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

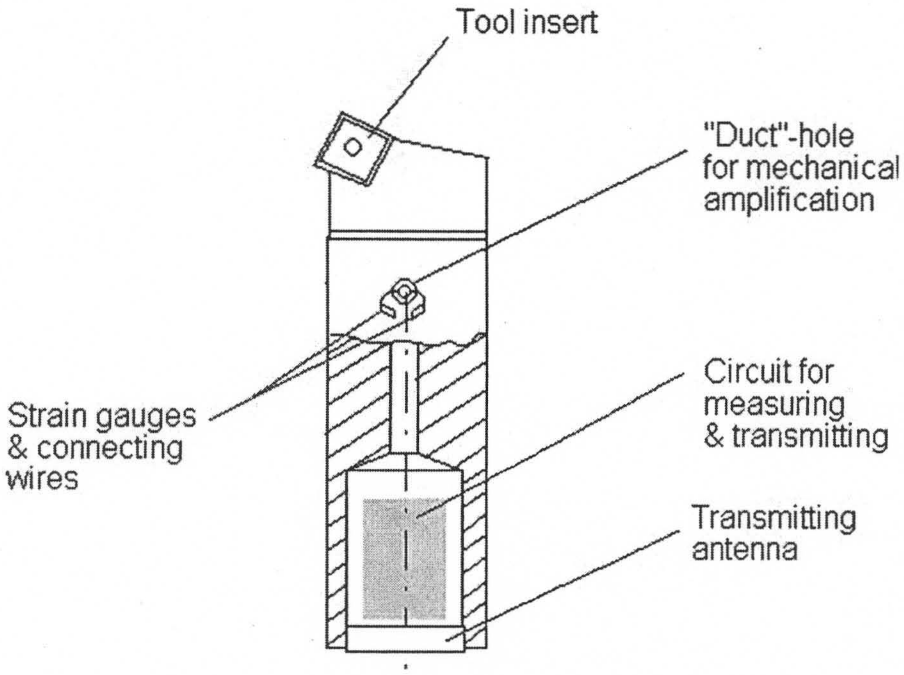

Figure 2.1: The tool holder by Santochi et al. (1996) uses strain gauges to measure cutting force.

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

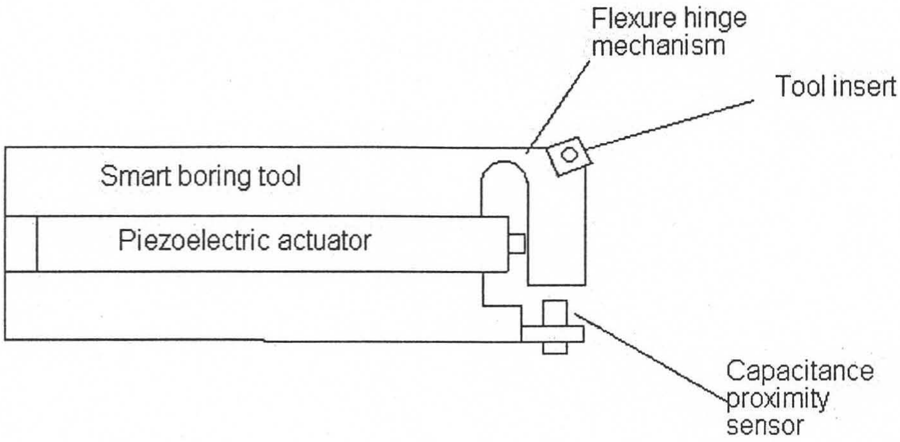

Figure 2.2: The smart tool produced by Min et al. (2002).

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.

