Table of Contents

TABLE OF CONTENTS

1. Introduction

2. Machinese
   2.1. Human Speech Recognition
   2.2. Machine Speech Recognition
      2.2.1. Metal Cutting Feature Space
      2.2.2. Comparing Human Speech and Machinese Recognition
      2.2.3. Analysis Approach

3. Preliminary Results
   3.1. Spectrograms

4. Discussion
LIST OF FIGURES

Figure 1-1: First Map of the IMW Controller (After Bourne 1987)
Figure 2-1: The Feature Space for Cutting Processes
Figure 3-1: Face Milling Geometry
Figure 3-2: Spectrogram of a single flute face mill on mild steel.
Figure 3-3: Spectrogram of a dual flute face mill on aluminum. One cutting insert is
good, the other badly worn.
Figure 3-4: Bandpass response of a single flute face mill on mild steel.
Figure 3-5: Bandpass response of a dual flute face mill on aluminum.
The application of human speech processing techniques to the machine shop may provide a new means to interpret the sounds created by the metal cutting process. Real-time signal processing in the frequency domain can identify those bandpass responses which indicate the health of the tools. When combined with knowledge of the tooling and the cutting path, spectrograms can verify the cutting phases and geometric features expected of a normal process.
1. Introduction

The goal of total automation of the machining process is currently being pursued in many industry and university laboratories. The Intelligent Machining Workstation (IMW) project (Bourne 1987) has identified several interdependent control elements which contribute to first part success. The sensing expert of the IMW (Figure 1-1) focuses on providing real-time externally-sensed information for the complete closed-loop automation of the machining process. Achieving this goal requires the implementation of an intelligent sensing system to integrate input data from multiple sensors of different types into a coherent sensory model.

Figure 1-1: First Map of the IMW Controller (After Bourne 1987)

To develop a sensory model, efforts have been made to find the relationships between several signal sources and various aspects of the cutting process, such as tool wear, tool breakage, surface finish and machine chatter. Chief among these sources, vibration signals, including low frequency vibrations of the machining system and high frequency acoustic emissions (AE) from the plastic deformation and fracture of working materials, appear to have the most potential for the diagnosis of metal cutting conditions.

Due to the wide frequency response of AE signals, one can selectively filter out certain bands, such as low frequency machine noises, and extract information from certain high frequency components. It has been indicated (Iwata 1977; Lee, M. 1987) that the signals between 100 kHz and 300 kHz are closely related to the condition of the tooling. In addition to acoustic emissions, low frequency vibration has also been selected to monitor machine chatter and tool wear. Although some substantial results have been achieved recently (e.g., Micheletti 1976, Tlusty 1983 and Yen 1983), the progress toward implementing a general-purpose in-process sensing system to diagnose the machining process has been hindered because the fundamental cause and effect relationships are still not clear. Without such an understanding of the chip forming process, the time domain analysis of the signals employed by most researchers in this area became highly system dependent. According to earlier studies (e.g., Iwata 1977; Kannatey-Asibu
1981; Lee, M. 1987), the total count of AE pulses over a certain threshold level has been shown to be a good index for predicting tool wear, if the threshold is properly chosen. It was also shown that the change of the average energy level of the AE signals could be another means for monitoring tool wear, if the system could be precisely calibrated. The precise calibration and proper selection of threshold values have proven to be too expensive and time-consuming for industrial implementation.

With all these concerns, we are not satisfied with a time domain analysis because its purely statistical approach loses insight of the system and is always system dependent. In contrast, the frequency representations often serve to place in evidence certain properties of the signal that may be obscure or at least less evident in the time domain representations. We believe that frequency analysis may present a unique and attractive opportunity for monitoring the details of a cutting process. In the past, frequency analyses may not have been pursued due to the time consuming nature of the FFT which was impossible to do in real time at these frequencies. It is now possible, however, equipped with a second-generation DSP microprocessor, advanced electronic technology and parallel algorithms, to implement a real-time frequency analysis system which is technically and economically feasible.

This paper will present our concepts and new ideas about multiple domain analysis of the metal cutting process.

2. Machinelse

From shop-floor experience, changes in machine sounds while cutting is always the first alarm system for the machine operator. An experienced machinist can detect irregular cutting noises and relate them to some cutting parameters instantly. We would like to determine the nature of these audio patterns extracted by the human operator and devise computer algorithms which might accomplish the same results. The very wide band nature of the acoustic signals available from machining might permit us to learn far more by listening with a computer-assisted ear. Researchers in human speech studies has been devoted to a seemingly similar problem for almost 50 years. Recently published work claims 97% reliability for speaker-independent, continuous speech. We would like to suggest the application of these approaches to the manufacturing domain: understanding the speech of the machine, or machinelse.

