Pattern Recognition System for Automatic Identification of Acoustic Sources

R. H. Cabell* and C. R. Fuller†
Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

An intelligent recognition system was designed using pattern recognition techniques to distinguish the noise signatures of five different types of acoustic sources. Information for classification was calculated from the power spectral density and autocorrelation taken from the output of a single microphone. The system included a training step where it learned to distinguish the sources and automatically select descriptive quantities for optimal classification performance. Information learned in training was stored and used to identify recordings of the five types of sources presented during testing. Results of testing indicate that the current optimal design could correctly identify 90% of the recordings. Identification of noise corrupted signatures and identification of recordings not used in training is discussed.

Nomenclature

x = pattern vector
$d(x)$ = general decision function
$d_i(x)$ = decision function of ith class
w = weight vector containing coefficients of decision function
ω_i = class i

Introduction

THE number of commercial air flights has been growing steadily in recent years, and with that growth comes an increased probability of annoyance from air traffic. Traffic noise measurements that establish noise guidelines and enforce guidelines can be time consuming and costly. Unsupervised measurements of air traffic noise are subject to interference from ground sources, which are not meant to be included in the analysis. An automatic identification system could simplify the monitoring of noise pollution by identifying the source and recording only relevant noise sources or pinpointing sources that violate noise thresholds.

An identification system is made up of a sensor for making physical measurements, a preprocessor, and a classifier. During training, the decision logic of the classifier is adjusted using pattern recognition techniques and a set of known training patterns. The information from training is then stored for later use in classifying unknown patterns.

Pattern recognition systems are used in a variety of problems, from optical character scanners to medical diagnosis systems.[1,2] Acoustic pattern recognition systems can be found in speech recognition, nondestructive evaluation,[3,4] and source identification.[5,6] Examples in nondestructive evaluation monitor acoustic emissions of reactor feeder pipes[3] and structural noise[4] to detect cracks or flaws in equipment without disturbing equipment operation. Source identification systems are usually limited in the number and variety of sources they can identify. One such system was designed to identify different helicopters[5] using analog techniques. A noise pollution identification system, described in Ref. 6, is limited to identifying only three types of sources.

Received May 9, 1989; revision received Jan. 2, 1990.

*Research Associate, Department of Mechanical Engineering. Member AIAA.

†Professor, Department of Mechanical Engineering. Member AIAA.

This paper describes a system designed to classify a noise signature as being either a jet plane, propeller plane, helicopter, train, or wind turbine.[7] This application evaluates the techniques on acoustic sources, where further work will be concerned with more sophisticated systems that identify the particular types of sources, i.e., the type of plane or helicopter. The pattern recognition theory is described, along with the variations of that theory tested for this application. Results are given as the percent of test sources correctly identified for each test classifier.

Pattern Recognition Theory

This section contains an overview of the pattern recognition theory. Terminology is taken from standard literature on the subject.[1,8] In this discussion, class refers to a group of patterns that have similar characteristics, whereas a classifier assigns unknown patterns to known classes or categories.

The techniques discussed in this paper are based on cluster generation and separation. Patterns are described by features, which are assembled into a feature vector. This feature vector locates a point in feature space corresponding to the pattern. Clusters of similar patterns are formed when the features are well chosen, and these clusters will be separate from clusters of dissimilar patterns. Decision logic can then be used to separate the clusters with a decision surface so that later patterns can be classified based on their position relative to that surface. Figure 1 illustrates two groups of patterns in two-dimensional space, separated by a decision surface.

Decision surfaces were used as decision logic for this application. A decision surface is described by a decision function $d(x)$. For a two class problem, a pattern x is classified as in Eq. (1),

$$x \in \omega_1 \text{ if } d(x) > \quad \text{or} \quad x \in \omega_2 \text{ if } d(x) < 0 \tag{1}$$

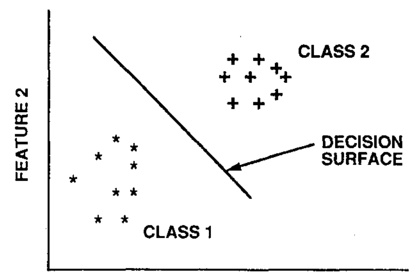

Fig. 1 Example of class clusters with decision surface.

