Development of an Automatic Classification System for the Cetaceans Using their Vocalizations

Prateek Murgai, Delhi Technological University

Mentors: Danelle E.Cline and John Ryan

Summer 2015

Keywords: MFCC, Odontocetes, Mysticetes, SVM, RFC

ABSTRACT

The paper presents the development of an automatic classification system for the cetaceans using their vocalizations. The system aims at classifying different whale species based on the vocalization acoustic signals(either raw or preprocessed) input to the system. An acoustic dataset from the whale acoustics lab at Scripps Institute of Oceanography is being employed to build the classification system, to develop its proof of concept. Currently we have extracted five different cetacean species from the dataset, two of them are Mysticetes (blue whale and fin whale) and the rest are Odontocetes (Cuvier's Beaked whale, Sperm whale and porpoise). Based on previous experiences with audio signal processing, thirty four acoustic specific features have been extracted for our classification task. Both time domain and frequency domain features are being employed as they present a complete story about the signal, also many features inspired from speech recognition and music classification applications are being used. Use of such a combination of features could provide new insights into the properties of cetacean vocalizations. For the classification purpose, we currently employ two classifiers separately, namely the Support Vector Machines and the Random Forest Classifier. The complete workflow of the development of the classification system is explained in the

paper. Results comparing both the classifiers are presented in the results section, also an introduction of the immediate future work on the soundscape is presented.

INTRODUCTION

Marine mammal occurrences are currently assessed using visual surveys or passive acoustic monitoring (PAM). Both methods are challenged by detection uncertainty: visual surveys are often hindered by poor sighting conditions, and missed detections due to short surfacing intervals, whereas the efficiency of the passive acoustic monitoring can be limited by variable calling rates, uncertainty in caller identity, and missed detections [14]. Whereas visual surveys are labor-intensive (i.e., expensive) and weather-dependent and are, therefore, limited to temporally sporadic sampling over short periods (days to weeks), acoustic recorders can sample continuously for periods ranging from hours to years, thus PAM systems have a distinct advantage over visual methods[13].

The single greatest drawback of passive acoustic monitoring is the large volume of raw acoustic data returned that requires analysis to generate reliable species detections [4,15]. Manual analysis entails visually inspecting spectrograms of acoustic data, aurally reviewing putative calls, and classifying and logging confirmed calls. This method is extremely labor-intensive, inefficient, and unrealistic for longer-duration acoustic recordings. Not surprisingly, the rise in the use of passive acoustic monitoring applications over the past decade has spurred the development of automated methods employing machine learning techniques to detect and classify calls. The overarching goal of this development effort is to significantly reduce the time required to derive detection information from acoustic recordings while maintaining a similar level of accuracy provided by a human analyst.

The advent of automatic classification of cetacean vocalizations using supervised machine learning techniques has been strongly motivated by conservation needs. Reliable and labeled cetacean call data is still very sparse a problem faced by us while developing the classification system in theory. Initially being devoid of any acoustic data from Monterey Bay due to deployment issues at the MARS observatory, an acoustic dataset

provided by the whale acoustics lab at Scripps Institute of Oceanography was employed, which included hydrophone recordings from seven different locations in the Pacific Ocean near California and different seasons. As the data was raw hydrophone recordings, annotations had to be used to extract calls of different cetaceans and organize them into different species. After this organization, it was found that some calls lasted for a time duration not suitable for processing (some odontocete calls were longer than thirty minutes in duration)so we limited our data extraction to a maximum of four minutes long calls. Currently our extracted data has five different species (fin whales, blue whales, Cuvier's Beaked whales, Porpoise and Sperm whales).

We are currently using thirty four features for our classification purpose, some features have been inspired from speech recognition and music applications, others from experience. There is a great overlap between the speech recognition, automatic music classification and cetacean vocalization classification tasks [7,8], still the use of such a combination of features from different applications for this specific purpose remains limited, thus the novelty. One of the most important features for our application are the MFCC's (Mel Frequency Cepstrum Coefficients) which have been used widely in speech and music recognition applications, and some bioacoustics applications [7,8,9,10,12]. Spectral entropy is another important feature we have employed that is derived from information theory[11] .For the purpose of feature extraction we use a sliding window technique with a window size of 50ms and an overlap of 50% (25ms). Windowing helps to preserve the information stored in small time windows whereas overlapping helps to maintain continuity between different audio windows. In the past many different classification and recognition techniques have been used for cetacean vocalization classification such as artificial neural networks [2,4,6] and spectrogram correlation[1,4] . For our application we have used the SVM and the random forest classifier for the classification purpose, not together but separately. A comparison of their performance is presented in the results section. Having doubts of over-fitting (due to a sparse data-set) we created new data artificially by superimposing random noise from the hydrophone recordings (technique inspired from Baidu Research Group [3]) from different locations onto the cetacean calls. We will see in the results section how that actually improves the performance of the classifiers.

COMPLETE SYSTEM OVERVIEW

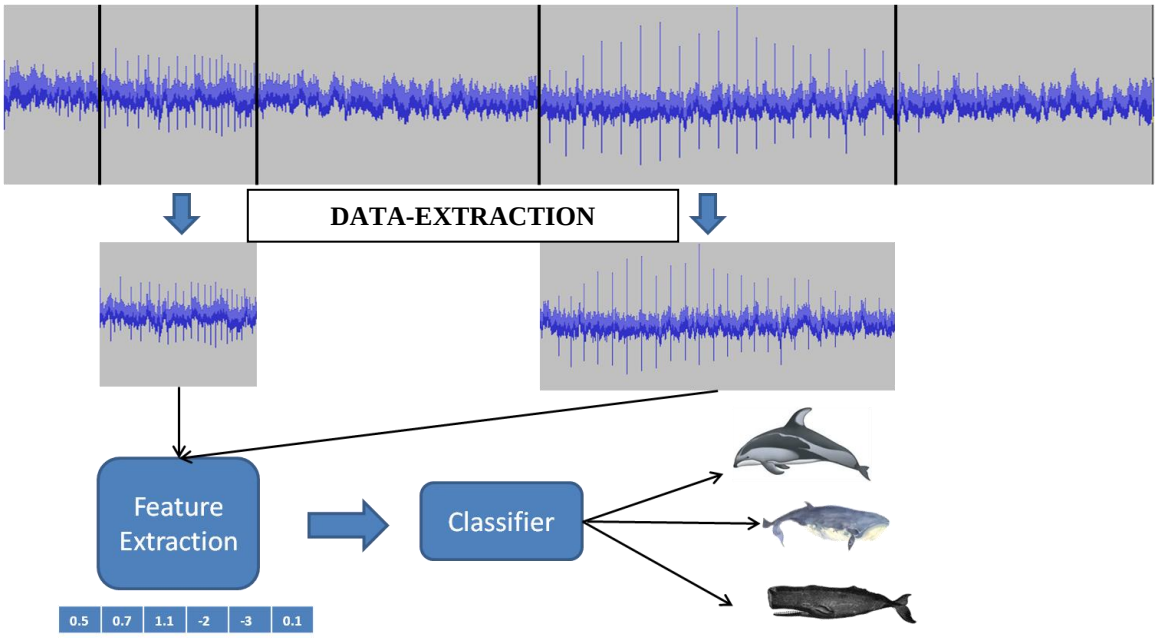


Fig.1Displays a complete three step classification system

The acoustic classification system comprises of broadly three steps:

- **Data Extraction –** To reduce a large data-set into a smaller data-set containing training data in a format that can be used by the classifier. Vocalizations for each of the five cetacean species are extracted from the hydrophone recordings using annotated files and saved separately for the classifier to access.
- **Feature Extraction-** A process of extracting specific acoustic features from the extracted data based on which the classification takes place. The system currently employs 34 features (both time and frequency domain) . This is probably one of the most important steps of a classification process.

