# Application of Linear Discriminant Analysis to Doppler Classification

**M. Jahangir**
QinetiQ
St Andrews Road, Malvern
WORCS, UK, WR14 3PS
United Kingdom

mjahangir@qinetiq.com

## *ABSTRACT*

*In this work the author demonstrated a robust and efficient method for implementing Doppler classification through the use of Linear Discriminant Analysis (LDA). LDAs were used to reduce dramatically the data dimensionality and thereby eliminate redundancy and improve the efficiency of the classifier. The performance was assessed on a three-class problem of personnel, tracked and wheeled vehicles. Real radar data from a ground based system were used in the design and testing of the classifier. The classifier algorithm was optimised by choosing the best set of features that maximised the performance and the bootstrap method was used to measure the confidence interval. It was shown that only the first few LDA features were relevant. At the very least these were shown to contain information regarding the frequency extent of target Doppler sidebands. The classifier was shown to be robust to changes in target viewing geometry and speed. Overall, good classification was achieved for personnel with some misclassification between tracked and wheeled vehicles.*

## 1.0 INTRODUCTION

MTI (Moving Target Indication) radars can provide an all-weather, day/night, surveillance capability. Such radar systems provide very efficient location information on moving targets but traditionally have limited recognition capability. Automatic recognition algorithms developed for imaging radars, which exploit target spatial information, are not applicable for MTI systems because they operate in a low resolution mode. However, there is potential for classification based on target Doppler signatures. The Doppler signatures are shifted in frequency in proportion to the target radial velocity. Movement or rotation of structures on a target may induce additional frequency modulations on the returned radar signal and generate sidebands about the Doppler frequency shift of the target's body. The signature characteristics of these Doppler sidebands provide a mechanism for classifying the target of interest.

The Doppler classifier models each target class as a multivariate Gaussian mixture distribution (GMD). The parameters of the GMD model are estimated using labelled training data. The input feature vectors are generated from the radar Doppler spectra. It is assumed that each Doppler spectrum provides an independent feature vector. Training uses multiple Doppler spectra per target class. Recognition is performed using a single Doppler spectrum (feature vector).

The size (and therefore the dimensionality) of the input feature vector depends upon the number of separate frequency bins in the Doppler spectra. Herein lies the limitation of a classification technique that uses the Doppler spectra directly for input feature vectors. Doppler spectra can comprise a large number of frequency bins (several tens, possibly hundreds) to cover sufficiently the full range of Doppler frequencies

*Paper presented at the RTO SET Symposium on "Target Identification and Recognition Using RF Systems", held in Oslo, Norway, 11-13 October 2004, and published in RTO-MP-SET-080.*

at enough resolution to be able to provide meaningful classification performance. High dimensionality leads to increased classifier complexity. There are more parameters to estimate per target model which results in an increased processing load. Reducing dimensionality makes the classification calculations quicker and saves on data storage space. Furthermore, the original set of variables may contain redundant and irrelevant information. Redundancy would result in the classifier having extra parameters over and above the minimum required to capture the structure within the data. For a finite training set this would lead to poorer estimation of the classifier parameters. Therefore, reducing the dimensionality could also improve classifier robustness.

Linear Discriminant Analysis (LDA) is a well established technique for obtaining a reduced-dimension representation of the data. LDA defines (a few) new variables as linear combinations of the original ones. Evidence from speech recognition has shown that the classification performance improves if features are extracted using LDA [1]. There is a key similarity between speech processing and Doppler processing *i.e.*, both use the spectrogram as the input measurement. LDA could potentially offer a good approach for reducing the number of variables in the Doppler spectra. The technique consists of transforming the Doppler spectra variables using linear combination into a set of features (the feature vector) that are mutually orthogonal. The individual features are assumed to be independent. The transformation is designed to maximise the between-class covariance and minimise the average within-class covariance. The transformed features are ranked in order of the class separability. In theory, the classification performance should increase monotonically as the number of features increases. This allows simple trade-offs to be made between complexity (number of features) and viability (classification performance).

