Cepstral Coefficient Feature Extraction for Active Sonar Classification

Boaz Suranyi and Binh Nguyen

Maritime Division, Defence Science and Technology Group, Australia

ABSTRACT

The effectiveness of acoustic classification is highly dependent on the features that can be extracted from a given signal. It is possible to employ a range of features simultaneously during classification to improve accuracy and redundancy, though certain features can degrade performance when used together. Mel Frequency Cepstral Coefficient (MFCC) and Linear Predictive Cepstral Coefficient (LPCC) algorithms provide a method for acoustic feature extraction through the processing of the short term power spectrum of a signal. Features generated using these coefficients may enhance classification performance in active sonar applications due to their robustness against background noise in low frequency bandwidths (20-2000Hz). This paper discusses the integration of these features into the Binary Classification Research Tool (BCRT); a research tool for the testing of feature and classifier performance. The analysis examines their compatibility with established feature sets as well as their overall potential in the area of sonar classification.

Through testing on a range of underwater signals, MFCC features were found to have strong isolated performance and to increase classification accuracy when combined with established feature sets. LPCC features had a poor performance in isolation but achieved the highest classification accuracy when combined with other feature sets.

1 INTRODUCTION

Active sonar provides a means for detecting and identifying underwater objects by emitting a sound pulse and analysing its echo return. Information about the object, defined as features, can be extracted from the return through signal processing and analysis. The goal of this process is to obtain informative and non-redundant data on the object to facilitate classification - the assignment of an object into a particular category.

1.1 Motivation

Lower frequencies of sound are absorbed less and can propagate further underwater. The signal to noise ratio (SNR) of an echo decreases with distance as the strength of the original signal compared to the level of background noise diminishes. Feature algorithms need to be robust against the effects of noise to facilitate sonar detection and classification at longer ranges. MFCC features may meet these requirements since “the method for extracting MFCC is robust to resist the disturbance of background noise in the auditory range (20-2000Hz)” (Wenbo et al, 2016). LPCC features are derived in a similar fashion to MFCC but differ in their methods of pre-emphasis and lack specific weighted filtering. LPCC are included in this study to benchmark the specific effects of the Mel filter as well as provide a more quantitative analysis of overall feature compatibility.

1.2 Background

MFCC and LPCC make use of the power cepstrum to obtain information on the rates of change in the spectral bands of a signal. Both methods include pre-emphasis techniques to more accurately “approximate areas of high energy concentration while smoothing out the fine harmonic structure of other less relevant spectral details” (Hermanxky, 1989). The Mel frequency is tailored to human cochlea, scaling the frequency of a signal to closely mimic the way human’s bracket and categorise frequencies of sound or “phonemes”. The human ear is more sensitive to changes in pitch for lower frequency sound and a Mel filter scales a signal to place higher emphasis

on this region. MFCC are used to great effect in the field of speech recognition. LPCC are also popular in this context despite lacking the filter bank processing - they extract features by calculating the smoothed Auto-Regressive power spectrum of a signal.

Mel Frequency Cepstral Coefficients were found to be effective features for the identification of radiated ship noise in a study by Zhang et al (Zhang, 2016), achieving identification rates of 85% and above for data with poor to neutral SNR (-10 dB to 0 dB). MFCC and PLPCC features among others were used to describe the sonar echoes of a range of surface vessels in testing performed by Korany (Korany 2012). The study concluded that an optimal number of Mel Frequency Cepstral Coefficients for maximised identification rate was 24-26. These tests involved the use of a single classifier, involved training and testing on the same data set, and provided no other features with which to benchmark performance. These concepts are built upon in this study to provide a quantitative assessment of cepstral coefficient performance in the field of sonar.

1.3 Structure

Section 2 describes the development of the MFCC and LPCC feature algorithms and their implementation using the Binary Classification Research Tool (BCRT). Section 3 details the classifiers and features used in this paper then outlines the testing procedure undertaken for the benchmarking of the new algorithms. Section 4 displays the results of this testing and discusses the associated trends and implications of the data, and Section 5 concludes the findings of the paper and presents areas for further development. Finally, an Appendix is included that presents expanded results from the testing.

2 ALGORITHM IMPLEMENTATION

Functions were created in Matlab to derive MFCC and LPCC and integrate them into the BCRT. The methods of calculating the coefficients can be broken into three stages: signal framing, filter bank processing and feature extraction. Both MFCC and LPCC utilise the same method of signal framing but LPCC do not use any filtering. Both methods differ slightly in their manipulation of the power spectrum to obtain features.

2.1 Signal Framing

An input signal file of time series snippets is broken into overlapping frames based on the sampling frequency and a specified frame size and overlap. This analysis uses a length of 25 ms and an overlap of 5ms for both MFCC and LPCC calculations and the signal is zero padded if required.

