Tool Wear Estimation from Acoustic Emissions: A Model Incorporating Wear-Rate

S. Varma and J.S. Baras
Center for Auditory and Acoustic Research
Univ. of Maryland at College Park
Maryland, USA
varma@eng.umd.edu, baras@isr.umd.edu

Abstract

Almost all prior work on modeling the dependence of acoustic emissions on tool wear have concentrated on the effect of wear-level on the sound. We give justification for including the wear-rate information contained in the sound to improve estimation of wear. A physically meaningful model is proposed which results in a Hidden Markov Model (HMM) whose states are a combination of the wear-level and rate and observations are the feature vectors extracted from the sound. We also present an efficient method for picking feature vectors that are most useful for the classification problem.

1. Introduction

Much work has been done in real-time monitoring of machinery to detect faults as and when they occur, rather than wait until the next maintenance period. This way, unnecessary maintenance as well as long runs in a faulty condition can be avoided. In the case of a cutting tool, trying to cut with a blunt tool can lead to the breakage of the tool and degradation of the job while pulling the tool off for frequent assessments is expensive in terms of the machinist's time. It is of interest to develop a method that can give an estimate of the wear from easily observable signals. The sound or vibration from the tool-post is one of the simplest signals to measure and it is rich in information relating to the current state of the tool.

Most previous work on estimating tool wear or damage from acoustic emissions has concentrated on using the power spectral density in various ways; the simplest approach being just the average power of the sound signal [8], [11]. A more sophisticated way of using the power spectrum is to compare the total power in various sub-bands [3], [6]. These simple approaches give surprisingly good results in many cases. One approach which uses a learning expert system with torque and thrust information, in addition to vibration data is given in [5].

In [9] the authors try to isolate high-energy transients from the sound signal; one of the assumptions being that transients would be good indicators of chipping or fracture. Another approach, influenced by speech processing, has been to model the dependence of the sound on the wear-level as a hierarchical HMM in multiple time-scales [1]. In a previous article [12], we have explored the applicability of biologically inspired filters to pick out appropriate feature vectors in multiple levels of detail which were then classified according to the wear by a multi-resolution tree structured classifier.

In all of the above work, it has been assumed that the only useful information contained in the sound is that of the wear-level. But it seems reasonable that the sound can also give information about the *wear-rate* at any instant. In particular, chipping is often accompanied by short time-scale transients [7] and chatter is characterized by chaotic vibrations [2].

2. How does the wear influence the sound?

There are two ways in which the wear of the tool can relate to the sound.

1. Different *wear-levels* result in different sounds.
2. Different sounds imply events that result in different *wear-rates*

There is a fundamental difference between these two phenomenon. The way the wear-level affects the sound is independent of the history of the tool. Whichever path the tool took to reach a particular wear-level, the effect on the sound is the same. Thus, if this was the only relationship between the sound and the wear, it would be possible to estimate the

1051-4651/02 $17.00 © 2002 IEEE

wear of the tool at any time by a short sample of the sound at that time. Classifiers without memory would be adequate.

The second relation is more subtle. Events such as chatter affect both the instantaneous wear-rate on the tool as well as the sound produced by the tool. It seems reasonable that large variations in the sound produced by the tool at a constant wear-level could be indicative of variations in the instantaneous *wear-rate*.

3. A mathematical formulation

From what was discussed in the previous section, it seems reasonable to propose that the sound at any time is a stochastic function of both the wear-level and the wear-rate at that time. Thus if we divide time into equal intervals and denote by r_t the wear-rate *during* time interval t and w_t the wear-level *at the end of* time interval t, then the sound produced during time t has a probability distribution that depends on (r_t, w_t). Furthermore, we have

$$w_t = w_0 + \sum_{\hat{t}=1}^{t} r_{\hat{t}} \qquad (1)$$

In this model we have three elements

1. r_t, the sequence of wear-rates for time t. For simplicity we assume that r_t can belong to one of R discrete values and is Markov.

2. w_t, the sequence of wear-levels for time t. Note that specification of the initial wear-level w_0 and a sequence of wear-rates r_t completely specifies w_t through Eq. 1.

