PATTERN RECOGNITION ANALYSIS OF ACOUSTIC EMISSION DATA FROM 7075-T651 ALUMINUM SIMULATED JOINT SPECIMENS

R.B. Melton, P.G. Doctor and D.S. Daly

Battelle, Pacific Northwest Laboratory
Richland, WA 99352

ABSTRACT

The objective of the work described in this paper is to develop signal analysis techniques that can automatically discriminate between non-critical acoustic emission (AE) from crack growth and acoustic noise signals, such as fretting, of fasteners. The ultimate application of this work is for in-flight AE monitoring of critical aircraft structures.

Fatigue crack growth experiments were performed with center notched plate specimens and simulated joint specimens of 7075-T651 aluminum. The experimental conditions were controlled such that acoustic signals were obtained from crack growth, crack interface rubbing, and from fastener fretting.

This paper reports the results of pattern recognition analysis of the signals using autocorrelation lags and statistical measures of the signals and their power spectra as features. The goal of the pattern recognition analysis was to isolate crack growth AE signals from the other acoustic data. The results indicate that autocorrelation lags are the most important features for discriminating these signals.

INTRODUCTION

At least year's Review of Progress we presented data analysis results for data from a plate specimen of 2024-T851 aluminum (Hutton, et al., 1981). We reported a success rate of between 83 and 90% in correctly classifying crack growth acoustic emissions (AE) and noise signals. Since that meeting we have generated data from two simulated joint specimens of 7075-T651 aluminum as described

by Hutton and Lemon (1982). This paper describes Battelle Northwest's analysis of the data from these two experiments, EXPB16 and EXPB17.

Our data analysis effort followed the general approach described previously (Doctor, et al., 1979; Hutton et al., 1981). Briefly, that approach is to analyze independent subsets of a large training set to determine variation in feature selection and decision rule performance. In the analysis of these sets of data we narrowed the approach by focusing on autocorrelation lags as features and by using 1000 data points from the middle of the waveforms rather than the entire 4096 points. The reasons for these changes are discussed in the section describing our analysis of experiment B17.

The first section of this paper describes our analysis of the data from experiment B17. This is followed by the analysis and results for the data from experiment B16. Finally, there is a discussion of the results and our plans for future work. For details on specimen geometry and transducer placement, refer to Hutton and Lemon (1982).

ANALYSIS AND RESULTS FOR EXPB17

Following our standard procedure we first screened the data from sensor P, the transducer closest to the bolt hole and crack, looking for saturated waveforms and waveforms with a low signal-to-noise ratio. During this procedure 45 waveforms were rejected for saturation and 22 for a signal-to-noise ratio below 1.5. In addition, 3 waveforms were discarded because of improper waveform numbers, 4 were discarded for having improper load positions, and the remaining calibration pulses were removed from the data set.

Based on the control of the experiment, the data that passed screening were split into three groups:

Prenotch Fretting (A05211-A05480)	230 events
Prelubrication Crack AE and Fretting (A05563-A05888)	320 events
Postlubrication Crack AE (A05896-A06153)	237 events

The data from the first and last groups were combined to form our training set. The data from the middle group were reserved as an evaluation set.

The first step of our analysis was to examine the load position relationships observed by Hutton and Lemon (1982) during the

experiment. They observed that the AE seemed to be separating on the load position curve by source of emission. We substantiated this observation by looking at load position histograms for our training set data. We found that crack growth AE occurred between load positions 0 and 33, fretting occurred between load positions 36 and 64, and events believed to be associated with crack interface rubbing occurred between load positions 72 and 90. This relationship can easily be seen by examining the plot of load position versus waveform number in Fig. 1. The circular points correspond to the fretting in the first part of the experiment. The triangles correspond to the mixture of crack AE and fretting from the middle part of the experiment, and the crosses correspond to the crack AE from the last part of the experiment. The data in the middle part of the experiment can be sorted by matching with the bands in the other parts of the experiment. The top two horizontal bands seem to be related to crack interface rubbing. This hypothesis is supported by the disappearance of these two bands when the mean load was increased late in the experiment.