The classification algorithm is developed for a three class problem based on personnel, wheeled vehicles and tracked vehicles. Section 2 gives an outline of the algorithm. It describes the pre-processing, the LDA feature extraction and the Doppler classification stages of the algorithm. The data sets used in this study are described in Section 3. Results are presented in Section 4. Section 5 summarises the conclusions.

## 2.0 CLASSIFICATION ALGORITHM

### 2.1 Pre-processing

The objective of Doppler classification is to classify an unknown target as belonging to one of a predefined set of classes based on the measured Doppler spectra. The Doppler spectra are obtained by Fourier transforming a sequence of samples obtained from a single range cell during the radar dwell. Figure 1 compares typical spectra of a wheeled vehicle, a tracked vehicle and a man jogging. The peak in the spectra corresponds to the Doppler shift due to the body of the target. The Doppler sidebands, if present, are due to any parts of the target which are moving independently of the main body at that moment. For the wheeled vehicle there are no Doppler sidebands visible. This can be contrasted with the much more complex, but asymmetrical, spectrum of the tracked vehicle, and this can again be distinguished from the more symmetrical spectrum of the walking man.

The information in the Doppler spectra, however, cannot be used directly for classification. This is because the Doppler radar signature is affected by certain factors such as the radar gain, noise level, *etc*., that are unrelated to the target class but can confuse the classification process. The data can be transformed so that the Doppler signatures are invariant to these factors. This process that is performed prior to classification is termed ‘pre-processing’.

The pre-processing aims to obtain a 2D spectrogram from a long sequence of temporal samples and process each individual spectrum to extract a target Doppler-profile that is independent of radar-calibration and target-velocity. The spectrogram is generated using a short-time Fourier transform. Clutter frequency bins are masked and those that contain noise only are clipped to a minimum value. The peak in each Doppler-profile is centred which makes the spectrum invariant to target velocity. Finally, the data are normalised with respect to received power and transformed using natural logarithms.

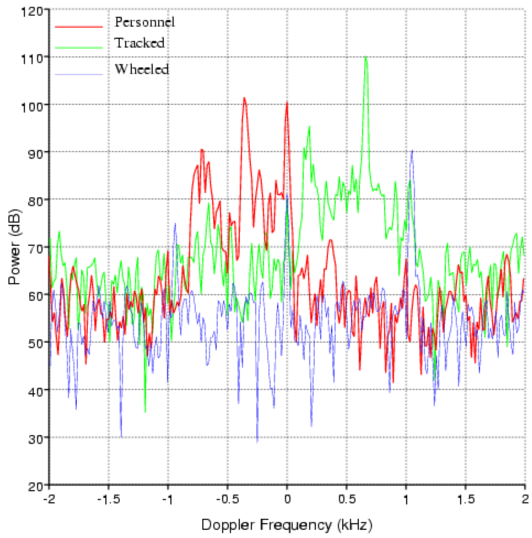


**Figure 1: Doppler spectra from land targets**

Figure 2 plots the Doppler signatures shown in Figure 1 following pre-processing. All the spectra have equal peak values and are centred on the same Doppler frequency. The pre-processed Doppler signatures are now invariant to changes in radar gain and target bulk velocity.

The pre-processing also partitions the data into five separate velocity bands based on the estimate of the target body velocity obtained using the peak in the Doppler spectrum. This is designed to enable the algorithm to model some aspects of the velocity dependent data attributes. A separate classifier is trained and tested for data from each velocity band.