Although there are different methods of generating decision functions, distance and likelihood methods are two of the more commonly used theories.[1,3] Distance functions use the location of an unknown pattern relative to known patterns for classification. If the unknown lies closer to members of class 1 than class 2, then it is classified into class 1. These methods require well-ordered clusters where proximity is a good measure of class membership. Likelihood functions use the statistical properties of each class to classify unknown patterns. A pattern vector is classified into the most likely class based on the class distributions. However, unless the patterns are noisy variations on a standard prototype, a classical distribution such as Gaussian, binomial, and Poisson may yield poor results.[9]

We decided the decision surfaces for this problem would be calculated using weighted linear combinations of the features. The equation of such a decision surface is

$$d(\boldsymbol{x}) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_{n+1} = \boldsymbol{w}^T \boldsymbol{x} \tag{2}$$

The decision function generates hyperplanes in n space. Using a decision surface as in Eq. (2) requires no assumption about the underlying class distributions or their shape in n space, but does assume the clusters are separable by a linear surface.

Classifier training requires finding the weighting coefficients using a training algorithm such that the clusters of training patterns are separated by the surface. Finding the best location of the decision surfaces is simple in a low-dimensional space, but as more features are added and the space becomes multidimensional it is impossible to visually place the surfaces. The training algorithm eliminates this problem by cycling through a set of training patterns until separation surfaces are found. If the training patterns are poorly chosen and are not representative of the class populations, then the surfaces will incorrectly classify class members in general. On the other hand, a well-chosen, representative set of training patterns can ensure good performance on other class members.

The selection of training patterns is important, but the selection of features that describe those patterns is just as important. Well-chosen features produce separated class clusters, simplifying the calculation of the decision functions.[9] Clusters that are close or overlapping may not be separable by a decision surface, and so the more distinct the clusters, the better the chance that a decision surface will separate them. The real power of a classifier, regardless of the type, comes from carefully chosen features.[10]

When a large list of features is generated, a selection process must be done to choose only the most descriptive features. Although it is usually simple to generate a list of features that might be useful for a given problem, classifier accuracy is reduced when the list is too large and includes redundancies. Too many features can impose artificial constraints on the data, causing training patterns to be classified correctly but real data to be classified incorrectly.[11] As a result, in most problems it is necessary to use a feature selection scheme to reduce a large list of features to a small list of good features.

System Design

This section describes the design of the system used to identify acoustic sources. Linear decision surfaces were used to separate the class clusters. These surfaces were calculated using a training algorithm and a training set of patterns, and a feature selection scheme was used to select a small set of good features from a larger list of features.

Classifier Design

The classifier used linear surfaces, described in Eq. (2), to separate the class clusters. For the case of multiple classes, the surfaces can be placed in different ways relative to the classes,[1] but we chose to use one decision function for each class. A pattern was classified by evaluating each decision function using Eq. (2). The resulting scalars were compared, and the pattern was assigned to the class whose decision function had the highest value. In equation form,

$$\boldsymbol{x} \in \omega_i \text{ if } d_i(\boldsymbol{x}) > d_j(\boldsymbol{x}) \qquad \text{for all } j, j \neq i \tag{3}$$

The equation of the surface separating classes i and j is found by equating their decision functions, $d_i(\boldsymbol{x}) = d_j(\boldsymbol{x})$.

Training Algorithm

The training algorithm was a gradient descent method, known as the perceptron algorithm.[1] The algorithm cycled through the training patterns, classifying each one using the stored coefficients, making adjustments to the coefficients when a pattern was misclassified. At the kth iteration, a pattern $\boldsymbol{x}(k)$ belonging to class ω_i was presented to the classifier. For M classes, each of the M decision functions were evaluated according to Eq. (2). The resulting scalar values were compared, and if

$$d_i[\boldsymbol{x}(k)] > d_j[\boldsymbol{x}(k)] \qquad j = 1,2,\ldots,M, j \neq i \tag{4}$$

then the decision functions classified $\boldsymbol{x}(k)$ correctly and no changes were made to the coefficients. Training continued with $\boldsymbol{x}(k+1)$, the next pattern in the training set. If, however, for some l,