The next step in our analysis was to generate features for pattern recognition. We followed the procedure established in our analysis of the data from plate specimens and generated statistical measures of the time domain waveform (including the standard deviation, skewness, and kurtosis), statistical measures of the power spectrum (including the first three moments), the frequency of

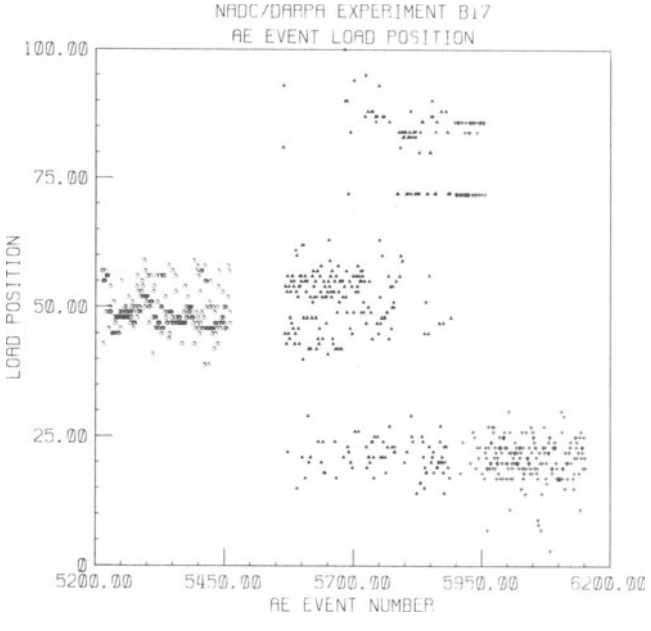

Fig. 1. Load position relationships for experiment B17.

maximum response, the level of response at that frequency, and the first fifty autocorrelation lags (Hutton, et al., 1981). These features were all computed using the first 4000 points of the sensor P waveforms. They were combined into one feature vector for each event for use in feature selection.

In the first feature selection and pattern recognition analysis all of the features were used. During this run features were selected from, and decision rules calculated for, 4 randomly chosen training subsets. Each subset contained 150 events from each of the two categories, crack growth AE and fretting. The events not used for feature selection and decision rule training were used as a blind test set in each case.

In this first analysis autocorrelation lags 46, 40 and 48 were chosen as the most important features. The data in the blind test set were correctly classified at rates from 96 to 100% for the fretting and from 97 to 99% for the crack AE.

In the next step of our analysis we considered all features except autocorrelations at lags greater than 25. We did this to evaluate the effect of using only high frequency information. We were motivated by the possibility that operational noise from an aircraft might extend above 100 KHz but cut off below 200 KHz. We are not, however, advocating excluding the information between 100 and 200 KHz permanently. A decision of that sort will be relevant only to specific applications. In this step we considered 5 training subsets in the same manner as before.

The most important features in this situation were autocorrelation lags 16, 19, and 6. These features were not as highly correlated to the two categories as the low frequency features, but still showed strong performance in classifying the blind test data. For the fretting, the classification rate was 96 to 100%, and for the crack AE, 94 to 98%.

In the first two analysis runs the statistical features from the time and frequency domain and the other frequency domain features were included in the analysis. We found, however, that these features did not show a strong correlation to our data categories in the feature selection process. With this in mind and with an eye toward hardware implementation, we decided to limit our future view to autocorrelation lags as features. We feel that the autocorrelation lags are more practical because of their computational simplicity as compared to the frequency domain features. In addition, they are relatively insensitive to amplitude scaling of the signals because their computation includes normalization by the variance of the signal.

We made another run, then, considering only autocorrelation

values at lag 25 and below to assess the impact of excluding the other features. In this step 2 independent subsets were formed. Not surprisingly, lags 16, 19 and 6 were again chosen as the most important features. The performance did not degrade significantly. In the blind data the fretting was correctly classified at a rate of 98% and the crack AE at 91 and 96%.

The results of this part of our analysis are well summarized by Fig. 2. It shows a cluster plot of the two best features when the entire training set is used for feature selection (i.e., no subset formed). These features are lag 40 and lag 45. The decision rule derived by training a linear discriminant function with these two features is also shown. Seventeen waveforms out of 467 are incorrectly classified for an overall performance of 96%. Also shown are two representative waveforms, one from each category. Notice that the waveforms are very similar, yet are very far apart in feature space. Another thing to notice about the waveforms is that there is a significant amount of noise recorded both before and after the acoustic event.