2.2 Filter bank Processing

The Mel filter is a triangular shaped filter bank created between a specified cut off and Nyquist frequency using methodology outlined by Lyons (Lyons 2012). The frequency limits are converted to Mel frequency using Equation 1 and *n* points (where *n* is the number of filters) are generated evenly spaced between them along the Mel scale.

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \tag{1}$$

To achieve the required weighted spacing, the Mel frequencies are converted back to conventional frequencies (Hertz) and rounded to the nearest FFT bin, resulting in *n*+2 points for the creation of the filter bank. Filter *n* begins when filter *n*-1 reaches its peak and filter *n* reaches its peak as filter *n*-1 returns to zero. This is described in Equation 2 with $H(k)$ defining the slopes of each filter, *n* is the filter number and *f*() is the list of Mel spaced frequencies calculated in Equation 1.

$$H_n(k) = \begin{cases} 0 & k < f(n-1) \\ \dfrac{k - f(n-1)}{f(n) - f(n-1)} & f(n-1) \le k \le f(n) \\ \dfrac{f(n+1) - k}{f(n+1) - f(n)} & f(n) \le k \le f(n+1) \\ 0 & k > f(n+1) \end{cases} \tag{2}$$

MFCC featues calculated in this analysis use 26 filters which was found to be optimal for feature extraction in previous studies (Korany, 2012), Figure 1 displays one such filter bank and a typical signal power spectrum.

The second and third rows of the Figure demonstrate the windowing effect of applying the filters. The emphasis placed on lower frequency bands can be clearly observed in the spacing of the triangular filters along the frequency axis.

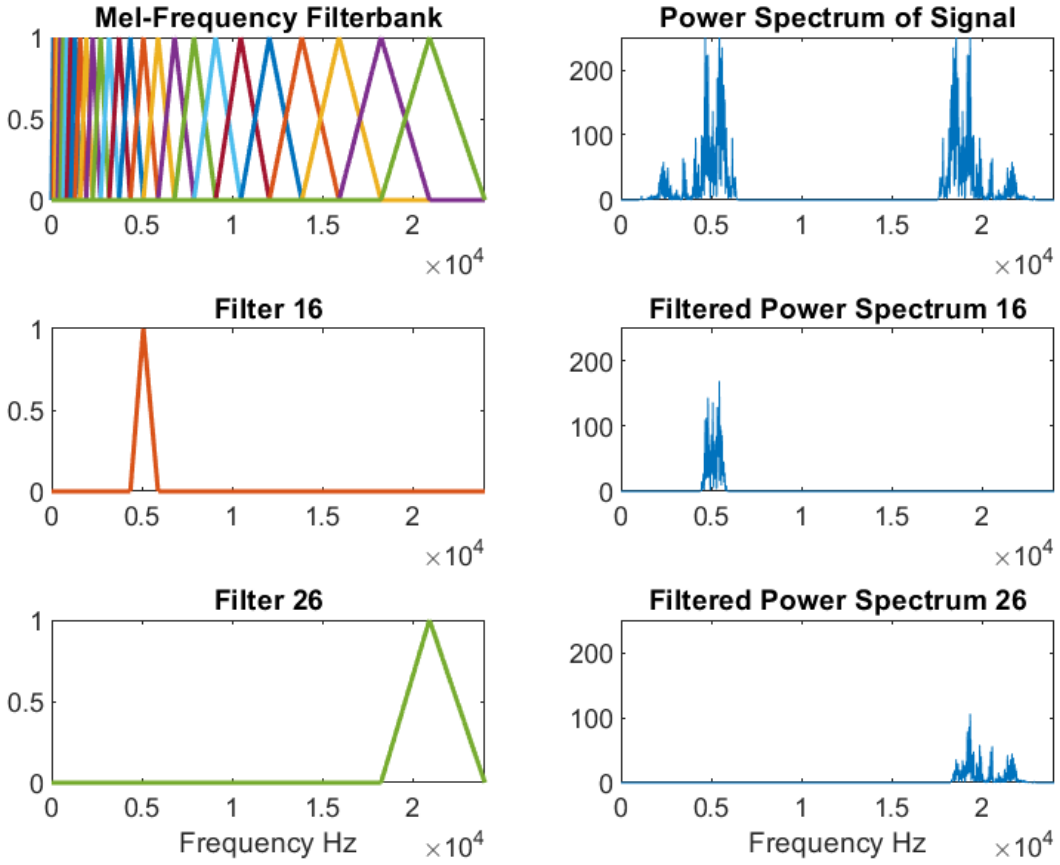

Figure 1: Example Mel Frequency filter bank and windowed power spectrums

2.3 Feature Extraction

MFCC: For each frame of the signal, a periodogram estimate of the power spectrum is calculated by taking the absolute value of the Discrete Fourier Transform (DFT) and squaring it. This process is shown in Equations 3 and 4 where $S(k)$ is the DFT of length *K*, $s(n)$ is the time domain signal and $h(n)$ is an *N*-sample long analysis window.

