PATTERN RECOGNITION CHARACTERIZATIONS OF MICROMECHANICAL AND MORPHOLOGICAL MATERIALS STATES VIA ANALYTICAL QUANTITATIVE ULTRASONICS

James H. Williams, Jr., and Samson S. Lee
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

If an ultrasonic signal is introduced into or generated within a structure, the state of the structure governs its propagation and detection. The assessment of the state of the structure can be affected by material properties, geometrical properties, environmental conditions, and measurement conditions. Because of the large number of possible properties and conditions, the quantitative ultrasonic determination of a specific microstructural or morphological state, independent of all other states, is difficult.

To complicate matters further, many nondestructive evaluation (NDE) parameters can be measured via ultrasonic interrogation as specified frequencies by using the Fourier transform (e.g., signal amplitude, signal duration, stress wave factor, and signal strength). Thus a large amount of data can be generated from a single ultrasonic measurement. Large multivariate data sets are difficult to decipher; thus, methods of summarizing and extracting relevant information are necessary.

One potential approach to the quantitative acquisition of discriminatory information that can isolate a single structural state is pattern recognition. The pattern recognition characterizations of micromechanical and morphological materials states via analytical quantitative ultrasonics are outlined in this paper. The concepts, terminology, and techniques of statistical pattern recognition are reviewed. Feature extraction and classification and states of the structure can be determined via a program of ultrasonic data generation.

INTRODUCTION

In acoustic-ultrasonic nondestructive evaluation (NDE), an ultrasonic stress wave is introduced into, or generated within, the interrogated structure and detected after it has propagated through the structure. Stress wave propagation is affected by the micromechanical and morphological materials states of the medium of propagation. Thus, acoustic-ultrasonic NDE involves the characterization of the tested structure on the basis of information contained in the detected stress wave signal.

The state of the structure, which governs stress wave propagation and detection, can be described by a broad range of properties and conditions, some of which are

Material properties:	elastic modulus, density, attenuation, velocity,...
Geometrical properties:	structural dimensions, discontinuities, microstructural and microstructural defect states, microstructural characteristic dimensions,...

Environmental conditions: mechanical loading, structural boundary conditions, residual stresses, temperature, absorbed moisture,...

Measurement conditions: location and size of transducers, sensitivity and frequency response of transducers, couplant, dynamic characteristics of electronic equipment,...

Even from this incomplete list, it is clear that the quantitative ultrasonic determination of a specific microstructural or morphological state, independent of all other states, is difficult.

To complicate matters further, many NDE parameters can be measured from an ultrasonic interrogation at specified frequencies by using the Fourier transform. A few of these are the maximum signal amplitude, signal duration, stress wave factor, and signal strength. Thus, a large amount of data can be generated from a single ultrasonic measurement. Large multivariate data sets are difficult to decipher; thus, methods of summarizing and extracting relevant information are necessary. Most often the summarizing and extraction are accomplished in an ad hoc qualitative manner.

One approach for the quantitative acquisition of discriminatory information that can often isolate a single structural state is pattern recognition. The objective of this study was to outline an approach for pattern recognition characterizations of micromechanical and morphological materials states via analytical quantitative ultrasonics. The concepts, terminology, and techniques of statistical pattern recognition are reviewed.

CLASSIFICATION BY PATTERN RECOGNITION

Determining the state of a sample via NDE by using pattern recognition techniques consists of three basic steps:

(1) Generating and processing NDE data
(2) Selecting significant features of the data
(3) Determining the sample state from the selected features

These three steps are illustrated in figure 1 as data generation, feature extraction, and classification, respectively.

DATA GENERATION

Data generation consists of ultrasonic NDE measurements that are expected to contain information for identifying the micromechanical and morphological states of a material or structure. Data generation may also involve data processing. Data processing involves signal conditioning or transformation of the collected data into various representations; an example of the latter is the acquisition of the frequency representation of a signal via its Fourier transform.