In the next part of our analysis we decided to reduce the number of waveform points used to calculate the autocorrelation lags. We felt that the pre- and post-event noise that was included in our calculations using 4000 points was in effect obscuring the information contained in the AE event. By limiting the number of points, we hoped to improve the performance of our pattern recognition by homing in on the AE events. We examined the time of peak response for the sensor P waveforms and concluded empirically that if we started at a time index of 850 and included the next 1000 points we would not miss any of the AE events. For the remainder of our analysis we used 1000 waveform data points and calculated only autocorrelation lags as features.

In our first analysis of the shortened waveforms, we used all 50 autocorrelation lags as features, and formed five independent subsets as before. In this case our most important features were autocorrelation lags 40 and 45. They are shown in a cluster plot in Fig. 3. Notice that as with the 4000 point waveforms, the data are well separated. Using the blind test data, the decision rule accuracy ranged from 99 to 100% for the fretting and from 97 to 100% for the crack AE. This performance is slightly improved over the same analysis with the longer waveforms.

In the second run in this part of our analysis we considered only the autocorrelation lags below lag 25 for the shortened waveforms. In this case the best features were lags 15, 19, and 11. This is a slightly different ranking than for the same analysis of the longer waveforms. The performance on the blind data ranged from 98 to 99% for the fretting and 92 to 97% for the crack AE.

