of terms in the polynomial function was of the same form for all photo interpretation tasks. For terms ordered by the reduction option of the linear regression program, the decrease of the standard deviation with increasing numbers of terms is quite rapid for small numbers of terms; however, after a small number of terms has been included in the polynomial, the addition of more terms causes little reduction in the standard deviation, and any reduction made is not statistically significant. The same phenomenon occurs for the maximum-magnitude error; however, it did not decrease as rapidly as the standard deviation, possibly because it was not being minimized directly but only as a consequence of the minimization of the standard deviation.

The form of the variation of the error discussed in item 4) if it is typical, raises two questions about the current work to develop better methods of feature selection.

1) The relatively sharp knee in the curve at the relatively small number of terms in many problems may obviate the need for sophisticated measurements selection techniques. The "brute-force" technique of evaluating all sets of measurements to find the best becomes feasible.

2) A standard, readily available linear regression program appears to do a creditable job of feature selection, which probably can be made slightly better by repeating the term-addition and term-reduction options several times. In the face of this performance, can any special-purpose algorithms be expected to do significantly better?

References

[1] F. W. Smith and M. H. Wright, "Automatic ship photo interpretation by the method of moments," Sylvania Electronic Systems, Mountain View, Calif., Tech. Memo. SESW-M1169, Nov. 1967.
[2] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. IT-8, Feb. 1962, pp. 179–187.
[3] F. L. Alt, "Digital pattern recognition by moments," in *Optical Character Recognition*, G. L. Fischer *et al.*, Eds. Washington, D. C.: Spartan, 1962, pp. 153–179.
[4] K. Udagawa, J. Toriwaki, and K. Sugino, "Normalization and recognition of two-dimensional patterns with linear distortion by moments," *Electron. Commun. Jap.*, vol. 47, June 1964, pp. 34–46.
[5] S. Moskowitz, "Terminal guidance by pattern recognition—a new approach," *IEEE Trans. Aeronaut. Navig. Electron.*, vol. ANE-11, Dec. 1964, pp. 254–265.
[6] F. W. Smith, "Automatic HF signal classification by the method of moments," Sylvania Electronic Systems, Mountain View, Calif., Tech. Memo. SESW-M1037, Mar. 1967.
[7] R. Bakis, N. M. Herbst, and G. Nagy, "An experimental study of machine recognition of hand-printed numerals," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, July 1968, pp. 119–132.
[8] R. G. Casey, "Moment normalization of handprinted characters," *IBM J. Res. Develop.*, vol. 14, Sept. 1970, pp. 548–557.
[9] H. D. Block, N. J. Nilsson, and R. O. Duda, "Determination and detection of features in patterns," in *Computer and Information Sciences*, J. T. Tou and R. H. Wilcox, Eds. Washington, D. C.: 1964, pp. 75–110.
[10] M. Minsky, "Steps toward artificial intelligence," *Proc. IRE*, vol. 49, Jan. 1961, pp. 8–30.
[11] E. Rosenblatt, *Principles of Neurodynamics: Perceptions and the Theory of Brain Mechanisms*. Washington, D. C.: Spartan, 1961.
[12] R. Casey and G. Nagy, "Recognition of printed Chinese characters," *IEEE Trans. Electron. Comput.*, vol. EC-15, Feb. 1966, pp. 91–101.
[13] E. Issacson and H. B. Keller, *Analysis of Numerical Method*. New York: Wiley, 1966, p. 34.
[14] M. A. Efroymson, "Multiple regression analysis," in *Mathematical Methods for Digital Computers*, A. Ralston and H. S. Wilf, Eds. New York: Wiley, 1965, pp. 191–203.
[15] W. M. Brown and L. J. Porcello, "An introduction to synthetic-aperture radar," *IEEE Spectrum*, vol. 6, Sept. 1969, pp. 52–62.

Pattern Recognition Signal Processing for Mechanical Diagnostics Signature Analysis

ROY L. HOFFMAN, MEMBER, IEEE, AND
KEINOSUKE FUKUNAGA, MEMBER, IEEE

Abstract—Signature analysis of small, complex, cyclic mechanisms is discussed. An envelope preprocessor for isolating signal events relating to mechanical impacts is developed. A recognition system based on second-order (normal) statistics is used and the use of statistical procedures for signature intepretation is presented. Experimental evidence is presented to support the validity of this approach to mechanical diagnostics signature analysis.