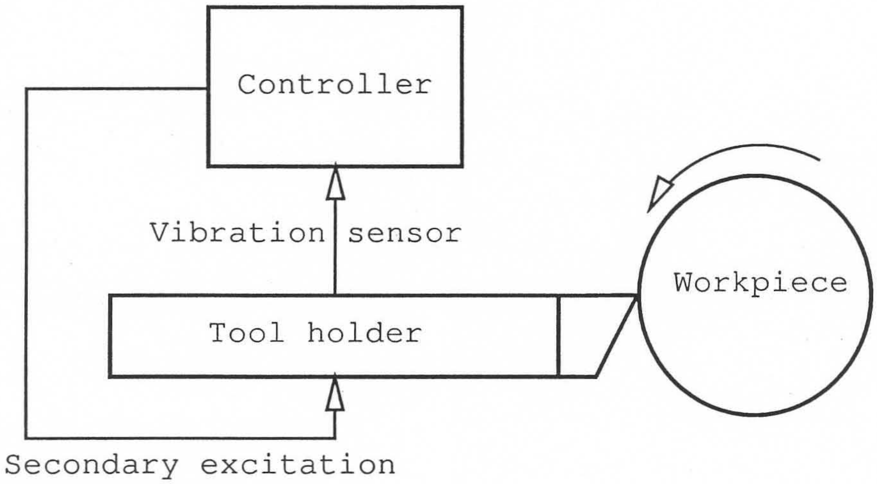

Figure 2.3: This is almost the generic setup for sensor/actuator tool holders. This is also the setup that Lägo et al. (2002), used.

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

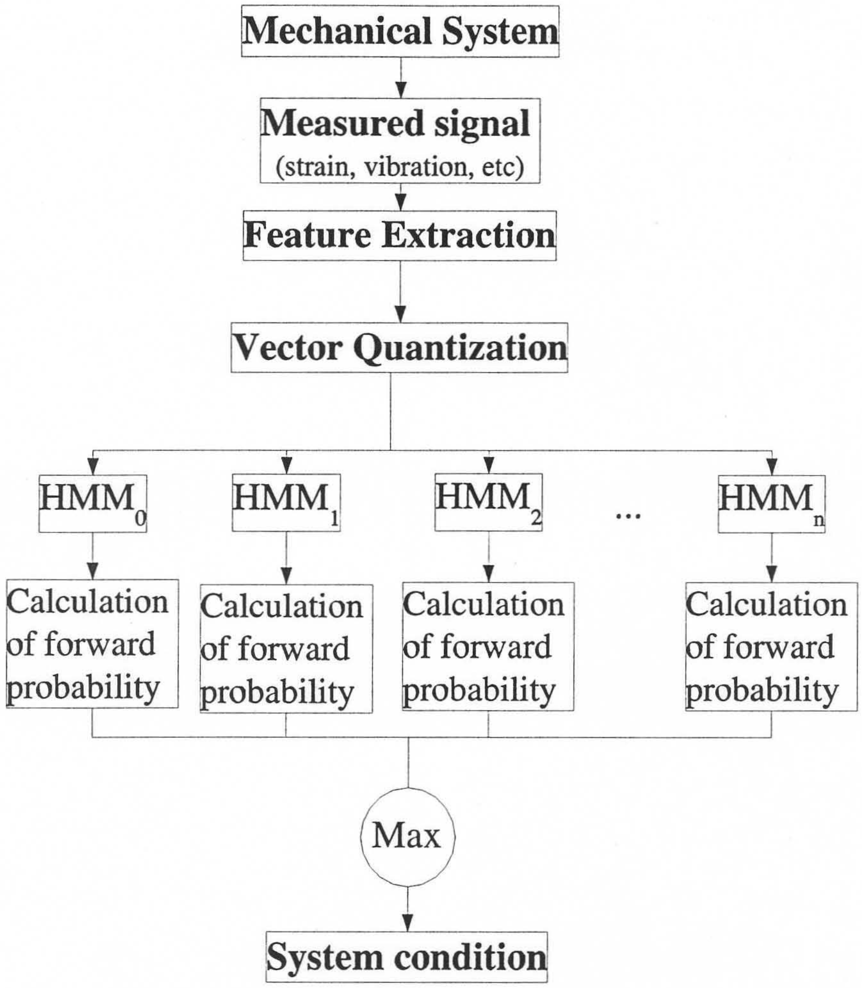

Figure 2.4: Hidden Markov model based fault diagnosis system based on scoring

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.