Sensor Integration Using Neural Networks for Intelligent Tool Condition Monitoring

S. Rangwala[1]
Graduate Research Assistant.
Department of Mechanical Engineering,
University of California at Berkeley

D. Dornfeld
Professor.
Mem. ASME
Department of Mechanical Engineering,
University of California at Berkeley,
Berkeley, CA 94720

A framework for intelligent sensors in unmanned machining is proposed. In the absence of human operators, the process monitoring function has to be performed with sensors and associated decision-making systems which are able to interpret incoming sensor information and decide on the appropriate control action. In this paper, neural networks are used to integrate information from multiple sensors (acoustic emission and force) in order to recognize the occurrence of tool wear in a turning operation. The superior learning and noise suppression abilities of these networks enable high success rates for recognizing tool wear under a range of machining conditions. The parallel computation ability of these networks offers the potential for constructing intelligent sensor systems that are able to learn, perform sensor fusion, recognize process abnormalities, and initiate control actions in real-time manufacturing environments.

Introduction

Successful automation of machining operations relies, to a great extent, on the ability to recognize process abnormalities and initiate corrective action. In the absence of human operators, this function has to be performed with sensors and associated decision-making systems which are able to interpret incoming sensor information and decide on the appropriate control action. According to Dornfeld (1986) an integrated system consisting of sensing elements, signal conditioning devices, signal processing algorithms, and signal interpretation and decision-making procedures constitutes an "intelligent sensor." Development of such sensor systems is a necessary requirement for successful automation of manufacturing processes characterized by noisy and unpredictable environments.

Intelligent sensor systems are expected to replace the knowledge, experience, and sensory and pattern recognition abilities of human operators. Successful implementations of these tasks depend on two factors: first, the quality of information generated by the monitoring sensors and second, the techniques used to process this information in order to make decisions. The first factor relates to the type and number of sensors used, and the signal/noise ratio of the information generated by these sensors. The second factor concerns the learning and decision-making procedures used to analyze this information in the context of the process state. Sensing strategies for unmanned machining should aim at integrating both these factors, thereby allowing for a sensor system design which possesses the ability to successfully mimic the sensory abilities and pattern recognition skills of human operators.

Metal cutting operations constitute a large percentage of current manufacturing activity (Barash, 1980). As a result, there is a strong thrust in research directed at automating the process. An important component of this research is aimed at developing reliable sensor technology for detecting factors such as chip form, tool condition, workpiece roughness, machine vibrations and bearing failure. Tool wear monitoring, which is the focus of the current work, is an area of active research, primarily because the condition of the tool exerts a strong influence on the surface finish and dimensional integrity of the workpiece and vibration levels of the machine tool. The development of reliable tool wear monitoring systems is also expected to reduce tool material costs and machine down times associated with tool change operations. Additionally, the availability of such systems is vital for implementing optimal strategies (such as adaptive control with optimization) in unmanned machining operations.

Several sensing strategies for tool wear detection have been proposed and evaluated in a number of review articles (Birla, 1980; Micheletti, Koenig, and Victor, 1976; Andrews and Tlusty, 1983; and Iwata, 1988). Each technique has its own advantages and drawbacks with the result that no single technique has proven to be completely reliable over a range of operating conditions. An important consideration is that under varying process conditions, the information required to make reliable decisions on the state of tool wear may simply not be available in a single sensor signal. Sensor integration is attractive since loss of sensitivity in one sensor domain can be offset by information from other sensors, enabling successful decision-making ability over a wider range of operating conditions.

In this paper, we present a technique for intelligent tool condition monitoring which employs information from multiple sensors. This information is integrated via a neural network, a

[1]Currently, Member of Technical Staff, AT&T Bell Laboratories, Solid State Technology Center, Breinigsville, PA 18031.

Contributed by the Production Engineering Division for publication in the JOURNAL OF ENGINEERING FOR INDUSTRY. Manuscript received April 12, 1989.

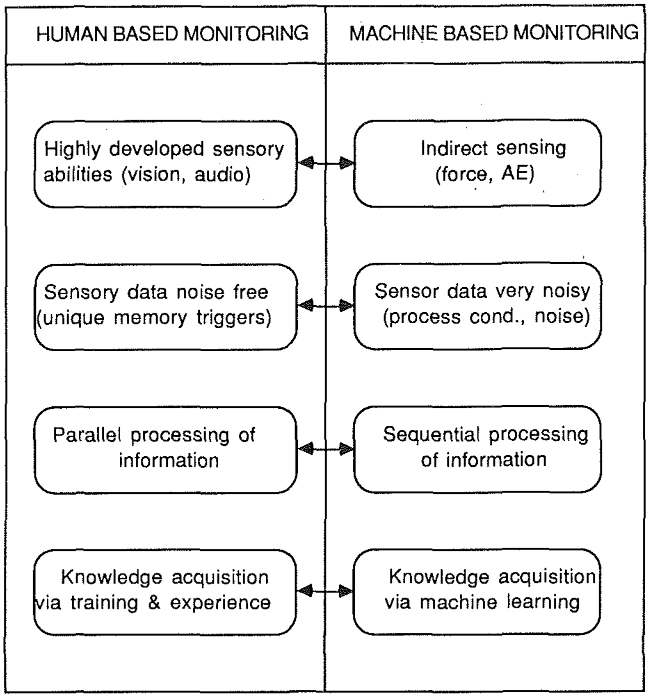

Fig. 1 Human and machine based pattern recognition

parallel computing architecture which can learn to recognize patterns of sensor information and associate them with decisions on the tool wear state. Initial efforts by Rangwala and Dornfeld (1987) demonstrated the feasibility of using neural networks for sensor integration in tool wear monitoring tasks. The networks were used as learning and pattern recognition devices, and were able to successfully associate sensor signal patterns with the appropriate decision on tool wear. Chryssolouris and Domroese (1988) performed simulations in order to study the learning capabilities of these networks. Based on the simulation results, they proposed the use of neural networks as the decision-making component in an intelligent tool condition monitoring system. As shown in this paper, neural networks are able to filter out noise in the sensor and this enhances their ability for successful pattern association tasks. These aspects are experimentally evaluated for tool wear monitoring in a turning operation, under a range of machining conditions.

Intelligent Sensor Systems for Tool Wear Monitoring

A human operator can detect whether a tool is fresh or worn by observing the machining operation and associating patterns of sensory cues with a decision on the tool state. The sensory information used to make this decision is usually of various types: visual (observation of chip color, presence of smoke, deteriorating surface finish of the workpiece), audio (sound generated by rubbing action of tool flank on workpiece), and olfactory (smell of smoke generated due to machining with a worn tool). Associating the sensory cues with tool wear depends to a great extent on the knowledge and experience of the operator. In many cases, information from a single sensor, say audio, may not be sufficient and visual information may also be necessary. The fact that human operators are very successful at the process monitoring task suggests that one possible method for designing computer-based monitoring systems is to model their learning and decision-making abilities after those of a human operator. The philosophy pursued in this paper is that an "intelligent sensor system" should be able to emulate as closely as possible, the learning, pattern recognition and sensor fusion abilities of human operators.