$$S(k) = \sum_{n=1}^{N} s(n)h(n)e^{-\frac{j2\pi kn}{N}} \tag{3}$$

$$P(k) = \frac{1}{N}|S(k)|^2 \tag{4}$$

The power spectrum $P(k)$ is then windowed by each Mel filter and the logs of the resulting output energies are taken. Finally, a Discrete Cosine Transform (DCT) is applied to these to obtain the final features - the Mel Frequency Cepstral Coefficients. There is coefficient for every filter in each frame of the signal.

LPCC: A forward linear predictor of order p is applied to each frame by computing the least squares solution to $Xa = b$ as shown in Equation 5, where x is the input signal in the time-amplitude domain.

$$X = \begin{bmatrix} x(1) & 0 & .. & 0 \\ x(2) & x(1) & .. & : \\ : & x(2) & .. & 0 \\ x(m) & : & : & x(1) \\ 0 & x(m) & .. & x(2) \\ : & : & : & : \\ 0 & .. & 0 & x(m) \end{bmatrix} \qquad a = \begin{bmatrix} 1 \\ a(2) \\ : \\ a(p+1) \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 0 \\ : \\ 0 \end{bmatrix} \tag{5}$$

The autoregressive estimate of the power spectrum is then calculated using the formula in Equation 6 where a_k are the linear prediction coefficients, *G* is the prediction error variance, and $f(n)$ are the final Linear Prediction Cepstral Coefficients.

$$f(n) = \begin{cases} 0 & n < 0 \\ \ln(G) & n = 0 \\ a_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) f(k) a_{n-k} & 0 < n \leq p \\ \sum_{k=n-p}^{n-1} \left(\frac{k}{n}\right) f(k) a_{n-k} & n > p \end{cases} \tag{6}$$

3 BENCHMARKING

3.1 Data Sets Used

The data used for classification training and analysis was 36 sets of recorded or synthesised acoustic echo returns, each set containing between 108 and 361 individual snippets. The data sets covered four classes; returns from artificial entities in the ocean such as wellheads categorised as Structure (9 sets total), returns from natural features in the ocean such as sandstone categorised as Terrain (9 sets total), scaled returns from generic ship models categorised as Vessel (6 sets total), and ambient noise data collected at sea or synthesised from experimental models categorised as Clutter (12 sets total).

The classification performance of the cepstral features was benchmarked by testing their ability to differentiate between each combination of signal category. Table 1 displays the testing outline.

Table 1: Class combinations for testing

	Class 1	Class 2
Test 1	Structure	Clutter
Test 2	Structure	Terrain
Test 3	Vessel	Clutter
Test 4	Vessel	Terrain
Test 5	Vessel	Structure

For each Test, every data set in Class 1 was tested against every data set in Class 2. For example, Test 1 is the average of 108 individual tests as there were 9 data sets in Class 1 each tested against the 12 data sets in Class 2.

3.2 Classifiers Used

Three classifiers were used for the training and classification of data in this study. Each classifier has several parameters which are discussed briefly below. For each of the three classifier categories below, the average performance among the parameter variations was used for the results in section 4.

Nearest Neighbours: The nearest neighbour rule is a simple classification algorithm that assigns each test point based the average class of the *k* nearest training points. The adaptive nearest neighbour classifier expands on this algorithm through the application of a weighted modifier, taking into account the "influence size" of

each training data point and adjusting its contribution accordingly (Ray 2017). A K Nearest Neighbour (KNN) and an Adaptive Nearest Neighbour (ANN) classifier were used in this testing.

Kernel Ridge Regression: Ridge Regression uses a linear least squares estimate to extrapolate co-linear data (Ray 2017). The kernel trick is applied to map the inner product of the regression onto a projected space allowing for well-defined separation of classes. The formation of this space is determined by the kernel used; this testing employed a KRR classifier using a Linear Kernel, Polynomial Kernel, and Gaussian Kernel.

Support Vector Machine: SVM classifiers operate by deriving a two dimensional hyperplane from the training data and using it to discriminate test data points into separate classes (Ray 2017). This plane or "support vector" is derived during training using a Linear, Polynomial or Gaussian Kernel. All three were used in this analysis.

3.3 Features Used

In this study, Cepstral Coefficient features have been benchmarked against the “Baseline” set of 16 established features derived from the statistical moments of the signal in the time and frequency domain. These features are outlined in Table 2 and have demonstrated effectiveness in sonar classification (Kouzoubov, Nguyen, Wood).