Index Terms—Diagnostic systems, envelope detection, mechanical diagnostics, pattern recognition, preprocessing, second-order statistics (or normal or Gaussian), signal analysis, signal averaging, signature analysis, time trends.

Introduction

Today there is widespread interest in diagnostic systems which can predict impending mechanical system failures [1]–[8]. Several studies have demonstrated that "secondary effects" such as sound, vibration, temperature, pressure, and other physical phenomena, exhibit telltale changes long before catastrophic mechanical failures occur and, therefore, can be used to predict and diagnose mechanical malfunctions [3]. The technology for monitoring secondary effects is called "signature analysis." The signature is the set of time-varying measurements of secondary effects. The analysis of these data is amenable to pattern recognition [4], [9], [10].

Previous investigations have shown how signature analysis applies to large mechanisms which in some respects are easier to diagnose. Sensors can usually be placed to detect the secondary effects of one component or mechanical subsystem, independent of similar emissions from other components or subsystems.

Small mechanical devices, in contrast, have many moving parts, in highly interrelated mechanical subsystems, in physical proximity, whose secondary effects cannot be independently sensed. Most sensor outputs are combinations of the secondary effects from several subsystems and successful signature analysis often is possible only when the secondary effects of the component of interest can be isolated through signal processing.

Signature Analysis

In this experimental study we found that the secondary effects of many cyclic components can be isolated by using time-signal averaging along with filtering and time gating [11], [12]. Time-signal averaging has been shown effective for reducing random (additive) noise in cyclic signals [3], [4], [7], [13]. Its applicability to isolating the emissions of concurrently operating mechanisms hinges on the interrelationships between signals from those mechanisms. If a

Manuscript received November 2, 1970; revised March 29, 1971. A preliminary version of this note was presented at the IEEE Symposium on Feature Extraction and Selection in Pattern Recognition, Argonne, Ill., October 5–7, 1970.

R. L. Hoffman is with IBM General Systems Division, Rochester, Minn. 55901.
K. Fukunaga is with the Department of Electrical Engineering, Purdue University, Lafayette, Ind. 47907.

set of conditions that we termed "signal-event independence" were present, then time-signal averaging proved useful.

Signal-event independence can be related to three signal parameters: time, amplitude, and frequency, and does not necessarily imply statistical independence. Mechanical subfunctions can be treated as time independent if: 1) their signals are uncorrelated or 2) the events of interest are in different time frames or 3) are at different cyclic rates. Amplitude independent implies that: 4) the amplitude of interfering functions is small with respect to desired signals. Frequency independence suggests that: 5) the frequency spectrum of the interfering and desired signals is different. Each of these characteristics must be considered, through study or experiment, in order to develop effective signal preprocessing, feature extraction, and recognition functions.

By analysis and experiment, we found that many functional failures in cyclic equipment appear in sensor signals as events relating to departures from the norm of mechanical impacts between parts in the subsystems. Not all functional failures in all mechanisms are manifested in this way but, for example, the operation of a clutch mechanism can be observed by measuring the vibration or sound throughout its operating cycle. Through experiments on a laboratory model we found that high-signal (vibration, sound) levels occur when the clutch engages or disengages and that changes in these levels could be related to malfunctions in the mechanism.

Signal preprocessing functions were developed to maximize information extraction for signal events related to mechanical impacts. Of the signal parameters which can be used, time and amplitude of impact contain the most information for discrimination. Signal-frequency spectrum primarily reflects the impulse response of the signal transmission path which was assumed to be constant. The preprocessing system is shown in Fig. 1. This is an envelope detector whose output to the A/D convertor follows the envelope of the rectified, high-passed sensor output signal. Mechanical impacts, like electrical impulse functions, generate signals with some energy at all frequencies while continuous mechanical events generate high-energy signals over narrow-frequency bands usually in the audio or subaudio range. Thus the SNR for mechanical impact signal events can be improved by looking at the spectrum above the audio range. The filter cutoff frequencies shown were determined experimentally to minimize classification error on a specific failure function. The history of using simple rectification for detecting impulsive noise events should be noted [2], [3], [7].

The waveforms coming from the preprocessor were noisy. So, as the example (Fig. 2) indicates, time averaging was used to obtain the average waveforms for (laboratory controlled) normal (N) and abnormal (A) operation, noting that it is possible to have many operating states between these extremes. These waveforms were sampled and became the "features" for pattern recognition signal processing.

