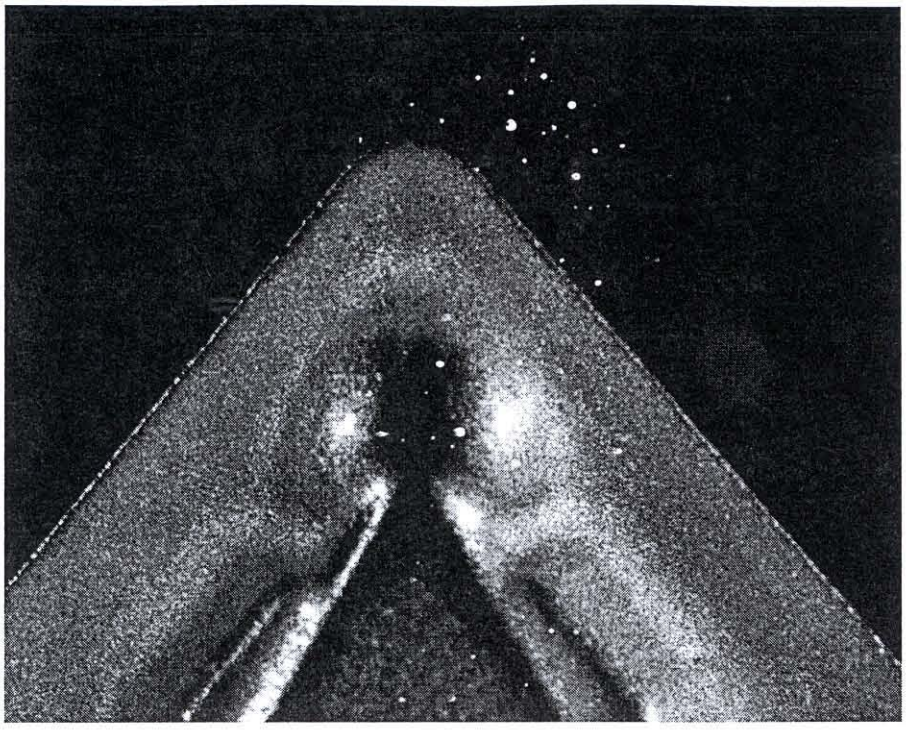

Figure 4.8: The nose of a new tool insert.

is not cold cut, it hardens and become almost unusable. This can also happen on the surface during cutting if the machining parameters are not correctly set. The mechanical properties can be seen in table 4.2.

Table 4.2: The mechanical properties of EN 19 steel.

Property	Limit	Elongation
Ultimate tension stress	1089 MPa	12 %
Yield stress	955 MPa	18 %

The steel was also oil quenched and tempered to the so-called T-condition, which had a Bernell hardness between $262 - 296BHN$. EN19 is a tough steel alloy and was chosen so that the "natural" life of the tool, in which we are interested would not be to long.

Workpieces were $300mm$, "roundbar" shafts with a $80mm$ diameter. The shafts were covered with a uncut material crust which had to be removed before the experiments could begin. After the removal of this crust the shaft had a diameter of approximately $75mm$.

CHAPTER 5

Results

The results from the cutting experiments are shown in this chapter. The techniques for signal processing shown in previous chapters are implemented and classification is done using HMM techniques. This is compared with results from a Bayesian classifier.

5.1 Wear progression

One of the problems with measuring in machining environments is the adverse conditions. To shield the delicate strain gauges an epoxy covering was applied over the gauges. This may sometimes have the effect that the strain gauge comes lose from the tool holder. This happens when the epoxy covering constrains the strain gauge during large strains and causes a complete tear of the strain gauge glue. The strain gauge is then completely lose from the tool holder. In such a case the measuring device becomes completely useless and has to be replaced.

Such a "release" of the strain gauge happened during this experiment. The photos from the microscope suggest that the tool has worn from $0mm$ to $o.1mm$ at its nose. This is a third of the usually allowable $0.3mm$ for flank wear and represents about a third of its useful life. It was decided to use the data from an incomplete tool life anyway. The rationale for this is, that if very accurate classification can be achieved at this stage, then surely better classification will be achieved with full tool life data.

If a full tool life is available then, using the same techniques, more wear levels can be appointed. This will ultimately point the system in the direction of a continuous wear estimator.

For the rest of the classification procedures, this wear level will be referred to as the "worn" condition.

5.2 Signal processing

5.2.1 The raw signal

Software for the recording of the signals were written in a such a manner that a software trigger can be set to initialise the recording process. This was set to start recording 1 second before cutting starts. Figure 5.1 shows a typical cutting signal in the feed direction. The little 1 second buffer in the beginning of signal can be seen. This part of the signal has its use to categorise noise of the system when no cutting is taking place. A noise filter can be implemented from information of this noise.

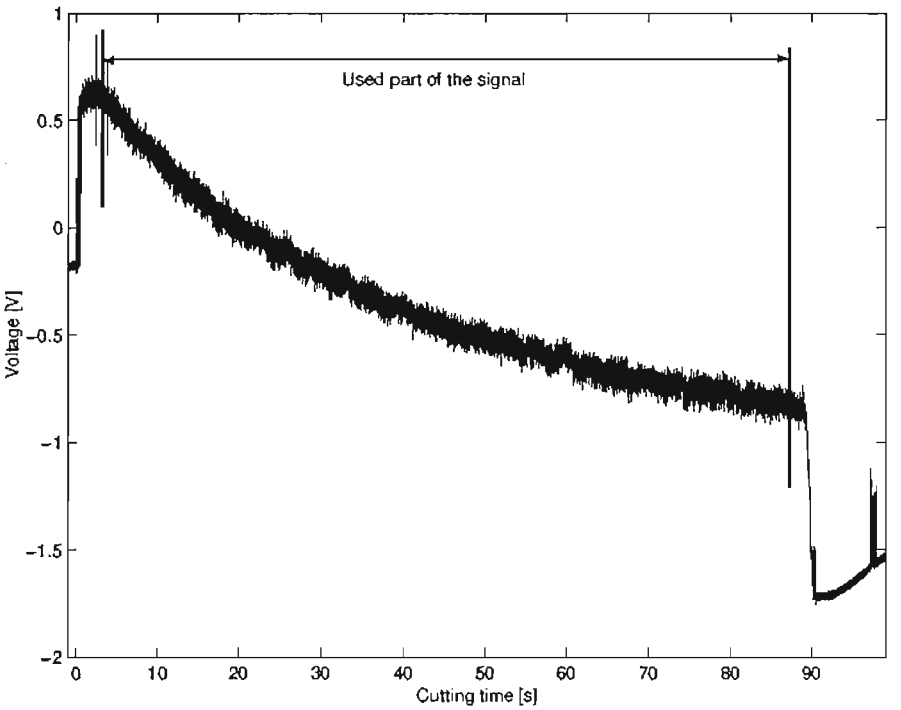

Figure 5.1: A typical cutting signal from the feed direction.

Since the cut is always made to be approximately the same length, a collar is formed on the workpiece. When the insert exits the workpiece, it "rubs" against the workpiece. The erratic nature of the last part of the signal is therefore caused by the exit procedure of the insert from the workpiece.

Another interesting observation from figure 5.1, is the effect of temperature at the tool tip on the signal. This effect can be seen as by the phenomenon that looks like an exponential decay on the signal. If recording of the signal was continued till long after the cut, the signal returned to a position very close to the original zero. This seemingly suggests a first order response which is indicative of temperature effects. This temperature response may be useful for the monitoring of tool wear in further studies,

but it is not nearly consistent enough. This "cutting temperature" is a strong function of the depth of the cut, and although it has been argued that inconsistencies in the data may prove the system to be more robust, the cutting temperature is considered to be too sensitive for practical use.

5.2.2 Segmentation and preparation

The signal shown in figure 5.1 is as such not yet very useful and still needs to be processed into a usable form. The first step was the removal of the temperature effects so that all digital drift effects are removed. This is done by removing a linear trend from the data so that the start and end of the signal are at the same voltage.

No use will be made of the transients at the beginning and end of the signal and they will also be removed. The transients may also contain valuable information but there are only two transients and their length is less than a thirtieth of the total signal length. The focus and emphasis will be on the continuous part of the cutting signal which is easier to monitor and segment.

Since the mean of the signal is very dependant on the depth of the cut it will also not be used for the processing since this features was very much influenced by the operators own expertise. This makes the segmentation and the removal of the temperature effects and the mean very easy. All signals will be segmented as shown into the useful parts as shown in figure 5.1. After this the signal is detrended thus removing the dominant linear trend. `Detrend` is a standard MATLAB function that is often used for processing of data for FFT analysis. This was done piece wise to ensure that the signal had a mean of zero. The remaining signal looks like figure 5.2. Figure 5.3 shows a magnified region of figure 5.2.

To show that these signals still carry information and are not just random noise signals, figure 5.4 is provided. This shows a scatter plot of two signals removed by some time. The plot has an oval shape which means that the variance of one of the signals has increased. This is also shown on the histograms plotted on the figure. These histograms have the tops of the bins connected to form a curve. They are also normalised in order to fit into the figure. These histogram have therefore no correlation with the figure axes. The two figures were normalised with the same factors. The figure was aimed to prove that there was still useful information captured in the signals.

To facilitate on-line monitoring the signal is, after detrending, segmented further again into smaller "snippets." It is on these snippets that feature extraction will be done. Each sample in the observation sequences is composed of the features of these snippets. Because frequency domain features are extracted, the snippets needs a certain length in order to contain a usable frequency resolution when the FFT is calculated. A length of $2^{11} = 2048$ was chosen, rather arbitrarily so that, with a sampling rate of $f_s = 20kHz$, a

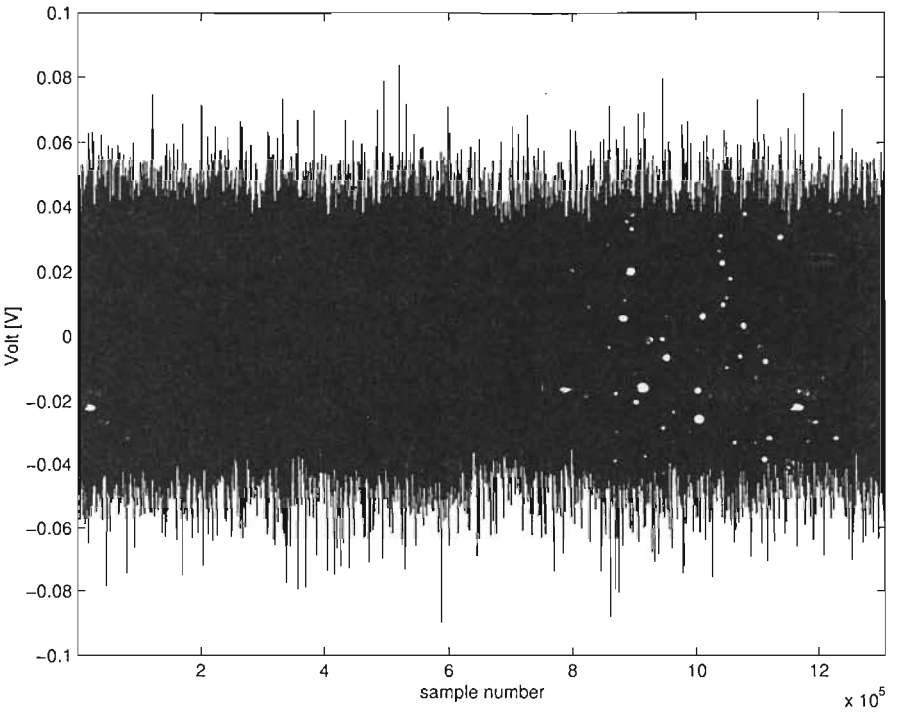

Figure 5.2: The final signal after segmentation and detrending.