$$d_l[\boldsymbol{x}(k)] > d_i[\boldsymbol{x}(k)]$$

then the decision functions classified $\boldsymbol{x}(k)$ incorrectly, and the weight coefficients of the ith and lth classes were adjusted according to

$$\boldsymbol{w}_l(k+1) = \boldsymbol{w}_l(k) - \boldsymbol{x}(k) \tag{5a}$$

$$\boldsymbol{w}_i(k+1) = \boldsymbol{w}_i(k) + \boldsymbol{x}(k) \tag{5b}$$

and the other weight vectors were left unchanged,

$$\boldsymbol{w}_j(k+1) = \boldsymbol{w}_j(k) \qquad j = 1,2,\ldots,M, j \neq i, j \neq l \tag{5c}$$

Equations (5a) and (5b) adjust the weight vectors by an amount equal to the pattern vector that was classified incorrectly. The algorithm is guaranteed to find the coefficients of the separation surfaces, provided the classes are separable.

If the classes are overlapping, all of the training patterns were never classified correctly, and so training was stopped when only a small percentage of patterns were classified incorrectly. The weight vectors were stored at the end of training for classifying unknowns.

Feature Selector

A feature selection scheme was included to select a small set of good features from a larger list. The selector eliminated the difficult task of prespecifying an optimal set of features for the problem. The feature selection algorithm was originally designed for a dynamic signature verification system, where a handwritten signature is classified as valid or forged.[11] It can be applied to any problem since it selects the features using the training patterns and the classifier designed for a particular problem. This section includes only a brief summary of the algorithm; interested readers are encouraged to see Ref. 11.

Sets of features were evaluated by constructing classifiers using the training patterns, then testing the classifiers using a different set of patterns, called an evaluation set. A good set of features resulted in a good classifier, which, in turn, had a low error rate on the evaluation patterns. Different combinations of features were tested, and the combination that had the lowest error rate on the evaluation set of patterns was saved as the best one. All possible combinations of features were not tested, and so the chosen set was not necessarily globally optimal. However, the amount of time required for a truly optimal search makes it impractical compared to the results of a suboptimal search such as the one just described.[2,11]

Although time consuming, the search routine was necessary to evaluate the features as sets in the classification system. A statistical test could be used to rank the features in order of discriminatory power, but the features must be evaluated as sets, not only as individual pieces of information.[11]

Structure of Classification System

An important aspect of any classification system is the number of features vs the number of classes. As the number of classes increases and the number of features remains constant, the features must become more descriptive, therefore, more precise. In a complex problem, the number of classes can become too large, requiring an impractical feature precision, resulting in classification errors.[10] In order to study this problem, a multilevel classification structure was used, shown in Fig. 2. This figure shows a hierarchical arrangement of four two-class classifiers ranging from general at the top to more specific at the bottom.[10,11] At any given level, a pattern is identified as belonging to one of two classes. The performance of this tree classifier was compared to that of a single level classifier, which identified an unknown as belonging to one of the five classes without going through intermediate classifications.

Each of the classifiers in the tree structure was trained with a different set of training patterns, depending on the sources to be identified. For example, the train-wind turbine classifier was trained with wind turbine and train recordings only, whereas the ground-aerospace classifier was trained with recordings of all five sources labeled as either ground or aerospace.

Testing of System on Acoustic Sources

The recognition system was tested using magnetic tape recordings of the sounds made by the sources. The signal from the tape recorder was used to simulate the output of a single microphone. This study evaluated recognition of the sources using acoustic information alone, and so location, speed, and seismic information were not considered.

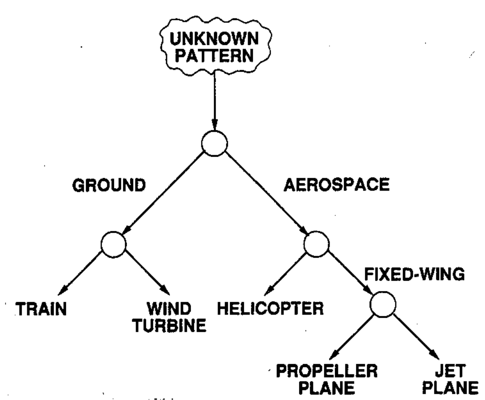

Fig. 2 Tree classification structure.[5]