The processed data are arranged in an ordered set called a pattern vector $\underline{z}$ as (ref. 1)

$$\underline{z} = \begin{Bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_m \end{Bmatrix} \tag{1}$$

whose components $z_1, z_2, \ldots, z_m$ may contain, for example, the maximum signal amplitude, signal duration, and signal strength at specified frequencies evaluated via the Fourier transform. The number of components in the pattern vector is at the discretion of the researcher. Usually only a few known or anticipated discriminatory components are retained for pattern recognition analysis. These are preferentially selected (for subsequent correlation with structural states) by using the feature selection schemes described in the next section.

FEATURE EXTRACTION

A subset of the pattern vector in equation (1) is selected and is called the feature vector $\underline{x}$ (refs. 1 and 2)

$$\underline{x} = \begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{Bmatrix} \qquad \text{where} \quad n \leq m \tag{2}$$

The attractiveness of dimensionality reduction from m to n is in simplifying the computational efforts necessary for classification.

The feature vector lies in a vector space called the feature space. Each component of the feature vector forms a dimension of the feature space. Thus, if the feature vector has n components, the feature space is n-dimensional.

The components of the subset in equation (2) are selected in a very particular way so as to contain the most significant discriminatory components of the pattern vector. Interpreted graphically, the feature vectors are selected from the pattern vector such that feature vectors from distinct sample states (material properties, geometrical properties, environmental conditions, and measurement conditions) form distinct clusters in the feature space as illustrated in figure 2. Figure 2 shows a two-dimensional feature space formed by using the i^{th} and j^{th} components of a feature vector (i.e., x_i and x_j, respectively). Ideally, the feature vectors from three distinct sample states form three distinct nonoverlapping clusters in this feature space. Those feature vectors of known sample states that are used to define the clusters are called training samples. All feature vectors corresponding to unknown sample states will be classified by comparing them with the training samples by using the techniques described in this study.

A suitable feature vector produces the maximum separation between clusters (intercluster separation) as measured relative to the cluster size (intercluster dimension). The cluster size is defined by the spread of the feature vectors within the same sample state. The spread may be defined, for example, by the covariance of the feature vectors in a cluster.

Various distances can be used to measure the intercluster separation. If the covariances of samples in all clusters are similar and can be represented by a pooled sample covariance, the Mahalanobis distance D^2 can be used (refs. 2 and 3). The Mahalanobis distance D^2 is defined as the square of the Euclidean distance between the sample cluster centroids (where the centroid is located at the mean vector of all feature vectors belonging to the cluster) normalized by an averaged cluster size (e.g., the covariance) (ref. 4). For example, the Mahalanobis D^2 between clusters A and B is (refs. 2 to 4)

$$\text{Mahalanobis } D^2 = (\bar{\underline{A}} - \bar{\underline{B}})^T \, \underline{S}_{\text{pooled}}^{-1} \, (\bar{\underline{A}} - \bar{\underline{B}}) \tag{3}$$

where $\bar{\underline{A}}$ is the mean sample feature vector of cluster A, $\bar{\underline{B}}$ is the mean sample feature vector of cluster B, and $\underline{S}_{\text{pooled}}$ is the pooled sample covariance matrix of A and B. The superscript T denotes matrix transposition. Specifically, $\bar{\underline{A}}$, $\bar{\underline{B}}$, and $\underline{S}_{\text{pooled}}$ are defined as (ref. 3)

$$\bar{\underline{A}} = \frac{1}{N_A} \sum_{i=1}^{N_A} \underline{a}_i \tag{4}$$

$$\bar{\underline{B}} = \frac{1}{N_B} \sum_{i=1}^{N_B} \underline{b}_i \tag{5}$$

$$\underline{S}_{\text{pooled}} = \frac{1}{N_A + N_B - 2} \left[\sum_{i=1}^{N_A} (\underline{a}_i - \bar{\underline{A}})(\underline{a}_i - \bar{\underline{A}})^T + \sum_{i=1}^{N_B} (\underline{b}_i - \bar{\underline{B}})(\underline{b}_i - \bar{\underline{B}})^T \right] \tag{6}$$

where $\underline{a}_i$ and $\underline{b}_i$ are the feature vectors corresponding to clusters A and B, respectively, and N_A and N_B are the numbers of feature vectors in clusters A and B, respectively.