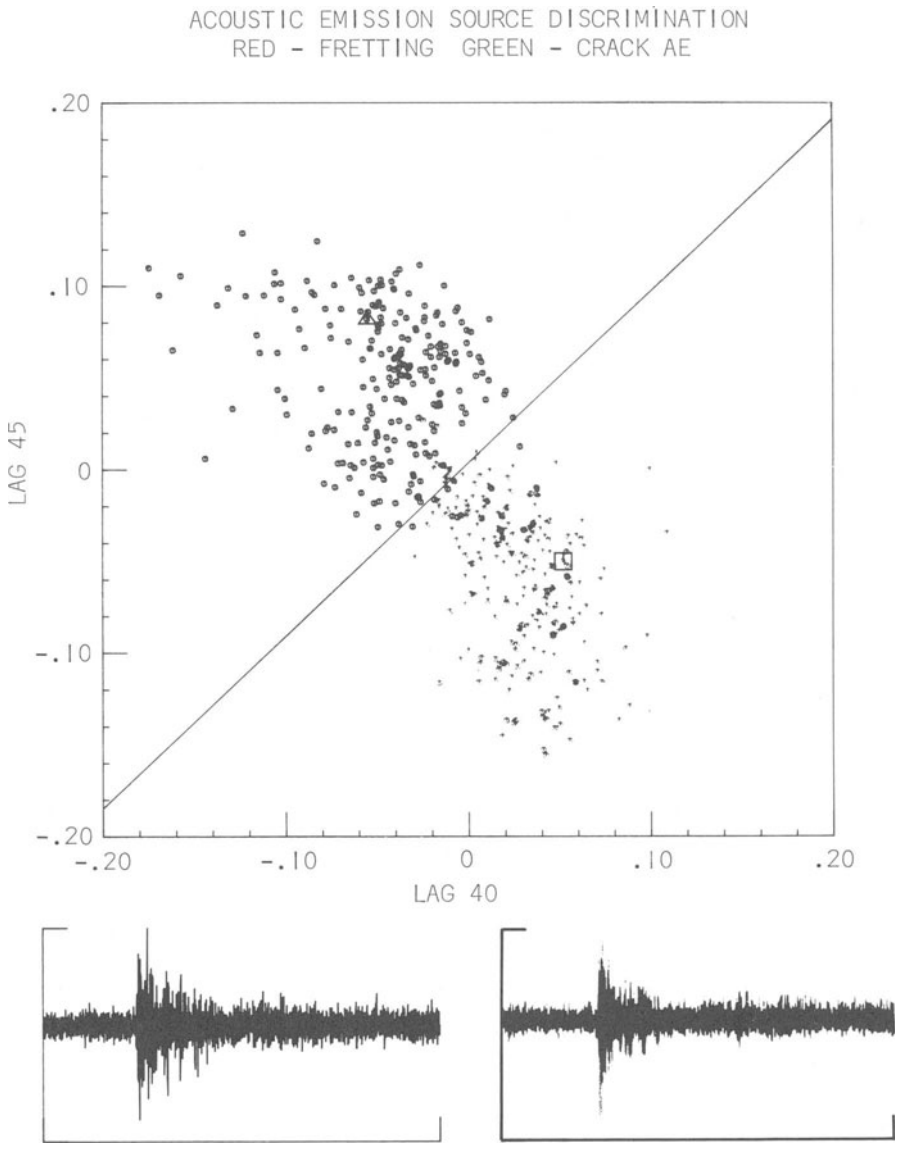

Fig. 2. Cluster plot of best two features for experiment B17 training set. The diagonal line represents the best decision rule. The upper cluster contains fretting data, the lower crack AE. The waveform on the left corresponds to the point marked by a triangle in the fretting. The waveform on the right corresponds to the point marked by a square in the crack AE cluster.

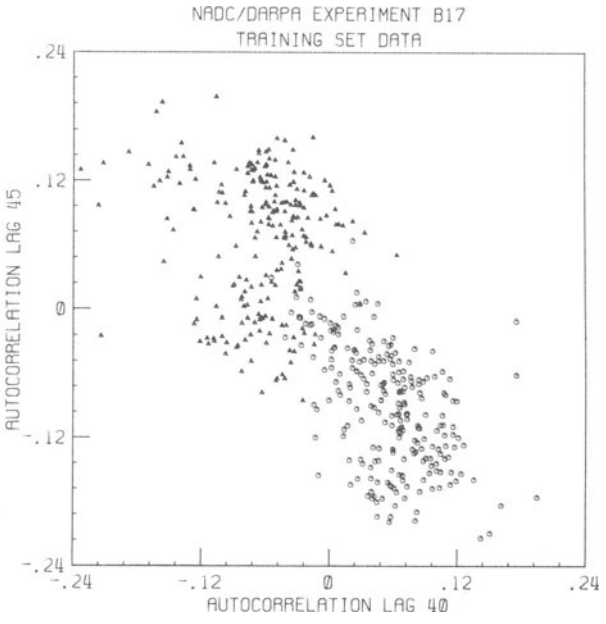

Fig. 3. Cluster plot of two best features for experiment B17 shortened waveforms.

Once again five subsets were considered. These results are slightly higher than the results for the longer waveforms.

In order to assess the robustness of a fixed set of features, we decided to use only lags 40 and 45 to train a decision rule for five independent subsets. This analysis gave good results. The range of performance was 95 to 99% for the fretting and 92 to 94% for the crack AE. This compares favorably with the previous analyses where seven or more features were generally selected for training the decision rule.

Having obtained good results in classifying the blind test data in our subsets, we decided to move ahead and use the 320 waveform evaluation set from the middle of the experiment to test a decision rule calculated using the entire training set. In determining this decision rule, all 50 autocorrelation lags were used as candidates. Eight of the 50 lags were selected as features for the decision rule. In order of importance they were lags 40, 46, 41, 33, 11, 19, 15 and 32.

The evaluation set data were assigned a category based on load position. Those events with a load position between 0 and 38 or 71 and 100 were assigned to the crack AE category while those between 39 and 70 were assigned to the fretting category. The

decision rule performance on this data was mixed. The classification rate was 96% for the fretting but only 71% for the crack AE. In examining the data more closely we found that early load position AE, crack growth AE, was classified correctly at a rate of 60% while the late load position AE, interface rubbing, was classified correctly at a rate of 82%. Further analysis showed that the misclassified crack growth AE was grouped together and came at the initiation of crack growth early in the second part of the experiment. Figure 4 is a cluster plot of the evaluation data. The poor performance on crack growth AE for these features is illustrated by the plot. Notice that the interface rubbing and the fretting form fairly well separated clusters, but that the crack growth AE is distributed throughout the region of feature space that is shown.