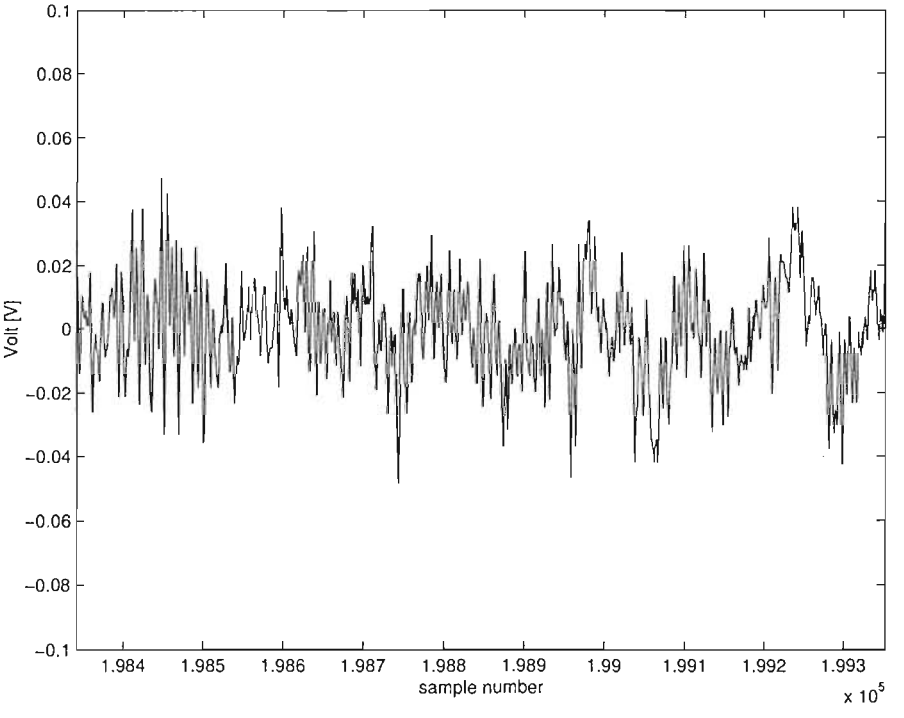

Figure 5.3: A magnified region of figure 5.2

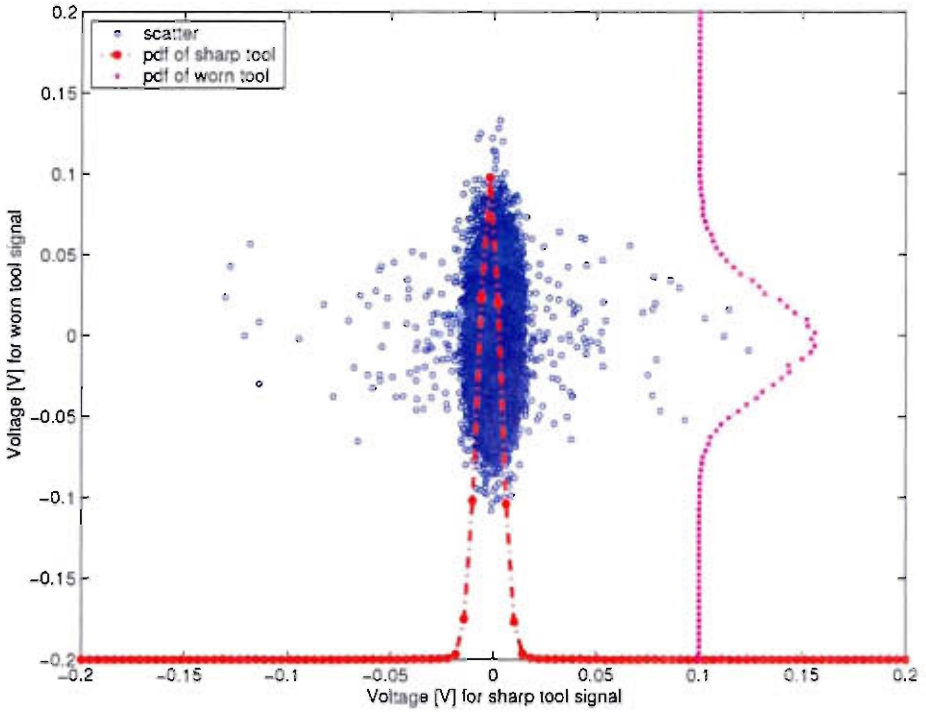

Figure 5.4: A scatter plot of two signal to show the increase in variance.

frequency resolution of $f = 9.76Hz$ can be achieved.

5.2.3 Critique on signal processing results and signal quality

It has been mentioned that the experimental conditions were not kept exactly constant. This was because of the nature of the machining process and the ability of the machine operator. Another interesting factor which has not been mentioned up to now, is that of the workpiece material. When the feature space is viewed, there are certain discontinuities in the signal. After the discontinuities the signal seems to follow a stationary trend until the next discontinuity. These regions each signify a new workpiece. The setup of the workpiece together with differences in chemical composition may be the cause of this.

Figure 5.5 show a pure noise signal produced by the machine. A normalised histogram is shown on top of the noise signal. Clearly the noise of a Gaussian nature and slightly skewed to the lower values. The signal to noise ratio can be calculated from this signal and the one in figure 5.1. Using the equation:

$$S = 20 \log_{10} \frac{RMS_{signal}}{RMS_{noise}} \tag{5.1}$$

In equation 5.1, S is the signal-to-noise ratio , in decibels, of the root mean squares of the noise and the signal. RMS_{noise} was calculated to be 0.005 and $RMS_{signal} = 0.0168$. S can then be calculated to be $S = 10.5dB$. This can be regarded a rather noisy signal if it

is considered that FM radio transmissions may have signal-to-noise ratios of $S = 50dB$. This all implies that for future work a noise filter might be applied to great advantage on this system.

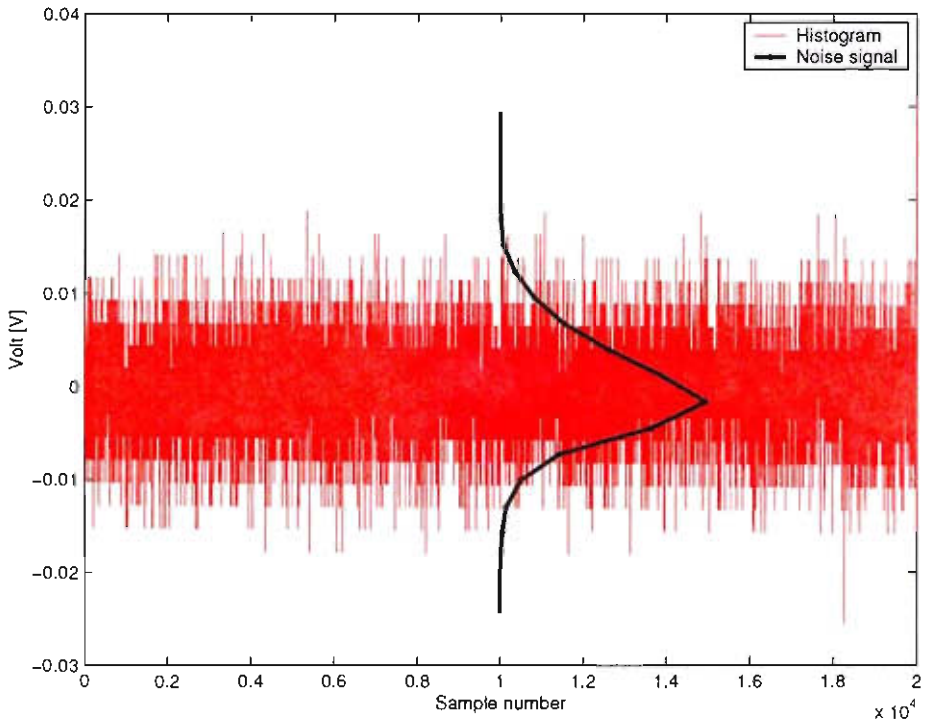

Figure 5.5: A noise signal from the system. Superimposed on the signal is a normalised histogram.

5.3 Feature selection and dimensional reduction

Features from the time domain

Figures 5.6 to 5.7 shows the features that were calculated from the snippets of the de-trended signals. A total of twelve features were extracted, six from the time domain and six from the frequency domain. The sample number is indicated on the x-axis of the feature-figures. Each of these samples represents a time interval for which the feature was calculated.

Figure 5.8 shows the PSDs of the cutting signals of one tool. The frequency axis is displayed up to the cut-off frequency of the anti-aliasing filter. This figure shows some peaks in the range below $300Hz$. These peaks are magnified in figure 5.9. This plot also shows a dotted line which is a sum of all the PSDs. (The sum is divided by a factor 100 in order to make it visible on this plot) This is helpful for finding regions where there is more energy present. Magnifying one of these regions shows the increase of one of the regions. The legend explains how the colour of the line is connected to the time "into" the

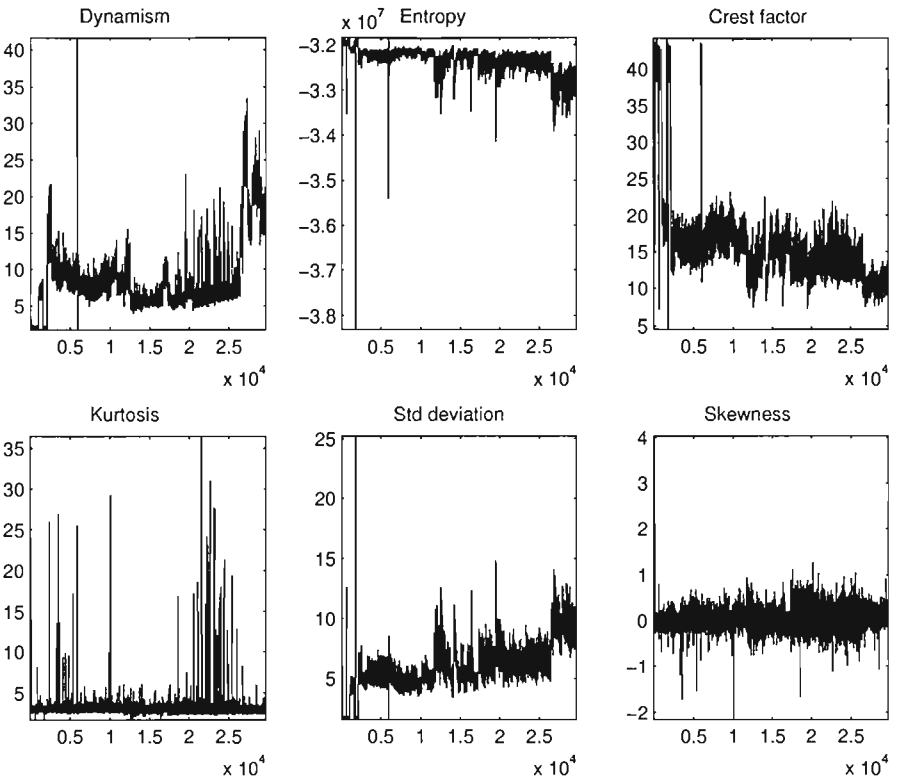

Figure 5.6: The time domain features extracted from the processed signals.

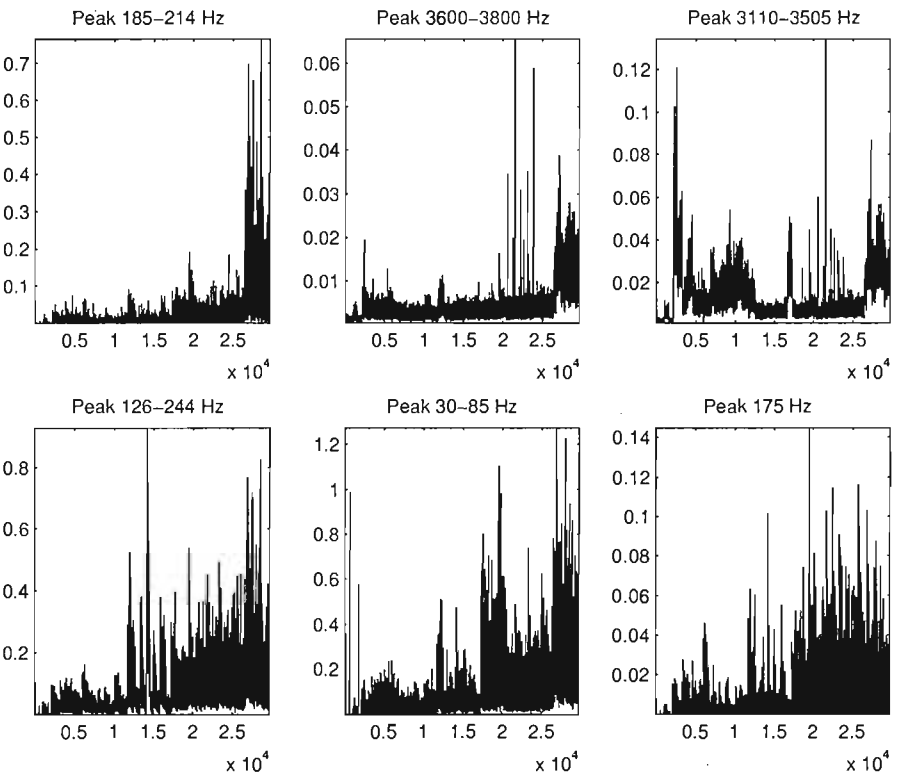

Figure 5.7: The frequency domain features extracted from the processed signals.

tool life. Peaks that may be used will then start out as a small dark hump and gradually "transform" via grey into a light grey peak. Peaks that shrink via this same process are also useful. Figure 5.10 represents these transformations of the peaks in figure 5.9.

