Final Report PFTR-1030-76-8
Contract No. NAS-8-31782
Report Date August 1976

DEVELOPMENT OF SIGNAL PROCESSING ALGORITHMS FOR ULTRASONIC DETECTION OF COAL SEAM INTERFACES

(NASA-CR-150024) DEVELOPMENT OF SIGNAL PROCESSING ALGORITHMS FOR ULTRASONIC DETECTION OF COAL SEAM INTERFACES Final Report (Perceptronics, Inc., Woodland Hills, Calif.) 43 p HC A03/MF A01 CSCL 08I G3/43

N77-10610

Unclas 08029

Denis D. Purcell
Moshe Ben-Bassat

Prepared For:

National Aeronautics and Space Administration
George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

PERCEPTRONICS
6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91364 • PHONE (213) 884-7470

Final Report PFTR-1030-76-8
Contract No. NAS-8-31782
Report Date August 1976

DEVELOPMENT OF SIGNAL PROCESSING ALGORITHMS FOR ULTRASONIC DETECTION OF COAL SEAM INTERFACES

Denis D. Purcell
Moshe Ben-Bassat

Prepared For:

National Aeronautics and Space Administration
George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

PERCEPTRONICS

COPY NO. 02

6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91364 • PHONE (213) 884-7470

This report was prepared by Perceptronics, Inc. under NAS-8-~~3170~~ 31782, "Development of Signal Processing Algorithms for Ultrasonic Detection of Coal Seam Interfaces", for the George C. Marshall Space Flight Center of the National Aeronautics and Space Administration.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 General

This report for NASA Contract NAS8-31782 describes the development of a pattern-recognition system for determining the thickness of coal remaining on the roof and floor of a coal seam. The system was developed to recognize reflected pulse echo signals that are generated by an acoustical transducer and reflected from the coal seam interface. The flexibility of the system, however, should enable it to identify pulse-echo signals generated by radar or other techniques -- the main difference being the specific features extracted from the recorded data as a basis for pattern recognition.

1.2 Background and Goals

It is difficult to interpret the pulse echo data conventionally due to signal attenuation and noise reflected by cracks and impurities. This is so because the desired information may not be present in only a small set of extracted features. Rather, it may reside in a relationship between many values and features which are useless when taken alone. Pattern recognition is capable of discovering and using such relationships when they are very complex and invisible to other techniques of examination. Our goal has been to specify feasible pattern-recognition algorithms which will permit application of acoustical pulse-echo techniques in the remote control of continuous mining machines.

Specific program objectives included the following:

(1) To determine the applicability and explore the feasibility of signal processing and adaptive pattern-recognition techniques for detection of coal thickness by acoustic pulse-echo signals.

(2) To realize feasible detection algorithms and evaluate their relative performance by computer in order to enhance the reliability of detection.

(3) To establish design specifications for implementation and interfacing to the acoustical sensing system.

(4) To provide guidelines for prototype construction and practical field utilization.

1.3 Summary of Accomplishments

A software system was developed which does the following:

(1) Reads and processes data samples.
(2) Extracts features from each sample.
(3) Uses training data to train a pattern recognizer.
(4) Classifies test data.

Many features can be extracted including Fourier values, power spectrum values, cross-correlation, cross-spectral density, time-domain maxima and minima, derivatives, etc. Pattern recognition algorithms include the following:

(1) Threshold Logic Machine (TLU) (See 2.4.1).[1] (Nilsson, 1965)

(2) Multiple-category classifier using discriminant functions (see 2.4.2).[1] (Nilsson, 1965)

[1]The TLU is really only a special case of a general discriminant function system. They are considered separately here because the context of their usage in the system is different.

(3) K-Nearest Neighbor classifier (see 2.4.3). (T.M. Cover and T.E. Hart, 1967; Young and Calvert, 1974; Duda and Hart, 1973.)

Success was achieved when we applied the system to ten acoustic data samples and nine radar data samples in two independent experiments. The K-Nearest Neighbor recogniiton algorithm was applied in both the acoustic experiment and the radar experiment. The acoustic samples were classified with 90% accuracy and the radar samples were classified with 89% accuracy.

2. SYSTEM SOFTWARE

2.1 Overview

A software system was developed which is capable of performing all of the tasks necessary to recognize recorded data samples on paper tape. In addition, options and parallel operations are available at three levels in the processing sequence. These levels are: (a) Pre-Processing of Signal, (b) Feature Extraction, and (c) Pattern Recognition. A functional diagram of the system is provided in Figure 1.

During the pre-processing phase, data samples are first read from paper tape and scaled according to coded scale parameters provided on the tape and, when necessary, according to coal-penetration energy. If there was any drift in the time scale during the earlier recording process, the scaled samples are then time calibrated.[1] A search window immediately following the front-surface pulse echo is then located either by cross-correlation with the transducer pulse or by a direct search (see Figure 2). This front surface echo is invariably strong and presents no difficulties in recognition. At this point a moving average can be used to smooth the search window values.

Feature extraction can be performed in two ways: (1) Features can be extracted from a smaller frame in the search window, just large enough to contain the coal-seam pulse echo, or (2) features can be derived from the complete search window. In the first case, the smaller frame is moved across the search window to generate a set of features for each possible position of the coal-seam echo. In the second case, only one set of features is derived.

---

[1]We found it necessary to calibrate the preliminary radar data samples.

