Classification of marine acoustic signals using Wavelets & Neural Networks

Paul SEEKINGS and John POTTER
Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore, 12a Kent Ridge Road, Singapore 119223.
paul@arl.nus.edu.sg

ABSTRACT

We describe a method to automatically classify Humpback whale (Megaptera Novaeangliae) song that offers improvements over matched spectrogram techniques currently widely employed. Humpback song is a useful training example for a range of ocean acoustic transient detection and classification problems because it consists of units of varying length, frequency range and type, from nearly tonal to highly transient. With any recognition system it is vital that the data is first segmented into appropriate units. This is nontrivial and often implemented manually. We have developed a segmentation using wavelet packet decompositions that also produces a 'feature vector' with which to classify the data using a neural network. The next step is to select the network architecture, where there are many good alternatives, including a principle component front end coupled to a back-propagation network and self-organising networks with Learning Vector Quantisation. Various architectures typically achieve 80% classification rates on a challenging variety of units. The approach has the added benefits of being shift invariant with respect to time, and somewhat tolerant of frequency and time stretching. Since the methods employed are not specific to whale song the approach can be usefully applied to other types of marine transient signals with minimum modification.

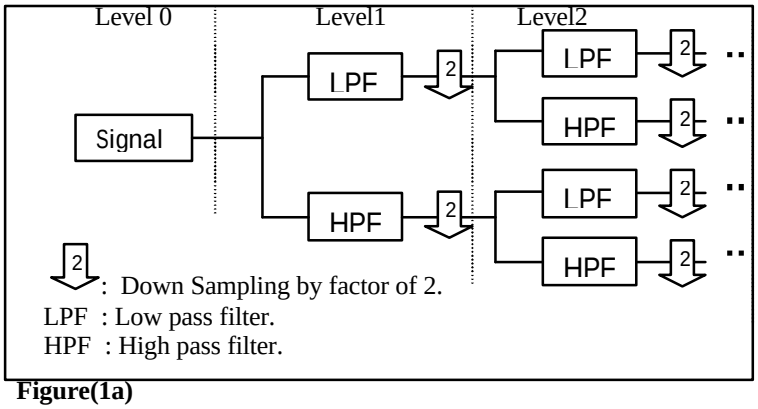
KEYWORDS: Whale song, Classification, Teager Energy, Neural Networks, Wavelets, Humpback whale.

INTRODUCTION

Most recognition systems require the input to have a high signal to noise ratio. However, marine acoustic signals often have a variable but generally low signal to noise ratio. Here Humpback whale (Megaptera Novaeangliae) song is used, which is often considered to consist of sequences of repeated stereotyped units [1]. The purpose of this research is to recognise and classify each unit that makes up the whale song. Each unit can consist of long tonals, or short pulses or frequency modulated signals. The information is thus neither succinctly expressed in either the time or frequency domain for all unit types. Discrete Windowed Fourier analysis can provide a time-frequency representation (spectrogram) but does not provide a natural way to condense the feature vector space for classification. Spectrogram matching techniques are also usually intolerant to time and frequency shifts and stretching. Wavelet analysis can overcome these problems and provides a richer analysis of the signal while achieving the time-shift invariance and feature vector compression required as a pre-processor for a neural network classifier.

WAVELET PACKET DECOMPOSITION

The wavelet packet transform is a simple extension to the wavelet transform that offers a richer decomposition by providing a multiply-redundant set of possible bases to represent the signal. [2]. The time domain signal is repeatedly down-sampled and split into frequency bands using low pass and high pass filters whose coefficients are determined by the mother wavelet function. This is shown in the schematic Figure(1a) below.



Figure(1a)