To use waveforms as features or pattern class references for signature analysis first requires a procedure for selecting

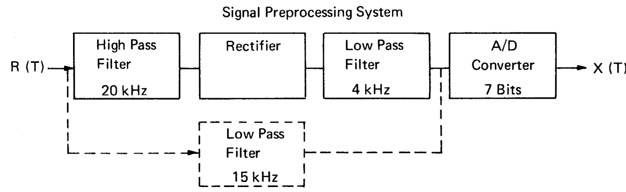

Fig. 1. Envelop preprocessor system.

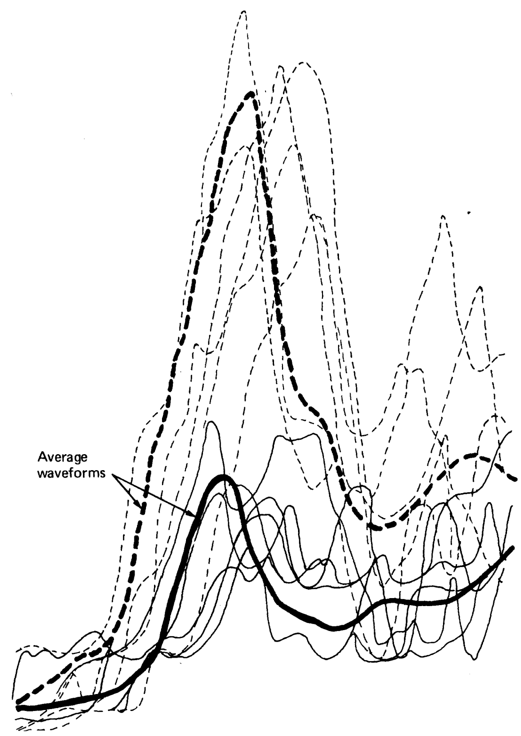

Fig. 2. Diagnostic signals.

the waveforms. Second, statistical properties must be developed so that an effective computer-based recognition system can be designed. To do this, assume that there are two active classes, normal (N) and abnormal (A). The extension to more classes is relatively straightforward [14]. Also assume that the noise $N(t)$ on $S(t)$ is additive and that second-order (normal) statistics are adequate for describing the process. Within this framework the whole rich tapestry of signal analysis and statistics as well as pattern recognition is available. Their power can be applied to understanding the process and selecting optimum recognition procedures.

Given the above assumptions, the selection of a recognition procedure is trivial since it is well known that the optimum decision (recognition) function for two-class normally distributed data is a parametric classification scheme based on the mean and covariance parameters of the data [14]. The use of second-order statistics and parametric pattern recognition often is criticized because of: 1) the applicability of normal statistics, and 2) the intermediate computations required (to obtain M and Σ), whereas adaptive nonparametric methods are computationally simpler. The first point addresses the characteristics of the physical process. The experimental results which follow seem to indicate that normal statistics apply to some signature analysis problems. The second point, while valid in an "on-line" operation, is less important in the recognition system design phase if computing resources are available. In fact, second-order parametric statistics are the basis of a number of procedures for characterizing data, feature extraction, feature selection, and improvement on the recognition system. Let us examine some of these aids.

Signature Interpretation Aids

The crucial test of a recognition system is usually how well it performs, as measured by recognition error rate. Typically, the recognition error rate is determined experimentally on a "test set" of patterns not in the "design set" [14]. If the error rate is below some minimum value the recognition algorithm is acceptable. However, if the error rate is too high, then it becomes difficult to ascertain whether the preprocessing, feature extraction, or assumptions made in designing the recognition functions are at fault or in fact, whether the problem is soluble. However, for problems which are adequately described by second-order statistics, it is possible to determine the recognition error analytically [15]. Further, by comparing analytic and experimental results, the validity of second-order statistical assumptions can be verified.

Although the probability of error is the best figure of merit for a recognition system, computing it analytically is somewhat difficult and experimentally all the data must be processed each time the probability of error is calculated. Thus, simpler criteria to measure the separability of two pattern classes are desirable. The divergence and Bhattacharyya distance are two separability measures whose properties are well understood for normal distributions. That is, recognition error can be obtained as a function of these criteria [14]–[17]. Only the divergence, which is defined as the expected value of the likelihood ratio of two distributions, will be discussed since the measures are similar. In general, the optimum decision boundary in the Bayes' sense is hyperquadratic. However, for simplicity, a linear boundary and equal *a priori* probabilities for both classes was assumed.