2.1. Human speech recognition

In the past fifty years, many researchers have been involved in human speech recognition. A wide horizon of research has been investigated to identify feasible strategies which spanned signal processing, pattern recognition, artificial intelligence, statistics, probability theory, information theory and linguistics. The early researches included: finding an appropriate physical model for human speech, reading phonemes by examining the resonant frequency trends on spectrograms, looking for the best representation of the speech signal and developing better statistical training and searching algorithms for human speech recognition.

Not until advances in digital computer hardware which appeared in the '70's have several feasible strategies been proposed and implemented. However, their efficiency and accuracy were achieved by sacrificing the effectiveness, i.e., by imposing one or more constraints. Four primary difficulties which constrained human speech recognition research were:

1. Speaker dependence

Speaker independence has been viewed as the most difficult constraint to overcome in human speech recognition. Most parametric representations of speech are highly speaker dependent because a set reference patterns suitable for one speaker may perform poorly for another.
2. Isolated words
Continuous speech is more difficult to recognize because words boundaries are unclear in continuous speech. Moreover the content words are often emphasized, while the function words are significantly neglected.

3. Small vocabulary
As the vocabulary size is increased, it becomes impossible to store and model each word separately. Instead, a subword representation must be defined and used. This always makes recognition more complicated.

4. Constrained grammar
The grammar involved in human speech recognition constrains uncertainties at each decision point. Less grammar required of input to the system raises the level of perplexity of it.

Research aimed at overcoming all these constraints has culminated in the SPHINX system (Lee, K.F. 1987) which employs several strategies for large-vocabulary speaker-independent continuous speech recognition. We will review these strategies briefly as background for the design of our research plan.

First, the system samples the speech at 16 kHz. The waveform is then blocked into frames. Each frame spans 360 samples, from which the system computes linear predictive coding (LPC) coefficients by autocorrelation. From these, a set of 12 LPC-derived cepstral coefficients is generated. The linear filtering of the cepstrum permits separation of the speech components representing the vocal tract response and the excitation source (Rabiner 1979)). This 12 dimension vector is then reduced to a symbolic code using a vector quantization approach which maps a real vector onto a discrete symbol. The symbols are completely described by a "codebook", which is a set of prototype vectors with the same dimension as the original cepstral vector. To perform mapping, the input vector is compared with each prototype vector and the vector is replaced by the most similar prototype's symbolic code. Initially, a Hidden Markov Modeling (HMM) of speech training is performed to build the codebook, or feature space. An HMM recognition algorithm is then used to extract the speech pattern from the codebook representation. This baseline system achieves a 30% recognition reliability.

To improve the accuracy, the SPHINX system includes knowledge engineering techniques based on human knowledge of speech properties. Instead of relying only on statistical training and pattern extraction, these knowledge based parameters are combined with the original vector to construct a multi-dimension symbolic code, thus increasing the robustness of the system dramatically. A clustering method has also been applied to separate the training data into logical groups, and a speaker prototype is generated from each cluster. Any new speaker will then be assigned to a specific cluster, thus overcoming the speaker dependency constraint. Furthermore, an adaptation algorithm has also been developed to adapt the existing parameters to the new speaker's characteristics through a small number of adaptation sentences. In addition, a subword unit representation of the speech is used to relieve the continuous-speech and large-vocabulary constraints. The reliability of the full system reaches 97%.

2.2. Machine Speech Recognition
The success of the foregoing research in human speech recognition is based largely upon models of the vocal system of human speakers, plus heuristic knowledge which is highly contextual. In the context of vibrations induced by specific machining parameters, we are inclined to ask what the machine is intending to say, or at least what we are interested in hearing from it. At the start, we will focus on trying to determine "normal" from "abnormal" speech.
2.2.1. Metal Cutting Feature Space

To guide the extraction of information from vibration signals, we identify those aspects expected from cutting processes in a feature space (Figure 2-1). The space of interest to automated machining, and the IMW in particular is as follows:

- Tool condition: tool is sharp, dull, worn, or broken.