Table 1 Source recordings for training and testing

Source type	Description	Total number of recordings
Jet	9 commercial airliners, 1 military jet	92
Propeller plane	10 types, takeoff and landing	28
Helicopter	1 light-duty helicopter	14
Train	1 empty coal train	18
Wind turbine	Downwind recordings	20

Training and Testing Data Base

Source recordings of sounds of jet planes, propeller planes, a helicopter, wind turbine, and train were collected to generate a data base for testing. All recordings were of typical operation of the sources; plane takeoff and landings, the helicopter in flight, train rolling by, and the wind turbine being driven by the wind. A list is given in Table 1.

Signal Processing

The frequency range for processing the recordings was selected to convey adequate recognition information on all the sources. Identification information is found below 5 kHz,[6] but for this study the range was limited to 2.5 kHz to allow resolution of the low-frequency wind turbine noise.

The recordings were digitized at 7812.5 Hz then split into segments of 4096 points for calculating the power spectral density (PSD) and autocorrelation. A window was not applied to the segments.

Calculated Features

The acoustic signatures of aerospace vehicles are usually produced by a combination of engine and body noise exciting a resonant structure in a periodic or semiperiodic manner. This type of noise is well described by the distribution of energy in the frequency domain.[12]

Representative spectra, such as those shown in Figs. 3a and 3b, were examined to generate a list of candidate features. The spectra in Fig. 3 illustrate the differences in energy density between the jet and the wind turbine. The energy of the jet is spread more evenly throughout the spectrum, compared to the wind turbine. Most of the features were calculated to describe the relative amounts of energy in different bands of the spectrum in order to describe the differences apparent in Fig. 3.

The features that best described the differences were not prespecified; instead, the feature selector made this selection. The selector was given a list of 108 features, which we felt had

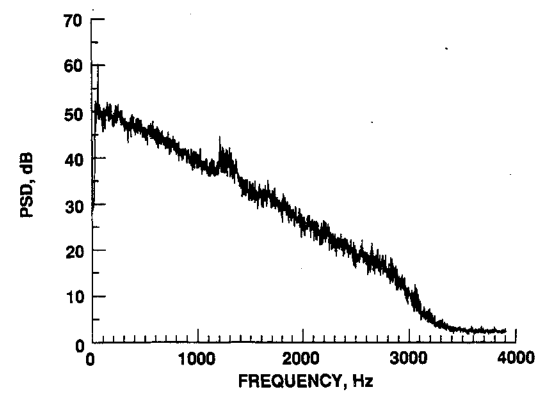

Fig. 3a Example of a power spectral density of jet plane noise.

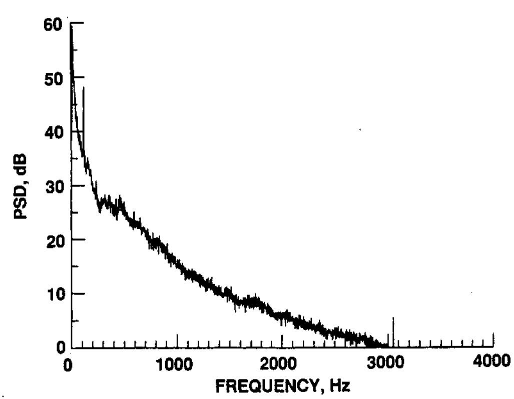

Fig. 3b Example of a power spectral density of wind turbine noise.

some use in classification, then the most useful features were selected using the method just described. Of the 108 initial features, 106 were calculated from the frequency domain, the remaining 2 from the autocorrelation. The frequency domain features were partial powers, accumulated power points, and ratios of partial powers.[3] The partial power features described the percent of total energy present in some small bandwidth of the spectrum. For example, one such feature might have described the percent of energy from 1000–2000 Hz present in the 1000–1100 Hz bandwidth. The accumulated power points specified the frequency at which a percentage of the power in a large band was located, when accumulated up from the bottom of the band. An example of one of these features is the frequency location of 75% of the energy in the 1000–2000 Hz band, where 0% energy is at 1000 hz and 100% energy is at 2000 Hz. The ratios of partial powers described the slope of the spectrum, in terms of energy. The two autocorrelation features were the mean value and variance, calculated over a short time interval.