Human pattern recognition is a highly developed and poorly understood characteristic, and the task of simulating it on a computer is a formidable one. The factors involved in human pattern recognition and how they may be mapped in order to develop computer-based pattern recognition capability is shown in Fig. 1. Information describing the process state in the form of sensor signals is required. Typically, indirect sensing techniques such as acoustic emission (AE) and cutting force are used. Ideally, the signal contains the necessary information required to discriminate fresh tool signals from worn tool ones. In many cases however, the sensor signals are extremely noisy and have to be processed in order to yield useful features which are highly sensitive to tool wear but insensitive to noise. The process of mapping the incoming sensor signals into usable features is called feature extraction. The final two components relate to a decision system which processes the incoming signal features and performs a pattern association task—in this case, associating the signal with a decision on the state of tool wear. This decision is usually a binary one (the tool is classified as fresh or worn). The processing could be done sequentially (make decisions sequentially, based on observation of each signal feature) or in parallel (make decisions based on simultaneous observation of all signal features). In order to make correct decisions, machine learning algorithms have to be provided. Such algorithms tune their learning parameters by observing sample signal features corresponding to fresh and worn tool cutting. The pattern recognition approach provides a framework for machine learning and knowledge synthesis in a manufacturing environment, based on observation of sensor data and with minimal human intervention. Additionally, such an approach allows for integration of information from multiple sources (such as different sensors), an aspect which is exploited in the current work.

Use of Multiple Sensors

In the current work, it was decided to use AE and cutting force information in order to develop an intelligent tool condition monitoring system. The primary and secondary shear zones are important sources of AE when cutting with a fresh tool. Kannatey-Asibu and Dornfeld (1981) have presented a comprehensive analysis for AE generation during orthogonal cutting with a fresh tool. In the presence of flank wear, the tool-work interface becomes an additional zone of AE generation due to intense friction between the tool and workpiece surfaces which move past one another at high relative velocities. The effects of tool wear on AE generation in the primary and secondary zones must also be considered. Kobayashi et al. (1960) conducted experiments with artificially ground worn tools and concluded that the presence of a flank land did not have an observable effect on the shear angle. This implies that flank wear does not affect the AE characteristics in the primary and secondary shear zones. However, the presence of crater wear effects the effective rake angle of the tool, and this could affect the generation of AE from the primary and secondary shear zones.

The root mean square (RMS) level of the AE signal (V_{RMS}) measures the total power level of the signal and has been found to be sensitive to the degree of flank wear in a turning operation. Experiments conducted by Lan (1983) for machining of SAE 4340 steel with carbide tools indicate that V_{RMS} increases with machining time due to increased flank wear. However, in cases where the crater wear is significant, V_{RMS} tends to decrease or remains constant. Since the presence of flank wear is expected to increase V_{RMS}, Lan concluded that the effect of crater wear is to cause a drop in V_{RMS}. The fact that V_{RMS} remains constant with increased tool wear due to opposing effects of flank and crater wear makes it difficult to

Table 1: Effect of Velocity & Wear on AE					
Condition	Total power	Low Freq. power	High Freq. power	Mean Frequency	Standard dev. of Frequency
Increased Velocity	+	+	+	-	+
Increased Wear	+	+	-	-	-

design an AE-based tool wear monitoring system which uses only information on the RMS level of the signal.

Emel and Kannatey-Asibu (1988) present experimental data which shows that the power spectrum is sensitive to tool wear and process conditions. Results for machining of AISI 1060 with carbide inserts (Rangwala, 1988) also indicated that the AE power spectrum was sensitive to the level of flank wear and process parameters such as the cutting velocity. Table 1, reproduced from Rangwala (1988), summarizes the qualitative effects of tool wear and cutting velocity on the AE power spectrum. The quantities f_m and σ are the mean and standard deviation of frequency of the power spectrum. The mean frequency, f_m divides the total power of the spectrum into two equal parts, whereas σ indicates the spread in power content around the mean frequency. Table 1 shows that an increase in flank wear and cutting velocity causes an increase in the low frequency (100–300 kHz) power of the AE signal. Other effects such as feed rate and depth of cut changes as well as chip tangling and chip breakage processes are also expected to affect the AE spectral characteristics. An important consideration for tool wear monitoring is that appropriate schemes should be used in order to identify spectral regions which show maximum sensitivity to tool wear under a range of process conditions.

The performance of an AE-based tool wear monitoring system can be enhanced by complementing the AE information with information from other sensors mounted on the machine tool (for example, force or power sensors). The magnitude of the cutting force is sensitive to the occurrence of tool wear in a turning operation (Andrews and Tlusty, 1983). According to Wright (1983), however, cutting force information by itself is inadequate for tool wear detection because its magnitude is also dependent on the cutting velocity. Another problem is that although flank wear tends to increase the cutting force, the accompanying crater wear tends to reduce it, so that the magnitude of the cutting force may not show any sensitivity to tool wear. Cook (1959) and Martin et al. (1974) have shown that the cutting force spectrum (which reflects the dynamic characteristics of the cutting force) is sensitive to tool flank wear. Vibrations in the direction of the cutting force are induced due to flank wear in high frequency regions ($>$5kHz) and lower frequency regions ($<$300Hz). The former is due to vibrations of the tool holder whereas the latter is attributed to workpiece vibrations. The force spectrum is also dependent on process variables such as cutting velocity and feed rate and oscillations in the shear angle during chip formation.

The AE and cutting force information relate to different effects of tool wear. Acoustic emission is sensitive to the microscopic activities (and the resulting stress waves) related to plastic deformation and friction in the cutting zone. The cutting force spectrum is sensitive to the vibrations induced in the tool and workpiece due to the effects of flank wear. The advantage of using AE and cutting force sensors is that they provide information relating to microscopic (stress waves) and macroscopic (vibrations) effects of tool wear. This helps provide better signal features to the pattern classifier, allowing a greater reliability in making decisions on the state of tool wear.

Background on Neural Networks

The human brain consists of a large number of interconnected neurons, each possessing very simple computational abilities. However, the interactions (through a dense system of connections of "synapses") between the neurons allows for parallel processing of information, which greatly enhances the speed of computation and causes a large amount of knowledge to be brought to bear in processing this information (Hinton and Fahlman, 1987). Synapses develop through learning processes in the brain; however, the exact mechanisms are not known at present. Artificial neural networks can be implemented in hardware by using amplifiers as the "neuron" element and resistive connections between the amplifiers represents the strength of the connection between the individual processors. These conductance values or "weights" (which can assume positive or negative values) are the learning parameters and encode knowledge in the system. Physically, the learning procedure involves changing of resistance values in response to the learning process, i.e., adaptive of programmable resistors are required. Also, to allow for negative weights, a method to implement negative resistances is required. Current research efforts are aimed at these implementations issues, for example, the development of programmable resistors (Spencer, 1986; Owen et al., 1982) and on implementation of vast arrays of interconnected processors in VLSI technology (Hubbard et al., 1986). A prototype of a VLSI network employing thin-film resistors has been demonstrated (Graf et al., 1986) for implementing a content addressable memory (CAM). A technique for using a neural network architecture in order to optimize a machining operation is presented in Rangwala and Dornfeld (1987).