The relationships between divergence and error rate for normal distributions were discussed by Marill and Green [18]. Fig. 3 shows the error boundaries they derived along with some of our experimental data to be discussed later. The lower boundary prevails when both classes have equal covariance matrices while the upper boundary is for the univariant case. This graph and the divergence may be used as follows. First, an initial experiment is run and both the divergence and probability of error are measured. After these data are plotted, such as in Fig. 3, one can judge the applicability of normal statistics. Also, the effects of changes in the signal processing technology can be studied without repeating the experiment. Three applications of using the divergence separability measure in this way are as follows.

1) The relationship between sampling frequency and the classification error can be determined by calculating divergence. Once the covariance matrices and mean vectors are calculated for well-sampled waveforms, the covariance

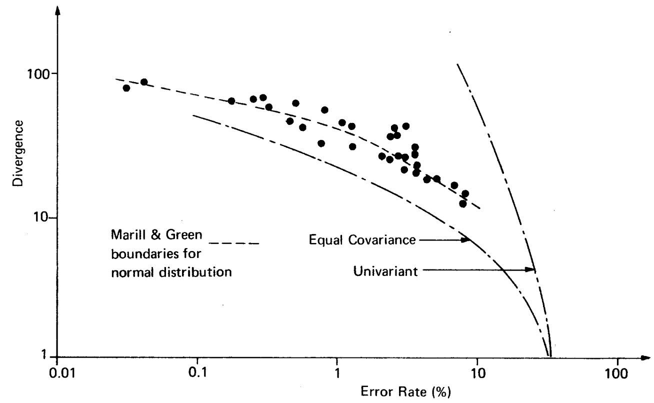

Fig. 3. Divergence versus error rate.

matrices and mean vectors (and hence divergence) for the same waveforms sampled at a lesser rate can be formed by selecting or combining the proper rows and columns from the original covariance matrices and mean vectors. By plotting divergence against sample rate, and noting the same rate where the divergence (separability) begins to decrease, the minimum adequate sample rate can be found.

2) The effect of measurement system bandwidth on recognition error can be studied by calculating divergence. Suppose that two signal processing systems are available. One of the distinguishing characteristics is the frequency spectrum from which they obtain information; audio versus ultrasonic range for example. The two systems have been tested and error rates calculated for each. Question: if both systems were applied simultaneously, would the error rate improve? This question can be answered by adding the divergences of the two systems since the information used by each system is independent of the other.

3) Trends of waveform parameters can be more easily observed by a single number statistic such as divergence. Statistics one might consider include short-term rms or short-term peak value, but for diagnostic purposes a single number statistic which measures separability is more effective. Divergence is one such statistic. Error rate could also be used. Time trending is covered more fully in the next section.

Tracking Time Trends

In the previous section, all available data were used to form one mean waveform and the deviations of individual waveforms from the mean were described by one covariance matrix. However, considering the data more carefully, it is noticed that randomness consists of two parts: one is a waveform-by-waveform variation and the other may be a rather slow migration of the mean waveform. Let us call the former "short-term statistics" and the latter "long time trends". The effects of short-term variation; that is, noise on individual samples, on our ability to estimate a waveform or its parameters can be reduced by averaging a number of time-contiguous samples. See Fig. 2 for example.

Long-time trends cannot in general be eliminated by averaging. Indeed, time trends of waveforms or waveform parameters, if they exist, should be observable by the diagnostic system. In mechanical diagnostics, observing time trends in the signals and relating these to events in the mechanism is an effective means of "failure prediction" [6]. Assuming that these trends are slowly varying, so that the mean and covariance parameters are stationary for a finite time, we can divide waveform instability into two problems: the measurement of short-term statistics, and the tracking of long-term trends.