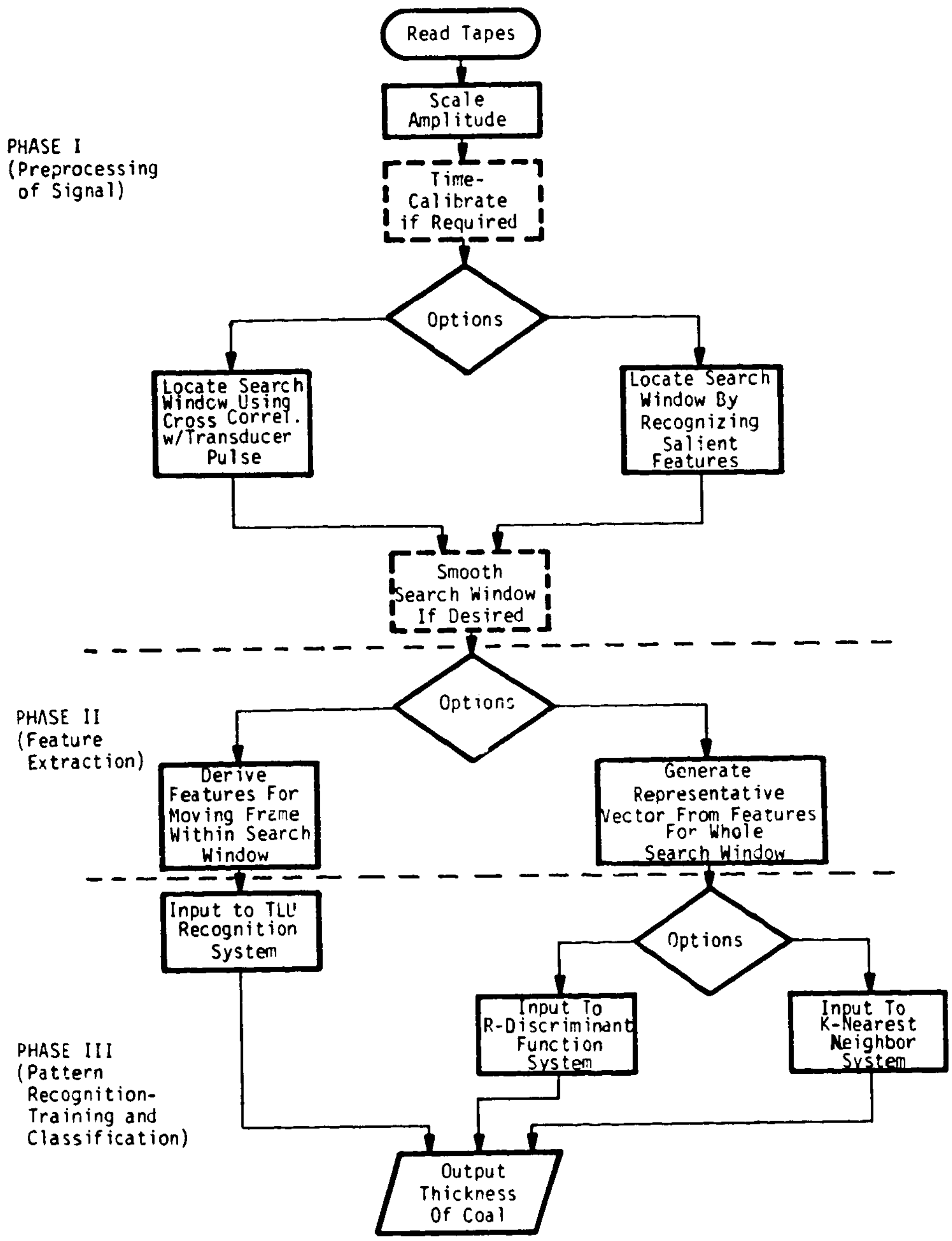


FIGURE 1
PATTERN RECOGNITION SYSTEM DEVELOPED FROM ACOUSTICAL AND PRELIMINARY RADAR SAMPLES

ORIGINAL PAGE IS
OF POOR QUALITY

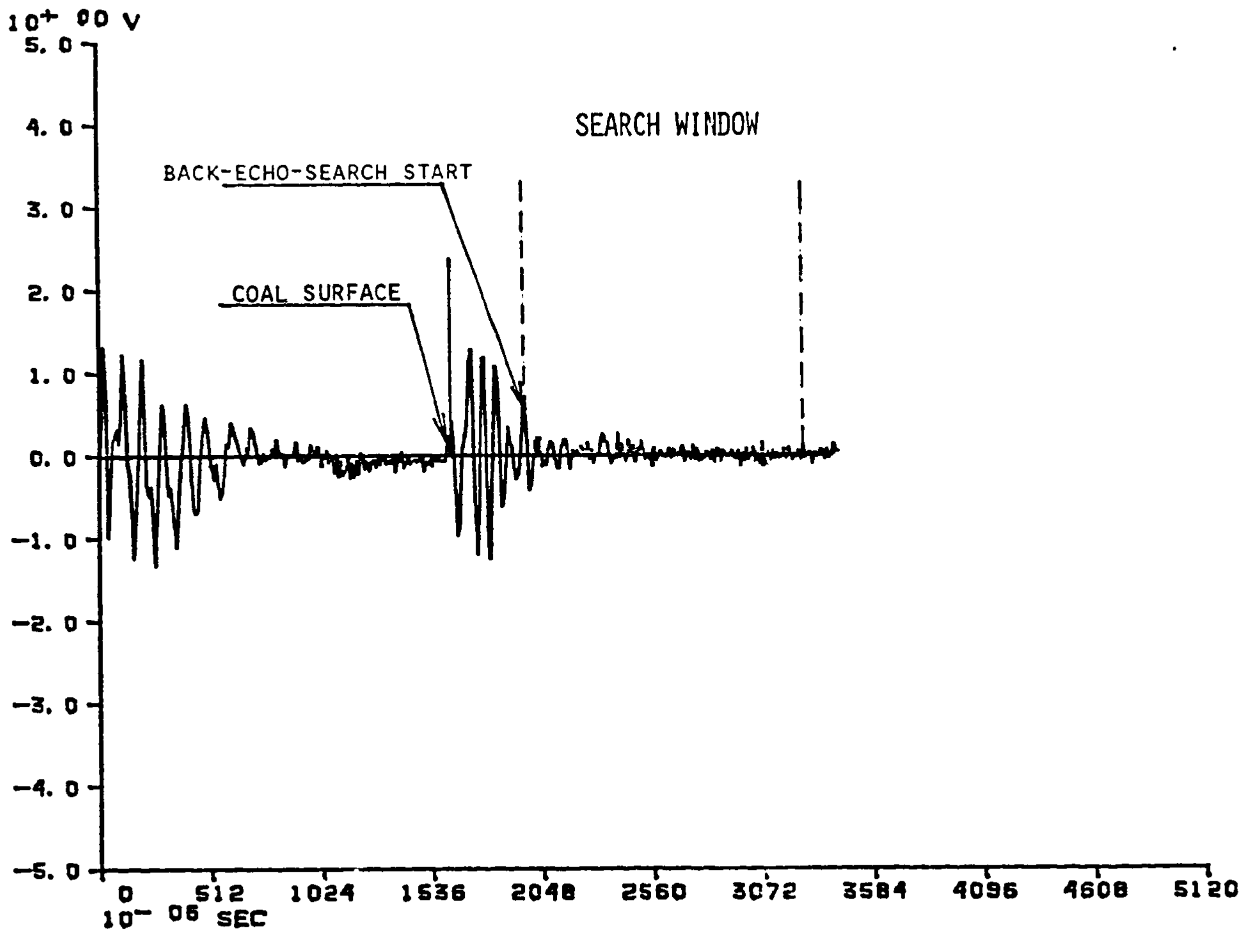


FIGURE 2. SEARCH WINDOW

Features derived may include raw time-scale values, special time-scale parameters including derivatives, maxima, minima, Fourier analysis, and power spectrum. With a given data set, only those features which prove most important during the pattern recognition phase are used.