If the covariances of the samples in all of the clusters vary significantly, the Chernoff distance can be used (ref. 2). The Chernoff distance between clusters A and B is (ref. 2)

$$\text{Chernoff distance} = \frac{1}{2} s (1 - s)(\bar{\underline{A}} - \bar{\underline{B}})^T \left[(1 - s) \underline{S}_A - s\underline{S}_B \right]^{-1} (\bar{\underline{A}} - \bar{\underline{B}}) + \frac{1}{2} \ln \frac{|(1 - s) \underline{S}_A + s\underline{S}_B|}{|\underline{S}_A|^{1-s} \, |\underline{S}_B|^s} \tag{7}$$

where s is a real number between zero and unity, $\underline{S}_A$ is the sample covariance of cluster A, and $\underline{S}_B$ is the sample covariance of cluster B. Specifically, $\underline{S}_A$ and $\underline{S}_B$ are defined as (ref. 3)

$$\underline{S}_A = \frac{1}{N_A - 1} \sum_{i=1}^{N_A} (\underline{a}_i - \bar{\underline{A}})(\underline{a}_i - \bar{\underline{A}})^T \tag{8}$$

$$\underline{S}_B = \frac{1}{N_B - 1} \sum_{i=1}^{N_B} (\underline{b}_i - \bar{\underline{B}})(\underline{b}_i - \bar{\underline{B}})^T \tag{9}$$

When $s = 1/2$, the Chernoff distance is known as the Bhattacharyya distance (ref. 2).

The feature vector is selected from the pattern vector by maximizing the resulting intercluster distance. This can be accomplished analytically (ref. 1) or by trial and error with a computer search scheme (ref. 2).

CLASSIFICATION

A sample of unknown state is classified by determining the most likely sample state based on its feature vector. This is mathematically represented by the use of discriminant functions (refs. 1 to 3).

The discriminant function g_k is determined from the training samples such that if $\underline{x}$ is a feature vector corresponding to sample state k,

$$g_k(\underline{x}) > g_j(\underline{x}) \qquad \text{for all } j \neq k \tag{10}$$

Equation (10) can be interpreted as follows: values of discriminant functions corresponding to all sample states can be evaluated by using a feature vector from sample state k. Then equation (10) states that the value of the discriminant function corresponding to sample state k is the largest among values of discriminant functions corresponding to all other sample states.

The discriminant functions illustrated in figure 3 are based on the feature space represented in figure 2. Figure 3(a) shows discriminant functions g_1, g_2, and g_3 along a line C-C in the feature space shown in figure 3(b). Within the region dominated by the cluster corresponding to sample state 1, g_1 is greater than both g_2 and g_3, as required by equation (10). The points where the discriminant functions corresponding to sample states #1 and #2 have the same value; that is,

$$g_1(\underline{x}) = g_2(\underline{x}) \tag{11}$$

is a point on the so-called decision surface (ref. 1) separating the clusters corresponding to sample states #1 and #2.

Figure 4 illustrates the decision surfaces separating the feature space into as many regions as there are distinct sample states. The sample whose feature vector lies within region k should be classified as belonging to region k. Thus, the feature vector of unknown sample state in figure 4 classifies as belonging to sample state #1. The main task in classification becomes the determination of discriminant functions or decision surfaces.

There are two approaches: (1) parametric and (2) nonparametric (refs. 1 and 2). Before discussing the methods of evaluating discriminant functions or decision surfaces, criteria for accessing the performance of the selected discriminant functions or decision surfaces will be described.

Assessing classification functions. - The ideal discriminant functions or decision surfaces minimize the sample misclassification rate. A misclassification occurs when a sample with sample state k is classified as having a sample state other than k. (Sometimes it is not simply the misclassification rate that is important; certain misclassifications are more costly than others. Thus, it may be the cost of misclassification that should be minimized (ref. 3).)

One procedure to estimate the misclassification rate is to split the total training samples into two portions. One portion is used to establish the discriminant functions, and the other portion is used as validation samples to evaluate the misclassification of the resulting discriminant functions (ref. 3).

Another procedure to estimate the misclassification rate is called Lachenbruch's leaving-one-out method (refs. 2 and 3). Specifically, one training sample is left out in forming the discriminant functions; the left-out feature vector is then classified with the resulting discriminant functions and any misclassification is noted. Each training sample is omitted in turn, and the misclassification rate (or cost of misclassification) is evaluated.