In this paper, a special class of neural networks called feedforward networks is used for tool wear monitoring in a machining operation. A software implementation of this network on a serial computer is used as the learning and decision-making component. The structure of this type of a network is shown in Fig. 2. There are three kinds of processing units in such networks: input layer units (or nodes) which accept patterns from the external world, output layer units which generate outputs to the external world and hidden units which do not directly interact with the external world. The role of the hidden nodes is to perform feature extraction on the patterns presented at the input layer. Essentially, this involves noise rejection in the raw sensor patterns, and produces new features with a higher signal to noise ratio. The networks perform pattern association tasks in which a pattern presented at the input layer of the network is associated with a pattern at the output layer. The output layer pattern is composed of the outputs of the nodes in the output layer. In these networks, information propagates from the bottom of the top layer, with connections existing only between processors in adjacent layers.

Let:

$w_{i,j,k}$ = weight between jth processor in $(k-1)$th layer and ith processor in the kth layer.

$net_{i,k}$ = input to ith node in the kth layer.

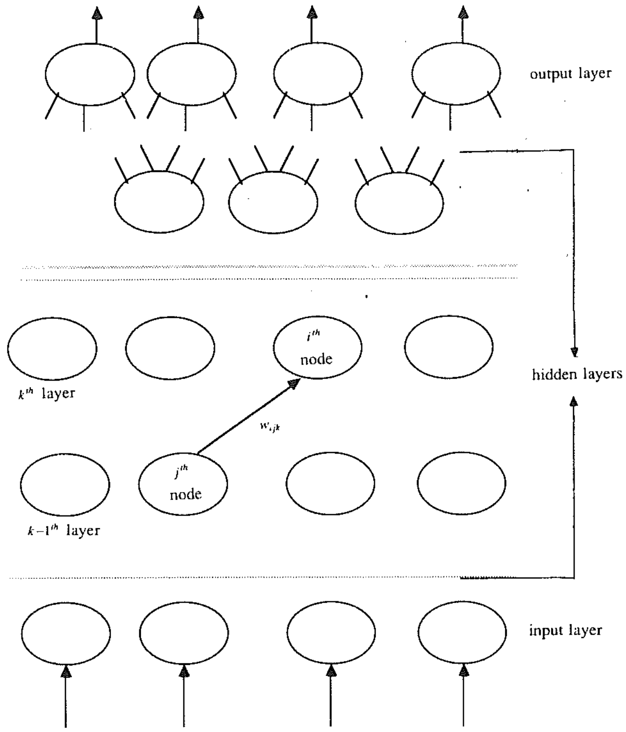

Fig. 2 Structure of a feedforward neural network

$out_{i,k}$ = output to ith node in the kth layer.
$t_{i,k}$ = threshold value associated with the ith node in the kth layer.

The input to a processor is given as:

$$net_{i,k} = \left[\sum_j w_{i,j,k} out_{j,k-1} \right] + t_{i,k} \tag{1}$$

The output of a given processor is assumed to be a sigmoid function of the input and can be expressed as:

$$out_{i,k} = f(net_{i,k}) = \frac{1}{1+e^{-net_{i,k}}} \tag{2}$$

The motivation for using the sigmoid function is that it resembles the firing characteristics of neurons (Hinton and Fahlman, 1987). Other functions (such as the threshold function) are also in use. The sigmoid and threshold functions are shown in Fig. 3.

Although it was realized by some of the earlier researchers (Rosenblatt, 1959; Minsky and Papert, 1969) that multilayered neural networks possessed attractive properties, one of the obstacles to their development was the absence of an efficient learning algorithm for training such networks. The generalized delta rule developed independently by Rumelhart and McClelland (1986) and Le Cun (1985) fills this gap and has been shown to work efficiently on pattern association tasks. This is a supervised learning procedure in which examples of input an output patterns (representing the patterns to be associated) are used to train the network. The rule consists of presenting an input pattern to the network, propagating activity among the various processors according to equation (1) and (2) and computing the pattern at the output nodes with the current set of learning parameters (thresholds and weights of the network). The actual output pattern is then compared to the desired output pattern and the error is calculated as follows:

$$E = \frac{1}{2} \sum_{i=1}^{q} (d_i - a_i)^2 \tag{3}$$

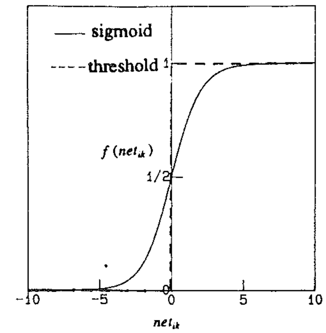

Fig. 3 Sigmoid and threshold functions

where d_i is the desired output at the ith output layer node, a_i is the actual output and q is the total number of nodes in the output layer. The procedure of calculating the error is repeated for all input-output patterns in the traning set and the individual errors are added, to compute the total error. This constitutes the forward pass through the network.

Next, the individual output node errors are propagated backward (from output to inut nodes). During this phase, the weights and thresholds are modified in such a way that the total error term in equation (3) is minimized. Computation of the error term with respect to each weight and threshold is accomplished using local information at each node, so that gradient caculations at each layer can be accomplished in parallel. Rumelhart and McClelland (1986) use the gradient information to adjust the weights and thresholds as follows:

$$\Delta w_{i,j,k} = -\eta \frac{\partial E}{\partial w_{i,j,k}} \tag{4}$$

$$\Delta t_{i,k} = -\beta \frac{\partial E}{\partial t_{i,k}} \tag{5}$$

where η and β are the step sizes in the minimization process.

A linear network is one in which the output of a processor is a linear function of its input, with the input to a processor defined as in equation (1). For a linear system, the error surface is bowl shaped and has a single minimum, so that convergence is guaranteed. In the present case, however, the error function is a nonlinear function of the learning parameters, so that any gradient descent scheme for error minimization is prone to termination in a local minimum. There is no guarantee that a gradient descent procedure will find a set of thresholds and weights so that the error term is zero. However, as pointed out by Rumelhart and McClelland (1986), this does not seem to present difficulties in practical implementations, since the number of hidden layers and number of nodes in each hidden layer can be so chosen that a set of weights and thresholds which drive the error to zero can usually be found. The final values of the thresholds and weights depend on the initial guesses for these values. The initial values of the learning parameters are randomized and usually lie between -1 and 1. Once such a network has been trained using a set of training patterns, it can be used to associate patterns presented at the input layer with appropriate patterns at the output layer. For example, for tool wear monitoring, the input pattern could represent sensory features whereas the output of the output layer nodes could be used to indicate whether the tool is fresh or worn.