According to (Iwata 1977) and (Dornfeld 1981, 1982), the acoustic emission (AE) signals between 100 kHz and 300 kHz are identified as the most important sources for tool wear monitoring. In addition, a sudden increase in AE amplitude has been observed by researchers, when the cutting edge actually failed.

- Working material

Although the working materials, including tool material and workpiece material, are given parameters predetermined by the planner and cutting expert of the IMW, we suspect that power spectrum analysis of AE signals will display some important characteristics of the working materials. Since different materials feature different chip morphology, the AE signals emitted from energy release may have unique resonant frequencies. Furthermore, the material strength can also be inferred from RMS value of the AE signal [Lan 1982]. These data could provide clues for analyzing the working material properties from AE signals.

![Feature Space Diagram](image_url)

Figure 2-1: The Feature Space for Cutting Processes
• Cutting history: cutter entrance, chip making, cutter exit

For the monitoring and diagnosis of a single cut, prominent signals appear at the very beginning and ending of the cut. Once the cutting edge hits the part, a significant impulse is emitted exciting a large number of system frequencies. We can easily detect these from an analog impulse detector. The chip formation process in milling is more unsteady due to chip breakage and geometric changes in the chip thickness. We ought to try to listen for chip breakage, since chip morphology is significantly influenced by cutting process parameters and is at present difficult to detect visually in real time. While the end moment of the cut is not as crucial with respect to tool stress as the start, we can readily detect it. The burnishing phase of certain milling cuts, e.g., slotting with an end mill, provides yet another characteristic set of vibrations superposed on relative silence.

A significant advantage of the machining context over the human speech context is that the timing of these "utterances" can be scheduled precisely at the planning stage and compared to the actual results in real time, as an indicator of normal system performance. For a machine tool which is not instrumented for spindle position, the HMM could determine the timing of these events after the fact.

• Chatter

Chatter instability is one of the most important conditions for real-time machine monitoring, since its occurrence seriously affects part quality and it cannot be reliably predicted at the planning stage. Researches to extract information from low frequency system responses have been underway for quite a long time (e.g., Tobias 1958). From our present concerns with precision machining, we are more interested in prediction than diagnosis. Several approaches have been proposed to predict machine chatter (e.g., Thusty 1986), but system-dependent parameters still prevent their real-time implementation. We suspect that the structural dynamics of the workpiece/machine system will be represented by its vibrational responses during cutting. Thus, the speech of the machine ought to extend through the low frequency audio band.

• Cutting speed

Even though the cutting speed is determined in advance by the planner, this parameter can be easily heard through the cepstrum analysis of low frequency signals. Such analysis is able to divide the system response into an impulse train whose frequency is the cutting speed and the resulting structural dynamics response.

• Width and depth of cut

These two cutting parameters are generally specified a priori and not subject to monitoring nor modification by adaptive control. Since tool wear has a strong geometric component (the cutting edge recedes from the original shape), small changes in the width and depth of cut should be detectable in-process. The expected temporal events related to tool contact with the workpiece can be derived prior to cutting and monitored with either spectrogram or bandpass time domain analysis.

• Finishing

The burnishing process which occurs when a tool passes over a finished section of the workpiece affords a particularly sensitive region for detecting tool geometry and could perhaps be used to estimate the condition of the finished surface.
2.2.2. Comparing Human Speech and Machinese Recognition

- Physical model

The human voice is generated by excitation of vocal cord on vocal tract. We speak a sentence by changing the shape of our vocal tract and recognize other's speech by extracting the patterns of vocal shapes. This understanding served as the basis of early research in human speech recognition. By analyzing the frequency response of human voice, researchers could locate the resonant frequencies of voice (formants), from which they could "read" the speech.

The machine "voice" is emitted from cutting excitations acting on the machining system. Its frequency response is much wider than human speech. We can roughly separate the signals to two categories—high frequency acoustic emission from plastic deformation processes and low frequency vibration signals from system responses and chip forming. The physical model of low frequency vibration is similar to vocal tract vibration. The geometrical configuration of the machine setup is the main factor effecting resonant frequencies of the response. Process signals with a potentially high information content are included in this area; however, most of the environmental noise also falls into this band. The high frequency model, on the other hand, is related to the micro-mechanism within working materials (dislocations and fractures), of which the causes and effects are still not yet clear.

- Vocabulary size

The vocabulary size of "machinese" is expected to be much smaller than that of human speech. But each cutting process speaks its own dialect. There is no official language for machinese.