All of these features describe the energy content of the spectrum. The partial powers describe small variations in a bandwidth, the accumulated powers describe more general variations, and the ratios describe the most general variations in energy distribution. The drawback to all of these descriptors is the redundancy among them. This redundancy was included on purpose because we preferred to give too much information rather than leave out information. As discussed in the results section, this redundancy may have reduced the performance of the feature selector.

The recordings were processed through an A/D converter then loaded into a computer for feature calculation and classification. Figure 4 shows the hardware setup used for processing the recordings. The recorded signal was monitored on the oscilloscope to trigger data acquisition, and the filters were used to reduce aliasing.

Training Patterns

Once the source recordings were reduced to feature vectors, they were assembled into training and testing sets of data. These sets were tailored to the classifier being trained. The single level classifier required only one training set containing all five sources, but the tree structure required four training sets for the four two-class classifiers.

Ideally, the training and testing sets would be made of different recordings to validate the performance on more general data,[1] but the limited number of source recordings meant the training and testing sets could not be totally independent. In particular, the number of helicopter, train, and wind turbine recordings meant the same recordings had to be used for training and testing. However, the training and testing sets were processed separately with different triggering times so they did not contain identical digitized segments.

There were enough plane recordings to test the system with ones not used in training. Ten different types of jets were available with multiple recordings of each one. The propeller planes were recorded on both takeoff and landing, so the classifier was trained with takeoff recordings only and then tested on its ability to identify both the takeoff and landing recordings.

Results

We used the percent of test sources identified correctly as a measure of the recognition performance. A recording was digitized, features calculated, decision functions evaluated, then identified, either correctly or incorrectly. Results were tabulated over all test sources, giving the recognition rate, which describes the percent of test sources correctly identified.

Four classifier variations were tested, consisting of two configurations of decision logic and two methods of feature calculation. A tree and single level arrangement of the decision logic were tested, along with two methods of calculating the spectral energy for the partial power features. The first

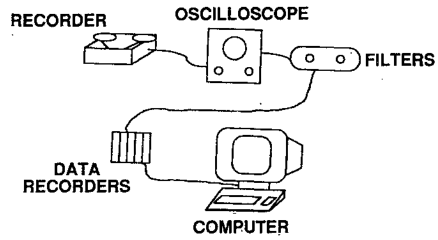

Fig. 4 Processing hardware.

method, called linear summation, summed the dB values linearly. For example, the average of 23 and 30 dB was calculated as $(23 + 30)/2$, giving 26.5 dB. The other method, called log summation, calculated the average using mathematically correct log summation. Using the same example as above, the average of 23 and 30 dB was calculated as 27.8 dB.

It was hypothesized that using partial powers, linear summation would be more useful for describing the wind turbine spectra than log summation. A large amount of energy at low frequencies would not affect linear summation features as drastically as log summation features. For example, a partial power feature might describe the percent of energy in the 0–1000 Hz band present in the 0–100 Hz band. Whereas the log summation feature for this band might have a value of 95% due to the large concentration of energy, a linear summation feature might have a value of only 50 or 60%. Only 5% of the energy is left to describe nine other bands using log summation features, whereas 40–50% of the energy is left using linear summation features.

Combining the different structures and feature calculations gave four basic classifier variations for testing. The variations are denoted by different symbols in the plots of results. In addition to the variations in structure and features, each basic variation was trained and tested using four to ten features. It was hoped that recognition rates would rise as more information was added. In subsequent sections of the paper, the term classifier is used to refer to each variation of structure and feature calculation with a given number of features. Hence, the tree structure with linear summation features produced seven classifiers: one with four features, another with five features, and so on, through 10 features.

It should be noted that feature selection for the different classifiers was done independently of one another. For instance, four features were selected for the tree structure/linear summation classifier. For the same classifier with five features, feature selection was begun anew, not from the four features just selected. Whether the five features were identical to the four plus an additional one was dependent on the feature selection algorithm.

Finally, note that a pattern would be classified into one of the five classes even if that pattern did not belong to any of the five classes. A car would be classified as either a jet, train, propeller plane, etc., depending on its feature vector. A more advanced system would have an unidentified choice for such cases to reduce misclassifications.