Suppose we take the first m waveforms, the second m waveforms, and so on until the original data set is broken up in k groups. Then these sets of m waveforms have the mean vectors $M_{i1} \cdot M_{i2}, \cdots$ and the covariance matrices $\Sigma_{i1}, \Sigma_{i2}, \cdots$ where $i = A$ or N for class A or N. The measurement of long time trends is concerned with the statistics of the averaged waveforms $M_{i1}, M_{i2}, \cdots$, which have their own mean vector and covariance matrix as

$$M_{Li} = 1/k \sum_{j=1}^{k} M_{ij} \tag{1}$$

$$\Sigma_{Li} = 1/k \sum_{j=1}^{k} M_{ij}M_{ij}^T - M_{Li}M_{Li}^T. \tag{2}$$

Short-term statistics are more concerned with Σ_{ij}, assuming that $\Sigma_{i1} \cong \Sigma_{i2} \cong \cdots$. Since the total mean vector M_i and the covariance matrix Σ_i can be expressed as

$$M_i = M_{Li} \tag{3}$$

$$\Sigma_i = \Sigma_{Li} + 1/k(\Sigma_{i1} + \Sigma_{i2} + \cdots + \Sigma_{ik}) \tag{4}$$

or

$$\Sigma_i \cong \Sigma_{Li} + \Sigma_{i1} \quad \text{for} \quad \Sigma_{i1} \cong \Sigma_{i2} \cong \cdots \cong \Sigma_{ik} \tag{5}$$

the divergence is bounded by

$$\text{div}(M_i, \Sigma_i) \leq \text{div} \leq \text{div}(M_{Li}, \Sigma_{Li}). \tag{6}$$

Thus, our problems are: 1) to find a proper m to separate short time statistics and long time statistics; 2) to find M_{Li} and Σ_{Li}; and 3) to check $\Sigma_{i1}, \Sigma_{i2}, \cdots, \Sigma_{ik}$ for approximate equality.

If one decides to use average waveforms for classification, the information of 3) is not needed. The divergence or the probability of error, when individual waveforms are used for classification, can be evaluated by combining the information of 3) with (3) and (4). The use of this approach is discussed further in the next section.

Experimental Study

Pattern recognition signal processing was applied to several mechanical subsystem failures sensed from signal events relating to mechanical impacts. The test vehicle was a small punching mechanism for 80-column IBM cards, which uses several start–stop mechanisms. A number of tests were performed and the effects of the signal processing, decision procedures, and analysis aids previously described were observed.

The first set of tests measured the effectivity of envelope preprocessing for mechanical impact subfunction failures. Fig. 4 shows recognition error rate with and without envelope preprocessing. The latter was obtained with the low-pass circuitry shown by the dashed lines in Fig. 1. Each point on the figure corresponds to a test. As shown, the minimum performance gain with envelope preprocessing is $> 2 \times$ reduction in error rate. On the average, the envelope performed $> 5 \times$ better (lower error rate) than conventional low-pass preprocessing.

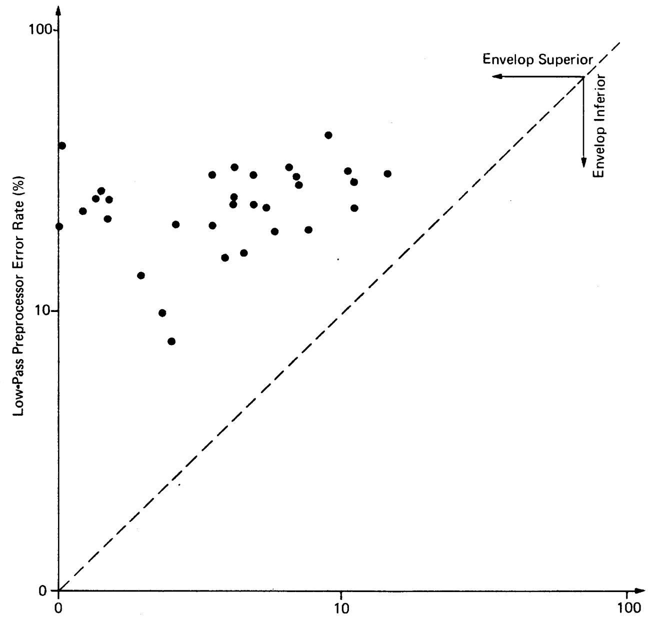

Fig. 4. Preprocessor performance.