Table 2: Baseline features, time and frequency domain

Baseline	
Time Domain	Frequency Domain
Shape Mean	Shape Mean
Shape Variance	Shape Variance
Shape Skewness	Shape Skewness
Shape Kurtosis	Shape Kurtosis
Amplitude Mean	Amplitude Mean
Amplitude Variance	Amplitude Variance
Amplitude Skewness	Amplitude Skewness
Amplitude Kurtosis	Amplitude Kurtosis

Testing was conducted on each feature set individually as well as each combination of feature set. The combination of MFCC and LPCC is referred to as CC All and is achieved by simply combining the two individual feature vectors before classification. The feature groups tested in this study were:

- Baseline
- MFCC
- LPCC
- CC All
- Baseline + MFCC
- Baseline + LPCC
- Baseline + CC All

3.3.1 Feature coefficient selection

Using the method described in Section 2 resulted in a large number of features; the number of filters times the number of frames for MFCC, and the number of the polynomial order times the number of frames for LPCC. Many of these features are redundant and can degrade performance. To improve identification rate and decrease processing time a single “optimal” feature was selected for each frame. Testing each class combination once and selecting the Maximum, Minimum, Average or Mean of the coefficients as a single feature for each frame gave the results in Figure 2.

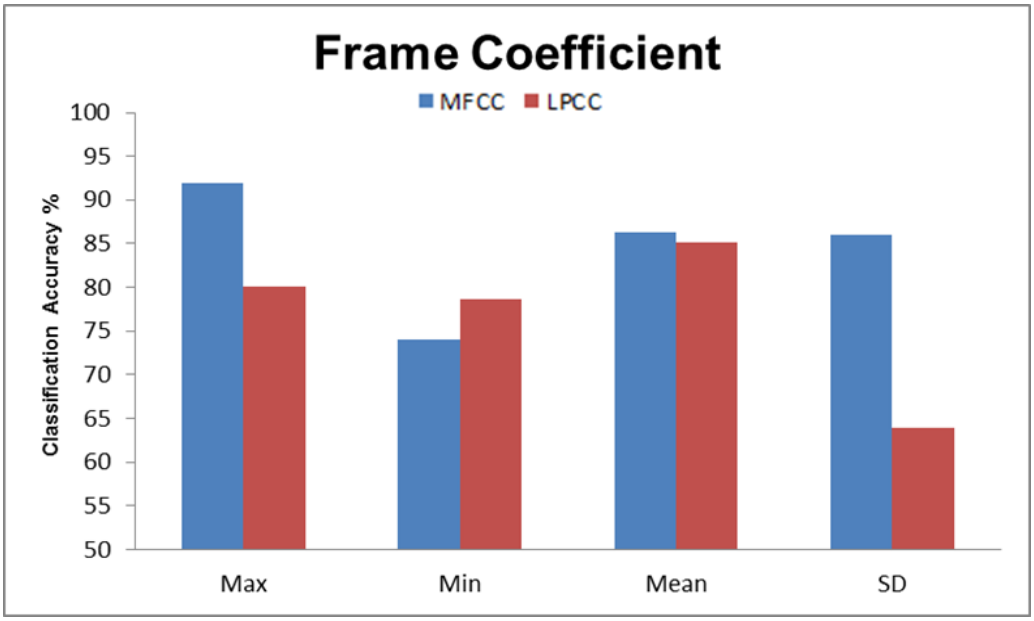

Figure 2: Optimal coefficient selection

For all subsequent tests the Maximum coefficient was used in each frame for MFCC and the Mean coefficient of each frame was used for LPCC.

3.3.2 Feature normalisation selection

Normalisation is an essential process during classification because it reduces the impact of outlying features that can degrade or invalidate the classification process. After testing feature groups 1 to 3 once for each class combination it was concluded that “mean-variance” normalisation would be used for the study, the results are shown in Figure 3, with the mean variance method achieving the highest accuracy for all feature groups.

Mean Variance normalisation is calculated as:

$$f(k) = \begin{cases} k - k_{mean}, & k_{SD} = 0 \\ \dfrac{k - k_{mean}}{k_{SD}}, & k_{SD} \neq 0 \end{cases} \tag{5}$$

Where k_{SD} is the standard devitaion of the features, k_{mean} is the average of the features and $f(k)$ is the normalised feature vector.