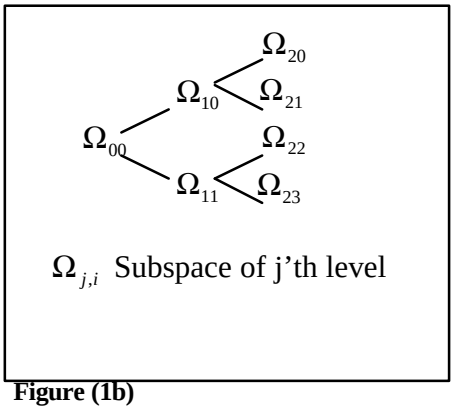
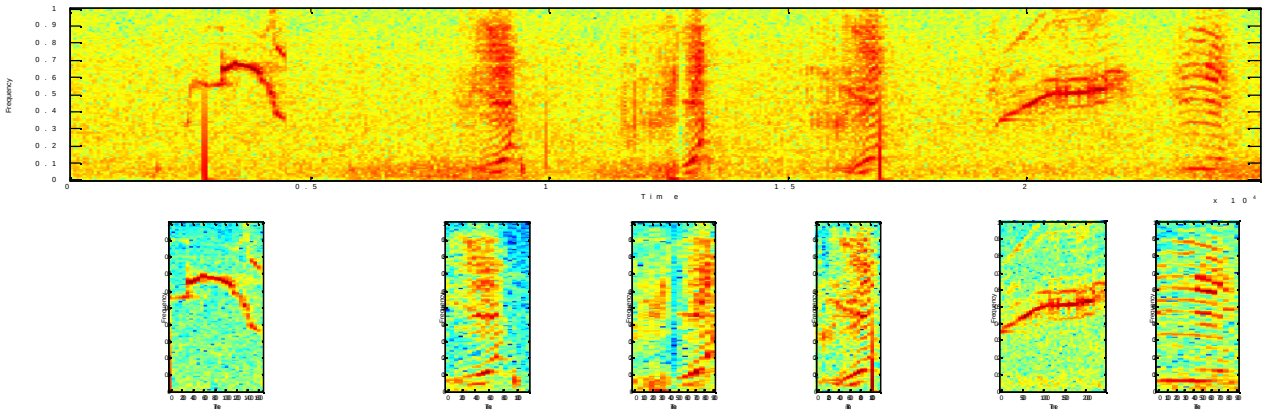


Figure (1b)

At each level the signal is being split into twice as many frequency bands. As the level increases we have more frequency resolution (more bands) but less time resolution due to the down sampling. Another way to represent the wavelet packet transform, as shown in Figure (1b), is in terms of the binary tree of the subspaces that result from the wavelet packet decomposition. Each pair of subspace (children) is contained in within a subspace from a higher level (parent). Each complete level contains a complete representation of the signal, hence the decomposition is highly redundant. A collection of subspaces can be used to reconstruct the signal as long all frequencies are represented, for example the subspaces $\Omega_{11}, \Omega_{20}, \Omega_{21}$ represent the signal across all frequencies.

There are whole families of wavelets. For this pattern recognition application were the intention is a real time system, the Daubechies Real Biorthogonal Most Selective (DRBMS) wavelet, [3], was chosen. It has the following properties that make it a very attractive choice: Time invariance; If the time series is time shifted then its wavelet packet coefficients are only time shifted. Fast computation; Daubechies wavelets have fractal-like self-similarity properties that lead to fast wavelet transform techniques. Sharp filter transition bands; Daubechies wavelets have very sharp transition bands which minimises edge effects between frequency bands.

SEGMENTATION



Figure(2) Showing the spectrogram (top) of section of whale song and the spectrograms of calls found by segmentation

Figure (2) shows the segmentation of a section of whale song provided by Cornell Bioacoustics Research Program. The recordings, sampled at 4Khz, were made in March 1994 from a pod located off the north coast of Kauai, Hawai. Wavelet packet decompositions (WPD) are taken of adjacent blocks of the whale song. The wave packet coefficients are thresholded using Donoho-Johnstone estimator, [4], which is optimised to remove noise. If the total energy in the wave packet decomposition is over a certain threshold then we regard that WPD as containing part of a whale call. The feature vectors are taken from adjacent WPD's that contain whale calls. It was observed that, the whale calls are between 0.5 and 5 seconds long, the inter-call duration could be as short as 0.25 Secs. The window length chosen was 512 samples long i.e. 0.12 seconds. This ensures that calls are reasonably tightly windowed, with only the minimum duration of noise at the beginning and end of the calls.

MANUAL CLASSIFICATION OF DATA

Each of the 1406 song units were manually classified into classes of similar calls using the spectrogram and by listening. Seventeen classes were found. The perceived class provided target vectors for neural networks and the testing of the final system.

FEATURE VECTOR

The feature vector consists of a Teager Cepstrum(TC) for each of the consecutive WPD's that contain part of a whale call. It should be noted that the WPD's have been calculated in the segmentation phase. To get a fixed length feature vector, a maximum of 8 TC's are used. Shorter calls have a shorter feature vector and are padded with zeros. Longer whale calls were observed to be tonal in nature and this is thought sufficient to classify them with the first 8 TC's. The duration of the call is a discriminatory factor and using this scheme is incorporated in the feature vector.