The second set of tests measured the effectivity of the recognition algorithms. The minimum error in a Bayes' sense was calculated by methods given in [15]. A common covariance hyperplane (called Σ hyperplane) and a first-order hyperplane (called I hyperplane) decision function were implemented. Performance statistics are shown in Fig. 5. The tests are arranged according to increasing minimum Bayes' error. As one would expect, if the data satisfied normal statistics (and it appears to do so), the minimum Bayes' error, Σ hyperplane, and I hyperplane rank 1, 2, 3 in error rate performance. For many tests the Σ hyperplane is nearly optimum.

The third set of tests show how divergence and error rate calculations correlate. The data are shown in Fig. 3 along with error boundaries calculated by Marill and Green [18]. The experimental data are distributed along and above the lower boundary which indicates that the covariance matrices of normal and abnormal waveforms in a given test are somewhat different. But, the relationship between error rate and divergence has the same tendency as the equal covariance case except being shifted toward higher error rates. From Fig. 3 it can be concluded that the error rate of these data can be reliably estimated from the divergence. Also, this indicates that second-order statistics are adequate to evaluate the distributions and set up the classification boundary.

The fourth set of tests show how divergence can be used to set sampling rates for A/D conversion. Average normal and abnormal waveforms are shown in Fig. 6(a) and (b) for two data sets. The frequency content of those signals, at least so far as the average waveforms show, appears quite

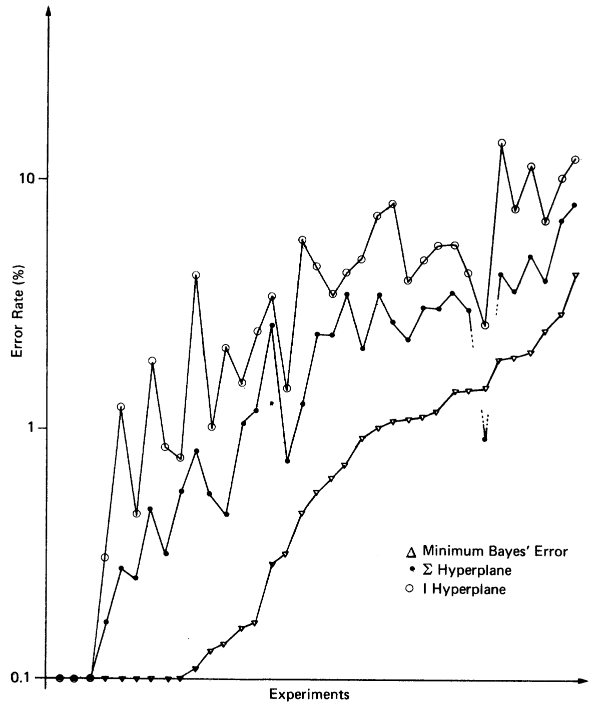

Fig. 5. Decision function performance.