The last analysis we did on the B17 data was to look more closely at the crack growth AE and interface rubbing AE. Under the assumption that our category identification by load position is correct, we combined the interface rubbing data from the evaluation and the training sets in order to increase our sample size. The data labeled as crack growth AE in the evaluation set were left as test data. This gave us a sample of 271 crack growth events and 124 interface rubbing events. Then a random subset was formed from the interface rubbing and crack growth AE in the training set such that 100 points from each were retained for training while the remaining points were combined with the crack growth AE in the evaluation set to form the test set.

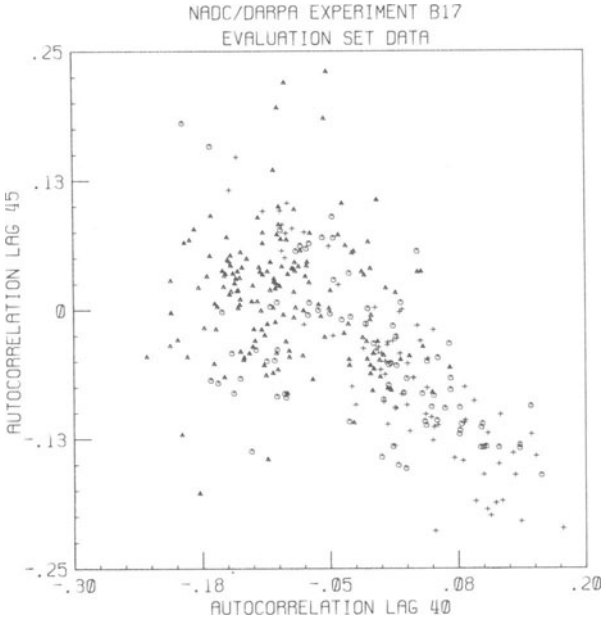

Fig. 4. Cluster plot of experiment B17 evaluation data.

In selecting features for training the decision rule several subsets were formed and lags 12, 16, 21, 45, 49 and 6 came up as the most important features. The performance of the decision rule on the test data was 94% for the crack growth AE and 92% for the interface rubbing.

ANALYSIS AND RESULTS FOR EXPB16

Once again following our standard procedure, we first screened the data from sensor P looking for saturated waveforms and waveforms with a low signal-to-noise ratio. During this procedure 44 waveforms were rejected for saturation and 73 for low signal-to-noise ratio. In addition, 26 waveforms were rejected for improper load positions, and the calibration pulses were discarded.

Based on the control of the experiment, the data were divided into three groups:

Prenotch Fretting (A04331-A04630)	213 events
Prelubrication Crack AE and Fretting (A04631-A04930)	292 events
Postlubrication Crack AE (A04941-A05191)	207 events

The data from the first and last groups were combined to form a training set. The data from the middle group were reserved as an evaluation set.

Following the screening and division of the data, we analyzed the relationship between load position and data type. Unlike the B17 data, there was no clear separation of data by load position. This can be seen by examining Fig. 5. At this point we were interested in using the B16 data to test the B17 decision rule, so we did not pursue this load position relationship.

We proceeded with the analysis by considering the B16 training set as a test set for the decision rule calculated from the entire B17 training set. The results of applying this decision rule to the B16 data were 86% correct classification for the fretting and 58% correct classification for the crack AE. Taking a closer look at these results, we found that 60 of the first 90 crack AE waveforms were misclassified while only 27 of the next 117 crack AE waveforms were missed. This is a lopsided distribution of error that is similar to what we found for the B17 evaluation set. This is contrasted by the errors on the fretting data which were scattered uniformly throughout the data. The poor performance is illustrated by the