The Teager energy cepstrum is used to obtain feature vectors for speech recognition in noisy environments. It has been shown, [5], that Teager energy gives a good measure of the energy of the signal in a particular sub-band in the presence of coloured noise in a wide variety of test signals. In our case, the Teager energy is calculated from the lowest level of the WPD, and for each frequency band.
It is defined as

$$e_l = \frac{1}{N}\left|\sum_{t=1}^{N} \Omega_{n,l},(t)^2 - \Omega_{n,l}(t-1)\Omega_{n,l}(t+1)\right| \qquad [6]$$

where $l = 0,...,2^n - 1$ and $N = \frac{N_s}{2^n}$. n is the lowest level of the decomposition, N_s is the length of signal hence N is the number of samples in each sub band. Each WPD had N_s =512, n=6, making N=16.

The log of Teager energy spectrum is then encoded using the discrete cosine transform (Teager Cepstrum).

$$TC(k) = \sum_{l=0}^{2^n-1} \log(e_l)\cos\left(\frac{k(l-0.5)p}{2^n}\right) \quad k = 1,...,12 \qquad [7]$$

Twelve points are used to encode the Teager energy spectrum. It was observed that using more points did not affect classification results.

NEURAL NETWORKS

In the classification stage, a feed-forward network trained with back-propagation learning. Principle component analysis was used to reduce the dimensionality of the input vector, [6], and hence speed up the training phase. No significant difference in classification rates was found if PCA was not used. As a comparison, a self-organising map with linear vector quantisation was training and tested. SOM-LVQ networks provide a 'feature map' of the training data [8]. LVQ rearranges the map and attempts to minimise intra class distance, but maximise the separation between classes. Using the same set of training data both network types gave very similar classification rates on the test data. Feed forward networks were favoured during development due to the quicker training times.

TRAINING AND TESTING

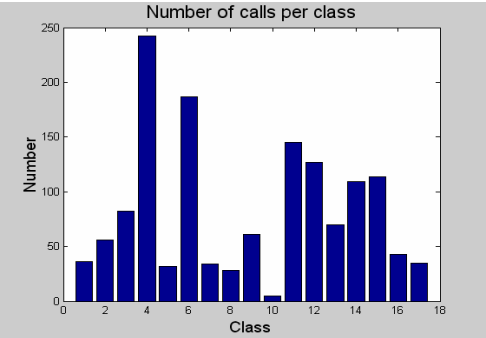


Figure (3) Showing the number of calls in each class.

The training set consisted of half the number of samples from each class that had less than 60 samples. For classes that have had more than 100 samples, 60 were chosen. Limiting the number of common calls was effective in stopping over training for those calls. The required number of calls for each class was taken at random from the set of manually classified calls. The remainder of the data was used to test the neural network. The training set consisted of 591 samples and test set consisted of 815 samples. Since the training/test data is chosen at random, different training sessions yielded different slightly different classification rates.

RESULTS

Using training /testing scheme described above the following results were achieved

Method	Network	Training data correctly classified	Test data correctly classified
Teager Cepstrum	BP	89%	86%
Teager Cepstrum	SOM-LVQ	91%	86%

CONCLUSION

The results show that high rates of classification can be achieved on a large set of test data. It is interesting is not much significant difference in classification rates between the two types of networks. This may imply the advantages that SOM-LVQ networks normally have over back propagation networks are invalidated since the feature vector naturally has good intra-class separation.

Having a good classification system for hump back whales call is important since it opens up other areas of research. It is hoped the system can be used to study the sequence of calls in a song, and the changes in songs from year to year. Since none of the methods used are specific to whales, and whales calls cover a wide range of signal types, then the methods employed could be used for other marine signals.

REFERENCES

1. Payne, R.S. and McVay, S, "Songs of Humpback whales", *Science vol.* 173 pp 585-597. (1971)
2. Mallat,S. *A wavelet tour of signal processing*, (Academic Press, 1998)
3. Cohen A., Daubechies I, and Feauveau J.C, "Biorthogonal bases of compactly supported wavelets", *Comm. Pure & Appl. Math* **45**, pp. 485-560, (1992).
4. Donoho, D., and Johnstone, I, "New minimax theorems, thresholding, and adaptation", Tech. Rep., Dept. of Statistics, Stanford University. (1992).
5. Jabloun, Firas, "Large vocabulary speech recognition in noisy environments ", M.SC Thesis, Bilkent University, (1998).
6. Diamantaras, K, *Principal Component Neural Networks (Theory and Applications*), (John Wiley & Sons, 1996).
7. Kohonen, T. *Self Organizing Maps*, (Springer Series in Information Sciences 1995).