There are three pattern-recognition algorithms available to the system:

(1) Threshold Logic Machine (TLU)
(2) Discriminant-Function System
(3) K-Nearest Neighbor System

The TLU is only used with the moving-frame system of feature extraction. This two-category classifier is first trained on the training data samples to identify the coal-seam echoes in these samples. After training, when the feature set for the specific frame containing a complete coal-seam echo is presented to the TLU, a "YES" response is returned. The TLU responds with a "NO" to the feature set for any frame in which the complete coal-seam echo is not present. Test data is subsequently presented to the TLU to locate coal-seam echoes and corresponding coal thicknesses.

The discriminant function system, used only with the full-search-window, feature-extraction technique, utilizes a set of discriminant functions representing a set of possible coal thicknesses. The number of discriminant functions is, therefore, a function of the range of coal thicknesses considered and the desired precision of the classification process. For instance, if it is desired that thicknesses between one inch and two inches be resolved to an accuracy of one-tenth of an inch, then eleven discriminant functions representing 1", 1.1", 1.2", etc., up to 2" are required. These discriminant functions are given the training data from known coal thicknesses and adjusted until they can accurately classify each member of the training set. Test data is then given to the discriminant functions to classify according to the corresponding coal thicknesses.

The K-Nearest Neighbor system is also used only with the full-search-window, feature-extraction technique. Again, there is a category assigned to each desired coal thickness within the desired range. However, rather than using discriminant functions to represent categories, each element in the training set of a particular thickness is used to represent that thickness. Thus, if there are five training samples for 1-3/8" coal, those five samples collectively represent the 1-3/8" category. A test sample is classified according to its proximity to the representatives of the various categories. The K-nearest representatives vote on the new sample's membership in a category. For instance, with a balanced training set, if K [illegible] 5 and the test sample is closest to representatives of the categories 1-1/2", 1-5/8", 1-5/8", 1-3/8", 1-5/8"; then the test sample would be associated with the 1-5/8" category.

2.2 Pre-Processing Signal

During the pre-processing phase, data samples are read into the computer, scaled, and, when necessary, calibrated; search windows are identified and, if desired, smoothed. Each of these activities is described in detail below. Phase 1 of Figure 1 diagrams this process.

2.2.1 Reading Tapes. The data samples sent us were on punched paper tape. The scale parameters and data were read into the computer using a specifically designed paper tape control program. In order to be sure that all data in a given set of samples was comparable in magnitude, the signals were unscaled according to the associated scale parameters.

2.2.2 Scaling and Calibration. The K-Nearest Neighbor technique is sensitive to any variation of energy content in the coal-seam echo; consequently, when using this technique, steps were taken to normalize the samples for energy content. This was done by scaling so that the maximum peaks in the front-surface echo of all data samples in a set were the same height. This procedure insures that the amount of signal energy actually penetrating the coal sample is reasonably constant. (Certainly some variation still occurs due to differences of reflectivity of the coal surface. Under the circumstances, however, it was the best procedure available. In a prototype system the amount of energy penetrating the coal should be kept constant.)

We developed the calibration procedure to handle drifts in the time scale in the preliminary radar data. A calibration value, the distance between the original pulse peaks in the radar signal, was used to stretch or shrink the time scale as needed.

2.2.3 Locating and Smoothing Search Windows. After the data samples have been processed for uniformity, it remains to identify an area of each data sample called a "search window" (see Figure 2). These search windows trail the front-surface echo by a fixed amount and are, therefore, aligned with one another.

This procedure involves locating the front-surface echo with high accuracy. This can be done using the cross-correlation of the data sample with the transducer pulse. The highest peak in the cross-correlation corresponds to the precise location of the front-surface echo. In the last group of acoustical samples we received, no transducer pulse record was available for cross-correlation. Consequently, we located the front-surface echo by a simple search for peak magnitude in the data sample. (A similar technique was used with the preliminary radar data. See 3.2.)

When a search window has been located, it may be smoothed using a moving average technique. This procedure if used, must be applied uniformly to all search windows over the set of samples involved.

2.3 Feature Extraction

The feature extraction phase is diagramed in Phase II of Figure 1.

2.3.1 Time and Frequency Parameters. The search windows provide a data base for feature extraction. Programs exist to derive the following parameters.

Time Domain.

(1) Selected raw amplitudes
(2) Maxima and minima
(3) Derivatives
(4) Maximum and minimum derivatives

Frequency Domain. (Cooley, J.W. and Tukey, J.W., 1965; Rosenfield, 1969; G.D. Berglund, 1969.)

(5) Fourier analysis
(6) Power spectrum
(7) Spectrogram snapshots
(8) Maxima and minima of power spectrum
(9) Derivatives of power spectrum
(10) Maxima and minima of derivatives of power spectrum

Other features such as cross-spectral density and cepstrum can easily be fitted into the current system if needed.

2.3.2 Moving Frame. The moving-frame, feature-extraction technique is diagramed in Figure 3. The moving frame is just as long as the expected coal-seam echo. By moving the frame across the search area in the manner of a template, the desired echo is sought. For each position of the frame, features such as those given in 2.3.1 are derived.

If the moving-frame technique is used, the TLU pattern-recognition system must be employed. The TLU, a two-category classification system, is taught to respond correctly with a Yes or No answer to the question of whether or not the current position of the frame contains the desired back echo.

2.3.3 Representative Vector. The simplest form of feature extraction is to use the entire search window itself as a feature vector. This technique proved adequate when recognizing the preliminary coal samples. However, the current system can derive any or all of the features mentioned in 2.3.1 and use them to represent the original data sample from which the search window was derived. This algorithm yields a single vector representing the data sample rather than a feature vector for each position of the moving frame discussed in 2.3.2.

Such a vector is then given to either the Discriminant-Function pattern-recognition system or the K-Nearest Neighbor system -- the objective being to train the system used to correctly classify a test vector according to the width of the coal from which its corresponding data sample was taken.