- **Classifier-** The extracted feature vector is fed into what we call as a classifier. It is a black box that executes mathematical models on the vocalization feature vectors to separate one class of species vocalization from another, Currently we are using two classifiers: Support Vector Machines (non-probabilistic) and Random Forest Classifier.

DATA-SET

The current data-set includes five whale species, they are divided into the following two broad groups

1. **Odontocetes**

Our current extracted data-set includes calls from three different odontocetes: **Cuvier's beaked whale, porpoise and sperm whale**. Odontocetes (also called as toothed whales) are a sub group of cetaceans which employ short click sounds for the purpose of echolocation. Most of their vocalizations are narrow-band high frequency (can go as high as 160 kHz), except sperm whales(use frequencies as low as 50Hz). A spectrogram of a Cuvier's beaked whale vocalization from the data-set is shown in Fig.2

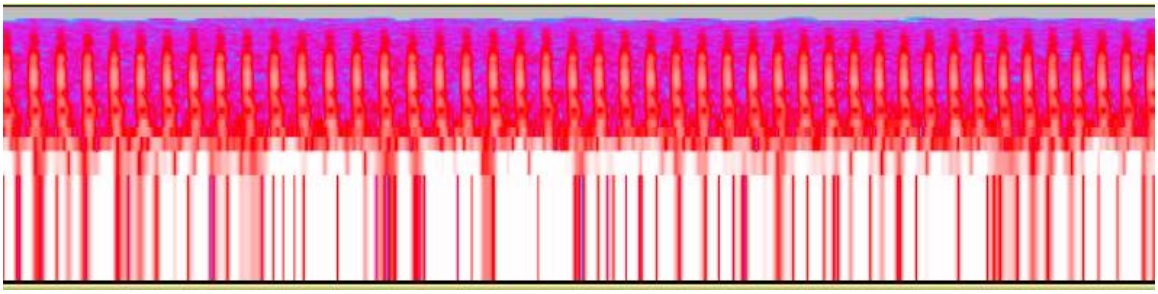


Fig.2. Spectrogram of an echo-locating Cuvier's beaked whale. The whale vocalizes at a peak frequency of around 10 kHz

Our observation from the extracted data-set is that many of the odontocete calls are of long duration (even going up to **1-2 hours** long). This presented a serious drawback as vocalizations that long cannot be processed for the classification system due to computational limitations. Due to this we were limited by the amount of data we could

practically use for our classification purpose. We currently have **400-500 odontocete calls** in total, their time duration varying from **3 seconds to 4 minutes**.

2. **Mysticetes**

   Mysticetes(also called as baleen whales) are a sub-group of cetaceans that use low-frequency vocalizations(as low as **30-50Hz**) to communicate over large distances. Our current data-set includes calls from two mysticetes: **blue whale and fin whale**. Both of these baleen whales vocalize at frequencies less than 100 Hz. A spectrogram of a blue whale vocalization from the data-set is shown in Fig.3

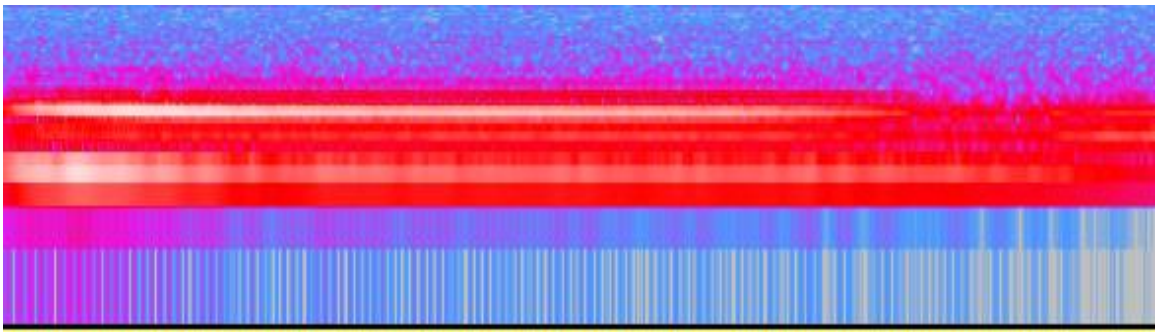


Fig.3. Spectrogram of a low frequency(40-50 Hz) vocalization of a blue whale

Observations from the hydrophone recordings show us that their vocalizations are of shorter duration as compared to odontocetes (as short as 1 sec) and thus easier to process. This was fortuitous for building our system as we could extract large number of mysticete calls and thus have a larger data-set to train our classifier. Currently we have **4500-5000 mysticete calls** in total with a time duration of **1-4 seconds** .

DATA EXTRACTION AND PROCESSING

Data is usually collected in forms and formats which are not suitable to be fed directly into a working architecture. Each system works on different data formats, thus it is important to organize them into a suitable form so that they can be employed with an existing system.

A similar problem was faced by us while designing this system. In fact the first few weeks of the internship were spent on extracting the correct data and organizing it into a format which is acceptable by the classifiers we wanted to use.

1. INITIAL DATA

The data-set from the Scripps Institute of Oceanography included raw hydrophone recordings from over seven locations in the Pacific Ocean from near California. The locations are as follows (for details regarding locations check : www.cetus.ucsd.edu) :

- CINMS-B
- CINMS-C
- DCPP-A
- DCPP-B
- DCPP-C
- SOCAL-E
- SOCAL-R

Each location has hydrophone data recorded over days in each of the three seasons(spring, summer and winter).

The recordings are separated for mysticetes and odontocetes. The **mysticete** recordings are continuous recordings of les than **40 minutes** , whereas the **odontocete** recordings are continuous recordings of longer durations , some greater than **100 hours**.

The data is recorded using **HARP's** (High Frequency Acoustic Recording Packages) and **ARP's** (Acoustic recording Packages) developed at Scripps.

2. EXTRACTED DATA

We cannot blindly feed in the hydrophone recordings to our classification system. A training set is imperative for the acoustic classifier to train itself for a number of classes of vocalizations. To build a training set we use the annotations provided with the hydrophone recordings to extract species specific data. The time annotations provided helped us extract vocalizations for certain cetacean species. Already mentioned before, we were able to extract practical vocalization data(that we could use for processing and was good enough in quantity) for five whale species.

3. DATA PREPROCESSING

The only data-preprocessing step we apply is the normalization of the feature vectors of the extracted vocalizations. Normalization helps to reduce the range of the values present in the data to a fixed range , making it suitable for a classifier to judge data on a fixed scale.

ACOUSTIC FEATURES

By definition, features are individual measurable properties of a phenomenon being observed. For our application, we employ acoustic features, which makes use of the 1-D signal in the time domain to extract time-domain signal such as Energy, Zero Crossing Rate, Entropy of Energy etc, and the 2-D spectrogram data to extract frequency domain features such as Spectral Spread, Spectral Entropy, Mel Frequency Cepstrum Coefficients (MFCC's), Chroma Vectors etc. In total we employ thirty four features.