Parametric methods. - If the probability distributions of the samples are known or can be assumed, parametric methods can be used to evaluate the discriminant functions (ref. 2). Denoting $P(k|\underline{x})$ as the conditional probability that a feature vector $\underline{x}$ belongs to sample state k, the discriminant function can be expressed as

$$g_k(\underline{x}) = P(k|\underline{x}) \tag{12}$$

Thus, for the feature vector $\underline{x}$ of unknown sample state, if

$$P(k|\underline{x}) > P(j|\underline{x}) \qquad \text{for all } j \neq k \tag{13}$$

then $\underline{x}$ is classified as belonging to sample state k. Equation (13) states that if the probability of a feature vector belonging to sample state k is greater than the probabilities of the feature vector belonging to all other sample states, the feature vector is classified as belonging to sample state k. Bayes' theorem is often used to evaluate the probabilities expressed in equation (13).

In any case, conditional probabilities of the samples must be known or must be assumed. Any remaining unknown parameters in the probabilities can be estimated by minimizing the misclassification rate.

Nonparametric methods. - If the probability distribution of the samples is not known or cannot be assumed, nonparametric methods must be used to evaluate discriminant functions or decision surfaces. There are many nonparametric classification schemes (refs. 1 and 2). Among these are the nearest-neighbor, nearest-centroid, Fisher, and kernel methods.

The nearest-neighbor method (ref. 2) consists of finding that training sample which lies closest to the unclassified feature vector and then classifying the unclassified feature vector to the same sample state as this nearest neighbor. The decision surface in this method lies equidistant between the boundaries of the clusters.

The nearest-centroid method (ref. 5) consists of finding the cluster whose mean feature vector (i.e., centroid) lies closest to the unclassified feature vector and then classifying the unclassified feature vector to the same sample state as this cluster. The decision surface in this method lies equidistant between the centroids of the clusters.

Fisher's method consists of assuming a linear discriminant function such that (refs. 1 and 3)

$$g_k(\underline{x}) = w_1^{(k)} x_1 + w_2^{(k)} x_2 + \ldots + w_n^{(k)} x_n + w_{n+1}^{(k)} \tag{14}$$

where $x_1, x_2, \ldots, x_n$ are the components of the feature vector $\underline{x}$ and the scalar coefficients $w_1^{(k)}, w_2^{(k)}, \ldots, w_{n+1}^{(k)}$ correspond to sample state k. The linear combination of components of the feature vector given in equation (14) is simple to calculate. For the special case where the covariances of the samples in all clusters can be assumed to be equal, analytical values for the scalar coefficients exist (ref. 3). The scalar coefficients $w_j^{(k)}$ are evalu- on the basis of the maximum separation between clusters and not the minimization of the misclassification rate.

The kernel method (ref. 6) consists of assuming kernel "potential" functions. Individual kernel functions are assumed to be centered at each training sample. The kernel function can be denoted by $K(\underline{x}, \underline{x}_i^{(k)})$, where $\underline{x}_i^{(k)}$ is the feature vector of the i^{th} training sample defining sample state k. Then the discriminant function is defined as the superposition over all training samples of the same sample state k as (ref. 1)

$$g_k(\underline{x}) = \frac{1}{N_k} \sum_{i=1}^{N_k} K\left[\underline{x}, \underline{x}_i^{(k)}\right] \tag{15}$$

where N_k is the number of training samples of the same sample state k.

In one-dimensional feature space, equation (15) can be written as (refs. 1 and 2)

$$g_k(x) = \frac{1}{N_k h_k} \sum_{i=1}^{N_k} K_o \left[\frac{x - x_i^{(k)}}{h_k} \right] \tag{16}$$

where K_o is the kernel shape, h_k is the kernel size, x is a position in the one-dimensional feature space, and $x_i^{(k)}$ is the position of the i^{th} training sample defining sample state k.

There are many forms for K_o in the literature. For example, the one-dimensional form of the exponential decay kernel can be written as

$$K_o \left[\frac{x - x_i^{(k)}}{h_k} \right] = \exp \left[- \left| \frac{x - x_i^{(k)}}{h_k} \right| \right] \tag{17}$$

The selection of the kernel shape and kernel size can be established by trial and error through a computer search scheme seeking to minimize the resulting misclassification rate.