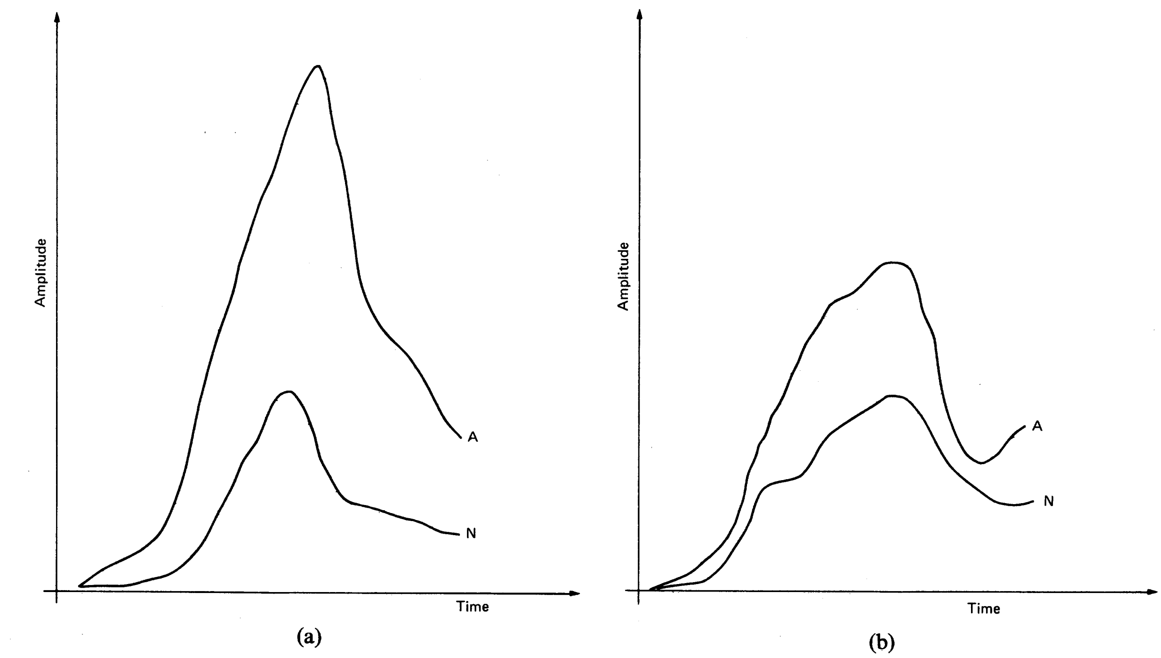

Fig. 6. (a) Average waveforms. (b) Average waveforms.

similar. Yet, when the sampling rate was decreased, the divergence decreased more quickly for the data shown in Fig. 6(b). Error rate tests also confirmed that higher sampling rates were necessary to extract adequate information from data similar to Fig. 6(b). See Fig. 7. This experiment was repeated for other data sets with similar results.

The last set of experiments involved the study and tracking of time trends. First of all, consider the individual waveforms in Fig. 2. The waveforms are highly distorted, probably due to variation of the mechanical impacts that produced them. However, the average of these waveforms over a short time period was found to be stable. We could plot the average waveforms and observe stability but this is not necessary. Stability can be observed from parameter plots such as peak value versus time or peak location versus time.

The number of waveforms to average (to eliminate short time variations) must be determined. This was done experimentally by measuring the variance (on the peak value for example) for progressively larger sets of m waveforms. For most available data, the variance stabilized when 20 waveforms were averaged. Thus, groups of 20 waveforms were

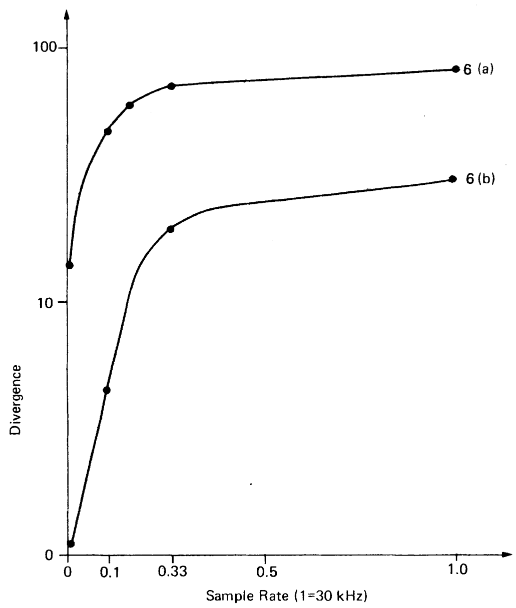

Fig. 7. Divergence versus sample rate.

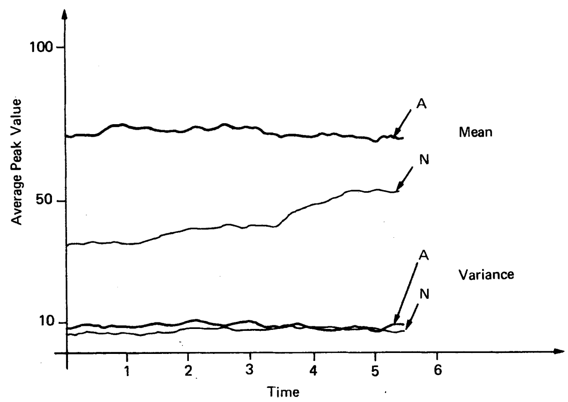

Fig. 8. Time trends.

used to generate the peak value versus time trend shown in Fig. 8. We notice, for this example, a distinct drift in mean value. This trend then can be observed and related to changes in the mechanism.