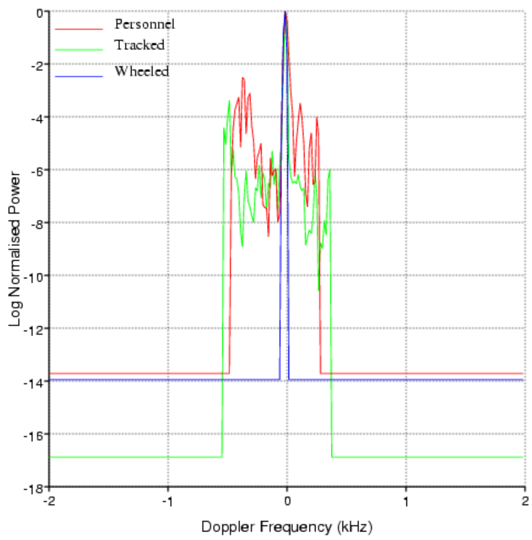


**Figure 2: Doppler spectra from land targets following pre-processing**

## 2.2 LDA feature extraction

The pre-processed Doppler spectra are put through the LDA data reduction process using the transformation

$$\mathbf{y} = \mathbf{A}^T \mathbf{x} \tag{1}$$

where $\mathbf{x}$ is the log-normalised Doppler spectrum with $p$ variables, $\mathbf{y}$ is the LDA transformed feature vector with $d$ variables and $\mathbf{A}$ is the $p \times d$ linear transformation matrix. The latter is the feature transformation matrix $\mathbf{A} = [\mathbf{a}_1 | \ldots | \mathbf{a}_d]$, where $\mathbf{a}_j$ are the eigenvectors of the generalised symmetric eigenvector equation [2]

$$\mathbf{S}_B \mathbf{a} = \lambda \mathbf{S}_W \mathbf{a} \tag{2}$$

The LDA process obtains the transformation that maximises the ratio of between class covariance to average within-class covariance. $\mathbf{S}_W$ is the average within-class covariance matrix given by:

$$\mathbf{S}_W = \sum_{i=1}^{C} \frac{n_i}{n} \hat{\mathbf{\Sigma}}_i \tag{3}$$

where $n_i$ is the number of measurements in the $i$-th class, $n$ the total number of measurements in the data set, $C$ the number of classes and $\hat{\mathbf{\Sigma}}_i$ is the sample covariance of class $i$ given by:

$$\hat{\mathbf{\Sigma}}_i = (1/n_i) \sum_{q=1}^{n_i} (\mathbf{x}_q - \mathbf{m}_i)(\mathbf{x}_q - \mathbf{m}_i)^T \tag{4}$$

where $\mathbf{x}_q$ and $\mathbf{m}_i$ are the measurement vector and the sample mean for the $i$-th class respectively. Each of these is a $p$-dimension vector. The latter is given by:

$$\mathbf{m}_i = (1/n_i) \sum_{q=1}^{n_i} \mathbf{x}_q \tag{5}$$

$\mathbf{S}_B$ is the between-class covariance matrix given by:

$$\mathbf{S}_B = \sum_{i=1}^{C} \frac{n_i}{n} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{6}$$

where $\mathbf{m}$ is the sample mean of the entire data.

The number of columns (eigenvectors) in the matrix $\mathbf{A}$ defines the size of the LDA feature vector $\mathbf{y}$. The upper limit for $d$ is the maximum number of non-zero eigenvalues for (2) given by:

$$d_{\max} = \min(p, C-1) \tag{7}$$

Since the eigenvalues for (2) are ordered in terms of class separability, in theory the classification performance should increase monotonically as the size of the LDA feature vector $\mathbf{y}$ is increased. The transformation matrix $\mathbf{A}$ is estimated using the same training data that is used for estimating the classifier parameters. As pre-processing partitions the data in to $V_b$ (=5) different velocity-bands a separate transformation matrix $\mathbf{A}_k$, where $k = 1, \ldots, V_b$ is estimated for each velocity-band. Furthermore, the estimation process requires that the data are class-labelled. One option would have been to use the three broad-class labels, personnel, tracked vehicles and wheeled vehicles. However, this would have limited $d_{\max}$ to just a maximum of two features. It was felt that this would not have been sufficient to fully exploit the structure in the data. For this reason a fine-class labelling mechanism was adopted to increase $C$ and thereby allow for a higher value for $d_{\max}$ for the transformed feature vector $\mathbf{y}$. The fine-class labelling was based on the target type, its aspect angle and its nominal speed. It may be possible, although this was not proven, that the fine-class categories have some physical justification.