There are strong similarities between perceptron type networks (Nilsson, 1965) and the multilayered, feedforward networks discussed above. In fact, multilayer networks have also been referred to as multilayered perceptrons (Nilsson, 1965). Perceptrons perform pattern classification by computing a weighted sum of sensor inputs and comparing it to a threshold value. In this case, there are only two layers of units with the

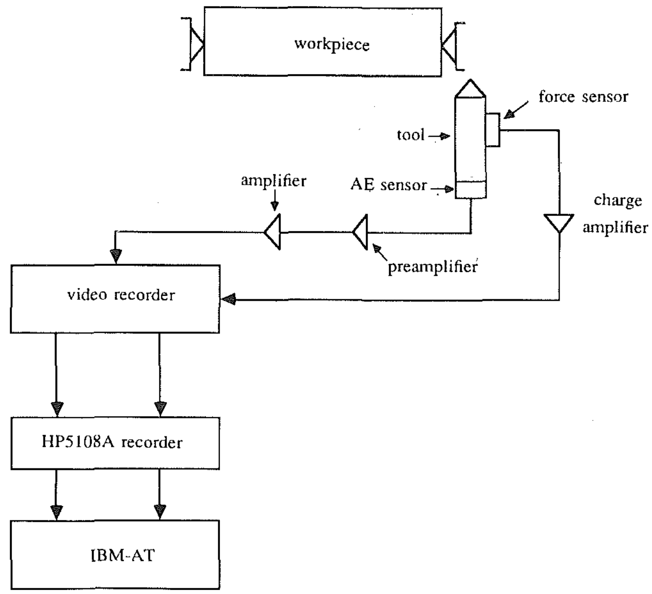

Fig. 4 Schematic of experimental set-up

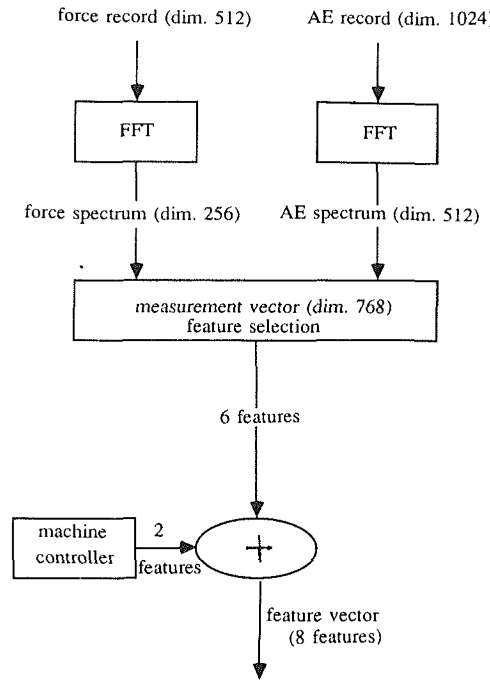

Fig. 5 Signal processing flow diagram

outer layer unit implementing a threshold type function such as that shown in Fig. 3. The perceptron implements a hyperplane in feature space, with the hyperplane surface representing the decision surface. Learning involves rotation of the hyperplane surface until all training samples lie on the correct side of the hyperplane. The learning procedure for feedforward neural networks forces the hidden nodes to perform feature extraction on the raw features, so that as information propagates through the network, noise is suppressed. The last two layers of the network essentially implement a perceptron, however, the features used in this case are the internal features which are relatively noise free.

Description of Experiments

To apply the machine learning approach discussed above, a series of machining tests were conducted on a Tree lathe. A schematic of the experimental setup is shown in Fig. 4. The work material was case hardened AISI 1060 bars (hardened workpieces were used in order to induce faster tool wear). A Kennametal TPGF-322 insert of grade K68 was used. The bars were of nominal diameter 2 in. (50.8 mm) and 12 in. (305 mm) in length.

An acoustic emission transducer (type D9201) was mounted on the tool shank. The tool shank was mounted in a fixture instrumented with a Kistler force dynamometer (type 9251A). The fixture was mounted in the tool turret. The AE sensor output was passed through a preamplifier (with a fixed gain of 40 dB) which high pass filters the incoming signal above 50 kHz. The preamplified signal was passed through an amplifier (5 dB gain) and recorded on the video channel of a modified Sony recorder. The cutting force signal was passed through a charge amplifier and recorded on the audio channel of the Sony recorder.

The process variables were varied in the following range:

Feed rate: 0.002 ipr-0.008 ipr (0.05 mm/rev – 0.20 mm/rev)
Depth: 0.01 in – 0.03 in (0.25 mm-0.75 mm)
Velocity: 278 sfpm-556 sfpm (85 m/min-170 m/min)

No signals were collected while machining the hardened layer (approximately 1.5-2 mm thick) of the workpiece. Signals were collected only when the workpiece diameter was 45 mm (1.75 in) diameter or less. The tool flank land was measured using an optical comparator. The procedure used was to measure the flank land width after every two passes through the soft section of the bar and after every pass through the hardened layer. The tool wear was recorded at flank wear levels of 0.1 mm (0.004 in), 0.125 mm (0.005 in), 0.25 mm (0.01 in), 0.5 mm (0.02 in) and 0.75 mm (0.03 in). Signals collected between these wear levels were ascribed to the wear value at the lower end of the interval. For example, signals collected between 0 and 0.1 mm flank wear were assumed to be generated due to cutting with a fresh tool. Between wear levels of 0.25 mm and 0.5 mm, no signals were recorded, although cutting proceeded. Signals collected during cutting below 0.25 mm flank wear were assumed to belong to fresh tool cutting, whereas signals associated with a flank wear level of 0.5 mm were assumed to belong to worn tool category.

During post-processing, the signals recorded on video tape were played back, filtered and digitized on a HP waveform recorder. The digitized AE signals has a record length of 1024 points, sampled at 5 Mhz, and the digitized force signals were of record length 512 points, sampled at 1 Khz. The sampled AE and force records were synchronized as closely as possible using the tape counter number as a reference. A total of 65 samples of fresh tool cutting and 58 samples of worn tool were collected for purposes of training and testing.