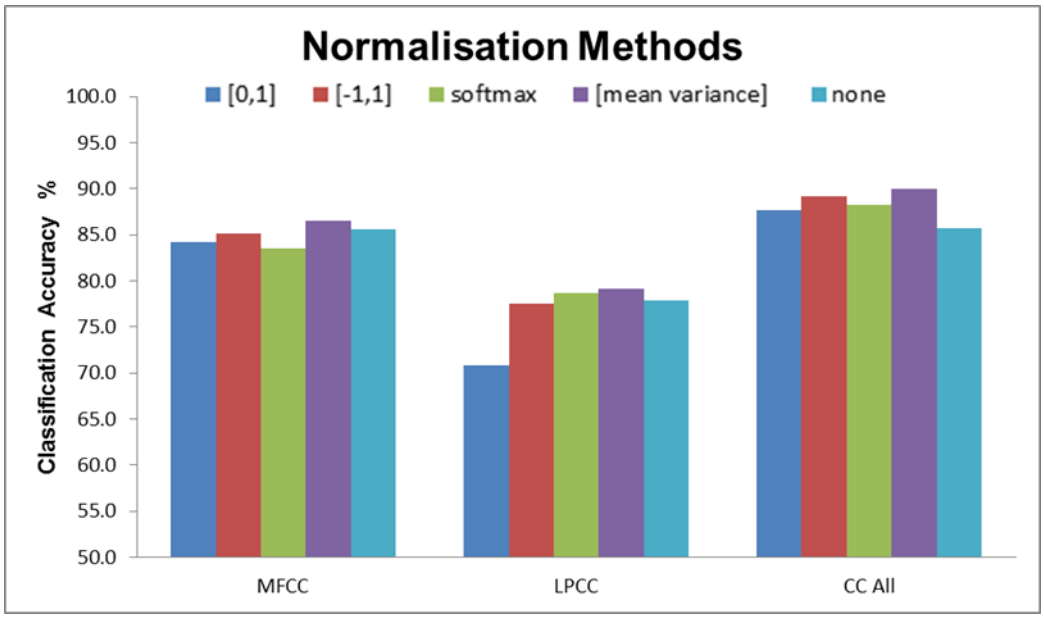

Figure 3: Optimal feature normalisation method

3.4 Binary Classification Research Tool

Figure 4 displays the classification process in BCRT using MFCC features on two data sets for Test 1, Structure vs Clutter. Features from each class are shown on the right with the highlighted cyan and green stars denoting the specific features extracted from the time-amplitude snippet on the left. There are 118 and 169 snippets respectively for each of the signals in this test resulting in 118 features per frame for Class 1 and 169 features per frame for Class 2. Without the selection of optimal features conducted in Section 3.3.1, 4,394 features per frame would be generated for the example below. A significant majority of these would be redundant and including them extends processing time and degrades the performance of the classification algorithms.

