TECHNICAL RESEARCH REPORT

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

by S. Varma, J.S. Baras

CAAR TR 2001-2
(ISR TR 2001-49)

CAAR

**Center for Auditory
and Acoustic Research**

Web site http://www.isr.umd.edu/CAAR/

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

S. Varma, J.S. Baras

Institute for Systems Research, Univ. of Maryland at College Park
E-mail: `varma@eng.umd.edu, baras@isr.umd.edu`

Abstract

Almost all prior work on modelling the dependance 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 plausible model is proposed which results in a Hidden Markov Model (HMM) whose hidden states are 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 are 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 density spectrum 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].

Another approach is presented in [9] where the author tries 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 dependance of the sound on the wear-level as a heirarchical 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 indepandent of the history of the tool. Whichever path the tool took to reach this 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 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 plausible 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. Sequence $\{r_t\}$ which is 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. $\{w_t\}$ which is the sequence of wear levels for time t. Note that specification of w_0 and a sequence of r_t completely specifies a sequence of w_t through Eq.1.

3. $\{x_t\}$, with $x_t \in \mathbb{R}^d$ is the sequence of feature vectors observed. x_t is distributed according to a probability distribution $P_{r_t, w_t}(x_t)$ that depends on r_t and w_t. As the simplest first approximation, we assume that P is Normal with mean and variance depending on r_t and w_t.

This results in a *HIdden Markov Model* where the state is (w_t, r_t), the wear-level and wear-rate at time t and the observations x_t have a distribution that depends on the current state. To train such a model from observations we would use the Baum-Welch algorithm [10] to obtain a set of parameters that locally maximize the likelihood.

4. Choosing feature vectors

One problem in building classifiers is choosing feature vectors that adequately compress the information necessary for good classfication. We want to pick out components that are most useful for the classification from a (possibly large) 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 of vectors of class 1 from class 2

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

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

$$F = \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} (\mu_i - \mu_j)^2}{\sum_{m=1}^{K} \sigma_m^2} \qquad (3)$$

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 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 criterion (3). 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 (3) as

$$
\begin{aligned}
F(a) &= \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} (\mu_i^T a - \mu_j^T a)^2}{\sum_{m=1}^{K} a^T \sigma_m^2 a} \qquad (4) \\
&= \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} a^T (\mu_i - \mu_j)(\mu_i - \mu_j)^T a}{\sum_{m=1}^{K} a^T \sigma_m^2 a} \\
&= \frac{a^T A a}{a^T B a}
\end{aligned}
$$

where $A = \sum_{i=1}^{K} \sum_{j=1}^{K} (\mu_i - \mu_j)(\mu_i - \mu_j)^T$ and $B = \sum_{m=1}^{K} \sigma_m^2$.

Denoting $Ca = b$ where C is the invertible matrix such that $C^T C = B$

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

which attains its maximum value for b equal to the eigenvector corresponding to the largest eigenvalue of $C^{-1^T} A C^{-1}$. Denote by $e^b = \{e_1^b, e_2^b, \ldots, e_{K-1}^b\}$ the eigenvectors corresponding to the largest $K - 1$ eigenvalues arranged in descending order, where K is the number of classes. All other eigenvalues will be zero. Then $e^a = \{e_1^a, e_2^a, \ldots, e_{K-1}^a\} = \{C^{-1} e_1^b, C^{-1} e_2^b, \ldots, C^{-1} e_{K-1}^b\}$ is the set of weights a arranged in order of decreasing Fischer discriminant values.

The weight a can also been interpreted as a direction in the d-dimensional space of observation vectors. Then e_1^a is the direction that corresponds to a maximum in the Fischer discriminant. e_2^a is the direction orthogonal to e_1^a that gives the maximum and so on. Transforming x to

$\hat{x} = [x^T e_1^a, x^T e_2^a, \ldots, x^T e_{K-1}^a]^T$ gives us a feature vector where the information necessary for classification has been efficiently represented.