Recognition of All Sources

Figure 5 shows the recognition rates on all sources for each classifier as a function of the number of features. The symbols show the percent of test sources identified correctly. For example, the classifier with a single level structure, linear feature summation, constructed using four features, identified 66% of the sources correctly, and with five features identified 69% of the test sources correctly.

Recognition rates were similar for the tree structure classifiers and classifiers with a tree structure and log summation features. However, classifiers with single level structure/linear summation features had consistently lower recognition rates, rising slightly as the number of features increased. The best classifier was the tree structure/log summation with eight features, identifying over 90% of the test sources correctly.

None of the classifiers showed a consistent rise in recognition rate as the number of features was changed, indicating that the new features did not necessarily add more information. This was a result of the suboptimal feature selection process, which selected a good set of features, but not necessarily the best set of features. Feature selection was also hampered by the large number of highly correlated features because all were calculated from the same PSD. This combination of suboptimal selection and correlated features meant it was difficult for the algorithm to select some features as being better than other features. In these cases, the final set was related to the initial ordering of features in the candidate list.

Recognition of Jet Planes

The recognition rates on the jet recordings, shown in Fig. 6, were similar to the recognition rates on all sources. As with all sources, the tree structures and the single level/log summation classifiers had similar recognition rates. The tree structure classifiers with four features identified 95% of the test sources correctly, as did the single level/log summation classifier with nine features. Again, the single level/linear summation classifier had lower recognition rates than the other classifiers.

The jet class contained many different types of jets, including multiple recordings of nine different commercial jets and one military jet, all used in training and testing. The high recognition rates indicate that the system could identify variation if trained with that variation.

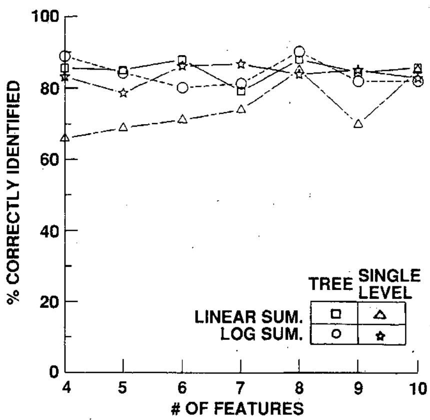

Fig. 5 Recognition of all sources.

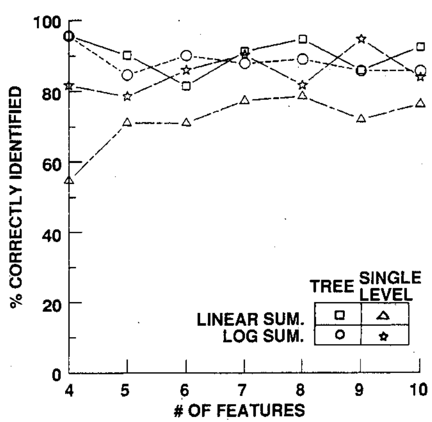

Fig. 6 Recognition of jet planes.

Recognition of Propeller Planes

The recognition of propeller planes, shown in Fig. 7, was worse than the jets, varying dramatically as the number of features changed. However, the single level/linear summation classifier did improve consistently as the number of features increased, up to 100% correct with 10 features.

Differences between the training and testing set of patterns may have caused the lower recognition rates. Unlike the jets, all class variation was not included in the training set, but was included in the testing set. For training, recordings of planes taking off were used, however, for testing, recordings of planes taking off and landing were used. The differences in sound between the two caused the system to misidentify some of the propeller planes. Table 2 shows the differences in recognition for the take-off recordings and the landing recordings. The recognition rates in Table 2 were obtained by summing the rates for each type of classifier across all numbers of features.