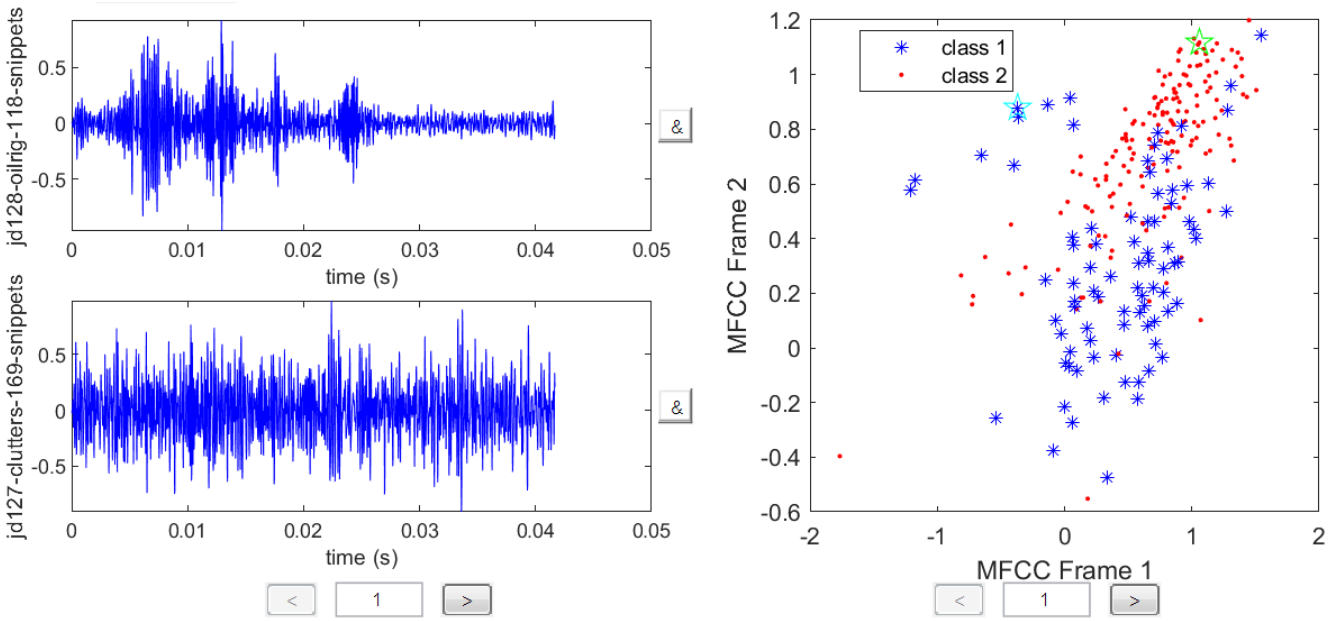

Figure 4: MFCC feature extraction and classification within BCRT

4 RESULTS

The goal of this study was to assess the robustness of the cepstral coefficient features and their ability to increase performance when combined with existing sets. When averaging classification performance across all test groups as shown in Figure 5 it is evident that adding cepstral features MFCC, LPCC, or CC All, to the Baseline set has a positive impact on accuracy. When used in isolation, MFCC and CC All were able to slightly outperform the Baseline set whereas LPCC achieved a significantly lower average accuracy.

When combined with Baseline, MFCC and LPCC achieved the highest overall accuracy across all classifiers. CC All had the strongest performance in isolation but the lowest combined performance of all the Baseline combination sets. This is likely a result of overfitting and indicates that the use of both cepstral feature sets in addition to the Baseline has exceeded the optimal number of features for training using this type of data.

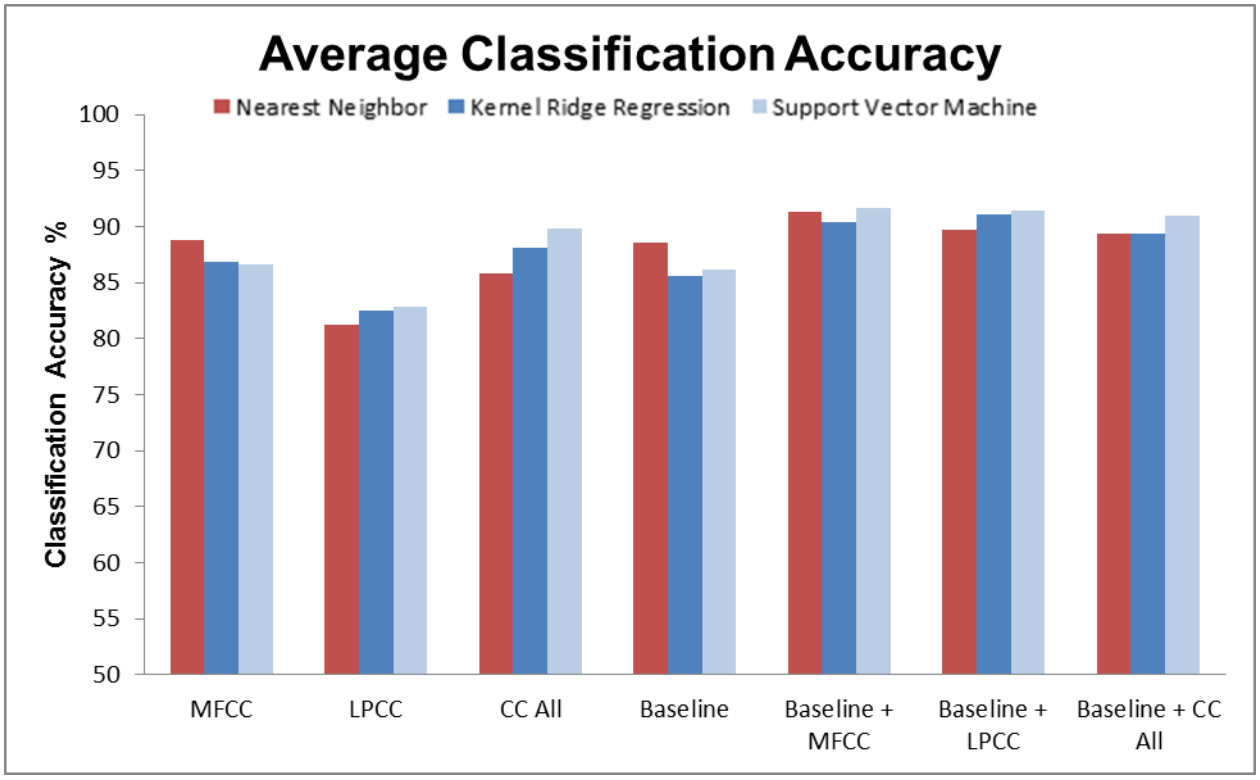

Figure 5: Classification accuracy averaged across all test groups

Figure 6 helps visualise the increase in accuracy achieved when combining each of the cepstral coefficient sets with the Baseline. MFCC features had largest impact on the performance of the Nearest Neighbour classifier, increasing accuracy by 2-2.75 percentage points. Both MFCC and LPCC increased performance of the KRR and SVM classifiers by around 5-5.5 percentage points whereas CC All provided the smallest increase to classification performance.