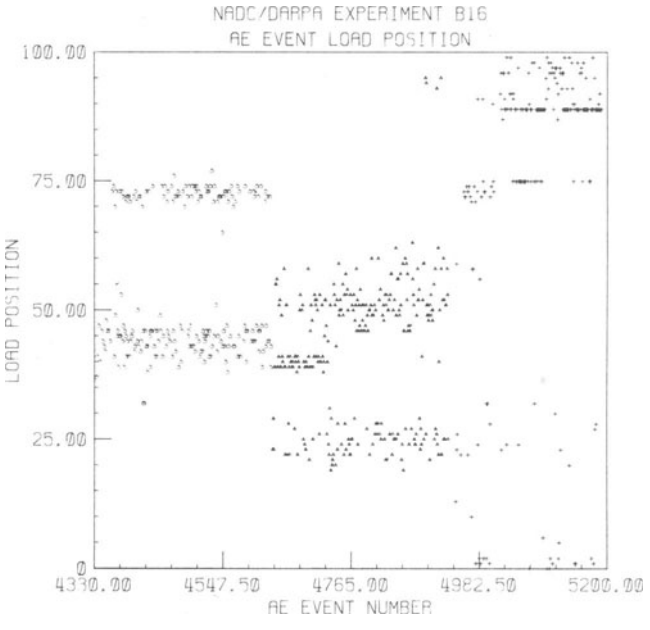

Fig. 5. Load position relationship for experiment B16.

cluster plot of Fig. 6. This is a plot of the B16 training set using the two best features from the B17 analysis. Note that the data do not form nice clusters, though some clustering is evident.

We decided to cross check the labeling of the data as to source by training a decision rule using the entire B16 training set. Our reasoning was that if a significant amount of the data was mislabeled then the decision rule would exhibit poor performance.

The results of this analysis indicated that the data were correctly labeled. The most important features were autocorrelation lags 6, 21, and 48. The classification results were 99% correct for the fretting and 97% for the crack AE. The cluster plot of Fig. 7 supports this result. When the B16 data are plotted with their best two features, the data cluster nicely.

We have not checked the performance of the B17 data using the B16 rule yet. Figure 8, however, shows a cluster plot of the B17 data using B16's two best features. Note that the data do not form good clusters. This leads us to suspect that the decision rule will not perform well. It should be noted, however, that we have in general been using more than two features in our decision rules and that the two-dimensional cluster plots do not necessarily tell the whole story.

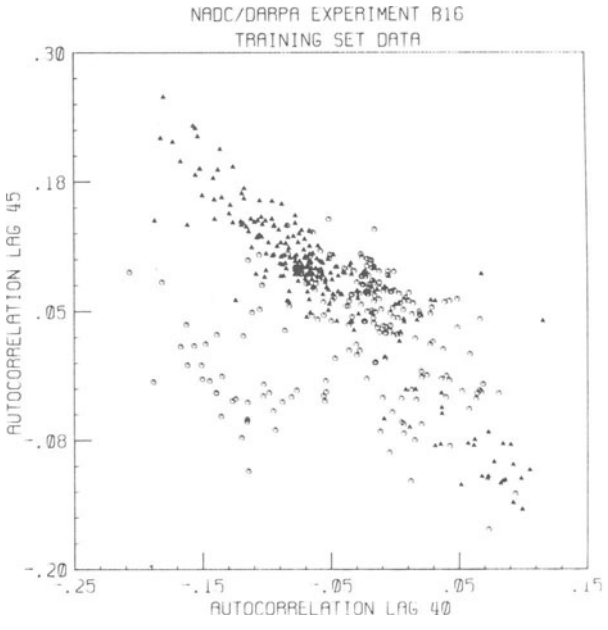

Fig. 6. Cluster plot of experiment B16 data using two best features from experiment B17.

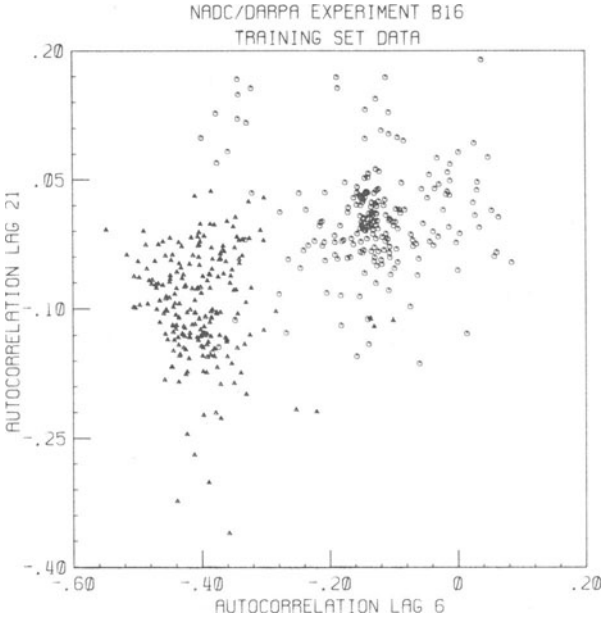

Fig. 7. Cluster plot of two best features for experiment B16 data.