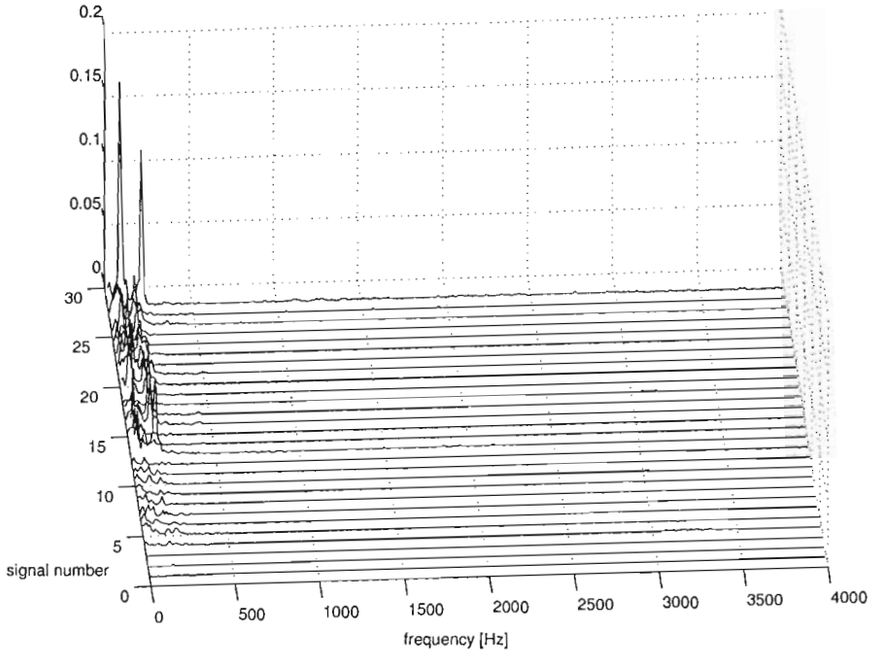

Figure 5.8: The PSDs of the cutting signals during the life of a tool.

5.3.1 The selection process

The selection process is shown in figure 5.11. Features were selected according to their relation to the theoretical tool wear function calculated by the correlation coefficient. The ideal trend function is then taken as a straight line with a slope of 40°. This slope was chosen arbitrarily and any positive figure may be used for this. This slope is chosen to be more steep than any of the slopes from the other features. This is then helpful to select the features that have most consistent correlation with the theoretical tool wear function. The effectiveness of the feature selection process is subsequently dependant on the assumption that the tool wear can be approximated by a straight line. This technique was proposed by Scheffer and Heyns (2001).

The sorted correlation coefficients are shown in table 5.1. Entropy and Crest factor were chosen from this list because of their obvious inverse relationship with the theoretical tool wear. It seems that only the skewness, the kurtosis and the $3110 - 3505Hz$ peak are unusable in this application because their correlation coefficients are a whole order less than that of the rest of the features. Selecting both negative and positive coefficients has the advantage of ensuring that the selected features contain minimal mutual information.

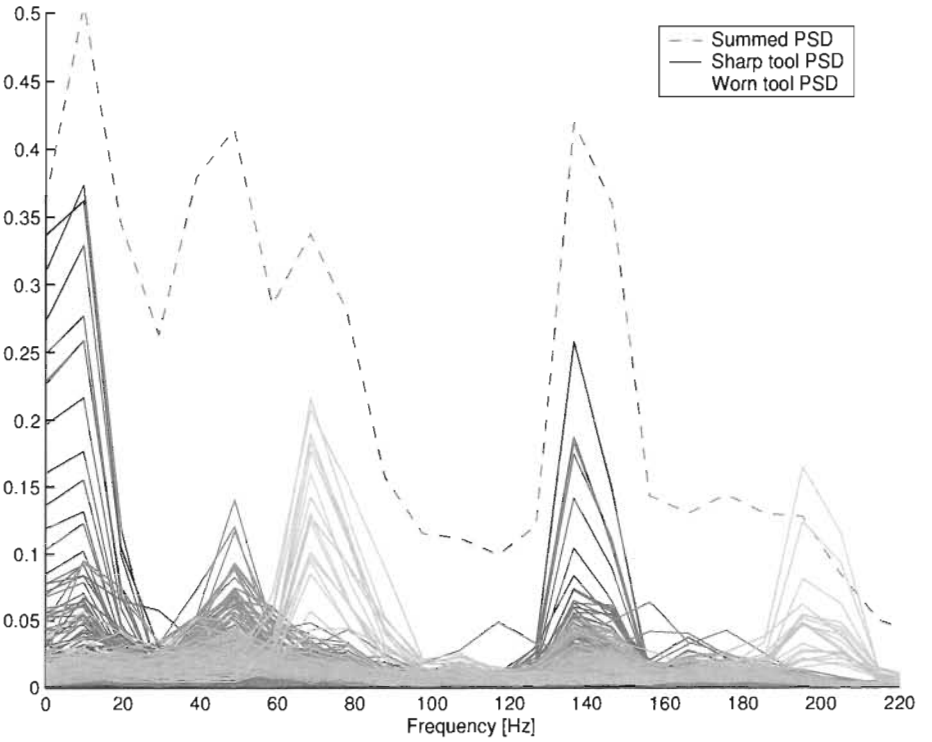

Figure 5.9: The magnified region and the summed PSDs.

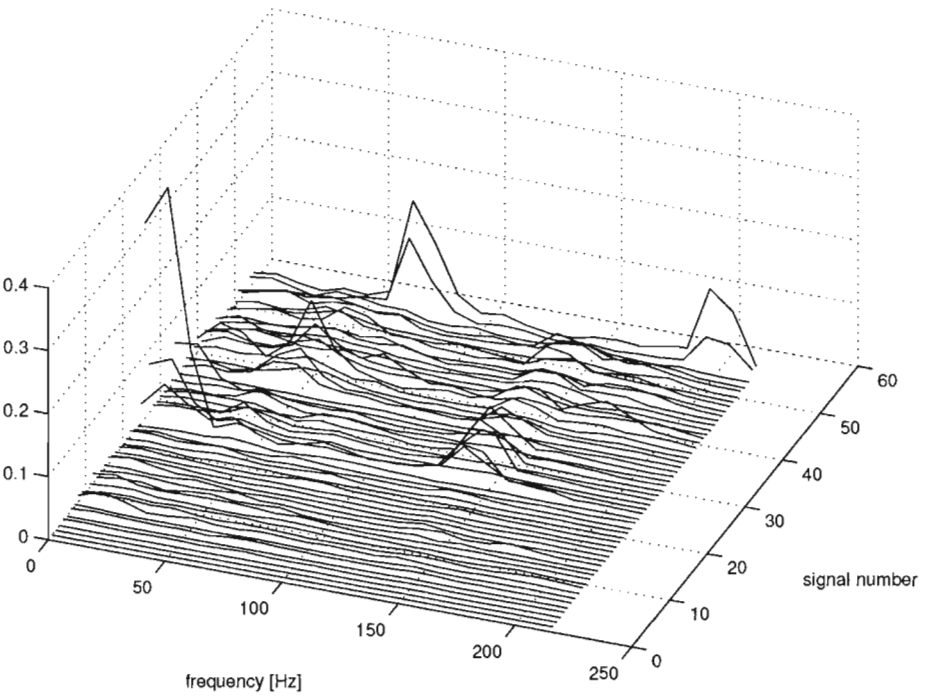

Figure 5.10: Another view of the progression of the PSD peaks.

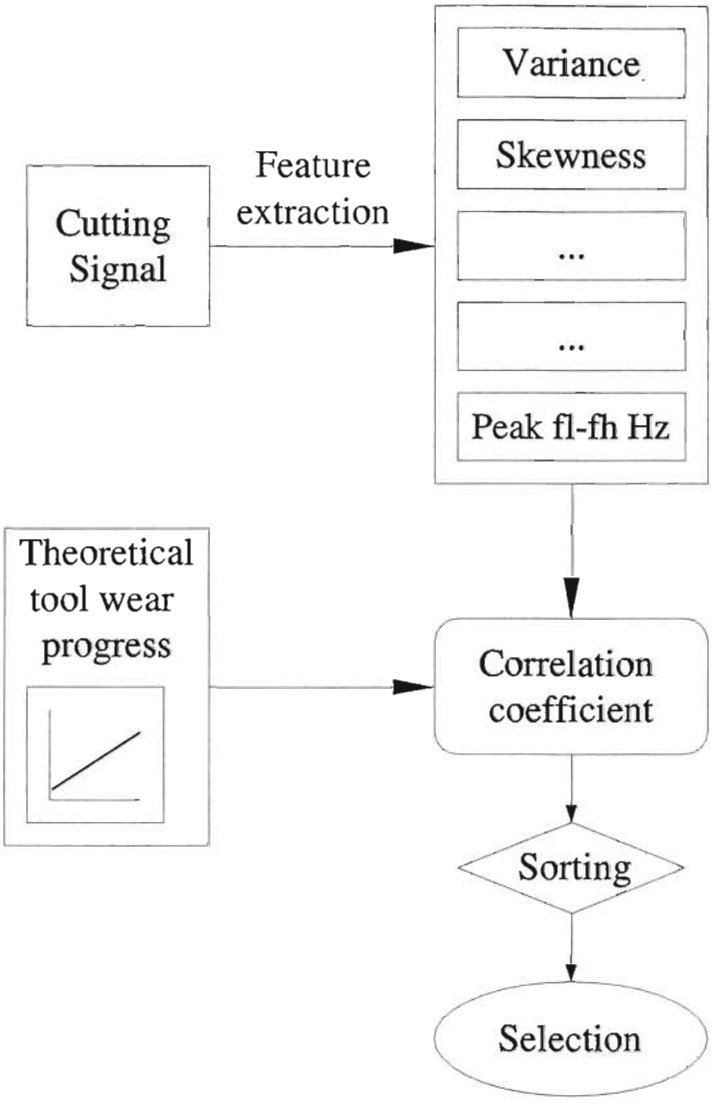

Figure 5.11: The selection of the features using the correlation coefficient.

Table 5.1: The sorted correlation coefficients.

Feature	Correlation Coefficient
Entropy	-0.651
Crest factor	-0.596
Kurtosis	-0.057
Peak 3110-3505 Hz	0.037
Skewness	0.085
Peak 175 Hz	0.379
Dynamism	0.405
Peak 185-214 Hz	0.445
Peak 126-244 Hz	0.516
Peak 3600-3800 Hz	0.540
Peak 30-85 Hz	0.601
Std deviation	0.704

All the other features, except for the above named three were used for the classification process. The final selection of features are listed in tabel 5.2.

Table 5.2: The selected features

Final features
Entropy
Crest factor
Peak 175 Hz
Dynamism
Peak 185-214 Hz
Peak 126-244 Hz
Peak 3600-3800 Hz
Peak 30-85 Hz
Std deviation

Preparation for the feature reduction

Before a feature reduction can be done a whitening transform is done on the data. After the whitening transform all the features in the data have a mean of 0 and a variance of 1. This transform is similar to the normalisation techniques used in neural networks. A normalisation is done on the data to ensure that the data is not biased toward one of the features.

Using the principal component decomposition, the selected features are combined into a single feature which will be used to train the HMMs. This feature is shown in figure 5.12. The final correlation coefficient of the universal feature is -0.695 which is very close to that of best feature. Table 5.3 shows the total variance explained by each

principal component after the decomposition. It can be seen from this that the first principal component explains more than 70% of the variance. On page 31 the use of only this one principal component is mentioned. The classification system is to be trained on data from this figure. The "noisy" figure shows why a threshold method for classification will produce many false alarms. The "spikes" seem to jump arbitrarily and may trigger a "worn-tool alarm." There is fortunately a trend that can been seen in the data. The lower values of the dimensionally reduced feature vector correlates with worn-tool conditions. This figure also shows why an advanced classification system is needed for machining data.

Table 5.3: The principal components and the amount of the total variance the represent.

Principal component no.	Percentage of total variance
1	73.57
2	10.53
3	7.17
4	3.81
5	2.45
6	2.45
6	1.07
7	0.84
8	0.50
9	0.02

The first third of the data was selected to be the first class and the last third to be the second class. These two thirds of the data was used to calculate the P-vector for the dimensional reduction. These same thirds will be used in the next section for the training and testing.

The last preparation before the HMMs are applied is the discretisation. Because discrete HMMs are used, it is necessary to discretize the data. It was decided to quantise the data into 150 levels. At this level there is still ample detail left in the universal feature. A lower discretisation level will have sharper decision boundaries, this will probably give better classification results. More advanced HMM techniques however use continuous PDFs. Using a high number of discretisation levels therefore will give a better indication how future and more advanced models may perform. A better platform for comparison for the performance between discrete and continuous models is also created.

Figure 5.13 shows a few training sequences for sharp and worn tools. These figures were already discretized and represent the final product that is fed to the HMM models.