2.3.4 Extraction Algorithms. By checking the accuracy with which a given pattern recognition system works for different combinations of features, a set of features can be extracted which relatively optimizes the performance of the classifier involved. Sometimes the number of these "optimal" features

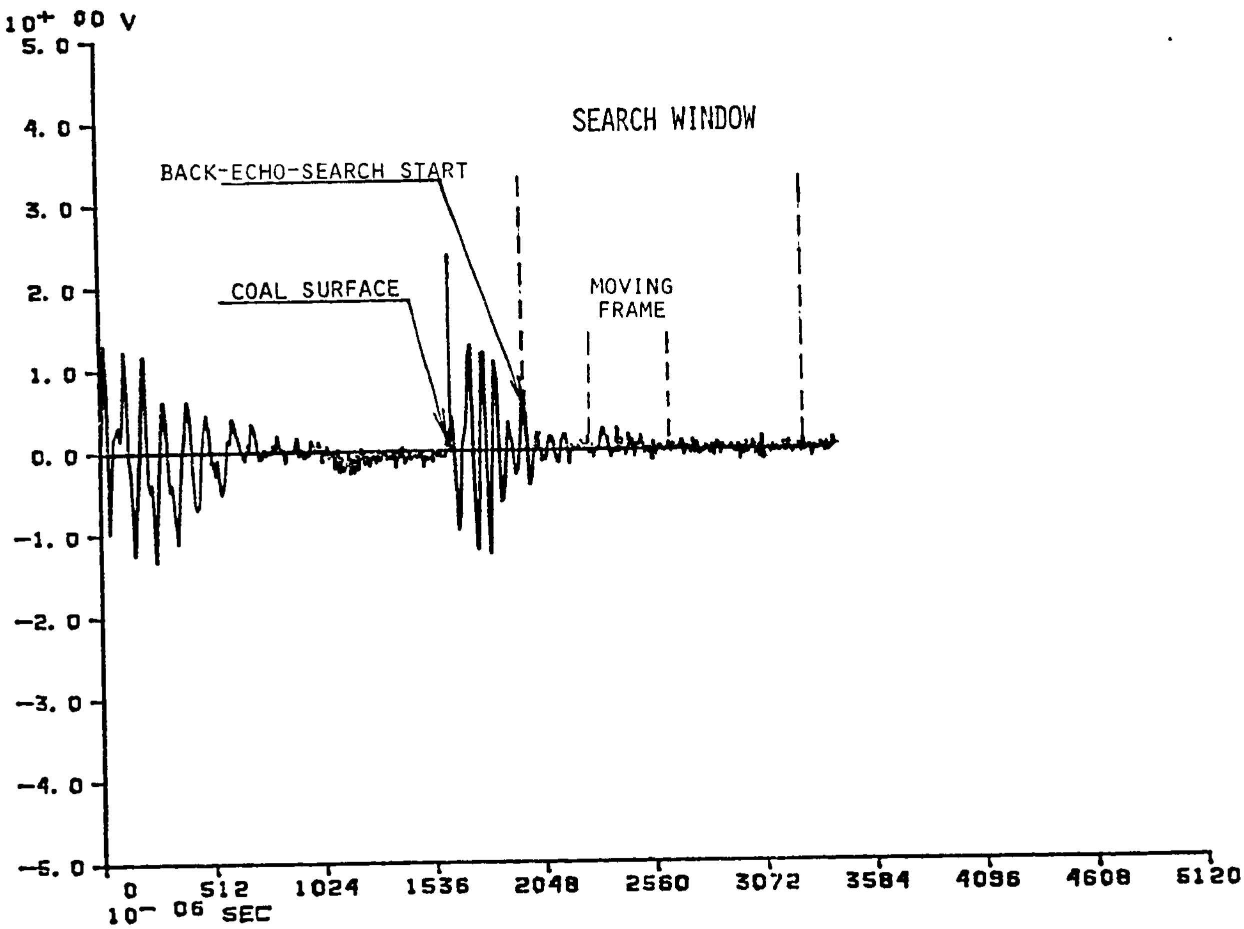


FIGURE 3. SEARCH WINDOW WITH MOVING FRAME

may be substantially less than the number of features first used. This makes subsequent pattern recognition simpler and, consequently, quicker for the given reference set of values.

2.4 Pattern Recognition (Training and Classification)

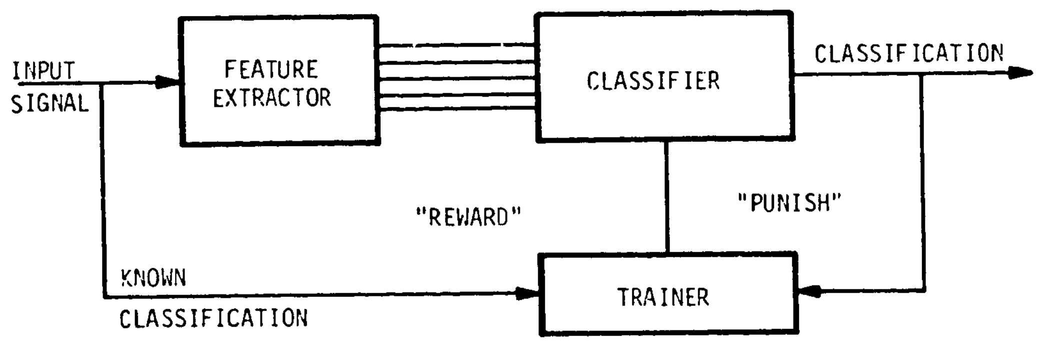
The Pattern Recognition section per se in the current system is diagramed in Phase III of Figure 1. Figure 4 shows a Pattern Recognition System of the TLU or R-discriminant function type. The TLU is discussed in 2.4.1 and the more general R-discriminant function system is discussed in 2.4.2.[1]

2.4.1 Threshold Logic Machine (TLU). Figure 5a diagrams a general TLU and Figure 5b diagrams a specific TLU (i.e., a TLU with a specific discriminant function). Basically a TLU is a single real-valued function g of a vector X. If g(X) is greater than 0, X is placed in category 1, if g(X) is less than or equal to 0, X is placed in category 2. The current system uses a linear or a quadric discriminant function. These functions have the following forms:

Linear

$$(1) \quad f(X) = a_1X_1 + a_2X_2 + \ldots + a_nX_n + a_{n+1}$$

[1]An R-discriminant function system is actually equivalent to a system of TLU's. The two are discussed separately because the single TLU is applied in a different context here than the general discriminant function system.