3. $x_t \in \mathbb{R}^d$, the sequence of observed feature vectors. x_t is distributed according to a probability distribution $P_{r_t,w_t}(x_t)$ that depends on r_t and w_t.

This results in a *Hidden Markov Model* where the hidden state is the couple (w_t, r_t) and the observation x_t has a distribution that depends on the current state. We will have more to say later about training and testing such a model given observed data.

4. Choosing feature vectors

One problem in building classifiers is choosing feature vectors that adequately compress the information necessary for good classification. We want to pick out components that are most useful for the classification from a set of observations while rejecting components that do not provide any useful information. The *Fischer discriminant* [4] is one way of doing this without actually building classifiers for all possible combinations of feature vectors.

Intuitively speaking, we should pick features such that vectors belonging to one class are separated as much as possible from those from another class. For scalar observations x, Fischer proposed the following measure of separation between observations for class 1 from class 2

$$F = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

where μ_1, μ_2 are the means of the observations belonging to class 1 and class 2 respectively and σ_1, σ_2 the variances. In the case of K classes, the above can be generalized to

$$F = \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} (\mu_i - \mu_j)^2}{\sum_{m=1}^{K} \sigma_m^2} \qquad (2)$$

Now consider an observation vector $x = [x_1, x_2, \ldots, x_d]^T$. Let $a = [a_1, a_2, \ldots, a_d]^T$ be a weight vector and let $y = x^T a$ be a scalar feature derived from x by a weighted combination of the components of x. We can ask what value of a will give a maximum value for the Fischer discriminant (2) for the feature y. Since $\mu_a = E\{x^T a\} = E\{x^T\}a = \mu^T a$ and $E\{(x^T a - \mu_a)^2\} = a^T E\{(x - \mu)(x - \mu)^T\}a$, we can write (2) as

$$F(a) = \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} a^T (\mu_i - \mu_j)(\mu_i - \mu_j)^T a}{\sum_{m=1}^{K} a^T \sigma_m^2 a}$$

Denoting $A = \sum_{i=1}^{K} \sum_{j=1}^{K} (\mu_i - \mu_j)(\mu_i - \mu_j)^T$, $B = \sum_{m=1}^{K} \sigma_m^2$ and $Ca = b$ where C is the invertible matrix such that $C^T C = B$, we get

$$F(b) = \frac{b^T {C^{-1}}^T A C^{-1} b}{b^T b}$$

which attains its maximum value for b equal to the eigenvector corresponding to the largest eigenvalue of ${C^{-1}}^T A C^{-1}$. Weight vector a can be obtained as $a = C^{-1} b$. Eigenvectors corresponding to a decreasing sequence of eigenvalues will give orthogonal weight vectors with successively decreasing value of the Fischer discriminant.

5. Initial classifier for wear

An initial classifier was built incorporating only wear-level information. The classification performance of this model is used as a base against which to measure the improvement in performance when we also include wear-rate information. This model also helps us to pick out features that correspond closely to wear-rate.

Acoustic emissions were measured from an accelerometer mounted on the tool spindle of a milling machine cutting titanium with a 0.5" tool. Data from 5 tools with a total of 12 wear measurements were used for the training set while data from a different group of 8 tools with 13 wear measurements were used for the testing set. Wear measurements vary from 0 thousandths of an inch (*thou.*) to 5 thou. The raw data was divided into frames, each corresponding to one revolution of the tool. The mean squared power in the frequencies from 0–24kHz was divided into 100 bins. The logarithm of the power in each frequency bin was very well fitted with a Normal distribution; i.e., the power is log-normal. The log of the power in the 100 frequency bins was used as observation vectors for each frame. We assumed a