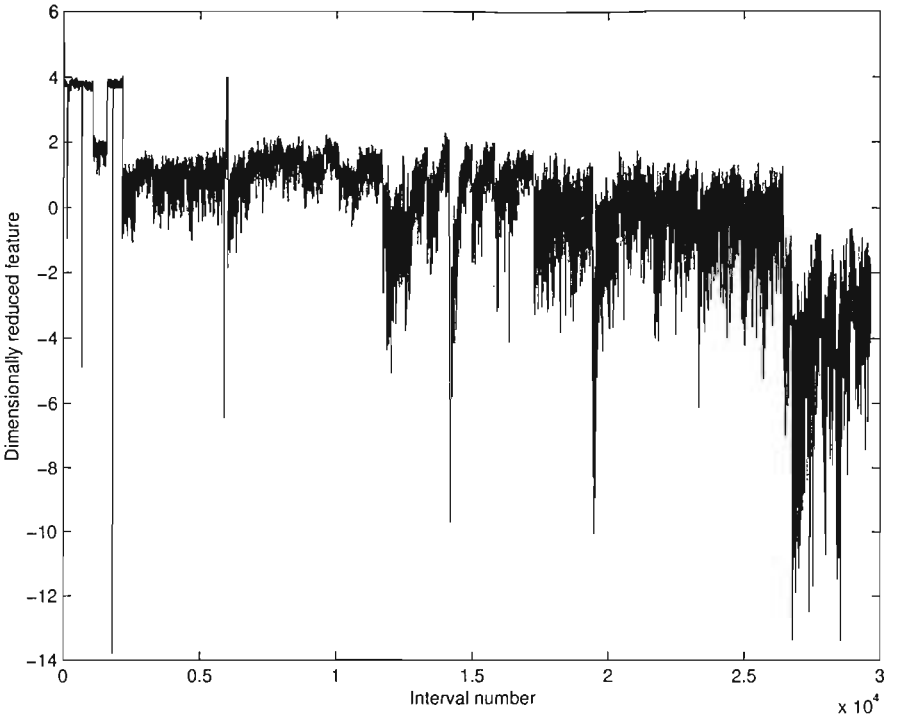

Figure 5.12: The final combined feature from which the training sequences for the HMM will be extracted.

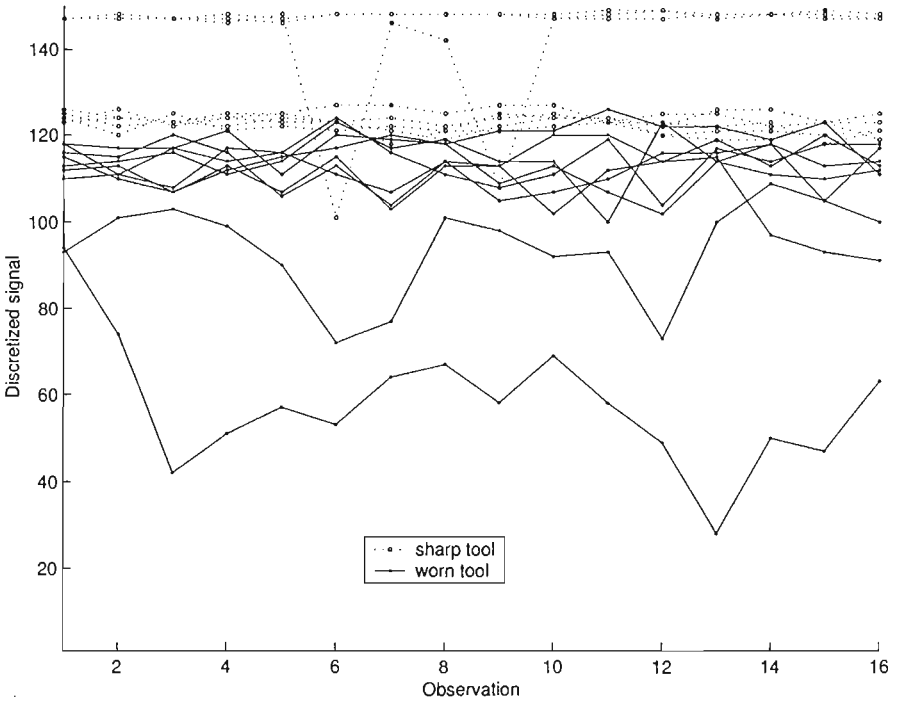

Figure 5.13: The training sequences after discretisation.

5.4 HMM training and classification

During the training, one HMM was trained for each of the identified classes. Each HMM was then trained on the data it is to be associated with. Afterwards the HMM were tested on mixed, unseen data. A HMM toolbox for MATLAB is used for the data classification techniques presented in this section.[1]

5.4.1 Selecting samples for training

Having already selected the classes to be recognised, it is necessary to select samples for training and for testing. From each class, one third of the samples are randomly selected and removed from the set. The remaining data is used for training. After training the models are tested with the remaining data. Figure 5.14 shows set of randomly selected training data samples. The different colours indicate different classes.

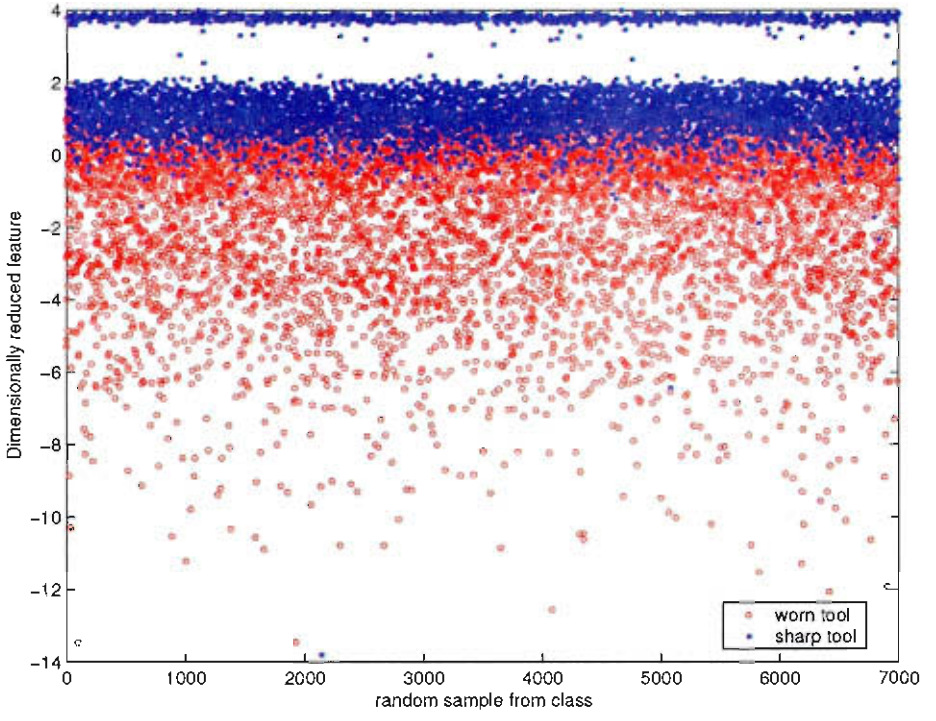

Figure 5.14: A training data set

The histograms in figure 5.15 show the areas of likelihood for samples in the different classes. Once again the tops of the bins were connected to form a curve. The second peak with the small variance in the histogram of the sharp tool, is an example of differences in workpiece composition and setup that affects the signal quality. From this figure it can

[1] Hidden Markov Model (HMM) Toolbox written by Kevin Murphy (1998). See http://www.ai.mit.edu/~murphyk/Software/hmm.html for details.

be seen that the PDFs of the sharp tools and the worn tools have a large overlap area. This will make recognition difficult.

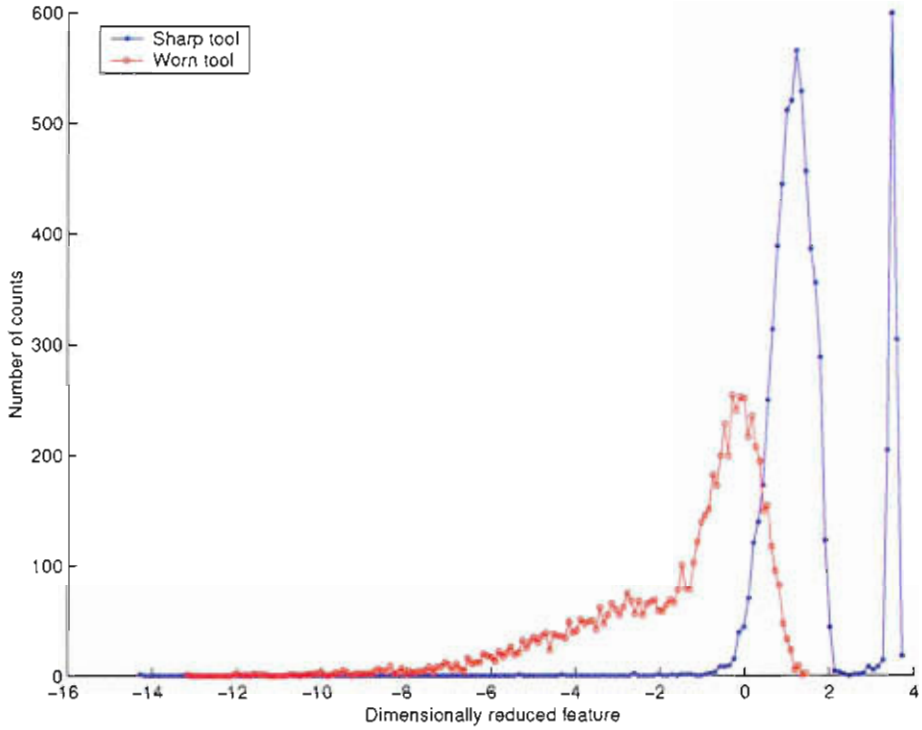

Figure 5.15: The histograms for the different classes

5.4.2 Condition for correct classification

At this point it is relevant to describe what is deemed to be a correct classification. After the HMM are trained they are tested. The testing entails that the HMMs are shown an unknown sequence. The sequences are similar to the ones shown in figure 5.13. The probability that each of the HMMs will produce the sequence is then calculated. The sequence is then classified in the class of the HMM with the highest probability. Since the testing data will have known labels (eg. the user has a prior knowledge of the class of the data), a correct classification will be when the HMM associated with the correct class has the highest probability.

5.4.3 The HMM topology

As with neural networks, there is no analytical way of predicting what HMM topology will produce the best results. An iterative procedure was followed where the whole training and testing procedure was repeated for an incrementally changing number of states. For each number of states, the training and testing was repeated five times and a mean was calculated. The results are shown in 5.16.

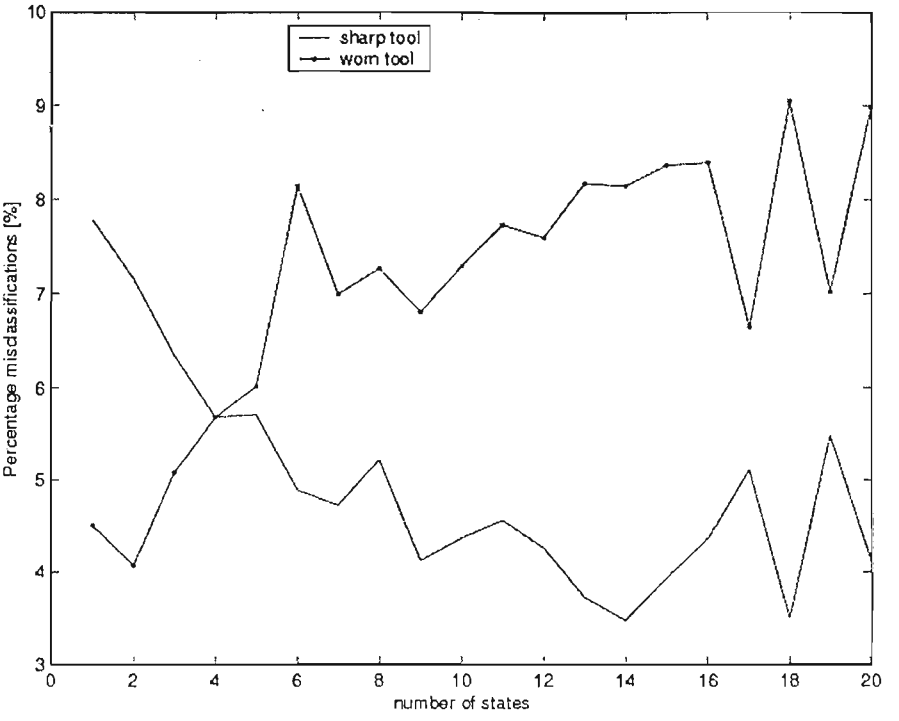

Figure 5.16: The number of states vs the recognition faults

From figure 5.16 the optimal number of states for each HMM with its associated class can be read off. This number will however be a trade off between complexity of the model and the performance. It was decided to select for:

- the HMM on worn tool data, the number of states as, 2
- the HMM on sharp tool data, the number of states as, 7.

With these parameters chosen, one can show more results of the HMM classification such as the forward probabilities.

5.4.4 Recognition and results

In order get an idea of the behaviour of the recognition of the HMMs, the test was repeated twenty with the chosen parameters. This is shown in figure 5.17. The mean of the performances are indicated on this figure. This figure shows that the behaviour is somewhat erratic, and can be ascribed to the quality of the data.