A signal can reveal very a different information from its time-domain and frequency domain analysis, thus it is necessary to employ features encompassing properties from both the domains. The features chosen are inspired from different applications of speech recognition(MFCC's) and music classification.

The novelty in this selection is that, such a combination of speech recognition, music classification and other acoustic features doesn't have a lot of literature review specially with marine mammal vocalization classification. Also this helps us to venture into an experiment which might fail, or succeed as experimentation results for such a combination of features is limited.

Here we describe two features in detail,

1. **Mel Frequency Cepstrum Coefficient**

Mel Frequency Cepstral Coefficients (MFCCs) are a feature widely used in automatic speech and speaker recognition. MFCC's collectively make up what we call as Mel Frequency Cepstrum(MFC). MFC's are a representation of the short-term power spectrum of a sound, based on a discrete cosine transform of a log power spectrum on a nonlinear Mel scale of frequency.

The following formula gives the conversion of frequency to Mel scale :

$$M(f) = 1125\ln(1 + f / 700)$$

The Mel scale relates perceived frequency, or pitch, of a pure tone to its actual measured frequency. Humans are much better at discerning small changes in pitch at low frequencies than they are at high frequencies. Incorporating this scale makes our features match more closely what humans hear.

Steps to calculate MFCC's :

1. Frame the signal into short frames. (**50 ms frames** for our application with **50%** overlap)
2. For each frame calculate the periodogram estimate of the power spectrum.
3. Apply the Mel filterbank to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filterbank energies.
5. Take the DCT of the log filterbank energies.
6. Keep DCT coefficients 2-13, discard the rest.

Thus in total we have **12 MFCC's** , thus that number of features from MFCC's itself

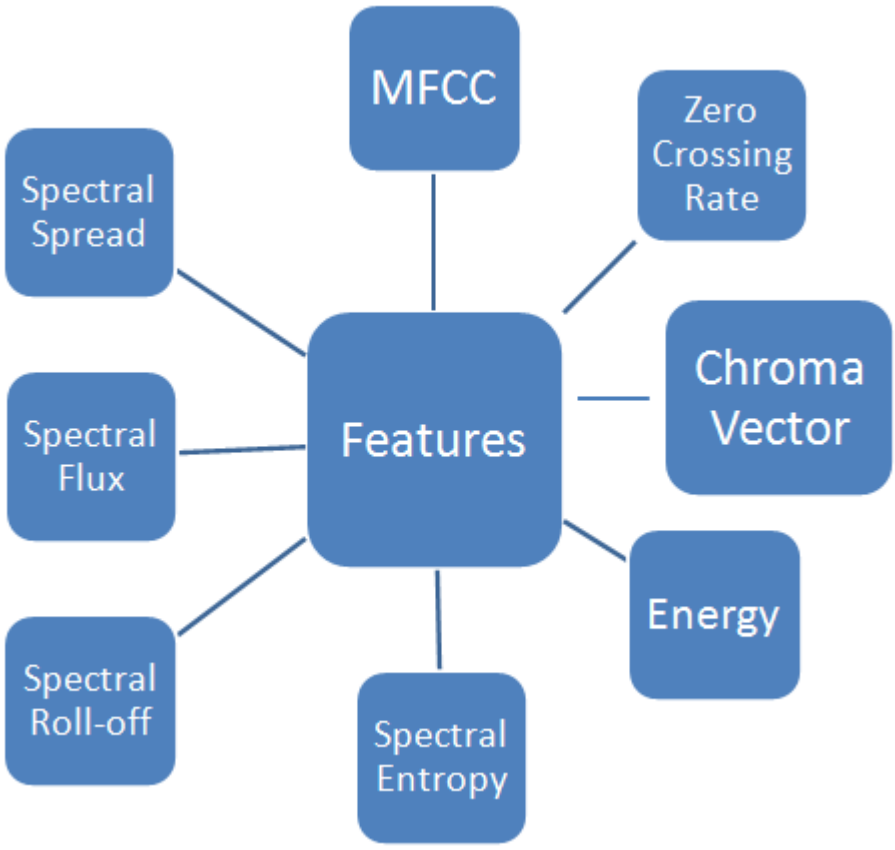


Fig.4. All the features employed for the classification, here only 8 are shown but in total we have 34 features as MFCC's and Chroma Vector each has several components

2. Chroma Vector

Chroma features are an interesting and powerful representation for audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave. Since, in music, notes exactly one

octave apart are perceived as particularly similar, knowing the distribution of chroma even without the absolute frequency (i.e. the original octave) can give useful information about the audio and may even reveal perceived musical similarity that is not apparent in the original spectra.

We investigate using chroma features, designed to reflect melodic and harmonic content and be invariant to type of audio/music. Chroma features contain information that is almost entirely independent of the spectral features.

Like MFCC's , **Chroma vectors** also have **12 components** , thus make up 12 features.

FEATURE EXTRACTION

Once we know which features to extract for our application, we need to find a way to extract these features from the signal itself. As each of these features on a basic level are simple mathematical formula's and each signal is a vector of numbers (1-D in time domain , 2-D in frequency domain), there can be two ways to extract these features :

Let the number of samples in a signal be N .

1) Apply the feature formula on the complete length of the signal $\boldsymbol{N}$, this would yield a single value , indicating the value of that specific feature over the length of the signal.

2) Define a buffer of size $\boldsymbol{M}$ where usually ($\boldsymbol{M} << \boldsymbol{N}$) , buffer the signal for that buffer length and calculate the feature over that buffered signal. Perform this method of buffering and feature measurement for each buffer over the whole signal length with a fixed percentage of overlap with the previous buffered signal. This method yields a 1-D feature vector with number of elements greater than $\frac{N}{M}$. We extract features using the second method for two reasons :

    1) It assumes the signal is statistically static within that small duration, reflecting the changes that occur in the features of the signal during each buffered

interval, thus yielding more information as compared to the first method. The overlap with a previous buffer helps maintain continuity over the feature vector.

2) Such buffering helps simulate a real time system. In a real time system continuous data is not present, rather it is buffered. Thus if such a system is ever made real-time, it is easier for us to make sense of this application in real-time.

   The first method, clearly doesn't have these advantages. Less information about the features are generated if we employ the first method, which is not favorable for a audio classification application.

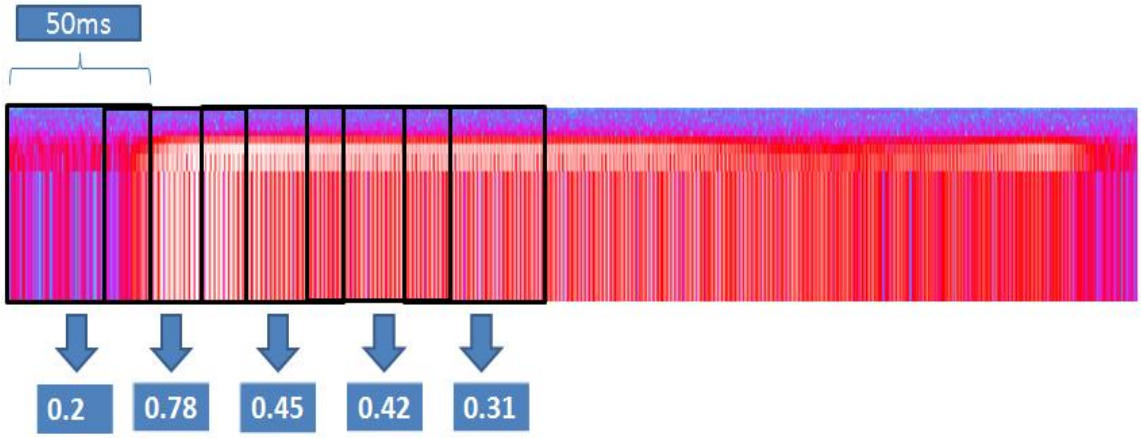


Fig.5. Displays the buffering of a spectrogram signal, each box is 50ms in duration. Bottom of the image shows the feature values for each buffer. Observe the overlap of each buffer with the previous one.