All classifier types did poorly on the recordings not used in training, although those using linear feature summation showed less of a drop in recognition rates. The energy shift due to the change in operating conditions affected the linear summation features less than the log summation features. If a

Table 2 Recognition of propeller planes: takeoff and landing

	Percent correct	
Classifiers	Takeoff	Landing
Tree, linear	81	66
Tree, log	91	37
Single level, linear	90	80
Single level, log	94	41

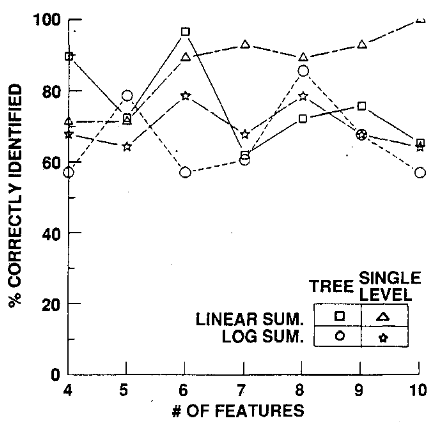

Fig. 7 Recognition of propeller planes.

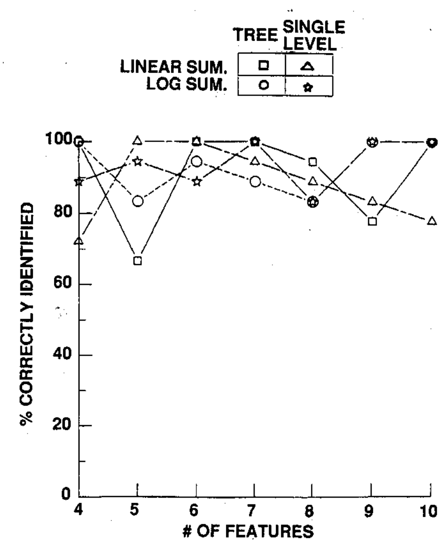

Fig. 8 Recognition of train.

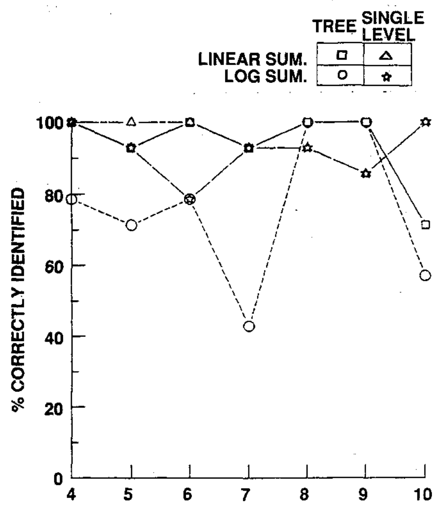

Fig. 9 Recognition of helicopter.

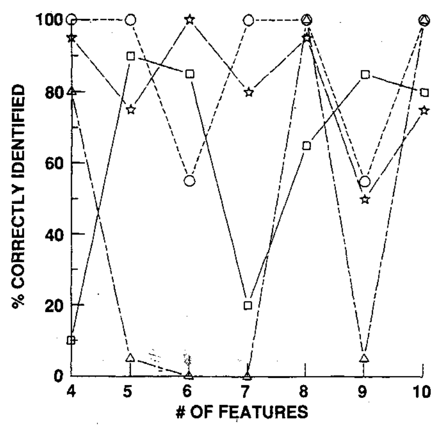

Fig. 10 Recognition of wind turbine.

tone shifted location slightly in the spectrum, the log summation features would change more dramatically than the linear summation features.

Recognition of Helicopter, Train, and Wind Turbine

The small number of recordings of the train, helicopter, and wind turbine negatively affected the recognition rates for these classes. A large number of recordings is desirable both for calculation of good decision surfaces and for testing the performance of the decision surfaces. The equation of a surface separating n points can be found quite readily when n is less than twice the dimensionality of the surface.[1] In a 10-dimensional feature space, 20 patterns or more should be used to generate each class cluster. This was not possible for the three classes mentioned previously, and so the decision surfaces may not classify correctly other members of these classes.

The recognition of the train recordings, shown in Fig. 8, was 100% for several classifiers with different numbers of features, dropping to an overall low of only 65%. In fact, in all but two cases, the classifiers identified over 75% of the trains correctly.

The recognition of the helicopter recordings, shown in Fig. 9, was similar to that of the trains, as several classifiers with different numbers of features identified 100% correctly. However, recognition rates varied more as the number of features changed, especially for the tree classifier using log summation features, ranging from 40% correct identification to 100%.