Figure 5.18 shows the probabilities that the HMMs will produce the testing data. The first half of the data is of class one and the second part is of class two. One then expects to see that the lines of the HMMs should cross in the middle somewhere. The probabilities of the HMMs are however a little more chaotic. The extreme dips in the data are caused by zeros in the probability density functions of the states. Since the data

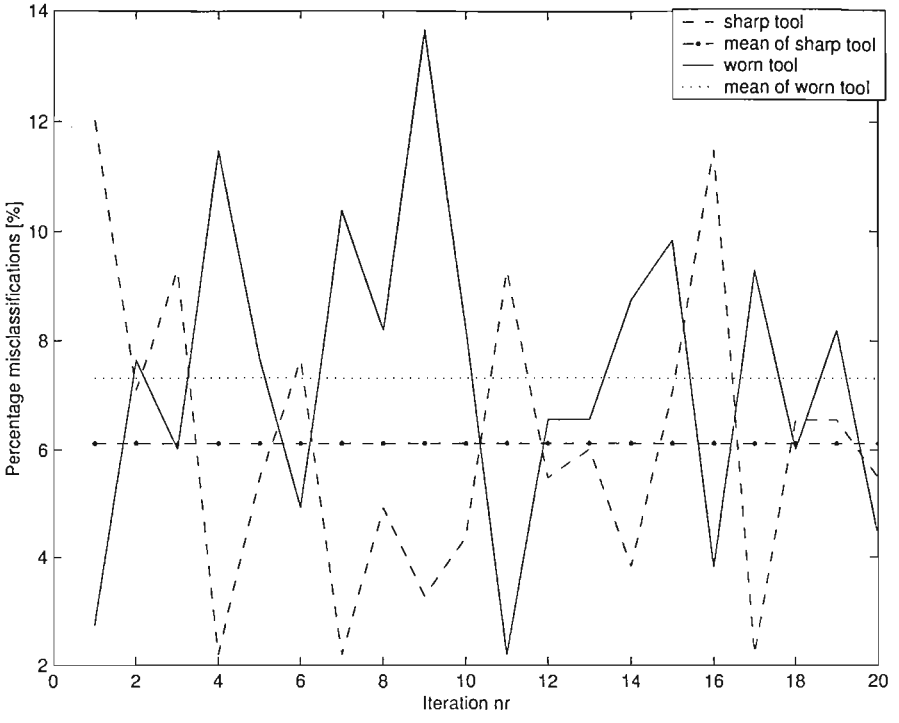

Figure 5.17: The behaviour of the classification performance.

is discretized it implies that the probability density functions are discretized as well. It may therefore happen that, in the PDF one of these "symbols" may have a probability of 0. The prediction probability of the HMM then drops to $-\infty$. A bound of -130 was put on this.

The outcome of this can be seen more clearly in figure 5.19. Once again in this figure, the data of the first class was shown in the first half and that second class in the second half. Correct classifications are therefore shown as red circles in the first half and the blue circles in the second half. On average this quantifies into:

- 6.5% incorrect classifications of sharp tools
- 7.5% incorrect classifications of worn tools

5.5 The Maximum Likelihood classifier

To create a basis for comparison, the recognition was also to be done with another maximum likelihood technique. The maximum likelihood was chosen for this. This type of classifier is easy to implement and usually very robust. The maximum likelihood classifier works by creating a decision boundary using the PDFs of the different classes. The training data in this case is used to fit Gaussian PDFs onto the data. Because of the smooth decision boundaries that this method creates, it is easy to predict the behaviour

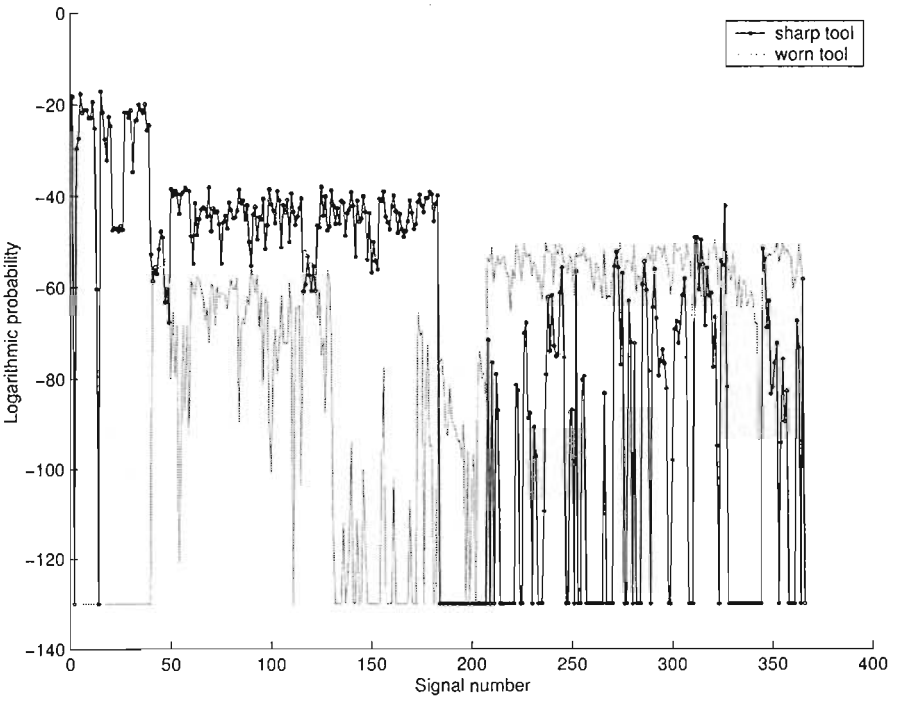

Figure 5.18: The prediction probabilities of the HMMs

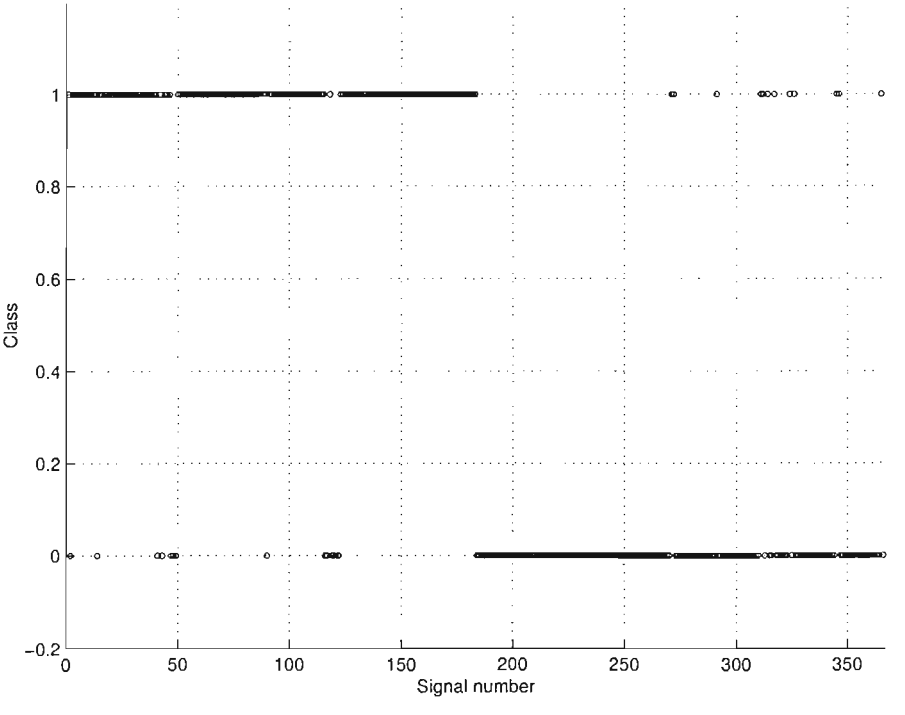

Figure 5.19: The classification results

of the system. The performance of this system is dependant on the quality of the fit of the Gaussian PDFs on the data classes.

Formally the maximum likelihood classifier works as follows: if $P_a(x)$ is the PDF of class a and similarly, $P_b(x)$ is the PDF for class b, then the decision boundary will be where:

$$P_a(x) - P_b(x) = 0 \tag{5.2}$$

Classification can then be done on any arbitrary value of x. If equation 5.2 is calculated for a value of x and the answer is a positive number, then x belongs the class a, otherwise class b. Figure 5.20 shows the PDFs for the two classes and the decision boundary.

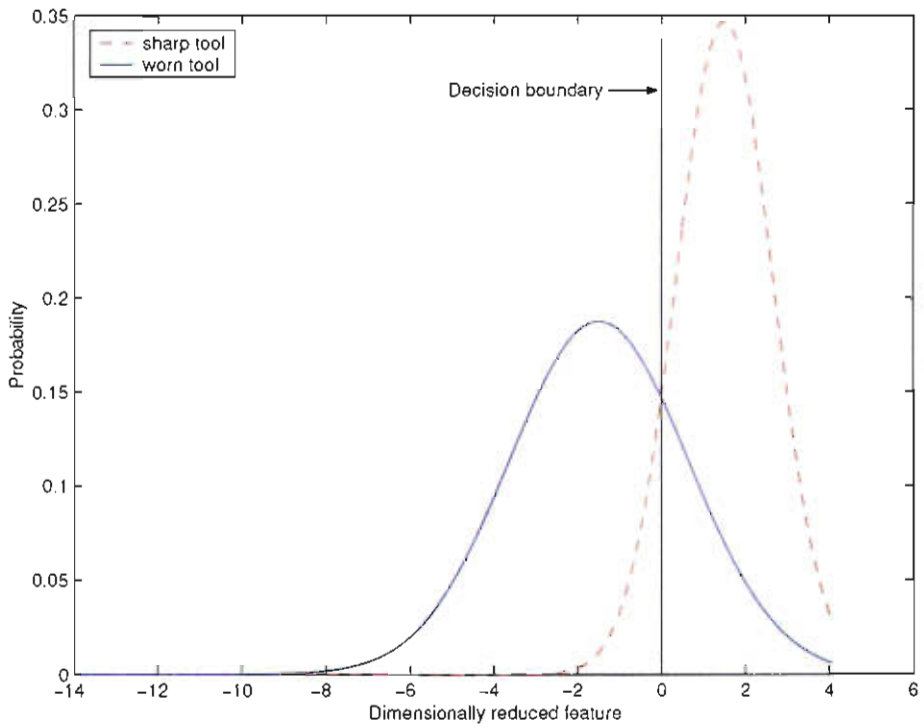

Figure 5.20: The Gaussian PDFs fitted onto the data and the decision boundary.

The training of the maximum likelihood classifier is done by simply drawing the histograms from a training data set with the defined classes. The decision boundary is then applied to a testing set. An example of this decision boundary, plotted on a training set in figure 5.21.

The training and testing is repeated a number of times and the performance and the behaviour of the performance is shown in figure 5.22.

From this figure, it can be seen that the performance is rather stable. The average performance for the maximum likelihood classifier turns out to be:

- $3,2\%$ incorrect classifications for a sharp tool
- $27,3\%$ incorrect classifications for a worn tool

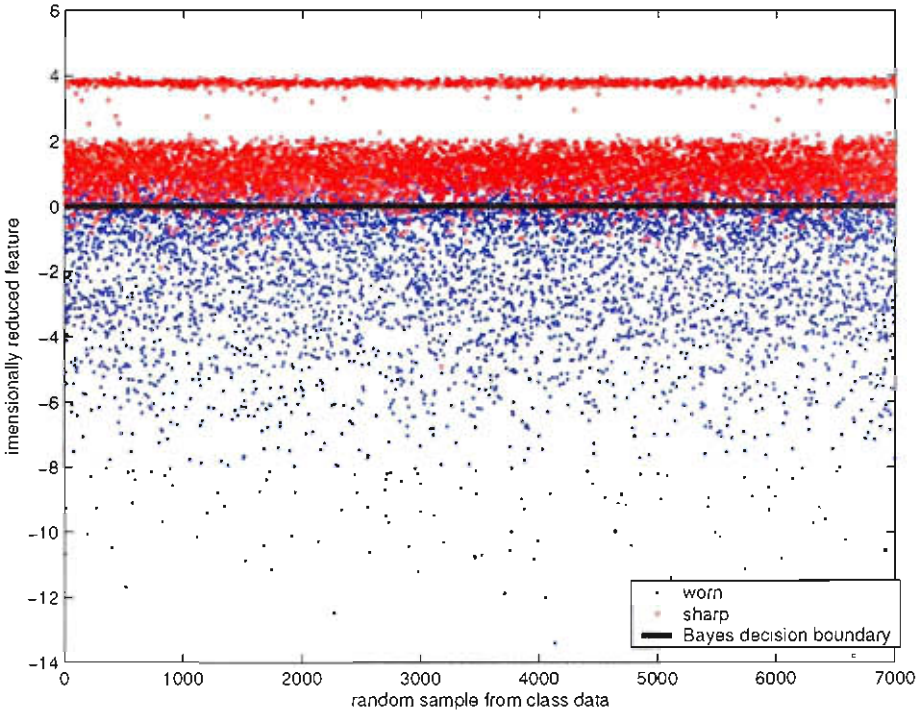

Figure 5.21: The training data with the decision boundary applied.

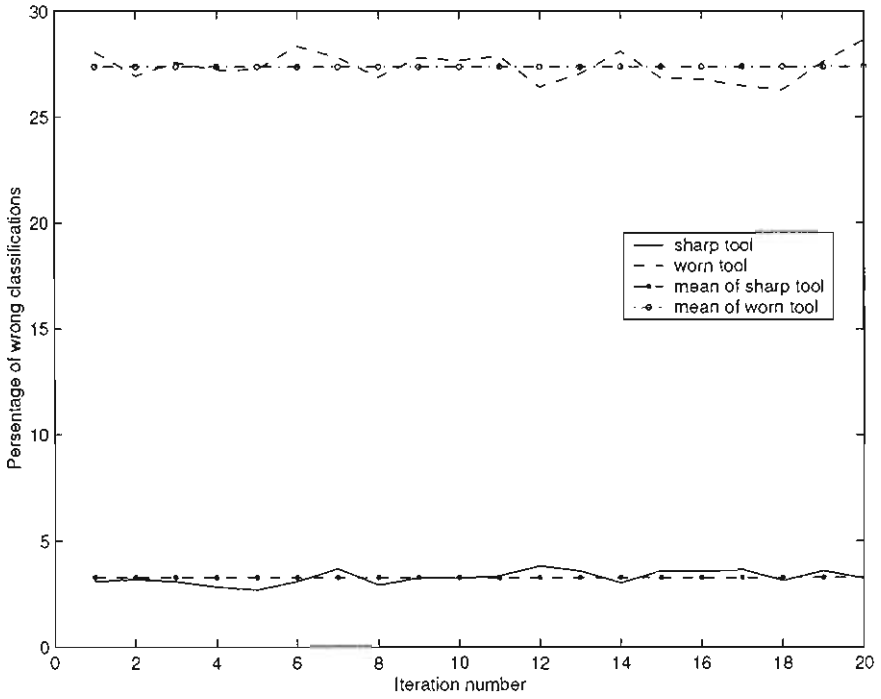

Figure 5.22: The performance of the maximum likelihood classifier over a number of iterations.

The low classification result for worn tool shows that a Gaussian PDF is not a very good approximation for the likelihood function of the worn tool. The reason for the low result is the high amount of overlap between the two PDFs, which is clear in figure 5.20. In figure 5.15 it can be seen that there is an amount of overlap between the histograms. This overlap is a lack of separation between classes. The observation sequences that the HMMs use is one way to overcome this problem because temporal characteristics are taken into account. Comparison between the two methods of classification shows that the HMM classification is not as influenced by this lack of separation.

5.6 Reduced dataset

The performance of any classification algorithm is dependant on the quality and the quantity of the data that is used to train the system. It has been argued that the performance of the HMM recognition system will improve if there is more data and better quality data available for training. In order to show that this was the case, the data set was reduced and the training and testing of the HMM classification system was repeated.

For this trail the first 75% of the data was used. Again the data was divided into three classes of which the last and first were used in the classification tests. Again two thirds of the data of each class were randomly selected to train the system and the last third was used to test the system.

A principal component decomposition was once again applied to the two classes to achieve separation and dimensional reduction. After the dimensional reduction, two histograms were drawn up of each class. This is shown is figure 5.23. It can be seen that less prominent separation is achieved between the two classes.

Once again the whole classification procedure was done exhaustively to find the "optimal" number of states for this application. According to figure 5.24 the optimal for this case is:

- 2 states for the worn tool data
- 8 states for the sharp tool data

It is evident that there is much less of a trend between classification performance and the number of states of the HMMs. Each data point on the graph represents the mean of an average of 5 classification iterations.

When these figures for optimal classification are applied for investigation into the behaviour of the classification test results, figure 5.25 is the result. As with the previous result, the performance behaviour of the HMM classification is rather erratic, much more than with the full data set.

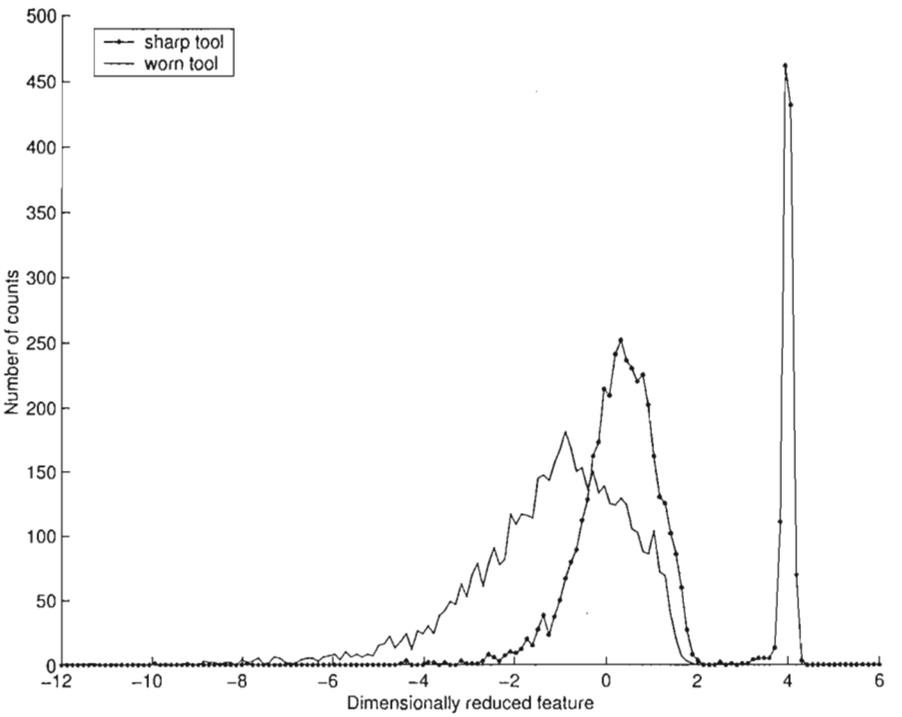

Figure 5.23: The histogram of the two classes in the reduced data set.

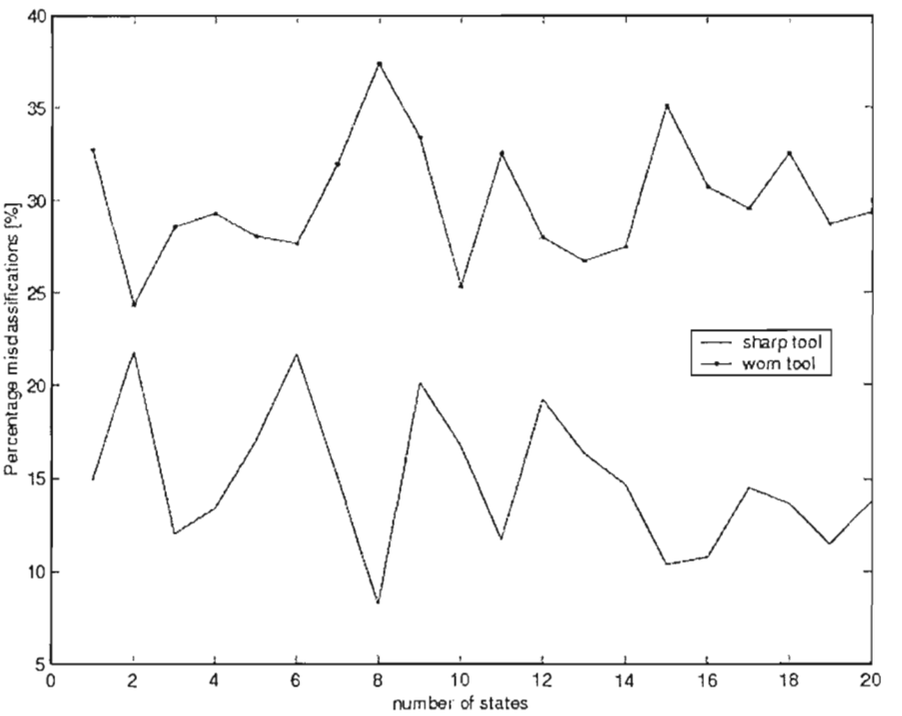

Figure 5.24: The classification performance as a function of the number of states.

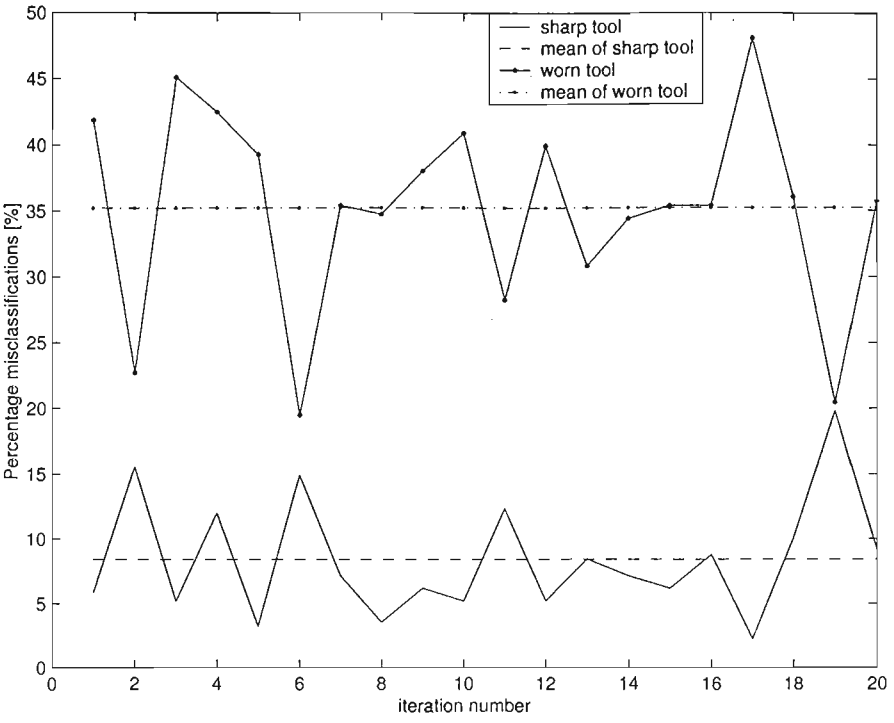

Figure 5.25: The behaviour of the classifications.

The average result of the classification for the reduced data set was:

- 35% incorrect classifications for worn tools
- 8% incorrect classifications for sharp tools

From this it shows that with less data there is lack of separation between classes which hampers the performance of the HMM classifier. This also validates the claim that more data will enable more, and better class allocation and separation, and ultimately better classification

CHAPTER 6

Conclusion

6.1 Review of results

In this study discrete hidden Markov models were trained for tool wear classification. The data was generated manually on a lathe with a cemented carbide tool insert cutting an EN19 steel alloy. Some of the parameters, such as the depth of cut and the feed rate was kept constant while the cutting speed had some variation. This variation was around $\pm 8\%$ of the cutting speed.

This has been the first work where a reduced feature space was used to train the HMM. The "essence" of the signals were compressed into a very relevant and robust single feature using the principal component analysis which lowers the dimensionality of the feature space. If more than one sensor signal is used, this single feature will be very robust indeed. Also in the case where more that one sensor is used the feature will probably give good results even if some of the other sensors fail.

The training and testing data was generated from one tool. The data encompassed only the first third of the life of the tool. This devalues the statistical integrity of the data. The samples for training and testing however, were selected randomly from the respective classes, and the performance of the classification is the average result of 20 iterations. This mitigates to a degree, the fact that only one tool was used. In cases where more than one tool will be used, the performance of the classification might be expected to be a little lower. It is however believed that comparable results will be achieved. Also if a total tool life might be used, even better separation between classes will be achieved and consequently better results will be achieved.

An HMM classification system was created from the data . This system scored the forward probabilities to create a binary, "sharp"/"worn" classification. Two HMMs were trained for the system, one corresponding to each wear state to be classified. The system

achieved a 91,5% correct classification of sharp tools and a 94,5% correct classification of worn tools. This was compared to a Bayesian classifier that achieved 96,8% correct classification for a sharp tool and a 72,7% correct classification for a worn tool.

In order to establish the relationship between the amount of available data for training and testing and the performance of the system, the whole recognition procedure was repeated with a reduced data set. In this reduced data set the separation between classes was very poor. The system achieved an average result of 65% correct classifications for worn tools and an 92% correct classifications for sharp tools. This is comparable with the performance of the Bayesian classifier. This investigation proves that with more data there will be more clearly identifiable classes which will make for a better classification system.

Optimal state topology of the HMMs were obtained by an exhaustive method. It was found that for this application a 7 state, ergodic HMM works well and for sharp tools and a 2 state, ergodic HMM works well for worn tools. The signals were discretized into 150 levels and this was kept constant through the experiments. Lowering this number might actually improve the performance of the system.

The HMMs achieved very good recognition considering that the tool has only reached a third of its total life and that the experimental results could not be kept very constant. Even through this, the HMMs achieved a robust recognition.

Finally, because the data used in this classification scheme is only limited to a part of a lifetime of one tool, it compromises the statistical integrity of the data and the results. Good generalisation of the classification scheme with the HMMs is therefore not guaranteed. The technique is however successfully demonstraded.