(Input signals are supplied from a training set)

FIGURE 4. TRAINABLE PATTERN CLASSIFIER

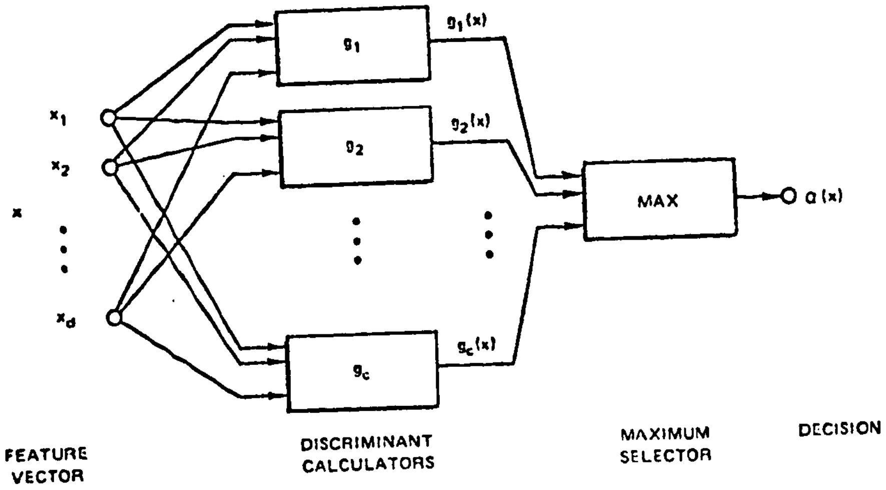
ORIGINAL PAGE IS
OF POOR QUALITY

AMPLITUDE (A) $\longrightarrow$ (a_1)
WIDTH (W) $\longrightarrow$ (a_2) $\longrightarrow \sum \rightarrow$ (g(A,W,C) $\rightarrow$ THRESHOLD $\rightarrow$ #1 / $\rightarrow$ #2
CENTRAL FREQUENCY (C) $\longrightarrow$ (a_3)

$$g(A,W,C) = a_1 A + a_2 W + a_3 C$$

IF $g(A,W,C) \leq$ THRESHOLD $\rightarrow$ #1
IF $g(A,W,C) <$ THRESHOLD $\rightarrow$ #2

FIGURE 5b. A TWO-CATEGORY CLASSIFIER EXAMPLE (LINEAR)



$$g_j = \sum_{I=1}^{I=c} x_i a_i \qquad \text{for linear } g_j\text{'s}$$

FIGURE 6. A PATTERN CLASSIFIER

Quadric

$$(2)\quad f(X) = a_1X_1^2 + a_2X_2^2 + \ldots + a_nX_n^2 +$$

$$a_{n+1}X_1X_2 + a_{n+2}X_1X_3 \ldots + a_{2n-1}X_1X_n +$$

$$a_{2n}X_2X_3 + a_{2n+1}X_2X_4 + \ldots + a_{3n-3}X_2X_n +$$

$$\vdots$$

$$+ a_{n \cdot (n+1)/2} X_{n-1}X_n +$$

$$a_{n(n+1)/2+1} \cdot X_1 + a_{n(n+1)/2+2} \cdot X_2 + \ldots$$

$$+ a_{n(n+3)/2} X_n +$$

$$a_{n(n+3)/2+1}$$

The coefficients of the discriminant function used are adjusted until the function accurately classifies the training vectors (see Figure 4). When the training is finished, the discriminant function is used to classify the test vectors.

The vectors fed to the TLU in the current system are feature vectors -- each representing a different position of the moving frame which slides over the search area for a data sample (see Figure 3). The TLU then classifies each feature vector as to whether or not the corresponding frame contains the complete coal-seam echo.

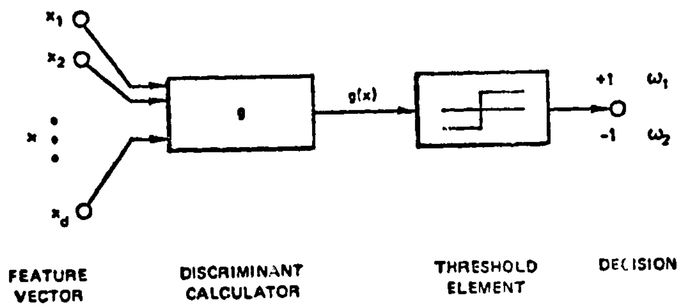
2.4.2 R-Discriminant Function Classifier. A pattern classifier with several discriminant functions is diagramed in Figure 6. The discriminant functions are of the same kind as those described in 2.4.1 and are trained in the same manner. The object of the training program is to produce coefficients in the various discriminant functions such that when presented with a feature vector X representing category i, the i'th discriminant function of X will be larger than the other discriminant functions of X. When fully trained, the discriminant function system is used to classify the test vectors.

The R-discriminant function classifier is used exclusively to classify feature vectors representing entire data samples. One decision is made to determine to what coal-width category the represented sample belongs.

2.4.3 K-Nearest Neighbor Classifier. A K-Nearest Neighbor Classifier is shown in Figures 7a and 7b. Although no discriminant functions are used, the K-Nearest Neighbor algorithm is applied in the same way as the R-Discriminant function technique. A single representative feature vector is used for each data sample and there is no moving frame as discussed in 2.4.1. Just as in the manner given in 2.4.2, each discrete coal width is represented by a separate category in the classifier.

The actual classification process, however, involves computing the distance between a new vector and all vectors in the training set. The closest K vectors in the training set then vote on the membership of the new vector. Since each vector in the training set belongs to a certain category, the new vector is assigned to that category with the largest number of close vectors. This constitutes a simple-majority voting technique and was adequate for our purposes.[1]

[1]A rejection rule can also be used to discard new vectors if their membership is not adequately clear-cut. This might occur, for instance, when no 2/3 majority vote was present (I.T. Tomek, 1976). Additionally, votes can be normalized by the distance of the corresponding training vector from the test vector (S.A. Dudani, 1976).



$$g(x) = g_1(x) - g_2(x)$$

x is in category 1 if $g(x) > 0$
x is in category 2 if $g(x) < 0$

$$g_j = \sum a_i x_i$$

FIGURE 5a. A TWO CATEGORY PATTERN CLASSIFIER

Input:

M = Number of possible classes

N = Number of preclassified patterns

$T = \{X^1, X^2, \ldots X^N\}$ training patterns

$L = \{\ell^1, \ell^2, \ldots, \ell^N\}$ labels, of training patterns

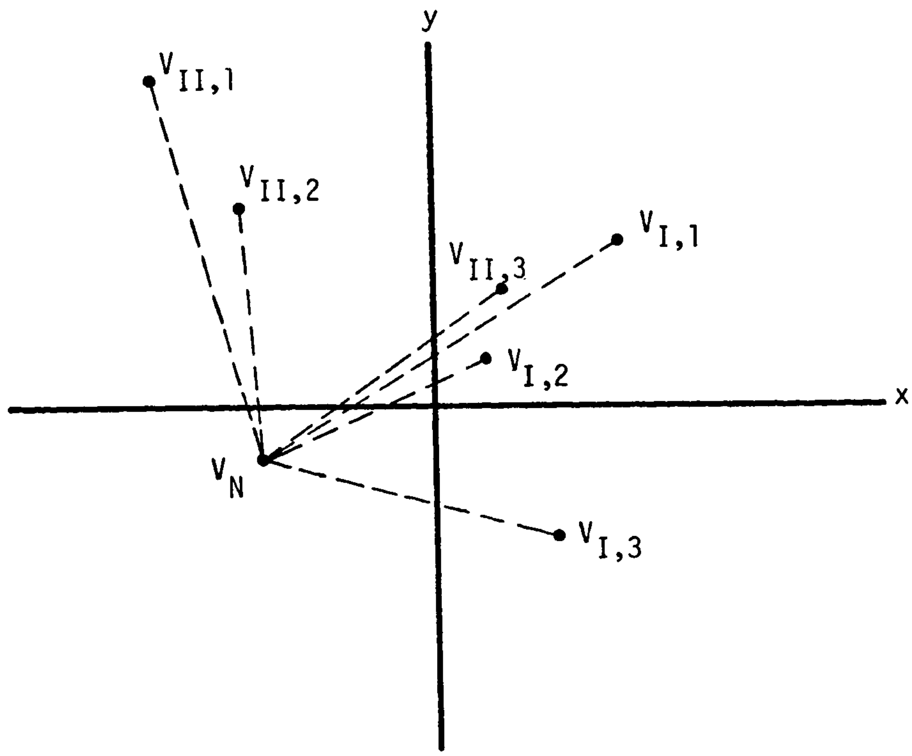
X = An unknown pattern

d = Distance function

Procedure:

1. Compute $d(X^j, X)$ for $j = 1, 2, \ldots, N$
2. Identify the k nearest neighbors $T_k = \{X^{j1}, \ldots, X^{jk}\}$ and their corresponding labels $L_k = \{\ell^{j1}, \ldots, \ell^{jk}\}$
3. Count N_i the occurrence of class i in L_k
4. Assign X to class c* where $N_{c*} = \max \{N_1, \ldots, N_m\}$

FIGURE 7a. K-NEAREST NEIGHBOR ALGORITHM



Vectors here have two coordinates and are represented by the corresponding points in Euclidean 2-space (E^2). If we consider the 3 nearest points, they are, respectively, $V_{I,2}$, $V_{II,2}$, and $V_{II,3}$.[1] Here I means that category I is represented and II means that category II is represented. Since the vote is 2 II's and 1 I, the new vector corresponding to point V_N is placed in category II.

---

[1]The simple two-dimensional distance metric is used here (i.e., $D(X,Y) = ((X_1-X_2)^2 + (Y_1-Y_2)^2)^{1/2}$.

FIGURE 7b. K-NEAREST NEIGHBOR DIAGRAM

Such a pattern classifier is "trained" simply by providing it with a set of training vectors. Unlike the discriminant function system, no adjustment of parameters is required, and the vectors themselves are used to represent the categories.

3. EXPERIMENTS PERFORMED

3.1 Original Acoustics Data

Three experiments were performed with the original acoustics data (i.e., the acoustics data received prior to early July). These three experiments involved the pattern recognition techniques described in 2.4.

It is appropriate here to indicate the data-base requirements of a successful pattern classification system. The R-discriminant function system and the K-nearest neighbor technique require a representative set of data samples *for each thickness category* to train the system adequately. In addition, the data acquisition and recording techniques must be uniform (e.g., the amount of energy penetrating the coal being tested must be about constant, artifacts from test apparatus used in recording must be absent or at least consistent from sample to sample). The data base with which we were dealing was very inadequate from these two standpoints.

Originally, we had hoped to obtain about six hundred uniform data samples over four thickness categories (i.e., 7/8", 1-1/8", 1-5/8", and 2-5/8"). In fact, we were sent sixteen data samples covering six categories. Fourteen of these samples had been smoothed with a 27KHZ filter, two had not. Ten of them represented an average of twenty signals, six were not averaged, and six of the samples were recorded at 100 MHZ, ten were recorded at 200 MHZ. The large variation in data acquisition and recording techniques represented by these samples made them virtually useless for training and testing purposes. However, we did the best we could with these data until the final acoustic samples were sent to us in early July.

The experiments described below were performed merely to illustrate the approaches and the methods from the very beginning to the very end including performance evaluations. They cannot be used as a scientific proof. However, the last two experiments indicate strong anticipation for a very high level of performance with an acceptably large data base. Moreover, a considerable part of the current software system which has been developed for this pilot study is still useful for a real system.

3.1.1 Using Six Discriminant Functions. Our first experiment used six discriminant functions applied to smoothed search windows representing fourteen of the sixteen acoustical samples available to us at that time. (Note: the two unfiltered samples were eliminated from consideration because visual examination of the graphs showed a great difference in the noise level present in these two cases.) The only features used were the smoothed search windows themselves (see Figure 1). The results of this experiment were inconclusive.

A system of six discriminant functions representing six thickness categories of from 7/8" to 2-5/8" were trained using ten samples from the set of twelve we had received last. Convergence was achieved (i.e., the functions successfully learned to recognize the ten training samples).[1] When the four remaining samples were presented to the pretrained system, two were correctly classified and two were incorrectly classified. A level of fifty percent success or better can be expected on a chance basis, however, 13.2% of the time. This experiment is summarized in Figure 8.

3.1.2 Using TLU. The second experiment represented an attempt to deal with the lack of uniformity in the data. A moving frame within the search window was used. The technique described in 2.4.1 has the advantage of providing rigorous training at identifying coal-seam echoes since each possible position of the frame over all data samples in the training set is used. All twelve of the latest acoustical samples we had received at

[1]This was anticipated since the number of training samples is almost equal to the number of discriminant functions.

EXPERIMENT NO.	DATA BASE	FEATURES	TECHNIQUE	RESULTS
1	14 Acoustical Samples	Smoothed Search Windows	Six Linear Discriminant Functions	50% Accuracy
2	12 Acoustical Samples	Smoothed Search Window, Maximum, Minimum; Derivatives, Max., Min.; PSD, Spectrogram Snapshot, Derivative, Max, Min.	TLU	No Convergence
3	12 Acoustical Samples	Smoothed Search Windows and PSD's	K-Nearest Neighbors	Failed
4	9 Coal-Ceiling Samples	Search Windows	K-Nearest Neighbors	89% Accurate
5	10 Acoustical Samples	PSD's for Search Windows	Nearest Neighbor	90% Accurate

FIGURE 8. EXPERIMENTAL RESULTS

that time were used in the training set (i.e., the two unfiltered samples were included in addition to the training set used in the first experiment).

