

HIDDEN MARKOV MODELS FOR MONITORING MACHINING TOOL-WEAR

Les Atlas, Mari Ostendorf,* and Gary D. Bernard***

*Department of Electrical Engineering
University of Washington
Box 352500
Seattle, WA 98195-2500

**Manufacturing Research & Development
Boeing Commercial Airplane Group

ABSTRACT

As summarized by Atlas, Bernard, and Narayanan [1], the sensing of acoustic vibrations can remotely estimate the state of wear at the tool edge. This form of monitoring offers the potential to characterize, in real time, the efficiency of metal removal processes such as drilling and milling. For example, information about sudden increases in tool wear, if manifest as a change in acoustic vibration, could be valuable to a machine operator. The nature of this monitoring problem has some similarities to automatic speech recognition. For example, there is significant tool-to-tool variation in details of vibration and lifetime. Also, the easy adaptability of monitoring systems across manufacturing processes is important. In this work we model the evolution of vibration signals with the same technique which has shown to be successful in speech recognition: hidden Markov models (HMMs). We focus on the monitoring of milling processes at three different time scales and show the how HMMs can give accurate wear prediction.

1. INTRODUCTION

To date, widespread use of machine monitoring systems has been limited by several factors, most notably by their inability to accommodate changes in manufacturing operations. Under current diagnostic approaches, each machine configuration requires the development and manual tuning of a complex diagnostic system by an expensive, highly specialized signal processing expert. This requirement is incompatible with modern manufacturing practices, where fast machine reconfigurations must be carried out to meet frequently changing product demands. The inflexibility and expense of the diagnostic systems present a significant barrier to future developments in high-quality flexible manufacturing. Our past year's work has contributed to reducing this barrier. Namely, we have developed a widely applicable and reconfigurable, real-time transient classification system that, in a large-scale manufacturing environment, predicts failures. Our system design is based upon the premises listed below.

Vibration and acoustic sensors provide much needed information: It is possible to extract early-warning prognostication information from vibration and acoustic sensors.

Advanced signal processing techniques are needed:

Conventional and commercially available signal processing techniques such as stationary parametric or non-parametric spectral analysis and short-time Fourier transforms are inadequate representations for monitoring and prognostication tasks. Instead, more advanced techniques, such as proper time-frequency distributions, are required.

Advanced classification techniques are required: Conventional and commercially available statistical estimation and classification techniques are inadequate for monitoring and prognostication tasks. Instead, more advanced techniques, such as time-frequency representations [1], for the determination of salient features, and hidden Markov models, for classification of nonstationary trends and events, are required.

Both parametric (mechanical-model driven) and non-parametric (data-trained) approaches need to be efficiently combined:

The integration of partial information about the mechanics of the systems being monitored is essential. However, adaptation and "training-by-example" can be used to refine this incomplete information. This approach is appropriate for hidden Markov models, where the state topology is based on mechanical models, and the state transition and emission probabilities are automatically estimated from data.

Adaptability is key: A cost-effective and useful system must be adaptable across a wide variety of applications.

2. RELATED PREVIOUS WORK

Over the past three decades, the manufacturing industry has realized the importance of automation, due to increased global competition and the potential economic benefits [2]. Consequently, a substantial amount of research has gone into the field of control of machining processes. Ulsoy et al [3,4] provide a detailed description of the state-of-the-art in machine tool automation through the 1970's and 1980's respectively. To summarize, the accomplishments of the 1970's had not solved the problems but clearly indicated the complexity of the machining system, including the non-linear, non-stationary and multivariate nature of the processes to be controlled, and, as a result, the 1980's saw increased research in the use of advanced methods for automating, monitoring and controlling manufacturing processes. In contrasting to the automation

emphasis of past work, it is our objective to set the stage for the next decades of progress by developing techniques which enhance machine operators' skill.

Najafi and Hakim [5] have compared various standard non-parametric spectral estimation techniques applied to machine vibration data. Ramirez and Thornhill [6] analyze the drilling forces signal and its spectra for use in monitoring circuit board manufacture. Narayanan, Bernard, Atlas and Fang [7] describe front-ends that capture features that preserve temporal resolution, and stress the need for a public-domain manufacturing database. Fang, Atlas and Bernard [8] have made use of quadratic energy detectors to accurately detect temporal events in acoustic emissions while drilling holes on a composite honeycomb sheet. Kittel and Hayes [2] present a symbolic signal representation for process monitoring signals. We believe that signal analysis is an area where our group has made substantial contributions by bringing into the field of manufacturing a variety of new non-stationary signal analysis techniques.