The wind turbine recordings had the lowest signal to noise ratio of all the classes, which may have accounted for the inconsistent recognition, shown in Fig. 10. The only noise generated by the wind turbine was blade passing noise in the presence of wind noise. This noise reduced the effectiveness of some sets of features, such as those used by the single level/linear summation classifier with six and seven features. However, the log summation features seemed to have been more resistant to the noise than the linear summation features. The recognition rates of the linear summation feature classifiers varied from $<10\%$ correct to 100%, whereas the rates of the log summation feature classifiers varied from a low of 50 to 100%.

Conclusions

A study was undertaken to determine the ability of an automatic pattern recognition system to identify acoustic sources. The system was trained with recordings of five source classes: jet planes, propeller planes, a helicopter, train, and wind turbine. The system was tested using similar recordings of the sources, classified either correctly or incorrectly. Conclusions on the performance of the system are as follows.

1) The linear discriminant functions identified up to 90% of the test sources correctly. Given a proper set of training data, these simple functions can be used for identifying acoustic sources.

2) The current feature selection algorithm did not select the globally optimal set of features, as evidenced by the widely varying recognition rates as the number of features changed. In addition, an optimal number of features could not be determined from the results of testing.

3) A class that contained significant variation, such as the jet planes, could be successfully identified if the training sets also contained that variation. If the training sets did not contain that variation then the recognition rate dropped. Propeller planes landing could not be identified as successfully as the same planes taking off because only take-off recordings had been used during training.

4) There was no discernible difference between the performance of the tree and single level structures on the test recordings. However, the arrangement of the tree structure has implications in the recognition of more classes.

5) The two methods of calculating features gave similar overall recognition performance, although each method did offer an advantage for particular cases of noise and variations in operating conditions.

Acknowledgments

The authors are grateful to the NASA Langley Research Center for its support for this research under Grant NAG-1-762.

References

[1]Tou, J. T., and Gonzalez, R. C., *Pattern Recognition Principles*, Addison Wesley, Reading, MA, 1974.

[2]Moser, J. M., and Aunon, J. J., "Classification and Detection of Single Evoked Brain Potentials Using Time-Frequency Amplitude Features," *IEEE Transactions on Biomedical Engineering*, Vol. BME-33, No. 12, 1986, pp. 1096–1106.

[3]Chan, R. W. Y., and Hay, R. D., "A Case Study of Sensitivity of Some Pattern Classifiers used in Sorting Acoustic Emission Signals," *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 3a, 1982, pp. 413–420.

[4]Horvath, P., and Cook, F. J., "Establishing Signal Processing and Pattern Recognition Techniques for Inflight Discrimination Between Crack-Growth Acoustic Emissions and Other Acoustic Waveforms," *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 3a, 1982, pp. 463–473.

[5]Hemdal, J. F., "An Investigation of Automatic Identification of Acoustic and Seismic Sources," Willow Run Laboratories Institute of Science and Technology, Ann Arbor, MI, CR No. 86, Feb. 1973.

[6]Moukas, P., Simson, J., and Norton-Wayne, L., "Automatic

Identification of Noise Pollution Sources," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-12, No. 5, 1982, pp. 622–634.

[7]Cabell, R. H., "The Automatic Identification of Aerospace Acoustic Sources," M.S. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, Feb. 1989.

[8]Bow, S., *Pattern Recognition: Applications to Large Data-Set Problems*, Dekker, New York, 1984.

[9]Bremermann, H. J., "What Mathematics Can and Cannot Do for Pattern Recognition," *Pattern Recognition in Biological and Technical Systems*, edited by O. Grusser and R. Klinke, Springer-Verlag, New York, 1971, pp. 31–45.

[10]Chandrasekaran, B., and Goel, A., "From Numbers to Symbols to Knowledge Structures: Artificial Intelligence Perspectives on the Classification Task," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, No. 3, 1988, pp. 415–424.

[11]Klassen, G. S., "A General Purpose Reduction Intensive Feature Selector for Pattern Classification," M.S. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, June 1986.

[12]Thomas, D. W., "Vehicle Sounds and Recognition," *Pattern Recognition: Ideas in Practice*, edited by B. G. Batchelor, Plenum, New York, 1978.