CONCLUSIONS

Ultrasonic nondestructive evaluation (NDE) of structural states has been considered in this study. Because of the large number of possible properties and conditions describing the state of the structure and the large number of ultrasonic NDE parameters that can be considered, pattern recognition techniques have been suggested for identifying structural states from discriminatory information.

An outline has been provided for the pattern recognition characterizations of micromechanical and morphological materials states via analytical quantitative ultrasonics. The concepts, terminology, and techniques of statistical pattern recognition have been reviewed.

Determining the state of a sample by NDE with pattern recognition techniques consists of ultrasonic NDE data generation, feature extraction, and classification. Ultrasonic data generation consists of ultrasonic NDE measurements that are expected to contain information capable of identifying the micromechanical and morphological states of the material or structure. The collected data are organized in a pattern vector.

By using samples of known states, called training samples, the significant discriminatory components of the pattern vector are retained as a feature vector. The feature vectors are extracted such that training samples of distinct sample states form distinct clusters in the feature space. Then classification is achieved by defining discriminant functions or decision surfaces based on the training samples by using parametric or nonparametric methods. The ideal classification scheme will minimize the resulting misclassification rate (or the cost of misclassification).

Thus, through a program of ultrasonic NDE data generation, feature extraction, and classification, the most likely materials states corresponding to an unknown sample can be determined. The pattern recognition techniques discussed in this study have broad applicability to various NDE procedures if samples of known states are available.

REFERENCES

1. H.C. Andrews, Introduction to Mathematical Techniques in Pattern Recognition, Wiley-Interscience, John Wiley & Sons, Inc., NY, 1972.

2. D.J. Hand, Discrimination and Classification, John Wiley & Sons, NY, 1981.

3. R.A. Johnson and D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1982.

4. D.F. Morrison, Multivariate Statistical Methods, McGraw Hill, NY, 1976.

5. M.R. Anderberg, Cluster Analysis for Applications, Academic Press, NY, 1973.

6. D.J. Hand, Kernel Discriminant Analysis, Research Studies Press, NY, 1982.

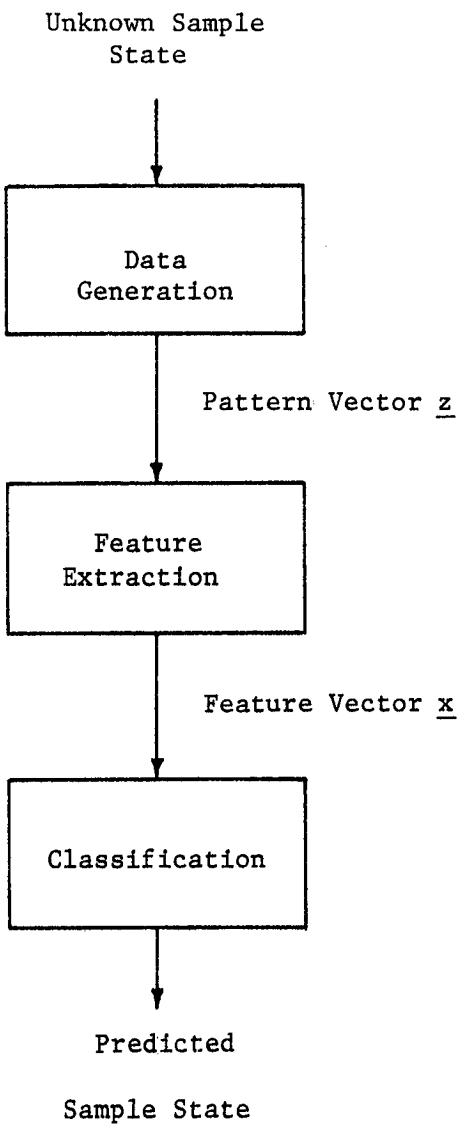

Fig. 1 Determination of state of unknown sample using pattern recognition techniques based on nondestructive evaluation data generation.