LPCC performed on par with MFCC when combined with the Baselines set, indicating that the benefits of the Mel Frequency filter bank are limited for these sets of data. There is a clear advantage to the use of power spectrum analysis for the feature extraction of underwater signals and it is likely that more optimal methods of filtering or pre-emphasis exist. The only distinct advantage of MFCC over LPCC is the stronger classification performance when used in isolation.

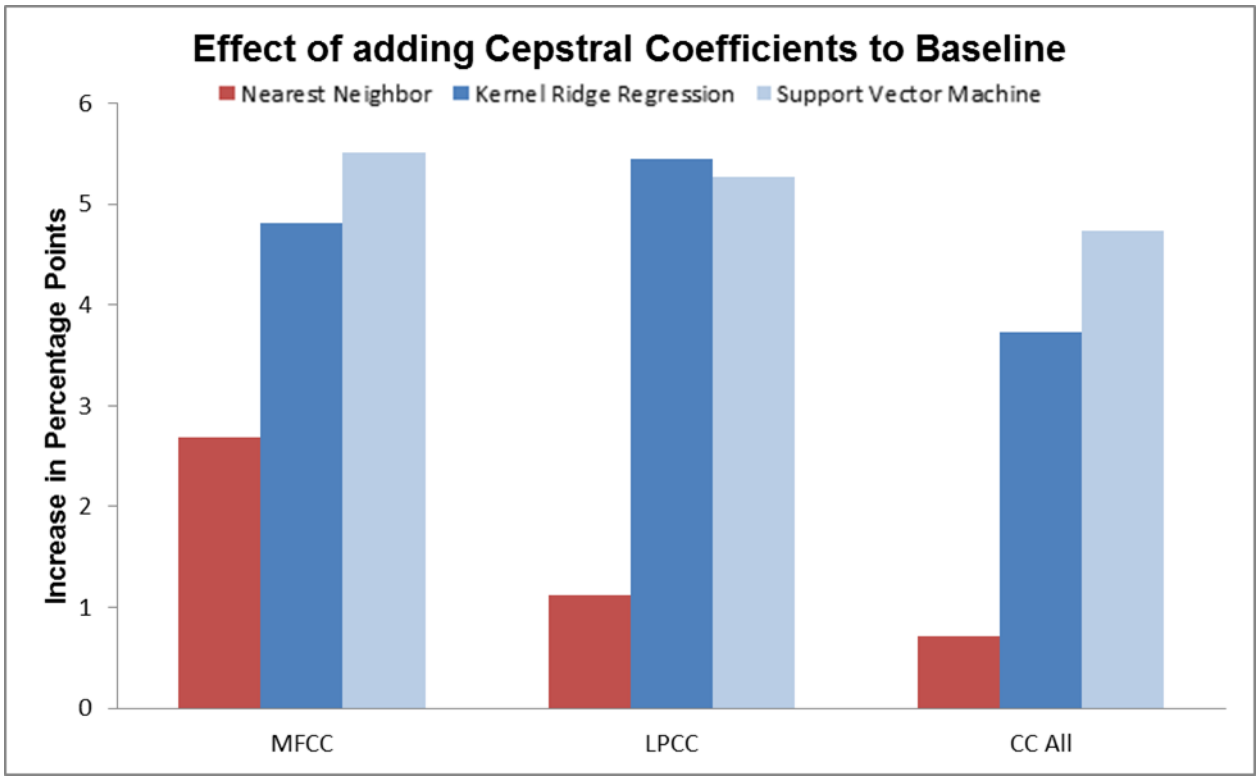

Figure 6: Cepstral coefficient feature performance when added to Baseline Set

An increase in accuracy was achieved in every test for each classifier regardless of which individual set had the stronger performance in isolation. This trait is vital and demonstrates the cepstral coefficients compatibility with the statistical moment features of the Baseline Set. This pattern can be examined more closely in the expanded results found in the Appendix.

Processing time

The external framing and subsequent Fourier processing of the signal used in the extraction of MFCC and LPCC features significantly increase the extraction time, resulting in a 50 fold increase in processing time over the Baseline. The additional processing through the Mel filter bank further prolongs MFCC extraction resulting in a 60 fold increase in processing time compared to the Baseline. The computational cost of the cepstral coefficient features is significant and the increase in accuracy they provide may not be justifiable in certain scenarios. These times could be reduced by further optimising the signal framing used in this analysis.