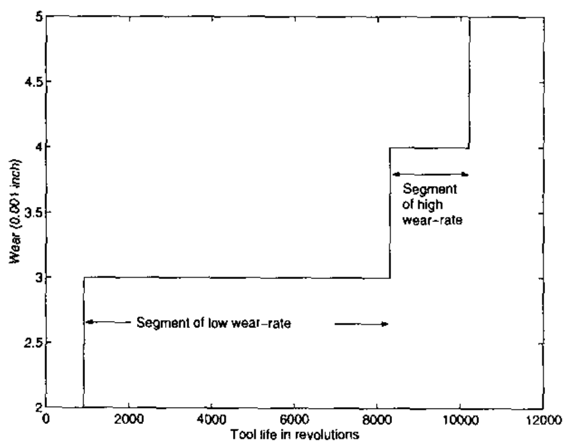

Figure 1. Wear prediction using only wear-level information for tool Ti1

linear increase in wear between successive wear measurements to initially separate the training vectors into 6 wear-level classes from 0–5 thousandths of an inch. Using these wear classes we computed the feature vector with the maximum Fischer discriminant as detailed above. Since wear increases monotonically, a left-to-right HMM was trained on the data to refine our model. The performance of this classifier is presented in Fig.1 and the first row of Table 1.

Table 1. Average absolute wear error

Type of classifier	Avg. error on training set	Avg. error on testing set
Using wear-level information only	0.46 thou.	0.42 thou.
Using wear-level and wear-rate information	0.33 thou.	0.37 thou.

Note: thou=thousandths of an inch

5.1. Wear-rate features

The number of time steps it takes to increase wear by 0.001 inch is a measure of the average wear-rate. We use the wear-level model to classify the training sequence and obtain segments as shown in Fig.1. The segments are divided into two sets; one with all the high wear segments and the other with all the low wear ones. The mean and variance of these two sets are used to find a feature that separates them maximally according to the Fischer discriminant. This feature was used as the wear-rate feature. Thus our feature vector is 4-dimensional with the first three components indicative of wear-level and a fourth component that corresponds to wear-rate. Fig.2 shows the weight vectors

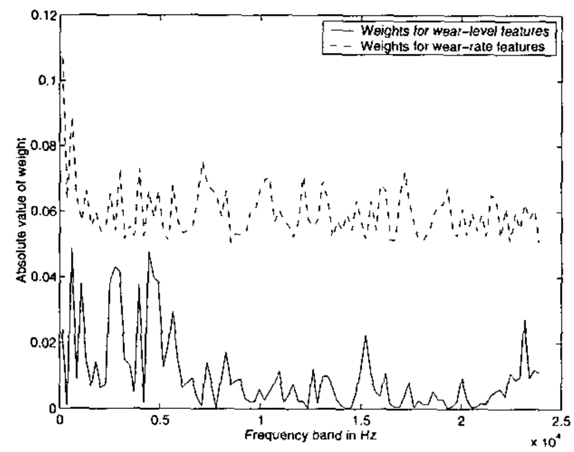

Figure 2. Weights in different frequency bands for best wear-level and wear-rate features

that correspond to maximum Fischer discriminant for wear-level and wear-rate. It is interesting to note that a few of the low frequency bands are the most indicative of wear-level while features that correspond to wear-rate are of considerably broader bandwidth. This confirms our intuition that short time-scale, broadband transients are the primary indicators of wear-rate.

6. Training and testing of combined model

Training uses the Baum-Welch algorithm [10] where, starting from an initial model, we calculate the expected values of the parameters given the observations. This gives an estimate for the parameters with a higher likelihood and iteratively repeating this step gives a sequence of models with monotonically increasing likelihood. This process converges to a model (set of parameters) that locally maximizes the likelihood.

To apply the Baum-Welch algorithm, we consider the set of all wear-rate sequences that would take us from w_0 to w_T in time 0 to T. Each sequence of wear-rates specify one unique sequence of wear-levels. Thus taking the set of all sequences of wear-rates and levels we can compute the *a-posteriori* probability of observing each of the states as well as the transition probabilities between any two states. We also obtain the conditional probabilities of obtaining the given observation sequence. The observation probabilities are assumed to be independent between the wear-rate and wear-level features. What this means is that the probability of observing a given wear-rate feature is independent of the wear-level feature that was observed.