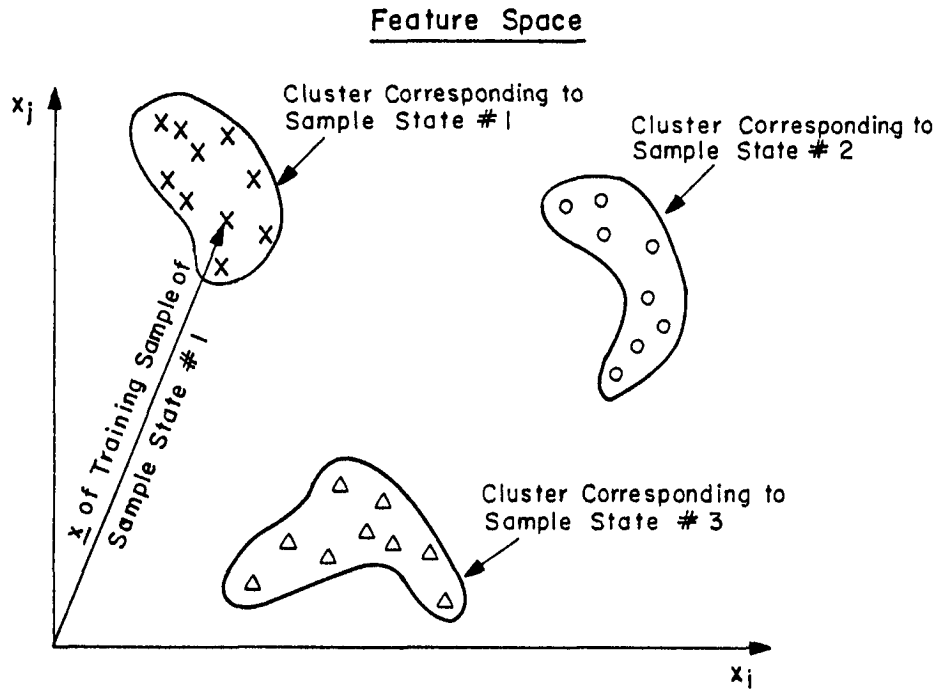

Fig. 2 Two-dimensional representation of feature space formed using i-th and j-th components of feature vectors, illustrating separate clusters of feature vectors $\underline{x}$ of training samples having distinct sample states.