A "maximal" set of features was selected so that if any combination of characteristics could discriminate between coal-seam echoes and non-coal-seam echoes, they would be available to the discriminant functions. Features extracted included smoothed frame values, maxima, minima, derivatives and their maxima and minima, power spectra, spectrogram snapshots, power-spectrum derivative, and various maxima and minima in the frequency domain.

No convergence was achieved with the training set -- indicating a condition of linear inseparability. Stated simply, this means that it is likely that none of the features derived provided consistent information about the location of the desired coal-seam echoes in the training set. A result summary is provided in Figure 8.

3.1.3 "Leave One Out" Method. Validation of a pattern classification system should be made by preclassified samples which were not used in the training set. With a large data base, we divide the available samples into two groups S_1 and S_2. S_1 is used for training and S_2 for testing. With a limited data base, the "leave one out" method is recommended (Lachenbruch, 1968).

This method involves separating one sample out as a test and using all other samples for training. After this test sample has been classified, it is placed back with the other samples and a new test sample removed leaving all others for training. This procedure is repeated until all samples have served as a test exactly once. The accuracy of the system over that limited data base can then be computed.

3.1.4 Using K-Nearest Neighbors. A last attempt was made with the preliminary acoustical data using the K-Nearest Neighbor technique (see 2.4.3). The advantage of this technique is for cases when discriminant functions with a high performance level cannot be identified. For a large set of training samples, it can also be theoretically proved to perform almost as well as the optimal Bayes classifier. Its disadvantage lies in its speed when many samples must be classified in a short period of time. Fortunately this is not our situation. In a real-time system, there would be adequate time to classify a sample while the transducer or radar transmitter were being moved to position it for a new sample.

The meaning of training is slightly different in a K-Nearest Neighbor system. Here we use the representative vectors themselves to define categories and no adjustment of parameters is required before testing can occur.

The features used were smoothed search windows and their power spectrums. A variety of distance metrics were used, including the standard Euclidean n-space metric, but under no circumstances was any success achieved. A summary of the experiment is provided in Figure 8.

3.2 Preliminary Radar Data

Having had little or no success with the acoustics data then available to us, we performed a fourth experiment with the preliminary radar data we had (see Figure 10). The experiment was identical to experiment number three except that the values had to be time-calibrated due to a time-drift during the recording process. Also, a direct search procedure was used to locate the front-surface echoes and, thereby, the search windows.

Success was achieved at last; nine of the coal-ceiling samples were used and using the remove-one-at-a-time technique explained in 3.1.3, eight were correctly classified.

The power spectrum proved useless for classifying these data and the unsmoothed search windows provided the results indicated above. Also, the standard Euclidean metric proved to be an adequate proximity measure. A summary is provided in Figure 8.

3.3 Final Acoustics Data

In early July, we received the final batch of acoustics data (see Figure 11). These samples had been taken under relatively uniform testing conditions. Four samples, however, were from a coal sample of variable thickness and were, therefore, not known to be associated with specific distance categories. In addition, two more samples appeared to have anomalous graphs within the search interval. Mr. Edward J. Drost suggested that this might conceivably be due to over-driving the tape recorder. Thus, of the sixteen samples sent us, we selected ten for testing.

We again used the K-Nearest Neighbor technique, only we let K be one for a simple nearest-neighbor technique. In this case, the test sample is associated with the category of its nearest neighbor. Nine out of ten of these samples were correctly classified when the power spectrum of the raw search interval was used as a representative feature vector.

Again, a standard Euclidean metric was adequate although the raw search windows themselves did not produce any results. It was only when we looked at the vector proximity in the *frequency domain* that the above results were discovered. This is, of course, in marked contrast to the radar experiment (see 3.2) where the power spectrums were useless but the raw search windows yielded successful results. A printed output with a metric exponent of 4.00[1] is shown in Figure 9, and Figure 8 summarizes the experiment.

[1]See Appendix A.

TRAINING SUPERSET:

1
2
3
5
6
7
8
9
11
12

METRIC EXPONENT IS 4.00.

REF PT.	NEAREST PSD
1	1 1/8
2	1 1/8
3	1 1/8
5	1 1/8
6	1 1/8
7	1 1/8
8	1 1/8
9	1 1/2
11	1 1/8
12	1 1/2

FIGURE 9. COMPUTER RESULTS PRINTOUT

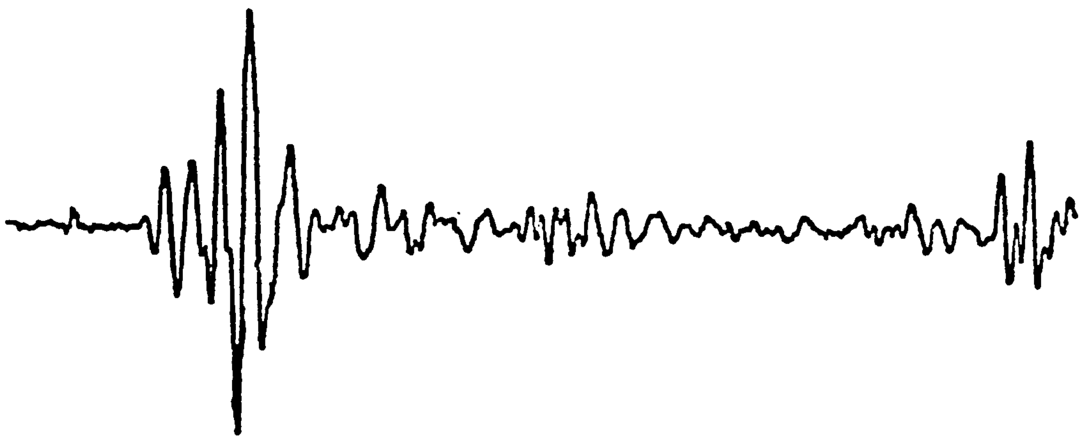


FIGURE 10. SAMPLE OF RADAR DATA (Experiment No. 4)