### 2.3 Doppler classifier

The LDA feature vectors are used as inputs to the classifier. A separate classifier is defined for each of the velocity-bands $V$. For $C$ broad classes the class membership is denoted by $\omega_{ik}$, $i \in \{1,\ldots,C\}, k \in \{1,\ldots,V\}$. For an unknown feature vector $\mathbf{y}_k$ the class membership will be one that maximises the posterior probability $P(\omega_{ik} \mid \mathbf{y}_k)$. According to Bayes' rule this is equivalent to:

$$P(\omega_{ik} \mid \mathbf{y}_k) = \frac{P(\omega_{ik})P(\mathbf{y}_k \mid \omega_{ik})}{\sum_i P(\omega_{ik})P(\mathbf{y}_k \mid \omega_{ik})} = \frac{P(\mathbf{y}_k \mid \omega_{ik})}{\sum_i P(\mathbf{y}_k \mid \omega_{ik})} \tag{8}$$

where $P(\mathbf{y}_k \mid \omega_{ik})$ is the probability of the feature-vector $\mathbf{y}_k$ from velocity-band $k$ arising from class $\omega_{ik}$, and $P(\omega_{ik})$ is the prior probability of class $\omega_{ik}$ being present. All the training classes were assumed to be equally likely. Thus class membership is based on the probability value $P(\mathbf{y}_k \mid \omega_{ik})$ calculated for each broad-class.

Each broad class probability was modelled as a multivariate Gaussian mixture model with a diagonal covariance matrix. The mixture distribution has the same dimensionality as the LDA feature vector. Four mixture components were used. The parameters of the model (mean, variance and weights) were estimated using training data. Performance was evaluated using independent test data.

## 2.0 DATA SET

Radar data from moving targets were collected using a J-band, horizontal polarisation, short range, ground based system using a 4 kHz pulse repetition frequency. The radar measurements were taken with the antenna pointing in a fixed direction and a control target moving through the radar swath at a specified aspect angle and speed. This constituted a single imaging run and the process was repeated for a number of different target types belonging to the three broad classes. The personnel data were obtained from a trial where two subjects were imaged walking and jogging either towards the radar or moving directly away from it. The vehicle data were obtained from a separate trial where three tracked and two wheeled vehicle types were imaged along 9 different aspect angles travelling at a nominal constant speed. This provided 53 different imaging runs from which data were extracted.

For each imaging run, a number of independent target signature files of four seconds dwell were generated by processing data from different locations along the range swath. The processed range resolution was chosen such that it was wider than the dimensions of the largest target in the data set. All the data files were pre-processed and partitioned into velocity bands. There was an uneven distribution of classes over the velocity bands. The lowest two velocity bands contained mainly personnel targets. All three target classes were represented in the next two highest velocity bands. Velocity band V (targets with velocity 12mph and above) on the other hand had only vehicle targets. The data files were given two different types of labels. Fine labels were used in the estimation of the LDA transformation matrix. A total of 53 fine-class labels were defined as summarised by Table 1. Only broad class labels were used in the training and testing of the classifier.

| Broad Class | Target Type | Aspect Angle | Speed | Total per broad class |
|---|---|---|---|---|
| Personnel | 2 | 2 | 2 | 8 |
| Tracked Vehicles | 3 | 9 | 1 | 27 |
| Wheeled Vehicles | 2 | 9 | 1 | 18 |
| | | | Total | 53 |