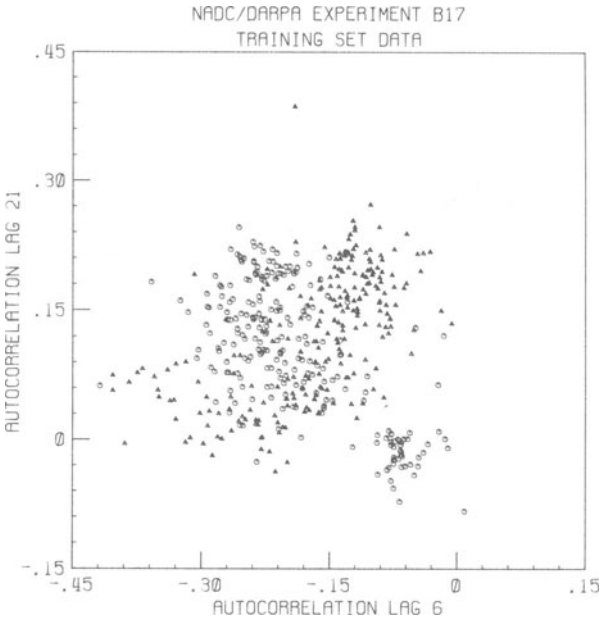

Fig. 8. Cluster plot of experiment B17 data using two best features from experiment B16.

DISCUSSION

In our analysis of the plate specimens we have shown that fretting and crack related AE can be separated for individual specimens. For EXPB17 we have also shown that two distinct types of AE which we believe correspond to crack growth and interface rubbing can be separated. In our limited cross-specimen analysis, the results were encouraging. For the fretting we obtained good results using the B17 decision rule on the B16 data. For the crack AE the story is more complicated and will receive more attention. We seem to be observing some type of time dependence in the crack AE data. It is not clear whether the source of this variation is related to the experiment or to the instrumentation used to conduct the experiment. There are various possibilities, one of which is crack length dependence. This could manifest itself as a change in the basic characteristics of the AE or as a geometrical change in signal paths.

FUTURE PLANS

As was mentioned agove, these results are very encouraging. On the individual joint specimens they are the best yet. The work is not complete, however. We feel that there are three main areas to cover in our future work. The first is to analyze data across

experiments to develop decision rules that perform well on more diverse data. In this process we expect to find a set of common features that do not give the best performance on individual data sets, but that give the best performance across data sets.

The second area involves the time dependence we noted in the crack AE data. This came out as a weak area for our decision rules and we will therefore be working to overcome this problem. We suspect that it is a matter of using a more representative sample for our training set. We will also be considering time dependence exhibited in the B16 load position plot of Fig. 5. Before we can make good use of the data from the middle part of the experiment, we must be able to assign a category to it.

The third area involves the development of more robust and practical decision rules. We plan to explore the use of decision rules that are dependent on the structure of our data. A rule of this type might give better performance in the presence of a class of data that was not available in the training set than a simple linear decision rule would.

REFERENCES

1. Doctor, P.G., Harrington, T.P. and Hutton, P.H., 1979, Pattern Recognition Methods for Acoustic Emission Analysis, NUREG/CR-0910, PNL-3052, Battelle, Pacific Northwest Laboratory, Richland, WA.
2. Hutton, P.H., Lemon, D.K., Melton, R.B. and Doctor, P.G., 1981, Develop in-flight acoustic emission monitoring of aircraft to detect fatigue crack growth, Review of Progress in Quantitative NDE, Vol. 1, D.O. Thompson and D.E. Chimenti, eds., Plenum Pub. Corp., New York, 1982.
3. Hutton, P.H. and Lemon, D.K., 1982, Development of acoustic emission methods for in-flight monitoring of aircraft structures, Review of Progress in Quantitative NDE, Vol. 2, D.O. Thompson and D.E. Chimenti, eds., Plenum Pub. Corp., New York, 1983.