5. Initial classifier for wear

Acoustic emissions were measured from an accelerometer mounted on the tool spindle. The raw data was divided into frames, each corresponding to one revolution of the tool. The energy in the frequencies from $0 - 24$kHz was divided into 100 bins. It was found that the logarithm of the power in each frequency bin was very well fitted with a Normal distribution; i.e., the power is lognormal. Thus the logarithm of the power in the 100 frequency bins was used as observation vectors for each frame. We assumed a
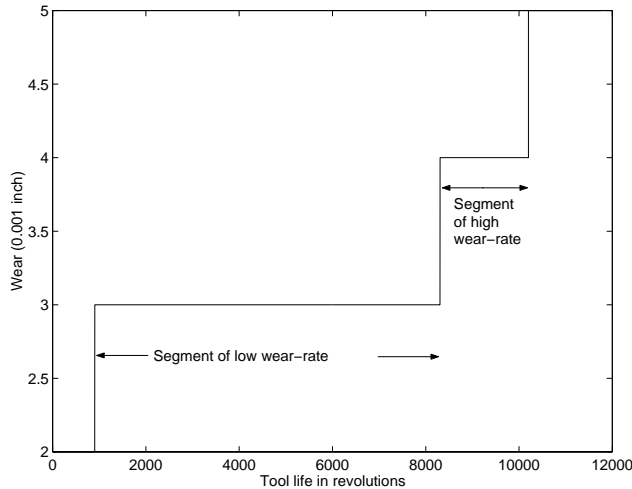

Figure 1. Wear prediction using only wear-level information for tool Ti1

linear increase in wear between 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 for 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 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. The performance of this classifier is presented in Table 1 and Fig.1.

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 produces the maximum in Fischer discriminant between them. 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
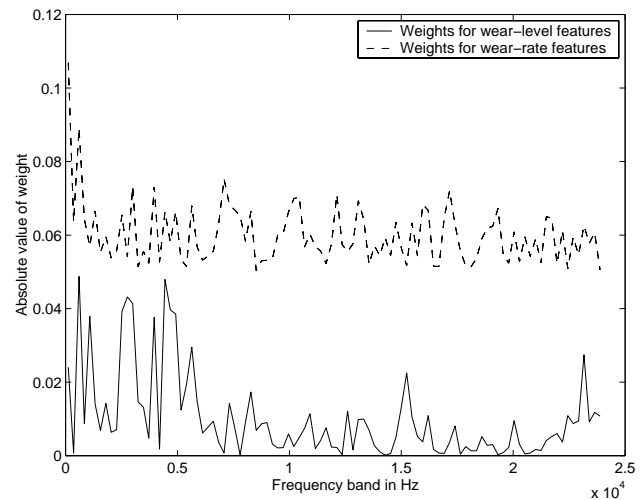

Figure 2. Weights in different frequency bands for best features for wear level and wear rate

Table 1. Average absolute wear error in thousandths of an inch

Type of classifier	Error on training set	Error on testing set
Using wear-level information only	0.46	0.42
Using wear-level and wear-rate information	0.33	0.37

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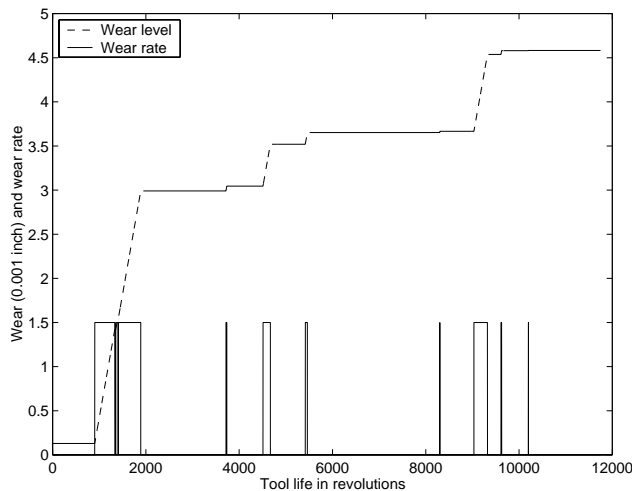
Figure 3. Wear prediction using wear-level and wear-rate information for tool Ti1

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 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. 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.

Once the model is trained, we can compute the state 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.

7. Results and conclusions

Fig.3 shows the maximum likelihood sequence of wear levels for one particular tool along with wear measurements. Table 1 shows the performance of the classifier incorporating wear-rate information in addition to wear-level

where it can be compared with the classification error using just the wear-level information. Although the number of wear measurements is not enough to make strong statements about any improvement in performance, we have enough evidence to support our approach.

References

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

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

[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. MATA-DOR Conf.*, April 1988, pp. 179–186.

[4] R. Fischer. "The case of multiple measurements in taxonomic problems". *Annals of Eugenics*, Vol. 7, part II, 1936, pp. 179–188.

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

[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.*, 1981, pp. 261–268.

[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*, June 1997, pp. 243–252.

[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.*, 1976, pp. 414–426.

[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.*, Vol. 45, No. 11, November 1997, pp. 2787–2798.

[10] L. R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE*, Vol. 77, No. 2, February 1989, pp. 257–286.

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

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