6.2 Suggestions on Improvements

The nature of HMMs lends itself very much to the detection of wear condition. It is however even better suited for discrete event detection. Events like excessive vibration of the workpiece (eg. where the workpiece might not be lined up) or self-excited vibration of the tool (eg. chatter), or breakage events, might be very well suited to be detected by HMMs. No work, to the knowledge of the author, has been done in the area of TCM using HMMs to detect these conditions. This may be an interesting future field to explore.

Measuring in the machining environment is exceedingly difficult and more work is needed to produce a cheap sensor integrated machining tool. A problem with strain gauges that are covered is that they might be torn off by the very same covering that is meant to protect it. This is a future area which could be explored to produce a covering for the strain gauges to not hinder their performance. Another option might be to embed the sensor into the tool holder itself.

A problem still relevant to neural network TCM systems is the automation of the

process of generation-and-selection features that are sensitive to tool wear. This problem is also relevant to systems that will use HMMs for the classification. More research in needed in this direction.

The techniques for using HMM for the classification of tool wear has been proved in this study, although the very simplest of the HMM family was used. Investigation into more complex models might improve the results. More configurations of the HMMs might also be used where Viterbi decoding could be used to determine the tool state.

APPENDIX A

Additional Theory on HMMs

Some more theory on HMM are presented in this chapter. The Baum-Welsch training procedure for HMMs is also presented in this chapter. Most of the information in this chapter is taken from (Rabiner, 1989).

A.1 Assumptions of the hidden Markov model

It was mentioned earlier that HMMs have a rich mathematical structure. This is again evident in the following assumptions. These assumptions are made to keep mathematical calculations tractable[1].

1. **The Markov assumption:** The Markov assumption is that the probability of transition from one state to another is only dependent upon the current state.

$$a_{ij} = P(q_{t+1} = j | q_t = i) \tag{A.1}$$

   This is actually called a first order HMM, where a second order HMM would be dependent upon the current state and the previous state. Naturally calculations become increasingly more complex.

2. **The stationarity assumption:** This implies that the state transition matrix is invariant with time. This means that :

$$P(q_{t_1+1} = j | q_{t_1} = i) = P(q_{t_2+1} = j | q_{t_2} = i) \tag{A.2}$$

   This holds for any t_1 and t_2.

[1]These definitions are from Narada Warakagoda website at http://jedlik.phy.bme.hu/ gerjanos/HMM/node3.htm

3. **The output Independence assumption:** This assumption means that the current observation is statistically independent of any previous observation. Let : $O = \{o_1, o_2, o_3, \ldots, o_n\}$ Then for a specific HMM model, λ:

$$P(O|q_1, q_2, q_3, \ldots, q_n, \lambda) = \prod_{t=1}^{T}(o_t|q_t, \lambda) \tag{A.3}$$

A.2 Training the hidden Markov model

Together with the forward procedure, another two parameters need to be introduced before the Baum-Welch re-estimation procedure can be explained. The first is the backward procedure which is very similar to the forward procedure. This is also calculated recursively and works as follows:

Define:

$$\beta_i(i) = P(O_{t+1}, O_{t+2}, \ldots O_T|q_t = S_i, \lambda) \tag{A.4}$$

We can solve for β inductively:

1. Initialisation:

$$\beta_t(i) = 1, \quad 1 \leq i \leq N \tag{A.5}$$

2. Induction:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}b_j(O_{t+1}\beta_{t+1}(j)), \qquad t = T-1, T-2, \ldots, 1, \quad 1 \leq i \leq N \tag{A.6}$$

In order to describe the rest of the procedure for re-estimation of the HMM parameters, we first define $\xi_t(i, j)$, the probability of being in state S_j at time t, and state S_j at time $t + 1$, given the model and the observation sequence e.g.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j|O, \lambda) \tag{A.7}$$

This parameter, ξ can now be written in terms of α and β

$$\xi(i, j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \tag{A.8}$$

A final quantity, γ needs to be defined. This is the probability of being in state S_i at time t, given the observation sequence O, and the model λ.

$$\gamma_t(i) = P(q_t = S_i|O, \lambda) \tag{A.9}$$

In terms of α and β this is:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(i)\beta_t(i)} \tag{A.10}$$

If γ is summed over t, the result is the expected number of transitions made from state S_j. Similarly, summing of $\xi_t(i,j)$ over t (from $t = 1$ to $t = T - 1$) can be interpreted as the expected number of transitions from state S_i to state S_j. Using the above formulae and the counting of event occurrences a method for re-estimation can be given:

$$\bar{\pi}_i = \text{expected number of times in state } \mathrm{S_i} \text{ at time } (\mathrm{t} = 1) \; = \; \gamma_1(\mathrm{i})$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{A.11}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^{T} \gamma_t(j) \qquad (\text{the number of times in state j and observing symbol } \mathrm{v_k})}{\sum_{t=1}^{T} \gamma_t(j)} \tag{A.12}$$

APPENDIX B

Training of the HMMs

Some results on the training of the HMM are given in this chapter. Convergence plots of the training is shown here.

The training algorithm of the HMM was set to allow no more than 10 iterations. Figure B.1 shows 20 convergence-curves from the training of HMM on sharp tool data and worn tool data. The Baum-Welch algorithm quickly converges to a local optimum. Because the initialisation of the state transition matrix and the state probabilities are done randomly, the end result differs slightly after each training. A basic trend can however be detected.

The y-axis of the figure denotes "Logarithmic likelihood", this is the calculated for the whole training batch. The differences between the training outcomes can be ascribed to the influence of the following:

- The data quality of the training samples. (which were selected randomly)
- The initialisation of the parameters. (which were also selected randomly)

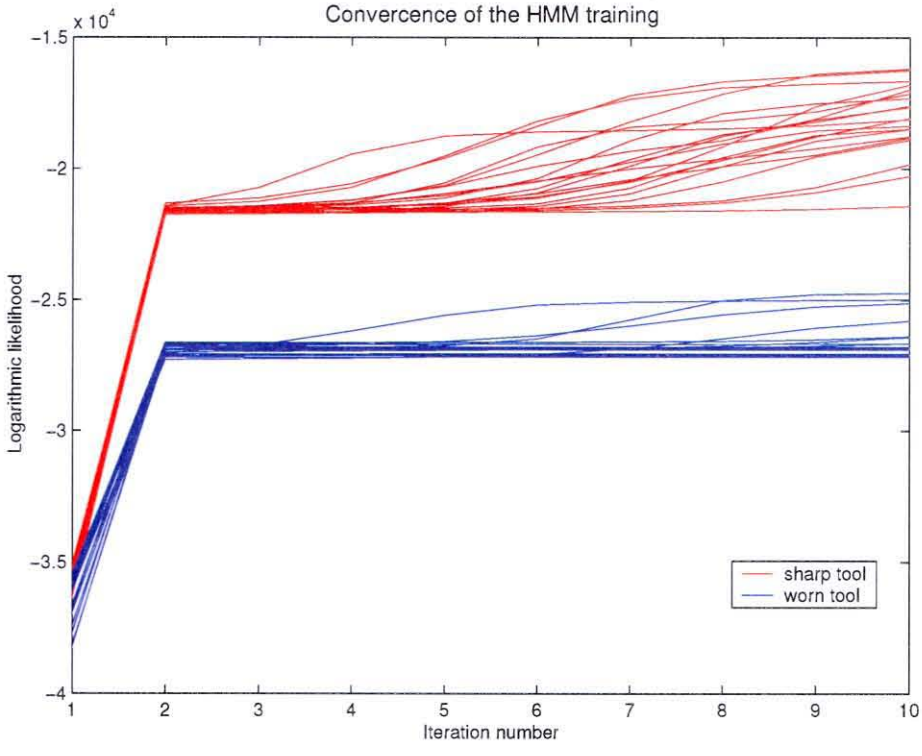

Figure B.1: Some convergence histories of the model training

APPENDIX C

Measurement of tool wear

This appendix will show the state of wear on the tool nose.

C.1 Nose wear

It was mentioned on page 39 the dominant wear mode that was detected in this experiment was nose wear. Nose wear is common at slow cutting speeds.

Tool wear usually has a slow initialisation phase followed by a an almost constant wear phase. The last phase is a very rapid wear growth followed by tool breakage.

In figures C.2 to C.3 the nose of the tool insert is shown. The angle from which is was taken is shown in figure C.1. The edge of the nose at the upper left corner is where nose wear would be seen. It can be seen that there is very little change between the two photos although they were separated by approximately 60 minutes of cutting time. This small change is an indication that the tool is still in the first phase of wear. A photo of a metal ruler calibrated in *millimetres* is shown is figure C.4. This photos was taken at the same magnification as the rest rest of the photos.

The scale that is presented in figure C.4 can also be used for figure 4.7 in the chapter on the experimental set up.

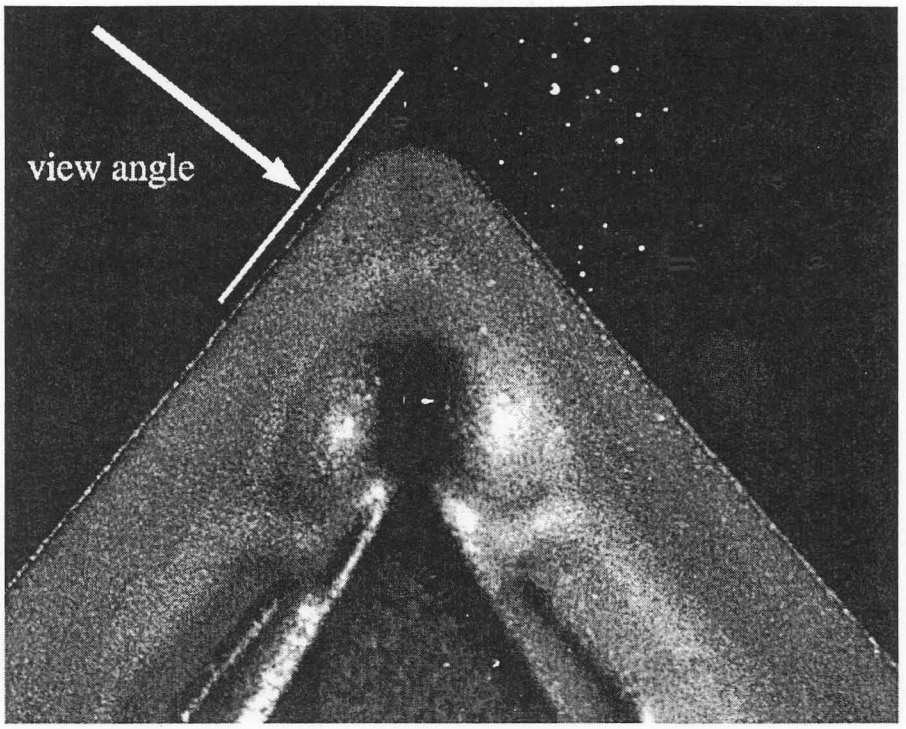

Figure C.1: The photo angle for figures C.2 and C.3.