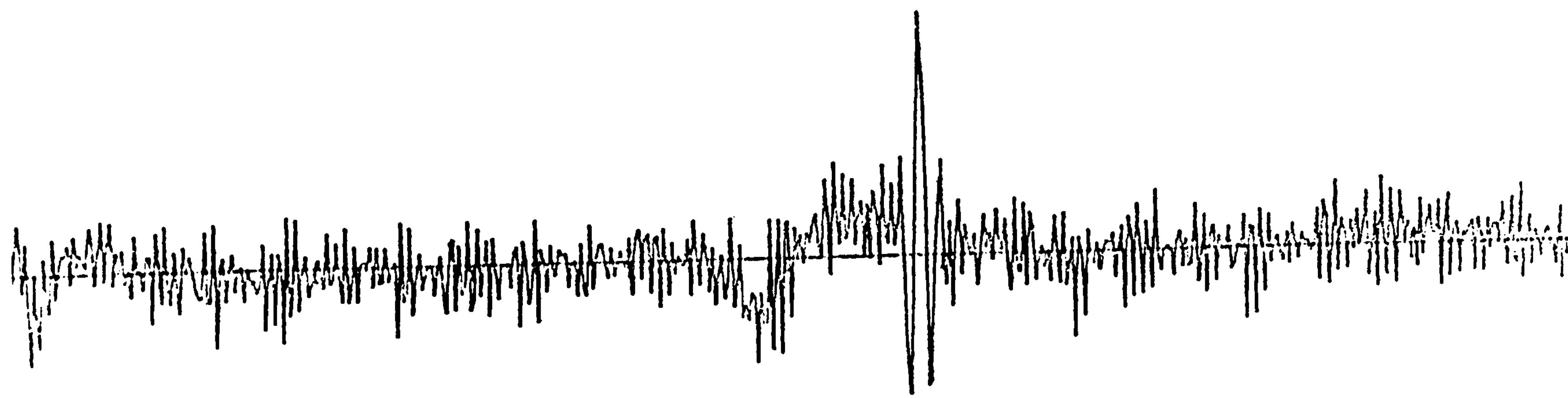


FIGURE 11. SAMPLE OF ACOUSTICAL DATA (Experiment No. 5)

4. EVALUATION OF RESULTS

4.1 Operational Capabilities

The need for uniformity in data acquisition and recording techniques cannot be over-stressed. Only the coal width itself should vary so that the desired information is represented with good consistency over the sample set. Our first three experiments were greatly hampered due to just such a lack of uniformity in collection and recording procedures.

In addition, an exhaustive and balanced training set should be provided for best results. If there are ten width categories, then each width category should be represented by a set of samples which covers the range of possibilities for coal of that width (e.g., variations in the consistency of the coal, or angle of the coal-seam interface, etc.). *This variation should not incorporate any change in test or recording conditions.*

Our experimental results suggest that the K-nearest neighbor technique, in combination with an adequate data base, can be used with a high degree of success to rapidly classify new acoustical or radar samples. Such a system would use a set of training samples for each width resolution within the desired range. Suppose, for instance, we want to know the width of coal to an accuracy of 1/10" and that the acoustical technique was limited to two inches in penetration. Then we would want to have twenty categories between .1" and 2.0" and an additional over-2.0" category. Each category would involve a set of samples in the training data that were exhaustive. In addition, about the same number of samples would be provided for each category. If five samples per category were adequate, then the training set would consist of one hundred five (105) samples.

Since the acoustic signals are not useful for coal widths above about 2", radar is a more promising approach. A practical system could not be expected to be limited to classifying coal less than about 2" in thickness.

4.2 Implementation of Real-Time System

4.2.1 Training. The example data base given at the end of 4.1 could be used as a training set for a K-Nearest Neighbor pattern-classification system. A microcomputer could be pre-programmed and fed these data samples. It would then generate a representative vector for each data sample and associate it with the corresponding thickness provided by the user. All such data could be read in from magnetic tape, paper tape, etc. The system would then be ready to operate on-line.

4.2.2 Operating. Once the microcomputer was trained it could be attached to the digging equipment. Each new analog signal from the transducer could be discretized and input directly to the microcomputer as a data sample to classify. The system would then extract the representative vector and classify the sample using its proximity to the other vectors in the system. The resultant width could be typed out immediately or saved for later dumping.

4.3 Suggested Avenues of Research

Subsequent experiments should be performed with radar or other promising non-acoustic data.

Although the K-Nearest Neighbor system seems to be the most promising for further research, the moving-frame TLU system should be checked out as well. The latter system requires a smaller training set and offers the possibility of continuous width-measurement read-outs.

In addition, it might just also be possible to achieve continuous read-out through the use of regression techniques. This avenue should also be explored.

Additionally, it is suggested that further work utilize samples taken in a coal mine rather than collected in a laboratory. The cracking and drying-out of coal could greatly influence the outcome of further experiments and, in any case, more realistic samples would provide more useful results. The larger the data base the better; we would prefer to work with hundreds of data samples rather than a maximum of sixteen.

Tests for the levels of significance of future experiments can be developed using methods developed from non-parametric tests (Gibbons, 1971).

5. CONCLUSIONS

A pattern recognition system can be constructed to detect coal t ckness using acoustic or radar pulse-echo signals.

As a result of our success with both acoustic and preliminary radar samples, the K-Nearest Neighbor technique appears to be the most promising. The moving-frame TLU system might be further explored, however, since it has yet to be applied to "good" data and would provide continuous depth read-out.

6. BIBLIOGRAPHY

1. Blackman, R.B. and Tukey, J.W., The Measurement of Power Spectra. Dover, New York, 1958.

2. Cooley, J.W. and Tukey, J.W. An Algorithm for the Machine Calculation of Complex Fourier Series. Mathematics of Computation, pp. 297-301, 1965.

3. Cover, T.M. and Hart, T.E. Nearest Neighbor Pattern Classification. IEEE Transactions on Information Theory, V. IT-13, pp. 21-27, 1967.

4. Duda, R.O. and Hart, P.E. Pattern Classification and Scene Analysis, Wiley, New York, 1973.

5. Dudani, S.A. The Distance Weighted K-Nearest Neighbor Rule. IEEE Transactions on Systems, Man, and Cybernetics, V. SMC-6, pp. 325-327, 1976.

6. Gibbons, J.D. Non-Parametric Statistical Inference. McGraw-Hill, New York, 1971.

7. Lachenbruch, P.A. An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis. Biometrics, V. 24, pp. 823-834, 1968.

8. Nilsson, N.J. Learning Machines. McGraw-Hill, New York, 1965.

9. Popoulis, A. Probability, Random Variables and Stochastic Processes. McGraw-Hills, New York, 1965.

10. Rosenfeld, A. Picture Processing by Computer. Academic Press, New York, 1969.

11. Tomek, I.T. A Generalization of the K-NN Rule. IEEE Transactions on Systems, Man, and Cybernetics, V. SMC-6, pp. 325-327, 1976.

12. Young, T.Y. and Calvert, T.W. Classification Estimation and Pattern Recognition. Elsevier, New York, 1974.

APPENDIX A

APPENDIX A

The general metric used in the K-Nearest Neighbor classifier is given below:

$$D(X,y) \quad = \quad (\sum_{\forall i} \; (x_i - y_i)^P) \; 1/p.$$

When p = 2.00, this is just the standard Euclidean metric.