**Table 1: Breakdown of fine-class categories for the entire database**

## 3.0 RESULTS

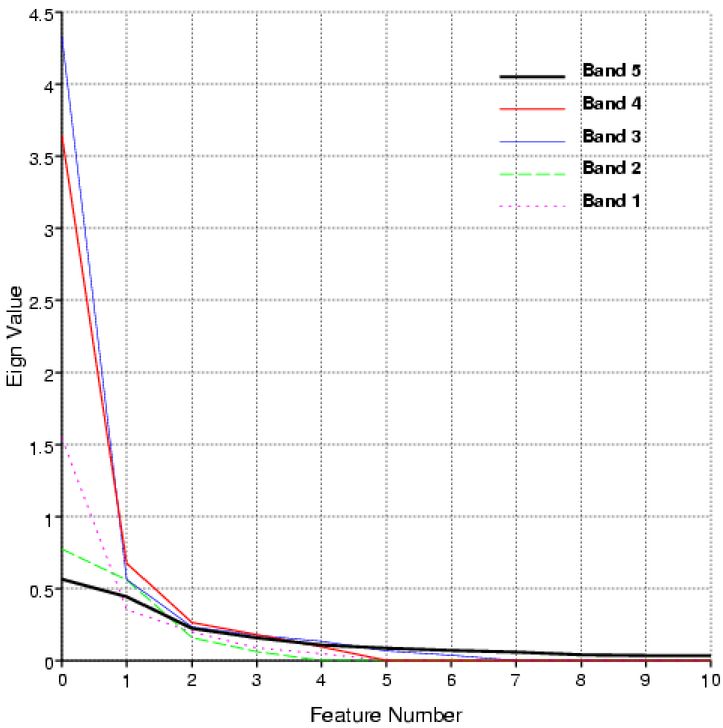


**Figure 3: LDA transformation matrix eigenvalues plotted for each velocity-band**

Figure 3 shows the comparison of the eigenvalues for the different velocity-band data. The eigenvalues provided some indication of the class separability. As the LDA theory stated, the eigenvalues were monotonically decreasing. Eigenvalues with values close to zero can be assumed to be irrelevant. Velocity-band I and II had data primarily just from the personnel class and therefore there was just one single dominant eigenvalue. Velocity-band III and IV also had a relatively high first eigenvalue. This suggested that the first eigenvector should provide good class separability. For velocity-band V there were no dominant eigenvalues, however, the first few eigenvalues were non-zero. This suggested that just a few features would probably be sufficient for optimum classification.

Further useful insight into the class separability can be obtained using 2-dimensional scatter plots of the feature vectors. Figure 4 compares the results obtained for plotting the first two LDA features for two different velocity-bands. The left-hand result is for velocity-band III which had data for all three target classes and the right hand result is velocity-band V that had only vehicle data. The feature values are labelled d0 and d1 respectively. Each point in the scatter plot is data from one feature vector.

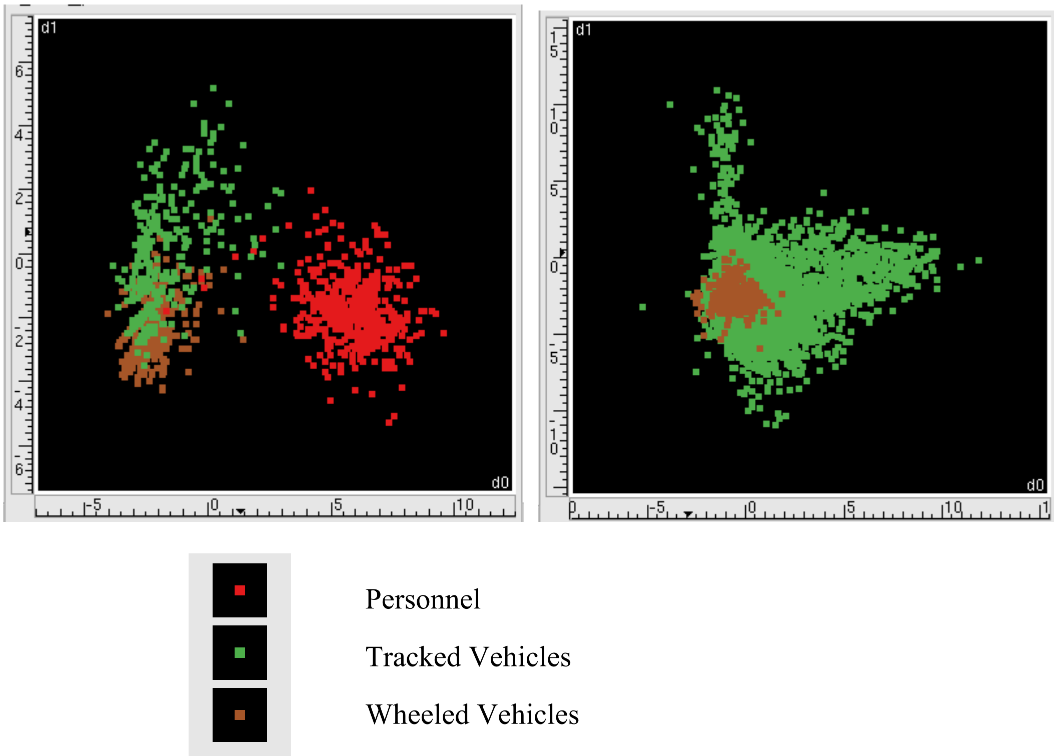


**Figure 4: Scatter plot for the first two LDA features. (left image) Velocity-band III (right image) Velocity-band V**

From Figure 4 it can be seen that for velocity-band III the personnel class separated completely from the vehicle classes. The vehicle classes also showed some degree of separation but there was some overlap between the tracked and wheeled vehicles. The same result for velocity-band V showed that there was a relatively small region of the feature space occupied by both tracked and wheeled classes. However, this was contrasted by a significantly larger region of the feature space that was occupied exclusively by the tracked class.

It is not trivial to interpret the eigenvectors in a physical manner. One possible method for determining what information is captured by an eigenvector (and therefore the LDA feature) is to look for evidence for any correlation between the LDA feature and ad hoc features that have a physical interpretation. The target Doppler sideband extent can be measured as an ad hoc feature. Empirical analysis showed that tracked vehicles tended to have broad extent whereas wheeled vehicles generally had a narrow Doppler extent. Figure 5 replots the scatter plot of the first two LDA features for velocity-band V highlighting data that has broad Doppler extent. It showed that a majority of the region, that separated the tracked from the wheeled class, was explained in terms of the Doppler extent. Thus the first two LDA features were capturing information regarding the Doppler sideband extent in some way. The LDA features, however, cannot exactly represent this ad hoc feature since the latter is a non-linear feature.

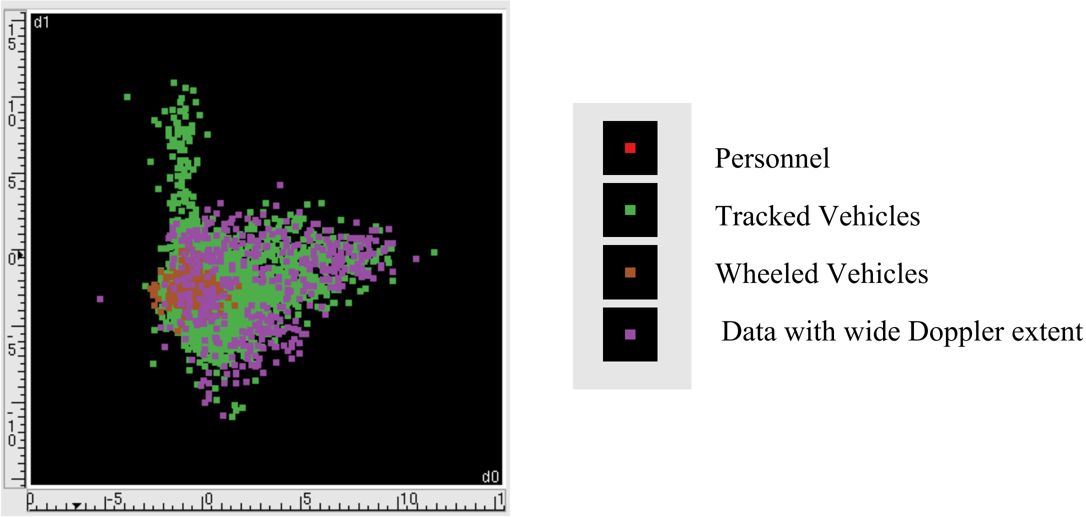


**Figure 5: Scatter plot for the first two LDA features for velocity-band V. Data points that corresponded to a wide Doppler extent are highlighted in purple**