For our application $\frac{M}{Fs} = 50ms$, with an overlap of **50% (25ms).**

Fs is the sampling rate of the recorded signal.

Feature Dimensionality Reduction.

From the previous section , we know that a single feature vector is 1-D.

0.51	0.18	0.14	0.001	0.004	0.63	0.22	0.23	0.45	0.33

Table 1. Displaying an example of a 1-D vector, in this case an example of energy vector

But when the 34 features over a time duration are calculated and are combined to get a complete feature vector we end up with 2-D vector .

Feature/Window	0-50ms	25-75 ms	50-100 ms	75-125 ms	100-150 ms	125-175 ms	150-200ms
Energy	0.51	0.18	0.001	0.004	0.63	0.22	0.23
Zero Crossing Rate	0.11	0.43	0.78	0.91	0.6	0.5	0.20
Spectral Centroid	0.001	0.002	0.1	0.5	0.003	0.45	0.67
Spectral Entropy	0.66	0.343	0.65	0.90	0.77	0.18	0.54

Table 2. A complete feature vector (2-D) which is a combination of all the features over a time duration.

Now as the classifiers take a 1-D feature vector as input, the above 2-D feature vector is inappropriate for feeding into the system. Thus we need to find a way to reduce this 2-D vector into a 1-D feature vector.

For our current system, two techniques are tried and either one of them is used as both reveal similar results upon classification. The two techniques are

1. **Simple Averaging** over the time duration, revealing a 1-D feature vector , with 34 rows(number of features) and 1 column .
2. Dimensionality reduction using **Principal Component Analysis**. Using the first principal component of the vector obtained after performing a singular value decomposition. This also reveals a 34x1 feature vector.

Both the methods give a similar performance, thus in most cases we employ the simple averaging method as it is computationally inexpensive.

CLASSIFIERS

With an overwhelming number of classifiers already at our disposal, it can be difficult sometimes to chose a specific classifier for the application. The choice of the classifier can depend on many different parameters such as **training data size(usually matters when data size is small), computational efficiency, classifier complexity, computational power** at hand. Even after having the knowledge of these parameters, it can be difficult to pick out a single classifier. Due to these choice issues, two very different classifiers were chosen for the acoustic classification purpose ,

A) SVM(Support Vector Machines)

Support Vector Machines is a non-probabilistic classifier, which is inherently a linear classifier (employs a hyperplane to separate training data into different categories) but can be used for non-linear classification by employing kernel trick.

A SVM constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

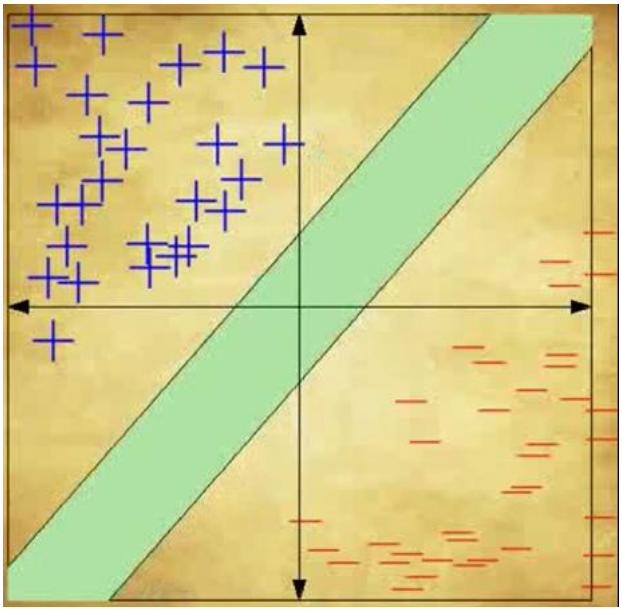


Fig.6 Displays the hyperplane (green), separating two classes (blue and red). The co-ordinate system represents a 2-D feature space

Advantages of using a SVM are follows : -

1) It is effective in high dimensional spaces.
2) Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
3) Executes much faster than most of the classifiers during both training and testing periods.
4) It is versatile, as you can employ application specific kernel functions. Kernels can be specified with the decision function.

Disadvantages of SVM are :-

1) Works really well with separated classes, but problems arise when a specific kernel function cannot be specified due the classification complexity.
2) As the number of classes go up, you could expect a significant drop in performance (we will see that in the results section).

A) Random Forest Classifier

Random forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forests correct for decision trees' habit of over-fitting to their training set.

The training algorithm for random forests applies the general technique of **bootstrap aggregating, or bagging, to tree learners.**

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if *bootstrap=True.* (In python)

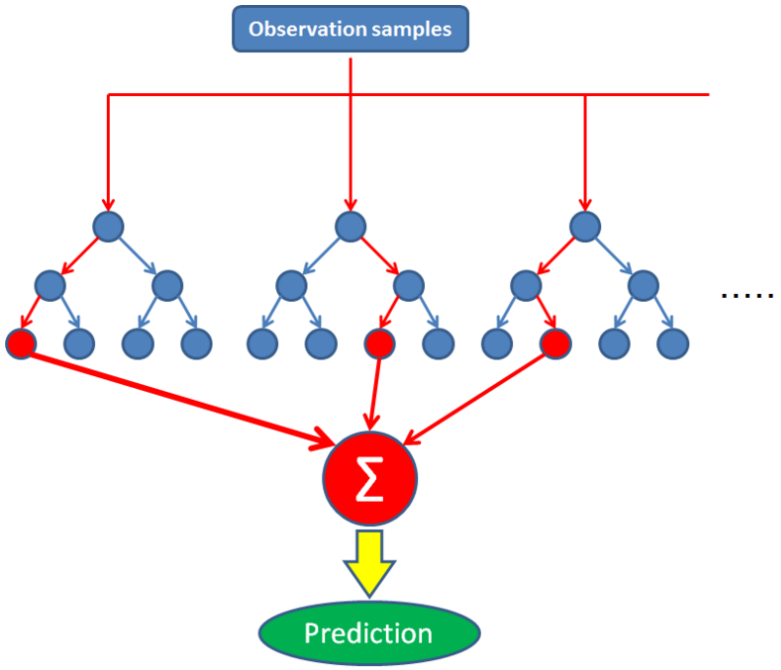


Fig. 7 An illustration of the decision trees employed by random forest classifiers.

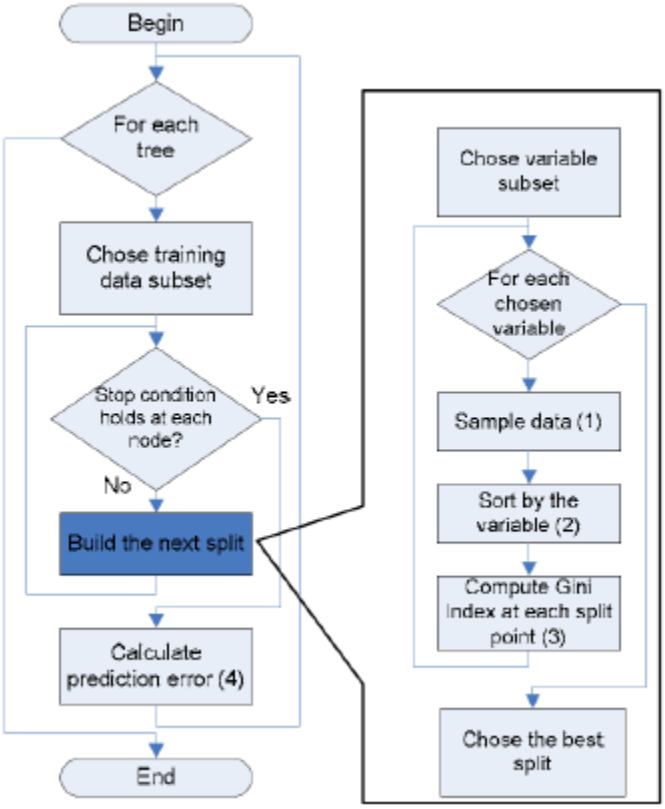


Fig.8 A basic algorithm flow of Random Forest Classifiers

Advantages for using RFC are :

1) It is one of the most accurate learning algorithms available. For many data sets, it produces a highly accurate classifier.
2) It runs efficiently on large databases.
3) It can handle thousands of input variables without variable deletion.
4) It gives estimates of what variables are important in the classification.

Disadvantages :

1) Random forests have been observed to over-fit for some datasets with noisy classification/regression tasks
2) Takes longer time to execute training and testing steps as compared to SVM's.

CLASSIFIER TESTING

This is the most important section, once we have our classifier architecture in place.

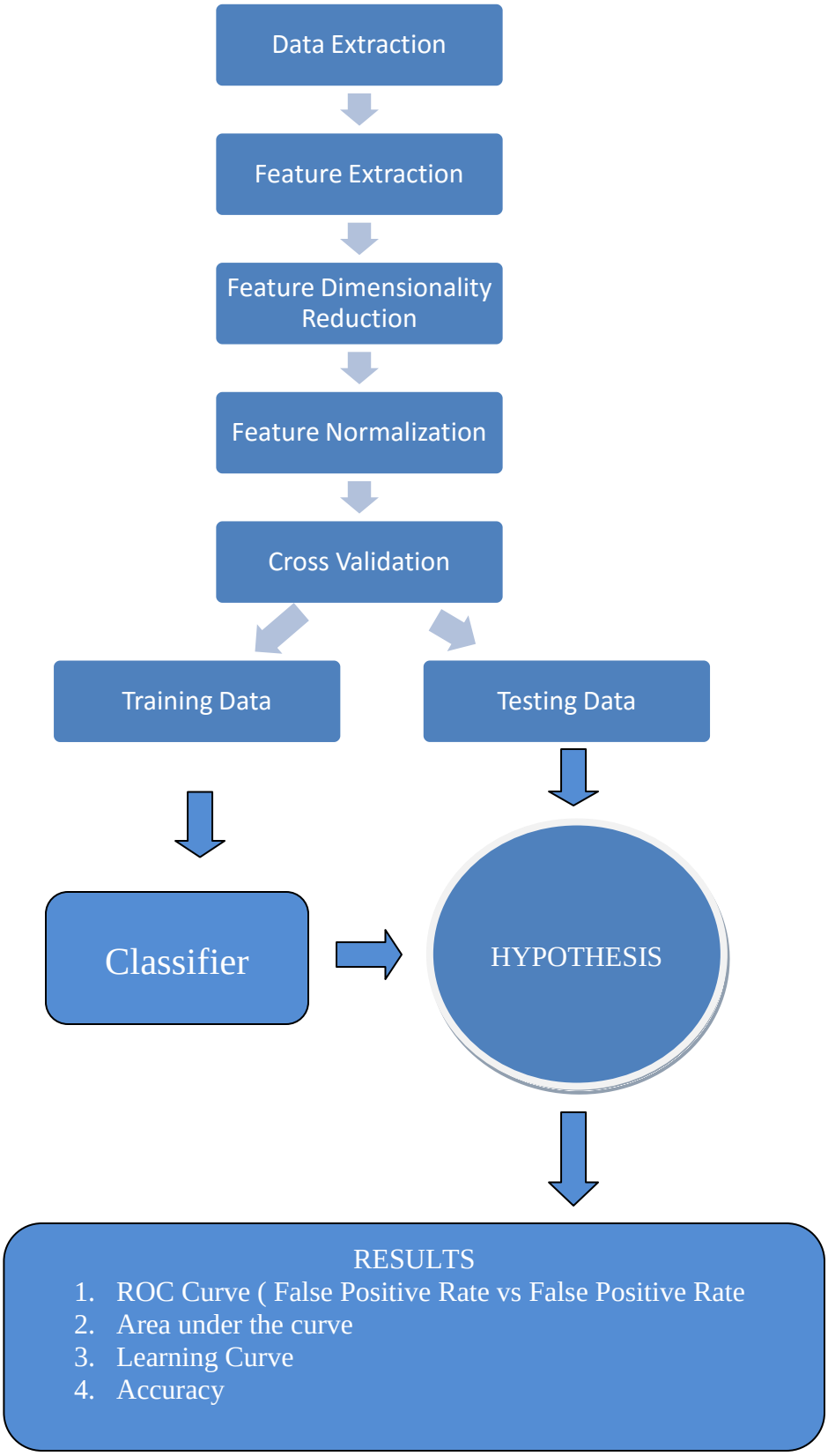
The two most important steps during classification testing and evaluation are

1) Cross Validation - To divide a data-set into a training and a testing set, when a separate testing set is not available.

2) Evaluation Parameters- One should know what evaluation parameters are being calculated to measure the performance of the classifier. For our application we measure :

    a) ROC Curve and Area under the curve
    b) Accuracy
    c) Learning Curves to study the variance and the bias

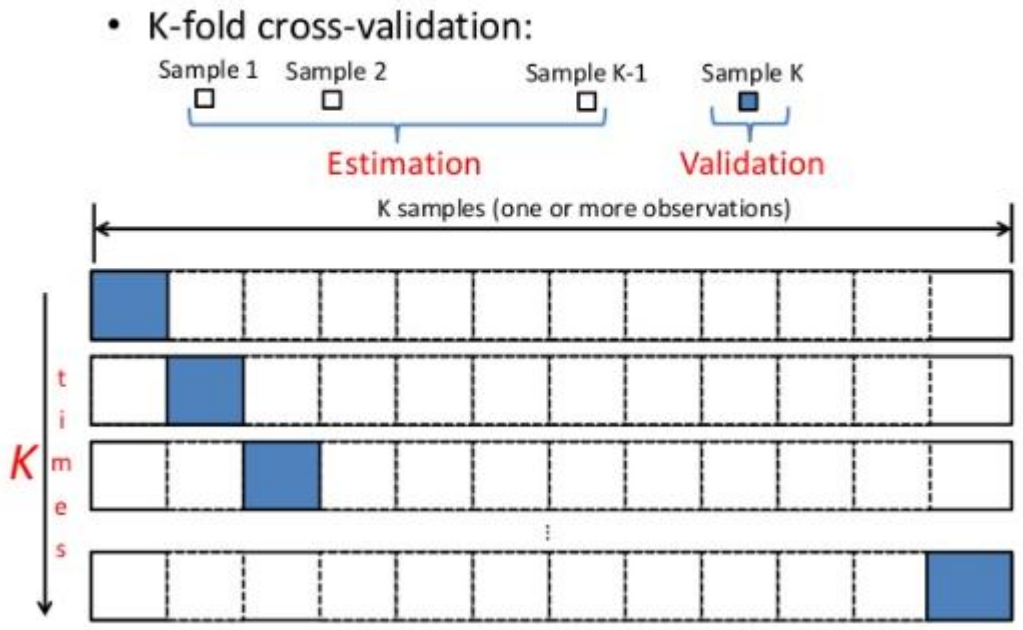
The following flowchart explaining the workflow of the classification system is given below :


Data Extraction
Feature Extraction
Feature Dimensionality Reduction
Feature Normalization
Cross Validation
Training Data
Testing Data
Classifier
HYPOTHESIS
RESULTS
1. ROC Curve (False Positive Rate vs False Positive Rate
2. Area under the curve
3. Learning Curve
4. Accuracy

Cross Validation

As the classifier requires a training set to build a hypothesis and then a testing data-set to predict the performance of the classifier, the given data (complete feature space) , needs to be divided into a training set and a test set when a separate test set is not available. Cross validation is a way to do so .

We employ a K-fold cross validation in python , the following image depicts the process :



The estimation samples consist of the training data , whereas the validation sample forms the test data. The parameter K is the number of iterations of different test data selections out of the complete data-set. The size of the test data is given as a fixed percentage of the total data. This method helps to validate the performance of a classifier using a single data-set itself (In case a separate test data-set is not available).

For our applications , the iterations were varied between **10-50** , and percentage of test data between **20-40 %.** This variation was just to test how the classifier performs at different parameters.

Once the data has been divided into a training data and a test data, the training data is fed into the classifier, the following parameters were used for testing purposes in python .

For **SVM** , we employ SVC python function within sci-kit learn , gamma is set at auto(optimal performance was at **0.000599**) and the **penalty factor C** is set at **100** . A **radial basis function (RBF**) kernel was used for the application.

For **Random Forest Classifier**, only the number of estimators was set equal to 120. Everything else was kept as default. For 120 estimators , the classifier gave an optimal performance. This was set by trial and error method

Average precision scores and **predicted probabilities** were used as decision function for the classification task, while evaluating the performance of the classifiers.

DATA AUGMENTATION – Artificial Data Creation

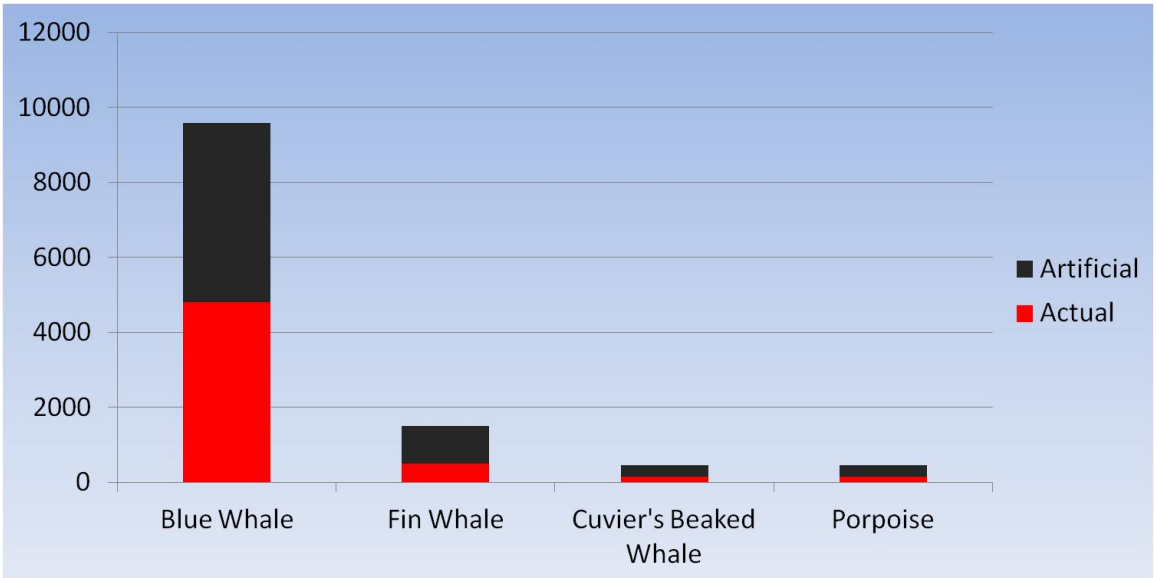


Fig.9 Displays the number of **cetacean calls** (y-axis) vs each cetacean species being classified

Inspired by a deep speech technique employed by the **Baidu Research group** , a step of data augmentation by creating artificial data from the existing data was employed to prevent the issue of over-fitting. As our data-set is small in size, there was a possibility that classifiers over-fit the data and yielded acceptable results. In addition, we were interested to see, if the artificial data would remove the high variance expected in case of

small training data-sets. The artificial data was created by superimposing random noise from hydrophone recordings from the seven different locations. This process was programmed to be completely random. By this process we created a noisy copy of the data-set, thus doubling our data-set. We will see in the results section how this data-augmentation step affects the system.

RESULTS

Our classification system is built to classify four cetacean species calls : blue whale and fin whale (low-frequency) , Cuvier's beaked whale and porpoise (high-frequency calls). Sperm whale classification is not inculcated into the system right now.

A more challenging problem is to classify species calls which lie within the same frequency range, i.e. it is much more difficult to classify two different low frequency calls or two different high frequency calls as compared to a high and a low frequency call.

Thus our initial tests are conducted by building a binary classifier to classify blue and fin whale vocalizations. Our results for both random forest and SVM are as follows :

(Blue and Fin Whale) Low Frequency Classifier- 34 features		
Metrics\Classifiers	Random Forest Classifier	Support Vector Machine
Accuracy	95.2%	95.1%
Average Precision	0.991	0.982

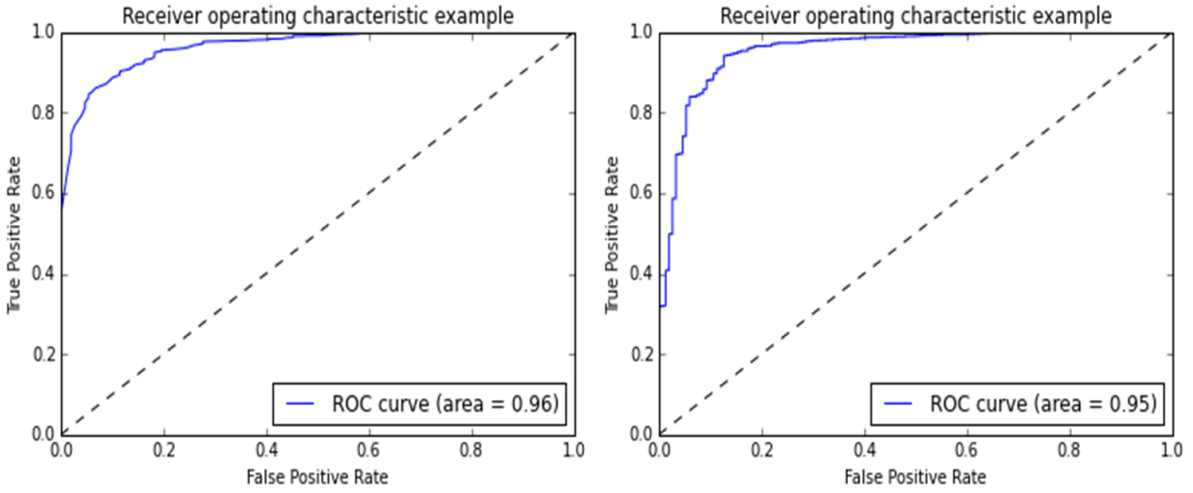


Fig. 10 Blue and fin whale classification results

The binary classification results reveal the following results :

1) **High true positive rate for a low false positive rate**
2) **High accuracy (~95%)**
3) **High area under the ROC curve**

All the properties observed above reveal that the classifier has performed really well in its task. The performance of both the classifiers is similar and this is expected as it's a fairly simple classification in terms of number of classes. The problem is complex in the sense that, the two calls overlap in the frequency domain but it seems that the features extracted have played really well in such a scenario.

On receiving good results for the binary classification, the system is tested for the four cetacean species calls, the results are as below

(Blue, Fin, C.Beaked, Porpoise) Classifier- 34 features		
Metrics\Classifiers	Random Forest Classifier	Support Vector Machine
Accuracy	93.2%	91.1%
Average Precision	0.985	0.961

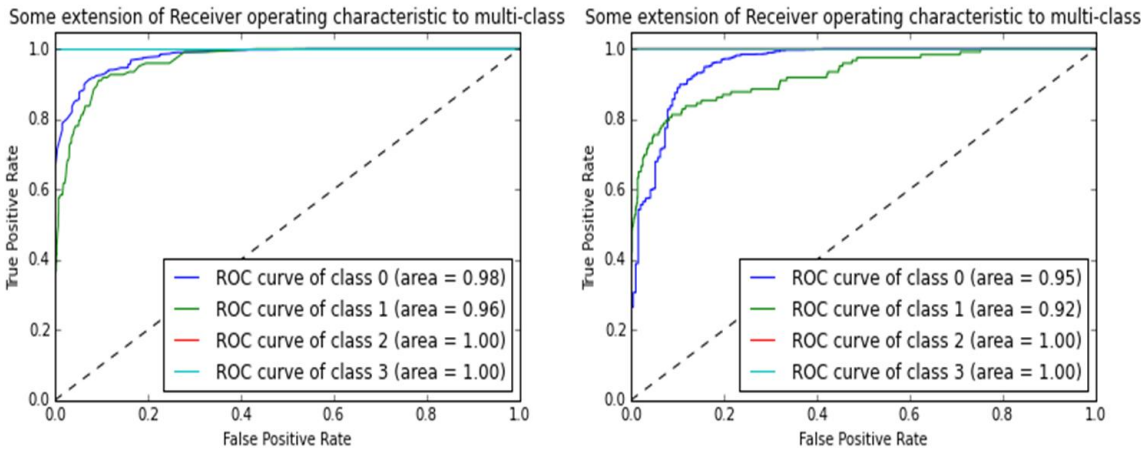


Fig 11. Classification results for four cetacean species calls.

From the above results, we observe that random forest classifier outperforms SVM to some extent, the difference is seen in the area under the curve values. But still the classifiers have performed really well.

Another experimentation was done by removing some **features** which were thought be be **redundant**, leaving us with 15 significant features. The results after removing those features are as follows :

(Blue, Fin, C.Beaked, Porpoise) Classifier- 15 features		
Metrics\Classifiers	Random Forest Classifier	Support Vector Machine
Accuracy	93%	87%
Average Precision	0.98	0.92

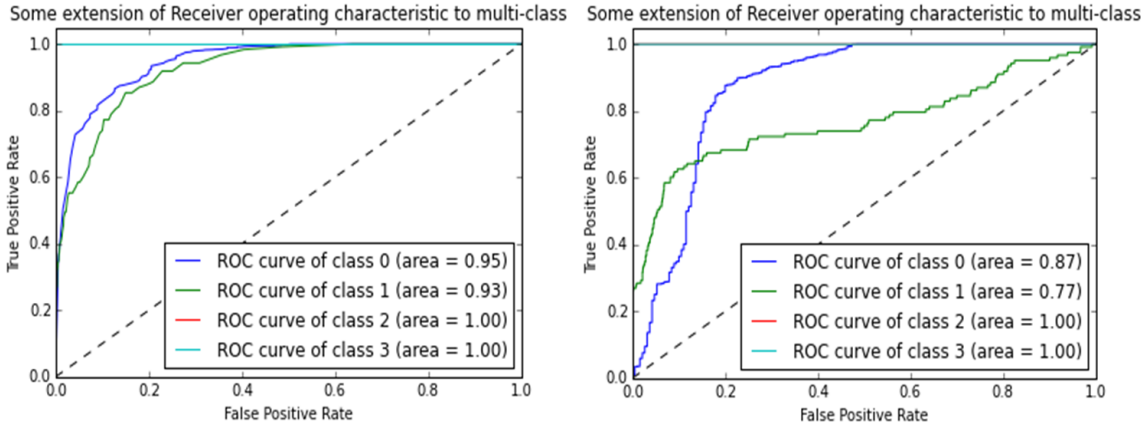


Fig.12 Classification results when thought-to-be redundant features are removed

There is a sudden drop in the performance of the SVM as seen in Fig.12, whereas the random forest classifier still performs really well and quite similar to when 34 features are employed. This shows to some degree that random forest classifier proves to be a better classifier than SVM in some specific situations like these. Using 34 features is not very efficient , thus in case of computational limitation where we cannot use too many features , random forest classifier might be a clear choice.

Finally we present the learning curves for both the original data and after data augmentation.

(Blue, Fin, C.Beaked, Porpoise) Classifier- Data Augmented		
Metrics\Classifiers	Actual Data	Actual + Artificial Data
Accuracy	93.2%	94.7%

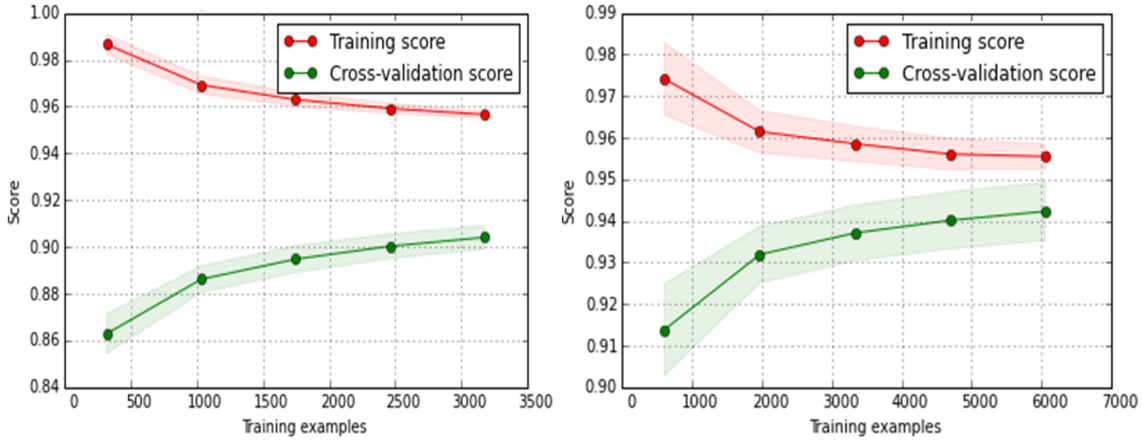


Fig.13 Learning curves before and after data augmentation

The above learning curves clearly show that , the variance in the data-set has reduced and accuracy increases as the number of data-samples are doubled , which is favorable for the system. Thus the process of creating artificial data might help with the overall performance of the classification system