Li, Ma, Hwang and Nickerson [9] have developed a pattern recognition analysis scheme based on features extracted from rolling element bearing vibrations using bispectral analysis. Zheng and Whitehouse [10] have observed that the moments of the Wigner distribution of sensor outputs are useful for detecting incipient chatter and characterizing changes in the workpiece. Due to the vast multitude of sensor signals encountered in this field, and due to the large variance of the sensor signal within the same class, we argue that an automatic feature selection procedure is necessary for systems to perform under a variety of conditions.

Choi, Wang and Dornfeld [1] have attempted to simulate the relationship between the input (cutting parameters) and output (sensor outputs) of the turning process using a neural network. Tansel and McLaughlin [12] have used a supervised neural network system for detection of tool breakage in milling operations. Sick has shown how local temporal variability can be captured with time-delay neural networks [13]. Heck and McClellan [14] present a multisensor continuous density left-to-right hidden Markov model based approach for tool wear detection and prediction. More recent results on systems which use hidden Markov models for prediction of end mill and tool wear can be found in McLaughlin *et al* [15] and Owsley *et al* [16].

3. BACKGROUND ON HIDDEN MARKOV MODELS

A hidden Markov model (HMM) is a stochastic model of a process that has piecewise stationary regions, where the time evolution of the non-stationary behavior can be characterized in terms of an unobserved discrete Markov chain. An HMM is specified by the set of discrete states, initial state probabilities, transition probabilities between states, and a set of state-dependent observation distributions. Standard algorithms exist for efficient search for the maximum likelihood state sequence (Viterbi algorithm) and for parameter estimation (Baum-Welch algorithm). A brief tutorial on HMMs for speech recognition can be found in Rabiner [17]; HMM system design issues for

large vocabulary recognition applications are discussed in Young [18]; and numerous books on speech processing and pattern recognition now cover this model.

Hidden Markov models are useful, in general, for problems where there are temporal dynamics. For tool monitoring, it may be useful to model observation dynamics at multiple levels, analogous to the different levels in speech recognition. At the highest level is the increasing wear of a tool which, though inherently continuous, can be represented at a quantized level with a left-to-right state topology, as first proposed by Heck and McClellan [14]. Detecting the wear "state" is more useful to a machinist than a binary decision of "worn" vs. "not worn," because it gives an indication of the remaining lifetime of a tool. Representing more than two states also allows for lower variance (and thus less confusable) models. The wear state can be thought of as analogous to the word in speech recognition. Depending on the type of cut, there may be multiple phases with different observation characteristics such as entry and exit into a work piece, which would be analogous to the specification of a word pronunciation in speech recognition. Lastly, there are dynamics over short time scales that are relevant for classification, such as within a transient in the tool monitoring process or within a phoneme in speech. HMMs can be simultaneously used at more than one of these levels, but so far most work has involved only one or two levels.

The analogy to speech recognition is not perfect, however, since there is no clear definition as yet for the fundamental "subword" units of the drilling process. Nor is it clear that such units could be learned automatically because of the chaotic nature of the wear process and the fact that the wear state is not inherently discrete as words are. However, one can represent temporal dynamics at this intermediate level by using *a priori* knowledge of the physical process of cutting to specify simple state progressions, in combination with mixture distribution models to automatically learn chaotically occurring events.

A related and perhaps more important difference between speech recognition and tool wear monitoring is that the amount of labeled data is sparse because it is so expensive to collect. This complicates training as well as evaluating classifiers.

4. HIDDEN MARKOV MODELS FOR PREDICTION OF WEAR

In this paper we focus on three different time scales of application of HMM's to modeling vibrations, as sensed by accelerometers, from end mill processes.

All experiments used training and testing data that were taken from an accelerometer mounted on the spindle housing of a numerically controlled machining center that was climb-cutting notches in hard metal. Data sets were segmented into passes, one pass being the period from the time the tool first touches the work piece to the time the tool cuts air after leaving the work piece.

We observed and modeled separately three time scales: the progression of the tool from sharp to worn; the dynamics

associated with entry, bulk machining, and exiting each pass; and very short, yet potentially meaningful, transients.

4.1 Results for Transients: The Shortest Time Scale

When looking at the accelerometer signal in detail, we found that short and energetic signals, which we call “transients,” occurred throughout the machining process. These transients were of very short duration (averaging roughly 0.0065 sec. in length) yet the rate of occurrence, energy, or durations of the transients did not correlate with tool wear. Our sample rate was high enough to show detailed time-frequency information for each transient. We also observed that this time-frequency information had a fair amount of complexity. We thus encoded [16] each transient as a vector quantized time-series, and used this encoding as our front-end representation for the HMMs.