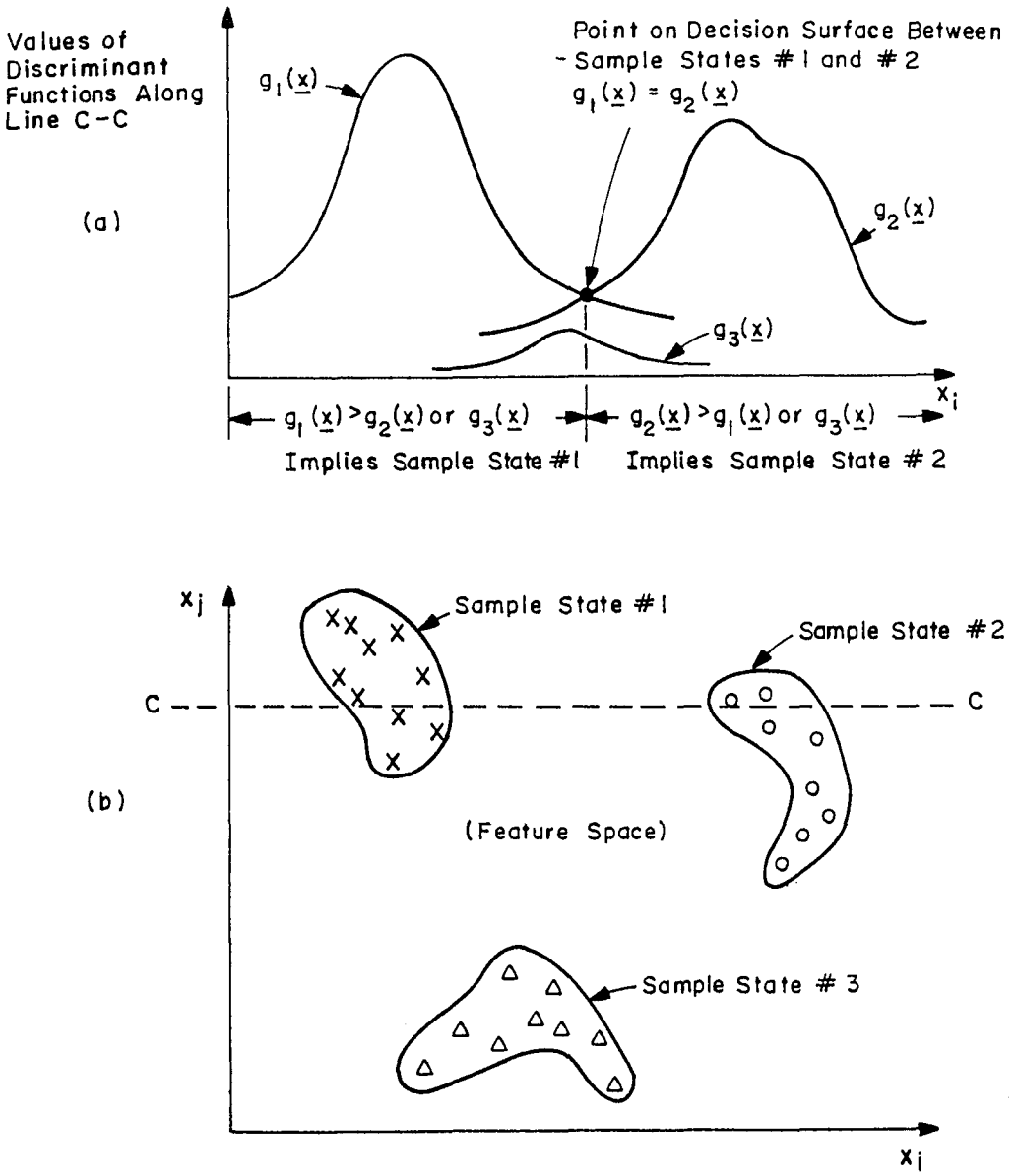

Fig. 3 Schematic illustrating (a) discriminant functions $g_k(\underline{x})$ along a line C-C in feature space shown in (b).

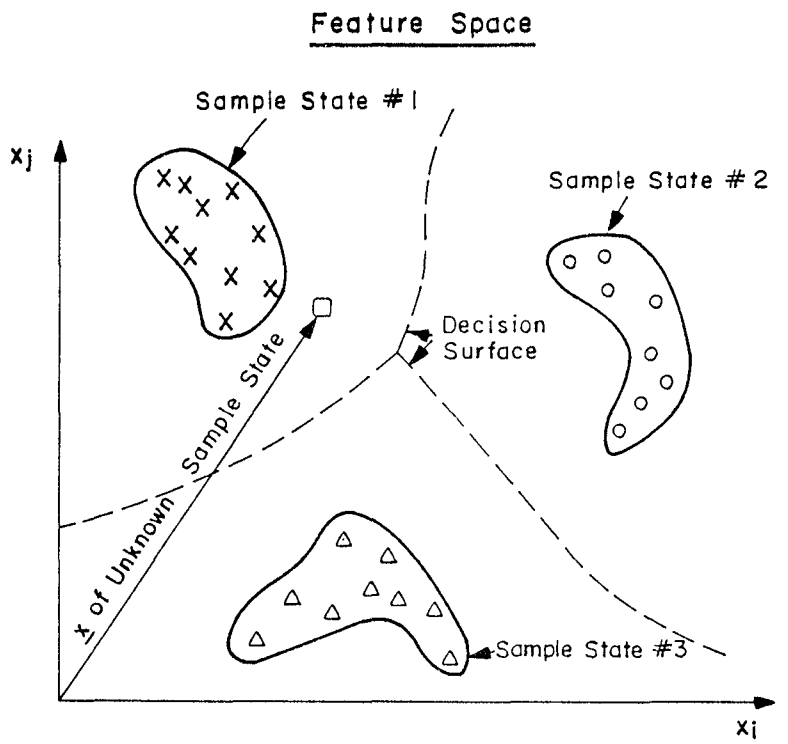

Fig. 4 Schematic illustrating decision surfaces separating feature space into regions corresponding to distinct sample states, and also illustrating classification of unknown sample shown as belonging to sample state #1.