Conclusions

Signature analysis, using second-order statistical data analysis for pattern recognition of time signals, for mechanical diagnostics of small mechanisms has been studied. A rationale for this method of analysis was established and one condition for successful diagnostics, namely signal-event independence, was discussed. As a result, preprocessing, feature extraction, and recognition algorithms were formulated and applied to diagnosing mechanical subfunction failures relating to mechanical impacts in a small cyclic mechanism. Experimental data tend to validate second-order statistical assumptions for the mechanisms under study. A significant bonus to those who develop diagnostic systems is the design aids and trend tracking techniques which become readily available when second-order statistics are used. The applicability of divergence and the separation of waveform instability into short- and long-term variations resulted from applying these tools.

References

[1] D. L. George, "Noise and vibration in Q.C. design, and maintenance," *Sound and Vibration*, Oct. 1967, pp. 21–26.
[2] A. Hannavy, "New detection devices help predict potential failure," *Prod. Eng.*, Mar. 27, 1967, pp. 37–44.
[3] F. J. Lavoid, "Signature analysis: Product early warning system," *Mach. Des.*, Jan. 23, 1969, pp. 151–160.
[4] J. Page "Recognition of patterns in jet engine vibration signals," *IEEE Comput., Conf. Dig.*, Sept. 1967, pp. 102–105.
[5] K. D. Roberson and B. D. Olson, "The value of maintenance," presented at the Federal Aviation Administration Reliability and Maintenance Symp., Oklahoma City, Okla., Nov. 8, 1967, Paper AD664382.
[6] R. A. Rogers, "Integration of airborne data recorders and ground-based computers for engine maintenance purposes," presented at the FAA R & M Symp., Oklahoma City, Okla., Nov. 8, 1967, Paper AD664383.
[7] R. B. Tatge, "Acoustical techniques for machinery diagnostics," presented at the Acoustical Society of America Conf., Ottawa, Canada, May 21–24, 1968.
[8] H. G. Zerkle, "Development and test of the engine analyzer system," Wright Patterson Air Force Base, Dayton, Ohio., Aug. 1967, Paper AD665158.
[9] P. W. Becker, *Recognition of Patterns Using the Frequencies of Occurrence of Binary Words.* Copenhagen, Denmark: Polyteknisk, 1968, 216 pps.
[10] T. Pavidis, "A linguistic model for waveform analysis," Princeton University, Princeton, N. J., Mar. 1969, Paper AD684360.
[11] R. H. Peterson and R. L. Hoffman, "A new technique for dynamic analysis of acoustical noise," *IBM J. Res. Develop.*, vol. 9, May 1965, pp. 205–209.
[12] V. T. Rhyne, "A digital system for enhancing the fetal electrocardiogram," *IEEE Trans. Bio-Med. Eng.*, vol. BME-16, Jan. 1969, pp. 80–86.
[13] C. R. Trimble, "What is signal averaging," *Hewlett-Packard J.*, vol. 19, Apr. 1968, pp. 2–16.
[14] N. J. Nilsson, *Learning Machines, Foundations of Trainable Pattern-Classifying Systems.* New York: McGraw-Hill, 1965.
[15] K. Fukunaga and T. F. Krile, "Calculation of Bayes' recognition error for two multivariate Gaussian distributions," *IEEE Trans. Comput.*, vol. C-18, Mar. 1969, pp. 220–229.
[16] ——, "A minimum distance feature selection criteria," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-14, Sept. 1968, pp. 780–782.
[17] T. Kailath, "The divergence and Bhattacharyya distance measures in signal detection," *IEEE Trans. Commun. Technol.*, vol. COM-15, Feb. 1967, pp. 52–60.
[18] T. Marill and D. M. Green, "On the effectiveness of receptors in recognition systems," *IEEE Trans. Inform. Theory*, vol. IT-9, Jan. 1963, pp. 11–17.

A Direct Method of Nonparametric Measurement Selection

A. WAYNE WHITNEY

Abstract—A direct method of measurement selection is proposed to determine the best subset of *d* measurements out of a set of *D* total measurements. The measurement subset evaluation procedure directly employs a nonparametric estimate of the probability of error given a finite design sample set. A suboptimum measurement subset search procedure is employed to reduce the number of subsets to be

Manuscript received October 29, 1970; revised March 26, 1971. This work was supported in part by the Rome Air Development Center, Griffiss AFB, Rome, N. Y., under Contract F 30602-69-C-0227. A preliminary version of this note was presented at the IEEE Symposium on Feature Extraction and Selection in Pattern Recognition, Argonne, Ill., October 5–7, 1970.

The author is with the General Electric Company, Syracuse, N. Y. 13201.