A separate classifier was implemented for each velocity band. The first two velocity bands had only data from the personnel class and therefore were excluded from the calculations. The data in each of the other three velocity bands were split into training and test sets using a 3-to-1 ratio. Performance results were averaged over all three velocity bands.

Figure 6 plots the percentage correct classification averaged over all three broad classes (personnel, tracked and wheeled vehicles) as the number of features was increased. Results are shown for two cases, (a) feature vectors based upon only LDA features (black curve), and (b) feature vectors that included Doppler sideband extent as an additional ad hoc feature (purple curve).

From the first result it can be seen that with just two LDA features near maximum performance was achieved. For six or higher number of LDA features the performance flattened out. This implies that the useful information is contained in just the first few features. A classifier with just six LDA features would give optimum performance. This equated to a considerable reduction in the data dimensionality and therefore the classifier complexity. Such improvements greatly enhance the viability of the classifier for real-time implementation.

With the addition of the Doppler extent feature, just the first two features alone provided the optimum performance. This pointed toward Doppler sideband extent being an important discriminating feature. It ties in with the observation from the feature analysis which showed a trend for tracked vehicles to have broad extent and wheeled vehicles to have narrow extent. It suggested that the LDA features are capturing the same information as in the Doppler extent of the sidebands albeit using more features. Unlike ad hoc features which are data specific and would often require lengthy and expansive data analysis, the LDA feature extraction process on the other hand would generalise for data with arbitrary attributes.

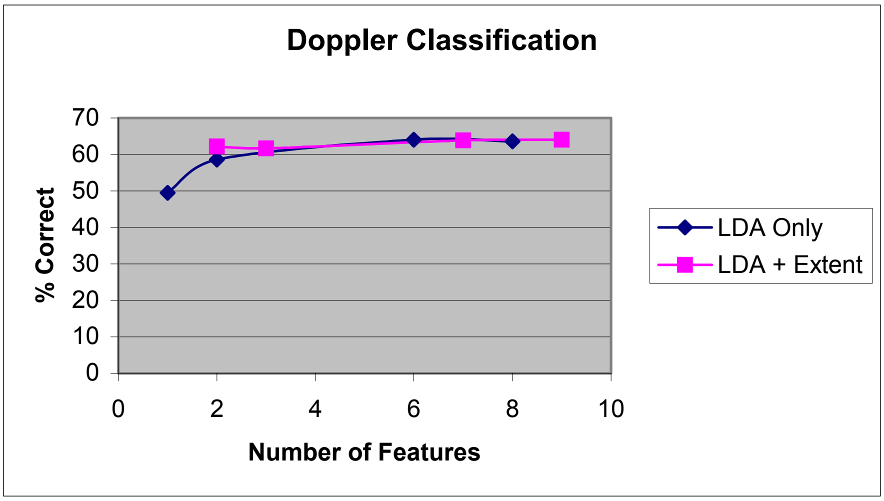


**Figure 6: Doppler classification as a function of number of features**

| OVERALL 62.7% [58.3,67.1] | | Classification Decision (%) | | |
|---|---|---|---|---|
| | | Personnel | Tracked | Wheeled |
| Actual Class | Personnel | 96.4 [99.8,93.0] | 2.7 [0.1,5.3] | 0.9 [0,2.9] |
| | Tracked | 0.8 [0.1,1.5] | 51.4 [41.7,61.1]] | 47.8 [38.2,57.4] |
| | Wheeled | 0.2 [0,0.6] | 21.4 [12.6,30.2] | 78.4 [69.7,87.1] |