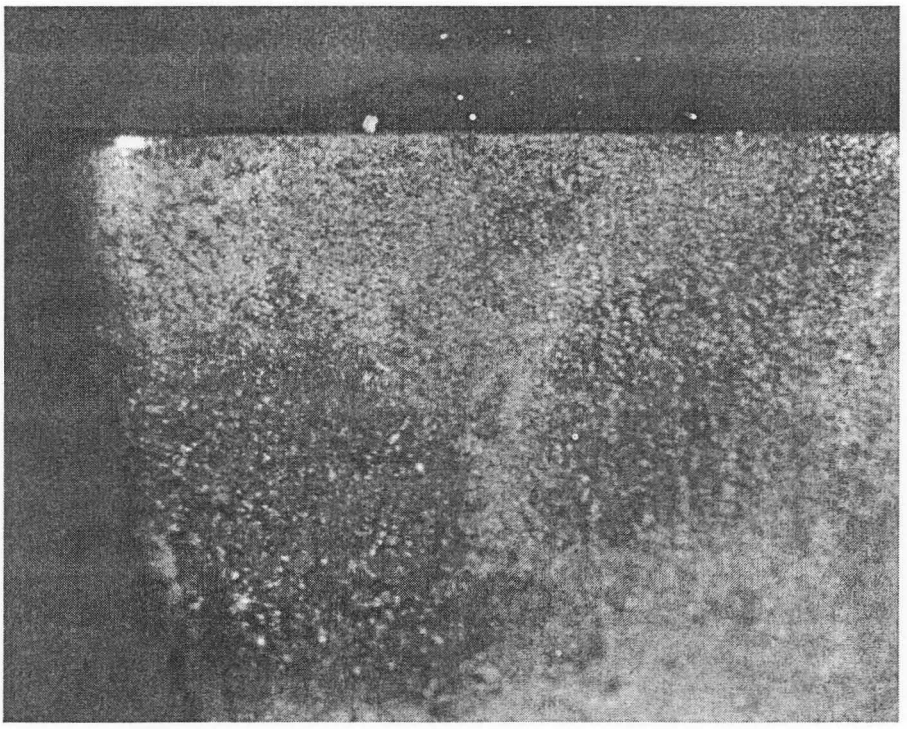

Figure C.2: Nose of a sharp tool

Figure C.3: Nose of a tool where wear has started

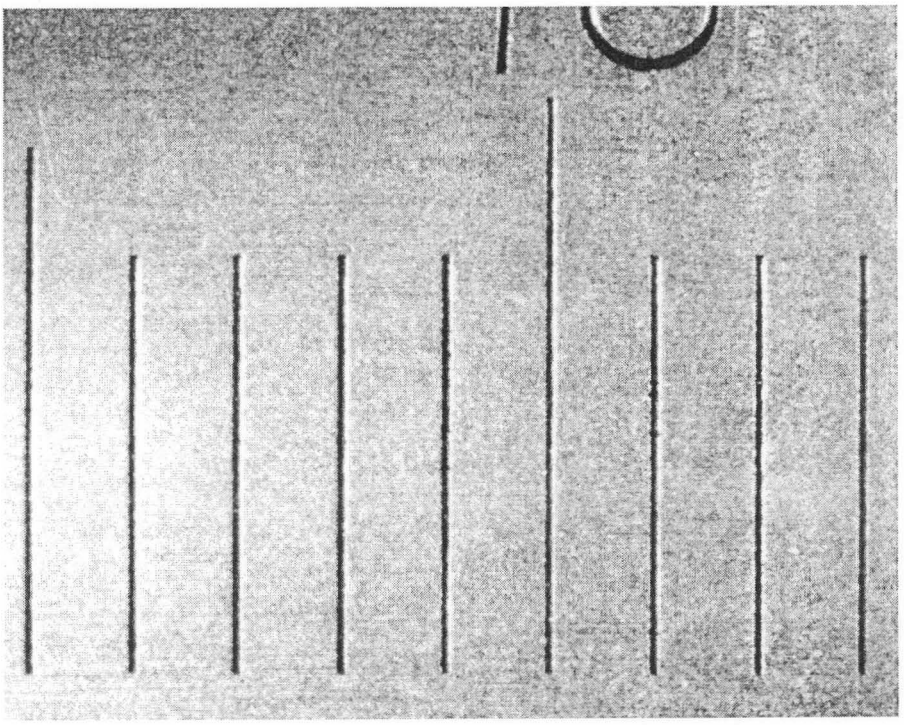

Figure C.4: A ruler calibrated in millimetres.

APPENDIX D

The setup

A few photos of the setup is shown in this chapter. Some more examples of cutting chips that were produced will also be shown.

The photos of the setup shown in figure 4.3 on page 34 will now be shown in this chapter. Figure D.1 shows the cutting tool in action. The cables from the strain gauges are contained in a shielded cable. This cable carries the wires to the strain gauge amplifiers. The strain gauge amplifier together with the anti-alias filters are housed in a metal

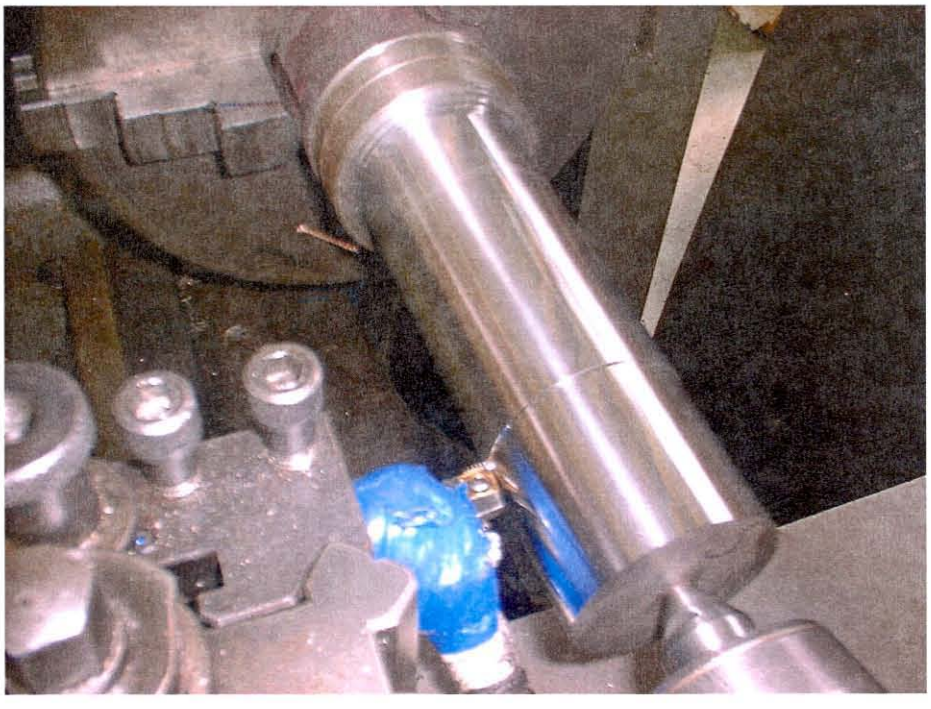

Figure D.1: The cutting tool in action.

container. This is shown in figure D.2. The output of the anti-alias filters are fed into the

PC via the National Instruments A/D card. The computer with the outside connectors for the output from the filters are shown in figure D.3.

Figure D.2: The housing for the strain gauges and filters

Figure D.3: The PC with the outside connectors shown in the upper right half

BIBLIOGRAPHY

Ajmera, J., McCowan, I. and Boulard, H. (2003) Speech and music segmentation using entropy and dynamism in a hmm classification framework, *Speech Communication, 40*, 351–363.

Allen, R. and Shi, H. (2001) Tool wear monitoring using a combination of cutting force and vibration signals, *International Journal of COMADEM, 4* (1), 26–32.

Atlas, L., Ostendorf, M. and Bernard, G. D. (2000) Hidden Markov models for monitoring machining tool wear, *Proceedings of the IEEE,* 0-7803-6293-4/00 3887–3890.

Balazinski, M., Czogala, E., Jemielniak, K. and Leski, J. (2002) Tool condition monitoring using artificial intelligence methods, *Engineering Applications of Artificial Intelligence, 15*, 73–80.

Bengio, Y. (1999) Markovian models for sequential data, *Neural computing surveys, 2*, 129–162.

Bicego, M., Murino, V. and Figueiredo, M. A. T. (2003) A sequential pruning strategy for the selection of the number of states in hidden Markov models, *Pattern Recognition Letters, 24*, 1395–1407.

Blimes, J. January (2002) What hmms can do, *UWEE Technical Report nr: UWEETR-2002-0003. URL: http://www/ee.washington.edu.*

Bunks, C., McCarthy, D. and Al-Ani, T. (2000) Condition-based maintenance of machines using hidden Markov models, *Mechanical Systems and Signal Processing, 14* (4), 597–612.

Byrne, G., Dornfeld, D., Ketteler, I. I. G., Konig, W. and Teti, R. (1995) Tool condition monitoring – the status of research and industrial application, *Annals of the CIRP, 44* 541–567.

Cho, D.-W., Lee, S. J. and Chu, C. N. (1999) The state of machining process monitoring research in korea, *International Journal of Machine Tools and Manufacture, 39*, 1697–1715.

Choudhury, S. K. and Kishore, K. K. (2000) Tool wear measurement in turning using force ratio, *International journal of Machine Tools and Manufacture, 40*, 899–909.

Dimla, D. E. (2000) Sensor signals for tool-wear monitoring in metal cutting operations - a review of methods, *International Journal of Machine Tools and Manufacture, 40*, 1073–1098.

Du, R. (1999) Signal understanding and tool condition monitoring, *Engineering Applications of Artificial Intelligence, 12*, 585–597.

Elliott, R., Aggoun, L. and Moore, J. (1995) Hidden Markov models: Estimation and control, *Applications of Mathematics, Springer-Verlag*, 1995 ISBN 0387943641.

Ertunc, H. M., Loparo, K. A. and Ocak, H. (2001) Tool wear condition monitoring in drilling operations using hidden Markov models (hmms), *International Journal of Machine Tools and Manufacture, 41*, 1363–1384.

Fugate, M. L., Sohn, H. and Farrar, C. R. (2000) Unsupervised learning methods for vibration-based damage detection, *Proceedings of the International Modal Analysis Conference, 18*, 652–659.

Ge, M., Du, R. and Xu, Y. (2003) Hidden Markov model based fault diagnosis for stamping processes, *Mechanical Systems and Signal Processing. This article was still in press at the time when this document was being compiled*

Ghasempoor, A., Jeswiet, J. and Moore, T. N. (1999) Real time implementation of on-line tool condition monitoring in turning, *International Journal of Machine Tools and Manufacture,* (39), 1883–1902.

Håkansson, L., Brandt, A., Lägo, T. L. and Cleasson, I. (2003) Modal analysis and operating deflection shapes of a boring bar, in *Proceedings of the International Modal Analysis Conference, 21.*

Jiang, C. Y., Zhang, Y. Z. and Xu, H. J. (1987) In-process monitoring of tool wear stage by the frequency band-energy method, *Annals of the CIRP, 36* (1), 45–48.

Kwon, K.-C. and Kim, J.-H. (1999) Accident identification in nuclear power plants using hidden Markov models, *Engineering Applications of Artificial Intelligence, 12*, 491–501.

Lägo, T. L., Olsson, S., Håkansson, L. and Cleasson, I. (2002) Design of an efficient chatter control system for turning and boring applications, in *Proceedings of the International Modal Analysis Conference, 20*, 4–

Lee, J. H., Kim, D. E. and Lee, S. J. (1998) Statistical analysis of cutting force ratios for flank-wear monitoring, *Journal of Materials Processing Technology, 78*, 104–114.

Lee, J. M., Kim, S.-J., Hwang, Y. and Song, C.-S. Pattern recognition of mechanical fault signal using hidden Markov model, in *International Congress on Sound and Vibration,10*, 4725–4730 Stockholm, Sweden July (2003).

Leem, C. S. and Dornfeld, D. A. (1996) Design and implementation of sensor-based tool-wear monitoring systems, *Mechanical Systems and Signal Processing, 10* (4), 439–458.

Li, C.-J. and Ulsoy, A. G. (1999) High-precision measurement of tool-tip displacement using strain gauges in precision flexible line boring, *Mechanical Systems and Signal Processing, 13* (4), 531–546.

Li, S. and Elbestawi, M. A. (1996) Fuzzy clustering for automated tool condition monitoring in machining, *Mechanical Systems and Signal Processing, 10* (5), 533–550.

Lim, G. H. (1993) Tool-wear monitoring in machine turning, *Journal of Materials Processing Technology, 51*, 25–36.

Miller, I. and Miller, M. (1999) John E Freund´s Mathematical Statistics, *Prentice Hall*, 1999, ISBN 0-13-1236132-X.

Min, B.-K., O´Neil, G., Koren, Y. and Pasek, Z. (2002) Cutting process diagnostics utilising a smart cutting tool, *Mechanical Systems and Signal Processing, 16* (2-3), 475–486.

Novak, A. and Wiklund, H. (1996) On-line prediction of the tool life, *Annals of the CIRP, 45* (1), 93–96.

Park, K. S. and Kim, S. H. (1998) Artificial intelligence approaches to determination of CNC machining parameters in manufacturing: a review, *Artificial Intelligence in Engineering, 12*, 127–134.

Rabiner, L. R. February (1989) A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE, 77* (2), 257–286.

Santochi, M., Dini, G. and Tantuss, G. (1996) A sensor-integrated tool for cutting force monitoring, *Annals of the CIRP, 46* (1), 46–52.

Scheffer, C. (2001) Monitoring of tool wear in turning operations using vibration measurements, Master's thesis, University of Pretoria,.

Scheffer, C. (2003) *Development of a wear monitoring system for turning tools using artificial intelligence*, PhD thesis, University of Pretoria,.

Scheffer, C. and Heyns, P. S. (2001) Wear monitoring in turning operations using vibration and strain measurements, *Mechanical Systems and Signal Processing, 15* (6), 1185–1202.

Scheffer, C., Katz, H., Heyns, P. S. and Klocke, F. (2003) Development of a tool wear-monitoring system for hard turning, *International Journal of Machine Tools and Manufacture, 43*, 973–985.

Sick, B. (2002) Online and indirect tool wear monitoring in turning with artificial neural networks: A review of more than a decade of research, *Mechanical Systems and Signal Processing, 16* (4), 487–546.

Silva, R. G., Reuben, R. L., Baker, K. J. and Wilcox, S. J. (1998) Tool wear monitoring of turning operations by neural network and expert system classification of a feature set generated from multiple sensors, *Mechanical Systems and Signal Processing, 12* (2), 319–332.

Wang, L., Mehrabi, M. G. and Kannatey-Asibu, E. August (2002) Hidden Markov model based tool wear monitoring and turning, *Journal of Manufacturing Science and Engineering, 124*, 651–658.