CONCLUSION AND FUTURE WORK

In this paper , an automatic acoustic classification system is presented, which can classify four specific cetacean species calls. The different components of the classification system namely, data-set, data extraction, feature extraction, data augmentation, classification and evaluation are described giving the exact specifications used for each component. Finally the performance of two different classifiers are compared in different classification

scenarios, also the effects of data-augmentation on the classification performance is presented.

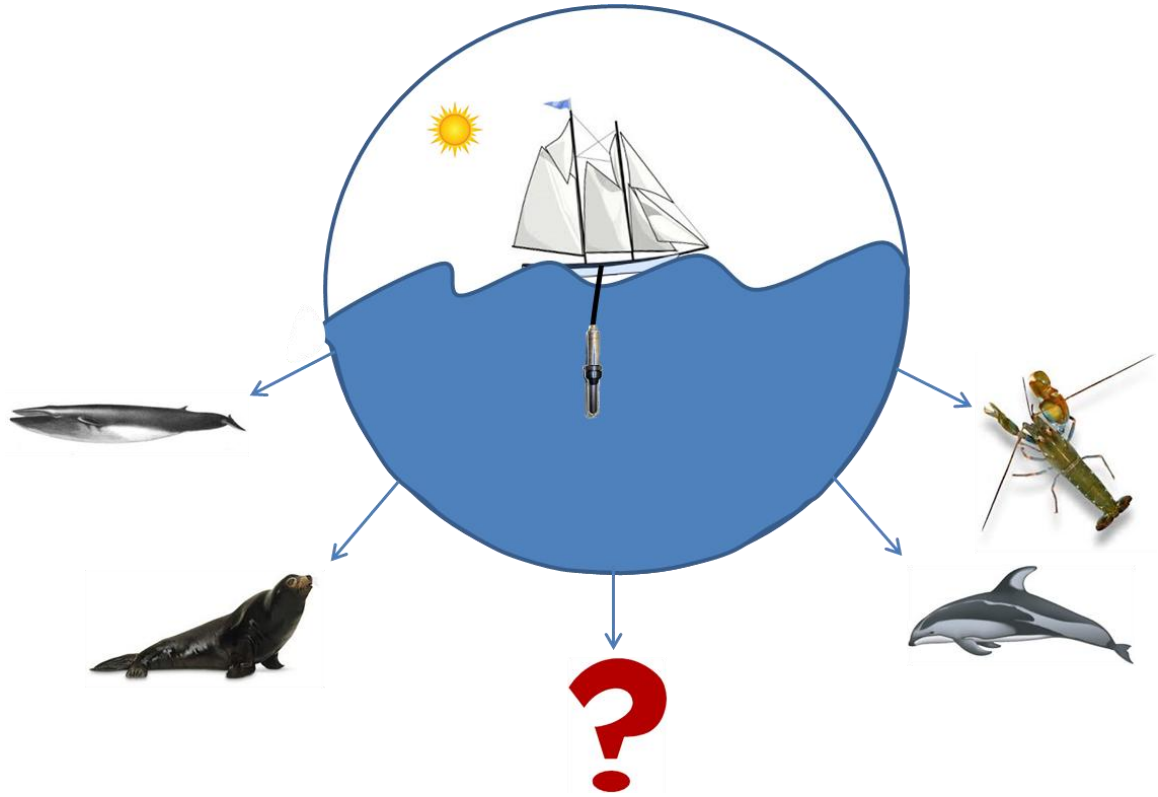


Fig.14 Depicting the idea of the soundscape

As a part of the future work we would like to work towards building a soundscape for the Monterey Bay. Our knowledge about the creatures which communicate using vocalizations is still limited, making it imperative to study these acoustic communication networks. Thus using the hydrophone deployed at MARS we would like to study different vocalizations (primarily cetaceans calls) over seasons and maybe integrate the classifier with the incoming data, once we know how we want to preprocess and store the data so that it is compatible with the classifier itself.

References

[1] D.K Mellinger and C.W Clark (2000) . Recognizing transient low-frequency whale sounds by spectrogram correlation. Acoustical Society of America [S0001-4966(00)01706-9]

[2] J.R. Potter , D.K Mellinger and C.W Clark (1994) . Marine mammal call discrimination using artificia1 neural networks. Acoustical Society of America, VOl 93 No.6

[3] A Hannun, C Case, J. Casper, B. Catanzaro, G Diamos, E Elsen, R Prenger, S Satheesh, S Sengupta, A Coates, A.Y Ng (2014) . Deep Speech: Scaling up end-to-end speech recognition Baidu Research.

[4] D.K Mellinger (2004). A comparison of methods for detecting right whale calls. Canadian Acoustics, 55- Vol.32 No.2

[5] D.K Mellinger, K.M Stafford, S.E Moore, R.P Dziak and H Matsumoto (2007). An overview of fixed passive acoustic observation methods for cetaceans. Oceanography, Vol.20, No.4

[6] D.K Mellinger (2008). A neural network for classifying clicks of Bainville's Beaked whales. Canadian Acoustics, 55- Vol.36 No.1

[7] F Pace , F Bernard , H Glotin, O Adam, and P White (2010). Subunit definition and analysis for humpback whale classification. Elsevier Journal of Applied Acoustics 1107-1112

[8] Cazau, D. , Xue, C. , Doh, Y , Glotin, H, and Adam, O (2013). Scattering representation for humpback whale vocalizations: applications to their detection, characterization and classification. 6th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics.

[9] G Lara , R. Miralles and A Carrión (2013). Right Whale activity detector and sound classificator using Mel-frequency Cepstral Coefficients. 6th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics.

[10] M.A. Roch, A Širović and S B Pickering(2013). Detection, Classification, and Localization of Cetaceans by groups at the Scripps Institution of Oceanography and San Diego State University. [Detection, Classification, Localization, and Density Estimation Workshop 2013.

[11] C Erbe and A.R King (2008). Automatic detection of marine mammals using information entropy. J. Acoust. Soc. Am. 124, 2833

[12] M Bittle and A Duncan (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. Australian Acoustical Society Proceedings of Acoustics 2013 – Victor Harbor.

[13] Moore, S. E., Stafford, K. M., Mellinger, D. K., and Hildebrand, J. A. (2006). "Listening for large whales in the offshore waters of Alaska," BioScience 56, 49–55.

[14] M.F Baumgartner and S.E Mussoline (2011). A generalized baleen whale call detection and classification system, J. Acoust. Soc. Am. 129 (5) , Pages: 2889–2902

[15] Van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., and Van Opzeeland, I. C. (2009). "Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales," Mar. Ecol. Prog. Ser. 395, 21–36