In order to test the hypothesis that these short transients encode information which relates to the life of the tool, we divided up 750 transients, which were observed over the life of the tool, into 3 sets of 75 training transients and 525 separate test transients. The training transients were taken from three regions of the tool life—early, middle, and late—and three distinct HMM’s were trained on 75 transients each. Thus, one HMM potentially models early, the second HMM models middle, and the third HMM models late transients.

The 525 test transients were matched against the three HMMs and the proportion assigned to each of the three different classes is plotted in Figure 1. This result was then cross-validated by repeating the experiment with a totally distinct new assignment of 225 training and 525 test transients.

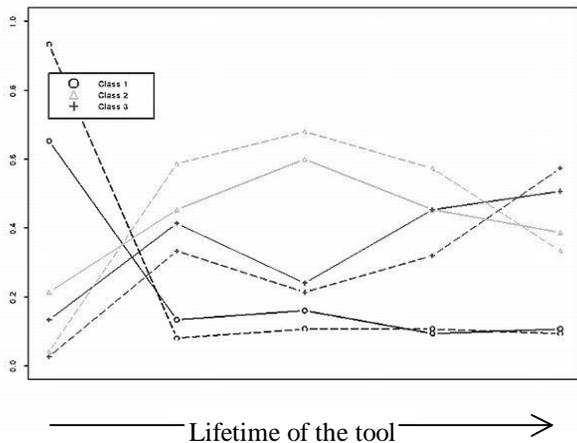


Figure 1. The vertical axis represents the proportion of transients assigned to the three different hidden Markov models over the lifetime of the tool. The solid plot is the result from the first experiment. The dashed plot represents the second experiment, which closely cross-validated the first.

This result shows that the HMMs clearly detected separate structure in early, middle, and late transients. However,

these trained HMM’s did not generalize well with other tools, implying that HMM retraining would be needed whenever tools were changed.

4.2 Results for Longer Time Scales

For these experiments we used HMMs to model dynamics on two levels: for capturing the progression of tool wear and for modeling the dynamics associated with different stages of a single cut. The output of the classifier is both an estimate of the actual wear on the primary cutting edge (quantized) and a confidence (actually, posterior probability) in the tool having exceeded an acceptable threshold of wear. The classifier was trained on data from six 1/2-inch cutters and tested on two independent sets of data, a different set of seven 1/2-inch cutters and a set of four 1-inch cutters recorded using a different accelerometer (to assess generalization).

Of the 87 passes recorded for the six cutters in the training set, only 12 passes were labeled with a measured wear level. To deal with the problem of sparsely labeled training data, we developed an approach of semi-supervised learning of training labels which makes use of prior knowledge that wear progresses monotonically. We progressively assign labels to unlabeled data, training on subsets of the tools using the current set of “labeled,” using constrained Viterbi decoding to label the rest, and assigning new labels when multiple models agree on a label. The Viterbi decoding is constrained to maintain the known wear labels in the sequence.

Evaluation of past classifiers has been based upon the accuracy of the “worn” vs. “not worn” labels assigned. Using this criterion, we were able to achieve excellent performance across changing cutting conditions, 92% and 94% accuracy for the two test sets respectively, even with a single sensor and using only simple RMS energy and energy derivative features.

While accuracy of binary wear prediction is a valuable metric, it provides only a limited assessment of the wear prediction capability. Predicting accuracy of the quantized wear state labels is impractical because of limited test data, so instead we assessed the classifier confidence that the tool is worn, i.e. the posterior probability that the tool is worn given the observations. The wear confidence prediction is analogous to word confidence prediction in speech recognition [19] and we borrow techniques from that literature. In particular, confidence prediction is assessed using the normalized cross entropy (NCE) measure

$$NCE = \frac{H(W) - H(W | X)}{H(W)}$$

where W is the binary wear variable, X is the observation sequence and the expectation in the entropy is based on the empirical distribution of the test data. The NCE measure indicates how much information is provided in the predicted posterior probability that the tool is worn relative to simply using the prior probability alone.

As it turns out, the HMM is a poor predictor of the posterior probability of wear in these experiments, even when the accuracy is high, because the fine time resolution used to capture short-time events leads to a long time sequence and a correspondingly over-confident HMM. With only one error, if the confidence of that decision is high, then the NCE measure can be negative. To address this problem, a second stage generalized linear model (GLM) is used to predict confidence of the binary wear state given the output wear level and optionally other information from the HMM. NCE scores for the two test sets were 0.66 and 0.26, respectively, which are very encouraging given the 0.20 scores typically seen in conversational speech recognition. Unfortunately, the GLM shifts the posterior estimates (conservatively) towards 0.5 and can sometimes result in an increase of errors, as seen here. However, because of the low dimensionality of the GLM (a single parameter in this case), it can easily be adapted to obtain improved performance under changing cutting conditions. Using this idea in a rotation paradigm for adapting and testing the GLM on the 1-inch test data, we recovered the lost performance (see Table 1).