Once the model is trained, we can compute the wear-rate sequence $\{r_t\}$ with the maximum likelihood for a given observation sequence. This is done through the Viterbi algorithm where we find the best (in terms of highest likelihood) sequence that ends in a particular state i at time t in a recursive manner for all i. A maximum likelihood estimate for the wear w_t at any time is thus possible through Eq. 1.

7. Results and conclusions

Fig.3 shows the maximum likelihood sequence of wear-levels for one particular tool along with wear measurements. The second row of Table 1 shows the performance of the classifier incorporating wear-rate and wear-level where it can be compared with the classification error using just the wear-level information. A significant improvement is obtained when the wear-rate is incorporated into the model. Current work concentrates on applying this methodology to different tool geometries (1 inch tool vs. 0.5 inch tool) and job materials (steel vs. titanium). Our goal is to develop classifiers that generalize to different tool and job configurations.

Acknowledgments

The authors would like to thank Dr. Gary Bernard of Boeing Co. for giving us access to the machining data. This research was funded in part by grant N-000149710501EE from the Office of Naval Research.

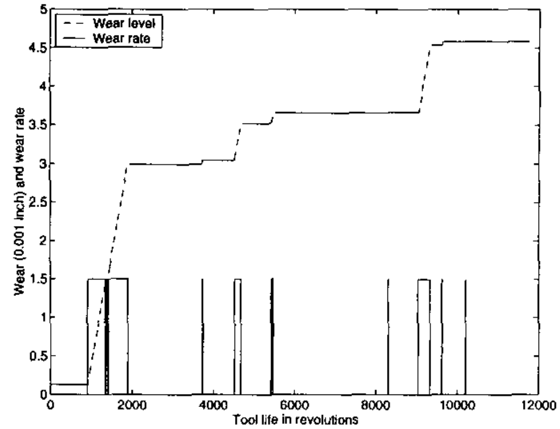

Figure 3. Wear prediction using wear-level and wear-rate information for tool Ti1

References

[1] L. Atlas, M. Ostendorf, and G. Bernard. Hidden Markov models for monitoring machining tool-wear. *Proc. of ICASSP*, pages 3887–3890, 2000.

[2] B. Berger, M. Rokni, and I. Minis. Complex dynamics in metal cutting. *Quarterly of App. Math.*, 51:601–612, 1993.

[3] M.-F. Chen, M. Natsuake, and Y. Ito. Preventive monitoring system for main spindle of machine tool - comparision of signal processing methods. In *Proc. 27th Inter. MATADOR Conf.*, pages 179–186, April 1988.

[4] R. Fischer. The case of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(II):179–188, 1936.

[5] S. Hong. Self-learning system for knowledge based diagnosis of drill condition in fully automatic manufacturing system. *Proc. SPIE*, 1707:195–206, 1992.

[6] I. Inasaki and S. Yonetsu. In-process detection of cutting tool damage by acoustic emission measurement. In *Proc. 22nd Int. Machine Tool Design and Res. Conf.*, pages 261–268, 1981.

[7] P. Loughlin and G. Bernard. The applicability of time-frequency analysis to machine- and process-monitoring. In *Proc. SAE Aerospace Manufacturing Technology and Exposition*, pages 243–252, June 1997.

[8] A. Noh and A. Kobayashi. Cutting tool failure detector. In *Adv. in Hard Material Tool Tech.:Proc. Inter. Conf. on Hard Materials Tool Tech.*, pages 414–426, 1976.

[9] L. Owsley, L. Atlas, and G. Bernard. Self-organizing feature maps and hidden Markov models for machine-tool monitoring. *IEEE Trans. on Sig. Proc.*, 45(11):2787–2798, November 1997.

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

[11] S. Shumsheruddin and J. Lawrence. In-process prediction of milling tool wear. In *Proc. 24th Int. Machine Tool Design and Res. Conf.*, pages 201–214, 1983.

[12] S. Varma, J. Baras, and S. Shamma. Biologically inspired acoustic wear analysis. In *Nonlinear Sig. and Image Proc. Conf.*, Baltimore, MD, June 2001.