**Table 2: Confusion Matrix of a Doppler classifier using 6 LDA features. Results averaged over 50 bootstrap replicates**

Table 2 provides the confusion matrix for the classifier with six LDA features. The results were generated using 50 bootstrap replicates. Bootstrap is a statistical inference technique, first proposed by Efron [3], which allows a confidence interval to be assigned to the estimated quantity. Table 2 lists the mean of the bootstrap replicates along with the 90% confidence interval shown in square brackets. The results for the confidence interval were only approximate since far more bootstrap replicates (>1000) would be required for a more accurate measure. Nevertheless, the results were useful in determining general performance trends. Earlier the selection of the six LDA feature classifier was based on results that were essentially a single bootstrap replicate. This choice is lent support by the estimate of the 90% confidence interval for this classifier. Since the performance of the other classifiers with fewer LDA features was outside this range it can be concluded that the choice of the optimum is statistically significant.

A per class comparison of the confusion matrix shows that just under half the tracked vehicles are misclassified as wheeled. This is not very surprising given the fact that a substantial proportion of the tracked vehicle data in the data set did not have the distinctive broad Doppler extent that differentiated it from wheeled vehicles. At this stage it can only be hypothesised that the confusion between the two vehicle classes is due to the absence of the track returns. The data were collected from vehicles that had skirts covering the tracks. This would make the moving parts of the tracks more likely to be visible when viewed front-to-back, and vice versa, but less so at oblique angles. The data supported this inference, with far fewer of the measurements taken for vehicles travelling at oblique angles to the radar reporting the presence of the broad Doppler extent. This was in contrast to tracked vehicles travelling either directly toward or away from the radar, for which the majority of the data had the broad Doppler extent present.

Unlike the two vehicle classes the personnel class separated very well. Some misclassification between personnel and tracked class may be expected since both possess broadening of the Doppler spectra. However, the manner in which the vehicle data were collected (constant velocity with aspect changing between measurements) meant that the personnel class was only being classified against vehicles that were travelling at very oblique angles. From the data it was observed that track returns were often absent when the vehicles were imaged at oblique angles. This may therefore explain the very good separation between the personnel and vehicle classes. More representative data that contains data from slowly moving vehicles with visible tracks would enable a better measure of the true performance.

## 4.0 CONCLUSION

For the three-class problem the classifier had no difficulty in recognising the personnel class but produced some degree of confusion between the wheeled and tracked classes. The classifier algorithm was optimised by choosing the best set of features that maximised the performance and the bootstrap method was used to measure the confidence interval. It was shown that only the first few LDA features were relevant for Doppler classification. At the very least these were shown to contain information regarding the frequency extent of target's Doppler sidebands.

The classifier was shown to be invariant to target aspect angle and speed and was able to model multiple target types. Models for additional classes that have distinct Doppler characteristics, like helicopters, can be easily incorporated into the algorithm. The LDA feature extraction represents a considerable reduction in data dimensionality and therefore is able to provide for very efficient implementation of the classification algorithm. The LDA based classifier, therefore, offers a very powerful tool for the automatic classification of moving targets from their Doppler signatures.

## ACKNOWLEDGEMENTS

The author would like to acknowledge the support of UK MOD for funding this research. He would also like to thank Dr Keith Ponting of 20 20 Speech Ltd in providing support with the development of the LDA classifier. Thanks also go to Ian Finley for the numerous helpful discussions with the author during the course of this work.

## REFERENCES

[1] Kumar N., 'Investigation of silicon-auditory models and generalization of linear discriminant analysis for improved speech recognition', Ph.D. dissertation, John Hopkins Univ., Baltimore, MD, 1997.

[2] Devijver P. A. and Kittler J., 'Pattern Recognition, A Statistical Approach', Prentice-Hall, Inc., London, 1982.

[3] Efron B., 'Boostrap Methods. Another Look at the Jackknife', *The Annals of Statistics*, Vol:7, 1979, pp1-26.