Test Set	% Accuracy	NCE
½ inch	92	0.66
1 inch	91	0.26
1 inch (adapt)	94	0.23

Table 1. Accuracy and NCE performance of systems for monitoring tool wear in a milling application, using an HMM followed by a generalized linear model (GLM).

Acknowledgments

We wish to thank our sponsor, the Office of Naval Research, via the Center for Auditory and Acoustics Research, a Multi-University Research Initiative, for their support of this project. We also thank Randy Fish of Boston University and Brad Gillespie of University of Washington for their help in this research.

References

[1] Atlas, L., G. D. Bernard and S. B. Narayanan, "Applications of Time-Frequency Analysis to Manufacturing Sensor Signals," *Proceedings of the IEEE*, Vol. 84, No. 9, pp. 1319-1329, September 1996.

[2] Kittel, W. A. and Hayes, M. H., "Monitoring rotating machine signals," *Proc. ICASSP '92*, San Francisco, CA, pp. 65-8 vol.5., 23-26 March 1992.

[3] Ulsoy, A. G. and Koren, Y., "Applications of adaptive control in machine tool process control," *IEEE Control Systems Magazine*, Vol. 9, No. 4, pp 33-37, 1989.

[4] Ulsoy, A G and Koren, Y., "Control of machining process," *Journal of Dynamic Systems, Measurement and Control*, vol. 115, pp 301-308, June 1993.

[5] Najafi, B. & Hakim, H., "Spectral Estimators for machine vibration analysis," *Mechanical Systems and Signal Processing*, vol. 6, no. 6, pp 551-574, 1992

[6] Ramirez, C.N. and Thornhill, R.J., "Drill wear monitoring in circuit board manufacturing using drilling forces and their spectra," *Transactions of the ASME. Journal of Electronic Packaging*. vol.114, no.3. pp. 342-348. Sept. 1992.

[7] Narayanan, S. B., Bernard, G. B., Atlas, L. E. and Fang, J. F., "Feature Representations for monitoring of tool wear," *Proc. ICASSP 1994*, Adelaide, South Australia.

[8] Fang, J., Atlas, L., Bernard, G., "Advantages of quadratic detectors for analysis of manufacturing sensor data," *IEEE-SP Symposium on Time Frequency and Time Scale Analysis*, Victoria B.C., 1992.

[9] Li, C. J., Ma, J., Hwang, B. and Nickerson, G. W., "Pattern recognition based bicoherence analysis of vibrations for bearing condition monitoring," *ASME, Sensors, Controls and monitoring issues*, 1991.

[10] Zheng, K. and Whitehouse, D. J., "The application of the Wigner distribution to machine tool monitoring," *Proc. Inst. Mech. Engrs*, Vol. 206, pp 249-264, 1992.

[11] Choi, G S, Wang, Z. and Dornfeld, D. A., "Adaptive optimal control of machining process using neural networks," *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 1567-1572, 1991.

[12] Tansel, I. N. and McLaughlin, C., "Monitoring of tool breakage with restricted coulomb energy type neural networks," *ASME, Sensors, Controls and Quality Issues*, Vol. 55, pp 37- 48, 1991.

[13] B. Sick, "Online tool wear monitoring in turning using time-delay neural networks." *Proc. Int. Conf. On Acoust., Speech and Signal Proc.*, v. 1, pp. 445-448, 1998.

[14] Heck, L. P. and McClellan, J. H., "Mechanical system monitoring using hidden Markov models," *Proc. 1991 International Conference on Acoustics, Speech and Signal Processing* (Cat. No.91CH2977.7). Toronto, Ont., Canada, pp. 1697-1700 vol. 3, 14-17, May 1991.

[15] J. McLaughlin, L. Owsley, L. Atlas, and G. Bernard, "Advances in real-time monitoring of acoustic emissions," in *Proceedings of the 1997 SAE Aerospace Meeting*, June 1997.

[16] L. Owsley, L. Atlas, and G. Bernard, "Self-organizing feature maps and hidden markov models for machine-tool monitoring," *IEEE Transactions on Signal Processing special issue on Neural Networks*, vol. 45, pp. 2787-98, November 1997.

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

[18] Young, S., "Large vocabulary continuous recognition: review," *IEEE Signal Processing Magazine*, v. 13, n. 5, pp. 45-57, 1996.

[19] M. Siu and H. Gish, "Evaluation of word confidence for speech recognition systems," *Computer Speech and Language*, v. 13, no. 4, pp. 299-318, 1999.