Table 3: Computational time

Feature Extraction Method	Average Time (s)
MFCC	25.95
LPCC	21.89
Baseline	0.44

5 CONCLUSIONS AND FUTURE WORK

Cepstral Coefficients of underwater signals were found to be effective features in the context of active sonar classification. MFCC and LPCC both increased classification performance between 3-5% when used in combination with an established Baseline feature set for a range of signals. However, this increase in accuracy comes at a significant cost in computational time. Future work in this area could include:

- Testing of other filter banks, specifically those orientated around distinct resonant frequencies.
- The development of an adaptive method for coefficient selection within each frame, tailored to the properties and changes occurring within each specific signal.
- Further experimentation with more data sets, specifically those with very low signal to noise ratios to represent weak or distant echoes.

REFERENCES

A. Kouzoubov, B. Nguyen, S. Wood, “Binary and Multiclass Classification Performance at Various Sonar Processing Steps”, 2011, Maritime Operations Division, Defence Science and Technology Organisation.

H. Hermanxky, “Perceptual Linear Predictive Analysis of Speech”, 27 November 1989, Speech Technology Laboratory, Panasonic Technologies California.

J. Lyons, “Mel Frequency Cepstral Coefficient (MFCC) tutorial’, Practical Cryptography, 2012, http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

N. Korany, M. Abd Elzaher and H.Khater, "Classification of Underwater Acoustic Signals Using Various Extraction Methods" Alexandria University Egypt, 2012.

S. Ray, “Essentials of Machine Learning Algorithms (with Python and R codes)”, 2017, Analytics Vidhya, https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

W. Wang, S. Li, J. Yang, Z. Liu, W. Zhou, “Feature Extraction of Underwater Target in Auditory Sensation Area based on MFCC”, 2016 IEEE/OES China Ocean Acoustics, 9-11 Jan. 2016.

Z. Zhongrui, W. Di, L. Zhang and H. Xue, "Feature Extraction of Underwater Target Signal Using Mel Frequency Cepstrum Coefficients Based on Acoustic Vector Sensor", 11th October 2016, College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China.

Appendix: Individual Test Results

Figure 7 presents the individual results of Tests 1 through 4.

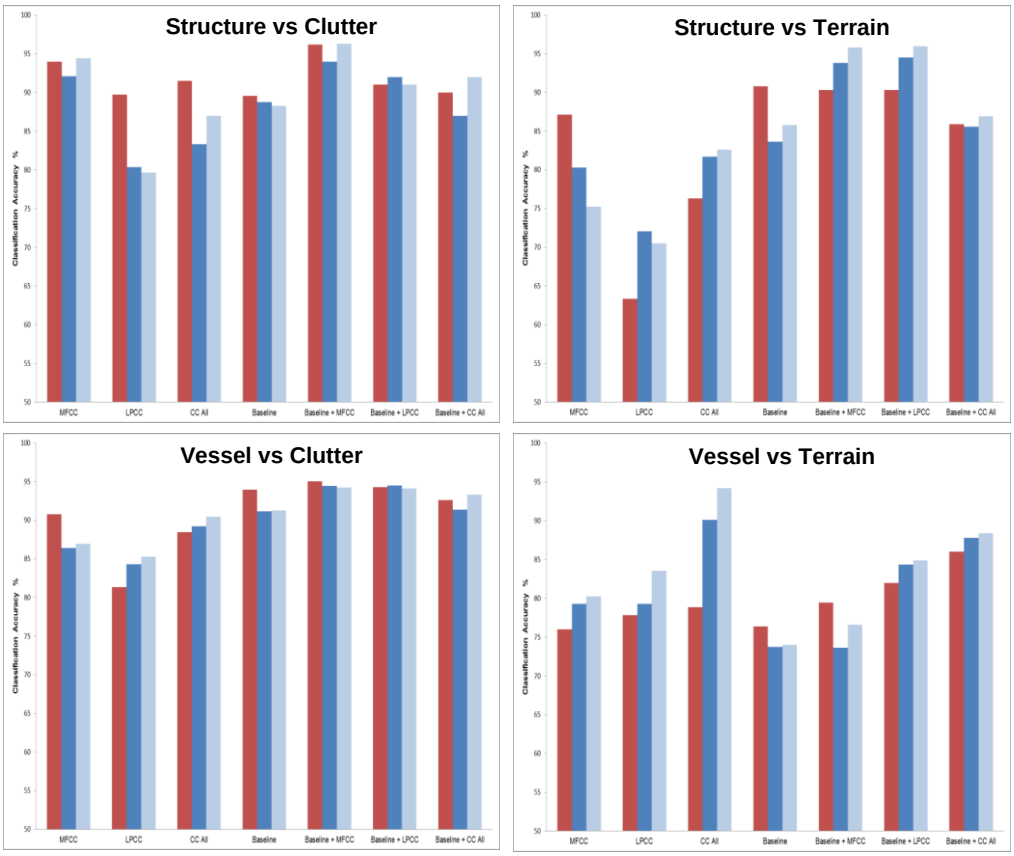

Figure 7: Classification accuracy for tests 1, 2, 3 and 4