- System-dependence and speaker-dependence

System-dependence is the primary obstacle preventing machinese recognition from real-time industrial application. Even with the same type of metal cutting process, the response would be very different depending on the calibration, system setup and the geometry. Since the human speech research faces the same problem, most of the strategies already employed to overcome this difficulty are potentially useful for us.

- Continuous speech and isolated machinese

As far as this aspect is concerned, machinese compares favorably in its regularity of excitation. Since the cutting parameters and geometry are already known before machining, we can anticipate most events and easily monitor them.

2.2.3. Analysis Approach

Within the IWM resides an expert system called the "cutting expert." This code is responsible for computing seven cutting parameters, viz.: cutting speed, feed rate, depth of cut, width of cut, tool material, workpiece material and lubrication. Additional geometric information about specific cuts is provided by the front-end expert system, the "planner."

Informed of these data and the constraints implied by the machining center used, we are beginning to extract the features of interest related to the cutting process from the following in-process measurements:

- vibration from accelerometers for low frequency signals between several Hz to 20K Hz, and
acoustic emission from AE transducers monitoring high frequency responses between 100K Hz to 300K Hz.

The approaches we propose to use to extract models of the cutting process based on human speech recognition techniques are summarized as follows:

1. Spectrogram analysis

The sound spectrograph was for many years the basic analysis tool in speech research. Taking advantage of the short-time Fourier transform, one could separate a continuous signal into multiple frames in the time domain, and take the FFT of each segment. The resulting spectrogram is a three-dimensional representation of the time-dependent spectrum in which the vertical dimension represents frequency and the horizontal dimension represents time. The spectrum magnitude is represented by the darkness of the marking on the paper. By investigating the spectrogram, one can monitor the change of frequency response continuously. Since this method is able to monitor time and frequency domain response simultaneously, it provides us with a good opportunity to gain more insights into system behavior.

2. Linear predictive coding analysis

Linear predictive coding (LPC) techniques have been employed broadly in system identification and estimation. While the real model of the system is changeable or unclear, one can use LPC to identify it. Once the predictor coefficients have been obtained, the system has been uniquely identified to the extent that it can be modelled as an all-pole linear system. Therefore, even if the physical basis for the AE signals is unknown, one can estimate the model through LPC. This technique obviates the time-variant system constraint. Moreover, we can directly derive cepstral coefficients and fundamental frequencies of the system from the LPC coefficients.

3. Cepstrum analysis

This approach has been very successful for human speech recognition. A continuous input waveform is first transformed to a frequency representation with a discrete Fourier transform (DFT). Taking the logarithm of the resulting DFT, one can separate the signals into different divisions. A simple filter implemented in this domain can extract the desired signal from background noise. Inverting the DFT components produces cepstral coefficients which serve as the primary parameters for speech training and recognition.

4. Noise rejection techniques

Low frequency system vibrations contain abundant signal energy. This feature has prompted researchers to investigate high frequency AE signals because most of the environmental (non-metal-cutting) noise also falls into this frequency band. Separating the vibration signals of the metal cutting process from other noise sources, would open up another possibly useful source of information. Since the machining process is both synchronous and repetitive, temporal averaging may assist in relating these signals to the health of the process. Noise rejection techniques developed for aircraft pilots [ref7 may well be applied to machining by implementing multiple sensors in different positions to cancel the unwanted noise sources.

5. Bandpass response analysis

Previous researches have indicated that several frequency bands are highly related to tooling conditions (Lan 1982), but are variable with the specific set-up. Once the system recognizes these bands through spectrogram analysis, a digital adaptive filter could be
designed on the host and implemented with modern DSP devices. Thus a relatively system-independent time domain analysis could be performed.

3. Preliminary Results

Our initial experiments have been performed using a face mill with replaceable indexable inserts and a nominal 100 kHz to 1 MHz AE transducer. Signals from two complete revolutions of the cutter were recorded while face milling according to the geometry shown in Figure 3-1. In the figure, the primary cutting phase is shown as the angle $\theta$, which depends on tool radius, $r$, and cutting width, $w$.