Signal Processing and Feature Selection

A schematic of the signal processing activity is shown in Fig. 5. The force time domain record is of length 512 (sampled at 1 kHz) and the AE time domain record length is 1024 (sampled at 5 Mhz). Using a Fast Fourier Transform (FFT) program yields the power spectrum representations of the time domain records. Consider the power spectrum as a vector whose components are the signal power at various discrete frequencies. The cutting force spectrum is of dimension 256 (256 discrete frequencies with a resolution of 2 Hz) and the AE spectrum is of dimension 512 (512 discrete frequencies with a 5 kHz resolution). Combining the AE and cutting force spectra yields a vector of dimension 768, each component of the vec-

Table 2: Set 1 Features

No.	Feature	J*
1	AE(88 kHz)	0.89
2	Force(43 Hz)	3.22
3	AE(161 kHz)	5.33
4	Force(10 Hz)	8.76
5	AE(122 kHz)	12.76
6	AE(151 kHz)	20.10

* based on 30 training samples

Table 3: Set 2 Features

No.	AE Features		Force Features	
	Feature	J*	Feature	J*
1	88 kHz	0.89	10 Hz	0.56
2	504 kHz	1.79	33 Hz	1.17
3	493 kHz	2.54	39 Hz	2.00

* based on 30 training samples

Table 4: Set 3 Features

No.	Feature	J*
1	AE(88 kHz)	0.89
2	AE(504 kHz)	1.79
3	AE(493 kHz)	2.54
4	AE(68 kHz)	3.31
5	AE(293 kHz)	4.7
6	AE(449 kHz)	5.86

* based on 30 training samples

tor representing the signal power at a discrete frequency in either the cutting force or the AE signal. This vector is referred to as the measurement vector.

Although valuable information may be contained in the entire measurement vector, from practical considerations, only a few of these components can be used for training and pattern association purposes. One important reason is that in training a pattern classifier such as the perceptron, the minimum number of training samples to be used is:

$$N = 2(d+1) \tag{6}$$

where N is the number of training samples and d is the number of features used. This is necessary in order to constrain the training procedure sufficiently, so that generalization behavior of the classifier is acceptable (Cover, 1965). Use of a high dimensional measurement vector would require a very large number of training patterns, and in most cases, this is impractical or expensive.

The approach for reducing the dimension of the measurement vector is to retain only those components of the spectra which show a high sensitivity to tool wear and low sensitivity to noise or process parameters. Considering that the measurement vector is D dimensional, the objective is to select d features which maximize a criterion representing the signal to noise ratio of the features. The selected features are the components of a d dimensional feature vector. The criterion function used in this case uses the concept of intraclass Euclidean distance measures and is duscussed in greater detail in Devijver and Kittler (1982). The criterion used in the current work is:

$$J = \mathrm{trace}(S_w^{-1} S_b) \tag{7}$$

where S_w is the within class scatter matrix and S_b is the between class scatter matrix of the d dimensional feature vector. S_w measures the scatter of data points within a class (fresh or worn) and S_b measures the distance between clusters representing fresh and worn tool data points in the d dimensional feature space. Intuitively, the value of J represents the signal/noise ratio of the feature vectors.

In the current work, 30 measurement vectors (equally divided between fresh and worn tool states) corresponding to various machining conditions were used to estimate S_w and S_b. The final d features were selected using the Sequential Forward Search (SFS) algorithm, developed by Whitney (1971). The algorithm works as follows: out of the D features in the measurement vector, select the one feature which maximizes J. Call this feature ζ_1. Next, pair each of the remaining D-1 features with ζ_1 and compute J according to equation (7) for each of these pair. The pair which maximizes J, is selected as the new feature set. This procedure is repeated until all d features have been selected. It should be pointed out that the SFS algorithm is suboptimal in the sense that it does not guarantee that the best feature set is selected. However, it is computationally viable and yields feature sets whose signal/noise ratio is reasonably close to the optimal case (Devijver and Kittler, 1982).

It was decided that 30 samples (equally divided between fresh and worn tool cutting) would be used for purposes of training. According to the criterion in Equation (6), the dimension of the feature vector was chosen to be 6.

Three feature sets were selected using the procedure discussed above. Set 1 features were selected using a combined measurement vector of the AE and force spectra. Application of the SFS algorithm yielded 4 AE and 2 force features in this case. Set 2 features were selected by considering the AE and force spectra as separate measurement vectors and selecting three features from each. In this case, the feature vector consists of three AE and three force features. Set 3 features were selected considering only the AE spectrum as the measurement vector. The selected features for each set and the corresponding J values as each additional feature is added are shown in Tables 2-4. It is seen that adding new features increases J, since additional features cause the distance between the mean values of fresh and worn tool clusters to increase. Note that these values of J are based on estimates of S_w and S_b, computed using 30 samples of the measurement vector.

Since measurement vectors corresponding to different process conditions are used, the selected features should show a low sensitivity to changes in process variables. However, some sensitivity may still be present, so that it makes sense to use the process conditions as additional features. Information such as the feed rate and cutting velocity is easily available from the machine controller and can be used as additional features. Depth of cut information is difficult to obtain online, and is not used as a feature. Thus a change in sensor feature values due to a change in the depth of cut has the effect of noise corrupting the sensor feature value.

Various design parameters affect the performance of the tool wear monitoring system. These include factors such as the number of training samples, the number of sensors and sensor features used, and the structure of the neural network. In the next section, we evaluate the effect of these factors on the performance of the tool wear monitoring system and arrive at a design which yields the best performance. Although the exact design will change for different situations we believe that the methodology presented in this paper will yield practical design strategies for implementing on-line process monitoring systems.

Discussion of Results

The perceptron training algorithm was used to train a linear classifier, using set 1 features. In order to see the effects of sensor fusion, perceptrons were also trained using set 2 and set 3 features. Unless otherwise specified, all training sets contained 30 samples, equally divided between fresh and worn tool cutting. The trained classifiers were then tested on the remaining 93 samples (of which 50 correspond to fresh tool cutting and the remaining to worn tool cutting). The results are shown in Table 5. Sets 1 and 2 yield comparable performance (88 percent and 87 percent classification success rates, respectively) whereas the performance of set 3 is lower (80 percent success rate). This indicates that feature sets composed of multiple sensor information provide better classification performance. To see the effect of process variables on the sensor features, the relative influence of changes in depth of cut and tool wear on the first AE feature (88 kHz) was studied. We consider one group of samples (group F) corresponding to fresh tool cutting and seven groups of samples (groups W1-W7) corresponding to worn tool cutting. The exact machining conditions for each sample group is shown in Table 6. The value of J for each fresh-worn sample group was calculated and is shown in Table 7. It is seen that in all 7 cases, the AE feature has zero sensitivity to tool wear (a J value of 0.001 or less is considered as zero sensitivity). Notice that group F samples correspond to a higher depth of cut than group W1-W7 samples. Thus, simply looking at the AE feature would cause increases in depth of cut to be mistaken as a "tool worn" condition, and hence lead to classification errors. The sensitivity of the force feature (10 Hz) to tool wear, is also shown in Table 7 and is seen to be reasonably high, irrespective of changes in depth of cut. Including the force feature would, in this case, reduce classifiction errors. Of course, additional AE features may also provide sensitivity to tool wear under these operational conditions (in fact, this is the motivation for using a large number of features). However, as a larger number of features from one sensor are used, the information provided by them becomes highly correlated, so that a loss of sensitivity to tool wear in one feature is accompanied by a loss of sensitivity in other features of the same sensor signal. This aspect is discussed in greater detail in Rangwala (1988).

Feature sets 1, 2, and 3 were used to train and test multilayered neural networks. Sensor feature values were normalized in order to prevent saturation of the sigmoid function. This was done by dividing the feature valve by its maximum value in the training set. Neural networks with a single hidden layer and three nodes in the hidden layer were used. The number of nodes in the input layer is equal to the number of input features, which in the current case is 8 (six sensor features and two process features). The output layer contains a single node, whose output level associates the current input pattern with a decision on tool wear. This yields a network with a 8-3-1 structure.

During the training phase, the target state of the output node was fixed at 0.01 for fresh tool patterns and 0.99 for worn tool patterns. The minimization of the error was achieved by using conjugate gradient optimization, which adjusts the weights and thresholds in a direction which minimizes the error. The weights and thresholds were initialized to uniformly distributed random values lying between -1 and 1. During the testing stage, a pattern presented at the input layer was associated with a "fresh tool" decision if the output node activity was between 0-0.5, else the pattern was associated with a worn tool state.

A threshold value is associated with all nodes in the input, hidden, and output layers. The role of the threshold is to compare the weighted sum of inputs to the node and generate an output which depends on the difference between this sum and node threshold. The threshold value thus acts as a filter for incoming signals. Theoretically, the learning procedure maps worn tool samples to an output node activity of 1 whereas fresh tool samples are associated with zero activity of the output node, so that the signal to noise ratio (measured by the value of the discriminant index, J) of the output node feature approaches infinity. In practice, this does not occur because the output node error does not converge exactly to zero; however, since it is sufficiently close to zero, each filtering step in the network is expected to suppress noise and increase the signal/noise ratio as the signal propagates through the network.

In order to observe the noise suppression behavior discussed above, the trained 8-3-1 network (set 1 features) was presented with all 123 samples and the variation of the J value of the features at each layer were calculated. This is shown in Fig. 6. The value of J increases at every layer, implying higher separability of the fresh and worn tool patterns at the decision layer. In order to observe the increase in J graphically, a two-dimensional feature space was plotted for each layer of the network. The first two sensor features of the network (AE spectral power at 88 kHz and force spectral power at 43 Hz) are shown in Fig. 7. The value of J at this stage, for this pair of features is 0.63. The input node feature pair corresponding to

Table 5: Perceptron performance for different feature sets

No.	Features	No. Misclass.	Success Rate (%)	J*
1	Set 1	11	88	1.02
2	Set 2	12	87	1.71
3	Set 3	18	80	0.72
* based on 123 samples				

Table 6: Description of cutting conditions

machining condition	tool status	feed ipr	vel. fpm	depth of cut (in)	flank land (in)	flank land (mm)
F1	fresh	0.008	450	0.03	0.000	0.00
W1	worn	0.008	450	0.02	0.020	0.50
W2	worn	0.008	450	0.02	0.030	0.75
W3	worn	0.005	370	0.02	0.020	0.50
W4	worn	0.008	278	0.02	0.020	0.50
W5	worn	0.004	370	0.02	0.030	0.75
W6	worn	0.002	278	0.02	0.030	0.75
W7	worn	0.005	556	0.01	0.030	0.75

Table 7: Signal/noise ratio for AE & force features

Feature	F1 & W1	F1 & W2	F1 & W3	F1 & W4	F1 & W5	F1 & W6	F1 & W7
AE (88 kHz)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Force (10 Hz)	0.2	0.1	0.3	12.0	1.0	0.4	0.5

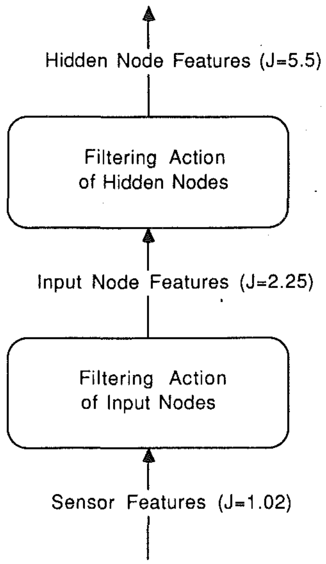

Fig. 6 Filtering action of 8-3-1 network

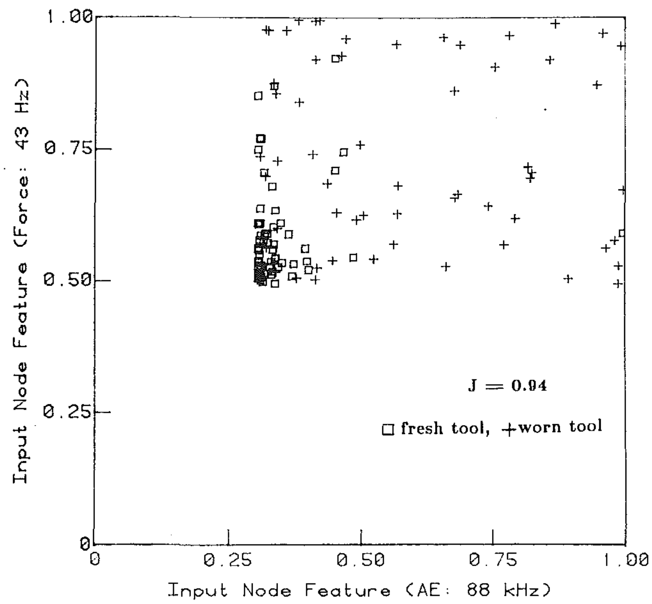

Fig. 8 Two-dimensional feature space [input node features]

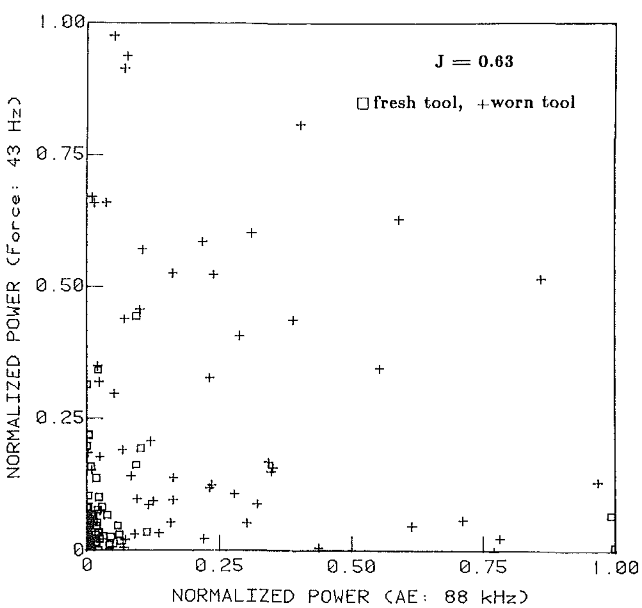

Fig. 7 Two-dimensional feature space [sensor features]

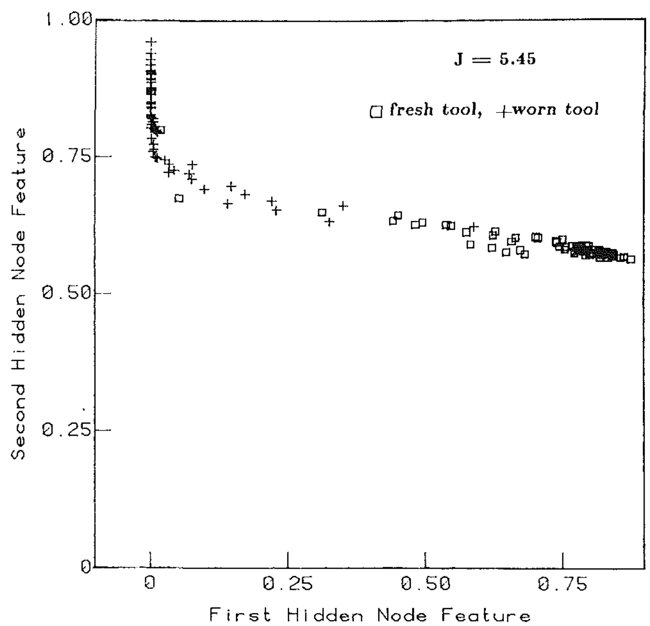

Fig. 9 Two-dimensional feature space [hidden node features]

these features has J value of 0.94. This pair of features is plotted in Fig. 8. Finally, Fig. 9 shows a plot of a pair of hidden node features with a $J = 5.45$. At this stage, the fresh and worn tool clusters are clearly separated, allowing for greater reliability in decision making.

The classification success rate of the 8-3-1 network used above, based on 93 test samples was found to be 94 percent. For comparison purposes, a perceptron network using the same normalized input features as the neural network was also trained and tested. The classification success rate in this case was found to be 88 percent (similar to that obtained with non-normalized features). The superior performance of the neural network is attributed to their noise suppression abilities.

It is of interest to see how the classification performance is affected when the number of features presented to the input layer is varied. To observe this, the least significant feature in the input feature vector was dropped sequentially. Process features were always included as part of the feature vector. The modified vectors were used to train and test the performance of a perceptron and a neural network. The successful classification rate as a function of the number of sensor features used at the input layer of the network is shown in Fig. 10. It is seen that an increase in the number of features used at the inpuit layer generally improves classification performance. For a given number of features, the performance of the neural network is seen to be superior to that of the perceptron. The effect of not using the process features is shown in Fig. 11, where it is seen that the performance is adversely affected when the process features are not included. In this case, it is possible that changes in process conditions are confused as being due to changes in tool wear state, so that the classification error rate increases. In this case also, the neural network performs better than the perceptron network. One aspect that is not fully explained is that although increase in the number of sensor features generally improves performance, in some cases, the use of an additional feature causes a deterioration in the performance level. A possible reason may be that the training and test data statistics for that

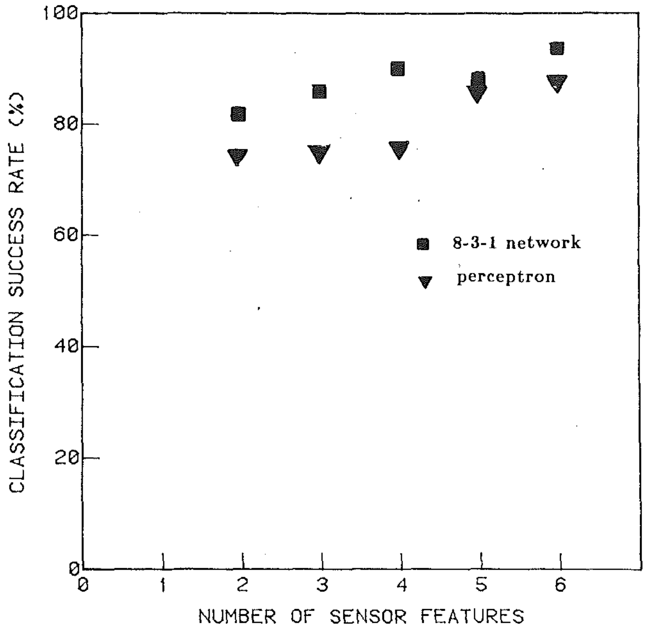

Fig. 10 Classification performance [with process features]

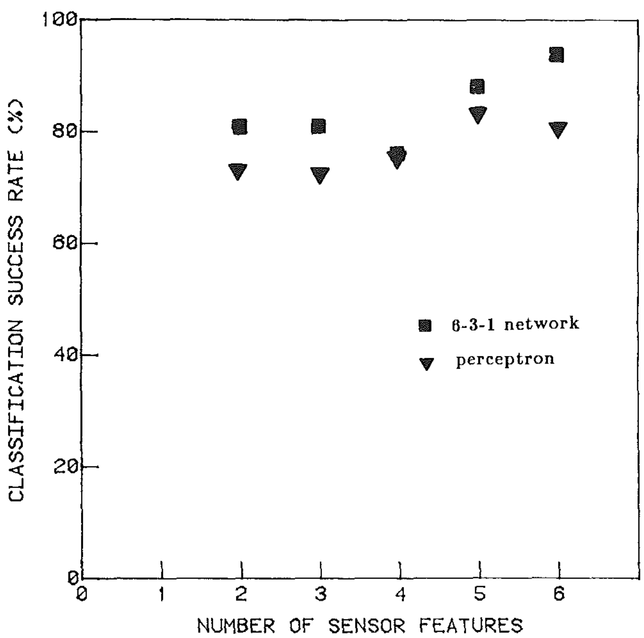

Fig. 11 Classification performance [without process features]

particular feature may be very different, so that the trained network may not be able to respond correctly to test data. Training anomalies may also contribute to this sort of behavior.

The performance of the 8-3-1 networks trained and tested using set 1, 2, 3 features is compared in Table 8. Set 2 features show the best performance (97 percent), and this correlates with the fact that the signal/noise ratio of the sensor features as well as the internal features is highest for set 2. On the other hand, set 3 features (which are very noisy) show poorer performance (84 percent). The data presented in Table 8 suggest that the noise suppression ability of a neural network is a function of the amount of noise present in the sensor features. Noisy sensor features cause the learning ability of the network to degrade, consequently, the internal features are noisy. On the other hand, noise suppression is enhanced when the incoming sensor features have a high signal/noise ratio.

In order to see if the initial values of the learning parameters

Table 8: Performance of neural network

Features	J value (sensor features)	J value (input node features)	J value (hidden node features)	Success rate (%)
Set 1	1.02	2.3	5.5	94
Set 2	1.70	3.6	11.5	97
Set 3	0.72	1.4	2.1	84

or network structure have significant effects, set 3 features were used to train different networks with various initial values. In all cases, the classification performance ranged from 80-85 percent. In some cases, the signal/noise ratio of the hidden node features was found to be lower than that of the input node features. This probably occurs because noise in the sensor features degrades the learning ability of the network. Consequently, even though feature extraction on training set samples increases the signal/noise ratio, performance deteriorates when new samples are propagated through the network.

Conclusions

A structure for an intelligent sensing system for tool wear monitoring was proposed and evaluated. The system utilized multiple sensors (force and acoustic emission) and a neural network as the learning and decision-making component. It was found that use of multiple sensors enhances performance, primarily because they provide independent information (with a higher signal/noise ratio (on the state of tool wear. Through appropriate feature selection procedures, the dependence of these features on changes in process conditions (such as feed rate or cutting velocity) can be minimized. This allows for greater reliability in making decisions on the state of tool wear under a range of machining conditions. The application of neural networks for sensor integration, learning and pattern recognition was demonstrated. The results presented in this paper show that these networks possess the ability for learning and noise suppression. As a result, they are able to transform noisy sensor signal patterns into patterns of information with a highly enhanced signal/noise ratio. This enables a high success rate (about 95 percent) for recognizing tool wear under a range of process conditions.

Acknowledgments

The assistance of Ms. Sheila Slavin in the preparation of the manuscript is appreciated. The authors wish to thank the references for their helpful suggestions.

References

Andrews, G. C., and Tlusty, J., 1983, "A Critical Review of Sensors for Unmanned Machining," *Annals of the CIRP*, Vol. 32, No. 2, pp. 563–572.

Barash, M., 1980, *Computer Integrated Manufacturing Systems*, L. Kops, ed., ASME, New York, pp. 37–50.

Birla, S. K., 1980, "Sensors for Adaptive Control and Machine Diagnostics," in *Machine Tool Task Force Study-Machine Tool Controls*, G. Sutton, ed., SME, Dearborn, Michigan, Vol. 4, pp. 7.12.1–7.12.70.

Chryssolouris, G., Domroese, M., 1988, "Sensor Integration for Tool Wear Estimation in Machining," *Proceedings of the Winter Annual Meeting of the ASME, Symposium on Sensors and Controls for Manufacturing*, pp. 115–123.

Cook, N., 1959, "Self Excited Vibrations in Metal Cutting," ASME Journal of Engineering for Industry, pp. 183–186.

Cover, T. M., 1965, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Transactions on Computers*, Vol. 14, pp. 326–334.

Devijver, P. A., and Kittler, J., 1959, *Pattern Recognition-A Statistical Approach*, Prentice Hall, New Jersey.

Dornfeld, D., 1986, "Acoustic Emission Monitoring for Untended Manufac-

turing," Japan USA Symposium on Flexible Automation, JAACE, Osaka, Japan.
Emel, E., and Kannatey-Asibu, E., 1988, "Tool Failure Monitoring in Turning by Pattern Recognition Analysis and AE Signals ASME JOURNAL OF ENGINEERING FOR INDUSTRY, Vol. 110 No. 2, pp. 137–145.
Graf, H., Jackel, L., Howard, R., Straughn, B., Denker, J., Hubbard, W., Tennant, D., and Schwartz, D., (1986), "VLSI Implementation of a Neural Network Memory with Several Hundreds of Neurons," *Proceedings of the American Institute of Physics Conference on Neural Networks for Computing*, pp. 182–187, New York.
Hinton, G., and Fahlman, S., 1987, "Connectionist Architectures for Artificial Intelligence," *IEEE Computer*, pp. 100–109.
Hopfield, J., 1982, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Sciences*, Vol. 79, pp. 2554–2558.
Hubbard, W., Schwartz, D., Denker, J., Graf, H., Howard, R., Jackel, L., Straughn, B., and Tennant, D., 1986, "Electronic Neural Networks," *Proceedings of the American Institute of Physics Conference on Neural Networks for Computing*, pp. 227–234, New York.
Iwata, K., 1988, "Sensing Technologies for Improving the Machine Tool Function," *Proc. 3rd International Machine Tool Engineers Conference*, Tokyo, Japan, pp. 87–109.
Kannatey-Asibu, E., and Dornfeld, D. A., 1981, "Quantitative Relationships for Acoustic Emission from Orthogonal Metal Cutting," ASME JOURNAL OF ENGINEERING FOR INDUSTRY, Vol. 103, pp. 330–340.
Kobayashi, S., and Thomsen, E., 1960, "The Role of Friction in Metal Cutting," ASME JOURNAL OF ENGINEERING FOR INDUSTRY, Vol. 82, pp. 324–332.
Lan, M., 1983, "Investigation of Tool Wear, Fracture and Chip Formation in Metal Cutting Using Acoustic Emission," Ph.D. Dissertation, University of California at Berkeley, Department of Mechanical Engineering.
Le Cun, Y., 1985, "A Learning Procedure for Asymmetric Threshold Networks," *Proceedings of Cognitiva*, Paris.
Martin, P., Mutels, B., and Drapier, J., 1974, "Influence of Lathe Tool Wear on the Vibrations Sustained in Cutting," *Proc. of the 15th MTDR Conference*, pp. 251–257, Birmingham, AL.
Micheletti, G. F., Koenig, W., and Victor, H. R., 1976, "In Process Tool Wear Sensors for Cutting Operations," *Annals of the CIRP*, Vol. 25, No. 2, pp. 483–494.
Minsky, M., and Papert, S., 1969, *Perceptrons*, MIT Press, Cambridge, MA.
Nilsson, N. J., 1965, *Learning Machines: Foundations of Trainable Pattern Classifying Systems*, McGraw-Hill, New York.
Owen, A. Le Comber, P., Sarraybarouse, G., and Spear, W., 1982, "New Amorphous-Silicon Electrically Programmable Nonvolatile Switching Device," *IEEE Proceedings*, Vol. 129, Part 1, No. 2, pp. 51–54.
Rangwala, S., 1988, "Machining Process Characterization and Intelligent Tool Condition Monitoring Using Acoustic Emission Signal Analysis," Ph.D. Dissertation, Department of Mechanical Engineering, University of California at Berkeley.
Rangwala, S., and Dornfeld, D., 1987, "Integration of Sensors via Neural Networks for Detection of Tool Wear States," *Proceedings of the Winter Annual Meeting of the ASME*, PED 25, pp. 109–120.
Rangwala, S., and Dornfeld, D., 1989, "Learning and Optimization of Machining Operations Using Computing Abilities of Neural Networks," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 2, pp. 299–314.
Rosenblatt, R., 1959, *Principles of Neurodynamics*, Spartan Books, New York.
Rumelhart, D., and McClelland, J., 1986, *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge, MA.
Spencer, E., 1986, "Programmable Bistable Switches and Resistors for Neural Networks," *Proceedings of the American Institute of Physics Conference on Neural Networks for Computing*, pp. 414–419, New York.
Whitney, A., 1971, "A Direct Method of Non-parametric Measurement Selection," *IEEE Transaction on Computers*, Vol. 20, pp. 1100–1103.
Wright, P. K., 1983, "Physical Models of Tool Wear for Adaptive Control in Flexible Machining Cells," in *Computer Integrated Manufacturing*, PED-Vol. 8, pp. 19–31, ASME M. R. Martinez and M. C. Leu, eds., New York.