![Figure 3-1: Face Milling Geometry](image)

3.1. Spectrograms

Case I: Mild Steel, single cutting edge

A spectrogram resulting from cutting mild steel with a single cutting edge face mill is shown in Figure 3-2. The energy appears to be concentrated in frequencies above 100 kHz due to the roll-off of the transducer. Although the distribution is otherwise rather uniform in frequency, two prominent intensity bands appear corresponding to the cutting and burnishing phases. In this case, the cutter was in contact with the work for only 72° of its rotation. The cutting time represented a 44° arc rather than the 36° arc expected from geometry. At this point, the spectrogram shows a distinct drop in energy across all frequencies, corresponding to the burnishing phase. A total contact arc of 72° is observed as expected. We can interpret these results by considering that the spindle and tool holder are not perfectly rigid, and that significant forces are produced during cutting which tend to separate the tool edge from the work. The extended cutting period is thus the result of elastic relaxation of the spindle and tool holder, which was subsequently verified by post-process measurement.

Case II: Aluminum, two cutting edges

An spectrogram resulting from cutting 6061-T6 aluminum with a dual cutting edge face mill is shown in Figure 3-3. One of the cutting edges was new, the other worn out. As with the steel specimen, the energy appears to be concentrated in frequencies above 100 kHz due to the roll-off of the transducer. However the worn tool (shown by the two narrow bands in successive cuts) displays a more uniform intensity on the spectrogram, indicating an increase in energy at lower frequencies. Although the nominal cutting phase occurs in a 90° arc of the tool in this case, the worn tool contacts the work for only about 45° of its rotation and shows no burnishing phase. This is due to the geometry of the workpiece produced by the previous pass of the fresh tool: having a larger overall profile, it removed more material. The intensity of the worn tool, however, is higher overall as expected from past experiments.
Figure 3-2: Spectrogram of a single flute face mill on mild steel.

Figure 3-3: Spectrogram of a dual cutting edge face mill on aluminum. One cutting insert is good, the other badly worn.
For the fresh tool, two prominent intensity bands appear corresponding to the cutting phase, and a lower intensity burnishing period. The cutting phase is marked by the initial removal of a relatively small amount of material followed by an effectively larger feed rate caused by the absence of cutting by the worn tool on the previous pass. The differences between worn and fresh tools is thus readily apparent. Since the rate of tool wear is typically accelerated at the end of tool life (Emel 1988), one could expect to find which tool insert of a group was beginning to wear out by comparison to the others. This approach has been verified further by experiment (Lee, M. 1987) which confirms that the likelihood of all cutting edges in a group wearing out simultaneously is very low.

3.2. Bandpass Response

The bandpass response corresponding to Case I is shown in Figure 3-4, taken between 140 and 160 kHz. The relative differences in amplitude between cutting and burnishing are preserved from cut to cut. The elastic relaxation time of 5 ms for the spindle and tool holder is another distinct consistent feature.

The bandpass response corresponding to Case II is shown in Figure 3-5, again taken between 140 and 160 kHz. The cutting time periods for both cutting edges are clearly distinguishable, and the relative differences in amplitude between the two cutting phases and burnishing are preserved from cut to cut. As expected, the lower cutting forces do not produce a pronounced elastic relaxation time feature on this plot.

Figure 3-4: Bandpass response of a single cutting edge face mill on mild steel.
Figure 3-5: Bandpass response of a dual cutting edge face mill on aluminum.

4. Discussion

- Since the dynamic response of the metal cutting process is highly system dependent, we do not anticipate an absolute reference to cover all systems. Instead, rules for signal interpretation and physical models seem more appropriate for future analysis.

- Although our final goal is real-time monitoring and diagnosing, processing rate is not necessarily critical: the linear system model of metal cutting is relatively time-invariant in the short term. However, monitoring trends would be very beneficial. This implies keeping a history of the cutter and the workpieces for diagnosis of future current cutting conditions.

- The cutting feature space is fortunately not large, but some features are elusive and ambiguous, such as tool breakage and the effects of structural dynamics. Therefore, feature classification and extraction will require close attention and verification. Techniques of human speech recognition would be applied to solve pattern matching problems in this level. A simple spectrogram has revealed several interesting features relative to tool wear. Refinement of the measuring and analysis techniques should yield further insight into this method of early detection of tool failure.
A working sensing expert should be smart enough to train itself from experience. Thus the "expert" should be applied to an arbitrary system in the shortest set-up time. The algorithm should first build a set of reference data in the representative feature space from a set of training signals and then classify each sensed signal with respect to a specific feature using the reference data and the constraints.

Combining the features thus extracted with information from other in-process experts, specifically the expected geometry of the cutting phases and their transitions, the sensing expert is expected to diagnose the health of the machining process in real-time.

References:


