Underwater Audio Event Detection, Identification and Classification Framework (AQUA)

by

Gorkem Cipli
B.Sc., Yeditepe University, 2004
M.Eng., Yildiz Technical University, 2007

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

Underwater Audio Event Detection, Identification and Classification Framework (AQUA)

by

Gorkem Cipli
B.Sc., Yeditepe University, 2004
M.Eng., Yildiz Technical University, 2007

Supervisory Committee

Prof. Dr. Peter F. Driessen, Supervisor
(University of Victoria, Department of Electrical and Computer Engineering)

Dr. Wyatt Page, Departmental Member
(University of Victoria, Department of Electrical and Computer Engineering)

Dr. Farook Sattar, Departmental Member
(University of Victoria, Department of Electrical and Computer Engineering)

Dr. George Tzanetakis., Outside Member
(University of Victoria, Department of Computer Science)

Supervisory Committee

Prof. Dr. Peter F. Driessen, Supervisor
(University of Victoria, Department of Electrical and Computer Engineering)

Dr. Wyatt Page, Departmental Member
(University of Victoria, Department of Electrical and Computer Engineering)

Dr. Farook Sattar, Departmental Member
(University of Victoria, Department of Electrical and Computer Engineering)

Dr. George Tzanetakis., Outside Member
(University of Victoria, Department of Computer Science)

ABSTRACT

An audio event detection and classification framework (AQUA) is developed for the North Pacific underwater acoustic research community. AQUA has been developed, tested, and verified on Ocean Networks Canada (ONC) hydrophone data. Ocean Networks Canada is an non-govermental organization collecting underwater passive acoustic data. AQUA enables the processing of a large acoustic database that grows at a rate of 5 GB per day. Novel algorithms to overcome challenges such as activity detection in broadband non-gaussian type noise have achieved accurate and high classification rates. The main AQUA modules are blind activity detector, denoiser and classifier. The AQUA algorithms yield promising classification results with accurate time stamps.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

Prof. Peter F. Driessen His incredible encouragement, mentoring, and support. Thanks for finding funds to achieve finishing my research. His feedback always made things better. Thanks for believing in me.

Dr. Farook Sattar His knowledge, creativity, mentoring and patience. Thanks for your endless support and never quitting on me.

My committee member, Dr. George Tzanetakis Providing valuable feedback on how to construct this framework.

Tom Dakin Not only his technical and financial support, but also his valuable motivation to continue my research. Thank you for believing in AQUA.

Kristen Kanes Her annotations and dataset support.

Ocean Networks Canada Their technical and financial support.

Ilker Manap His software support.

Mohamad El-Hage For hiring me at Blackberry and his encouregement on the patent filings. Thank you for believing in me.

Karl Scheffer, Gershom Birk For hiring me at PMC Sierra and for all positive support.

My Mom, Dad, and my sister Always being there for me and for giving me all the positive energy.

Allison Brock Her support on proof reading my dissertation and encouragement.

Jocelyn Farmer Taking care of my dog Whiskey and her positive support

Felicitas family and friends : Michael Shamus Murray, Lyle Harrison, Ben Scotney, Dustin Spencer, Colin Pate, Colin Hender, Kathleen Fawley, Kyle Rubin, Kyle James, Tori Davies, Elise Matzanke and John Fulton. Thank you for the amazing on-campus job with an amazing crew.

Friends: Erkan Ersan, Jonaton Reaume, Megan Saunders, Stephen Harrison
Their friendship, positive encouragement and support.

My puppy, Whiskey Keeping me healthy and my heart warm.

This document was typeset in TeXstudio 2.11.0 " on Windows 10. Simulation results were obtained from Python-based simulations using the NumPy and SciPy libraries. Most plots were generated using the MATLAB. Some plots were generated using Sox. Technical drawings were created using Inkscape, or Microsoft Visio.

Sometimes its the very people who no one imagines anything of who do the things no one can imagine.
The Imitation Game

DEDICATION

Dedicated to my mom

Chapter 1

Introduction

The focus of this work is the development and evaluation of algorithms to detect and classify marine mammal sounds in the Northern Pacific Ocean. To assist in our research efforts, Ocean Networks Canada (ONC) agreed to provide us with access to their underwater recording database. Ocean Networks Canada operates world-leading ocean observatories for the advancement of science and the benefit of Canada. The observatories collect data on physical, chemical, biological, and geological aspects of the ocean over long time periods, supporting research on complex Earth processes in ways not previously possible. The ONC shared the challenges that they, and other researchers, face in searching their acoustic database, because of the large amount of non-annotated data. In fact, this is an ongoing issue for underwater acousticians and acoustic biologists globally.

Along the west coast of Canada, from the Salish sea to Prince Rupert, there are six Non-Governmental Organizations (NGO) collecting underwater passive acoustic data. The primary reason for the collection of this data is cetacean research, however, underwater noise and its impact on marine mammals, fish, and invertebrates has become increasingly important in the last 5 years. ONC estimates these six organizations are collecting 0.26 PetaBytes of acoustic data per year, which is anticipated to grow as more hydrophones are added along the coast. The Department of Fisheries and Oceans and ON is collecting 100TB of data each year on the west coast of Canada. Commercial organizations, such as ports and fossil fuel companies, are also collecting a large amount of data to support environmental impact assessments. In addition, west coast First Nations, such as the Metlakatla and Gitgaat, are now involved with underwater acoustic monitoring to support their environmental stewardship programs.

At the June 2014 underwater workshop held at the Vancouver Aquarium, the primary impediment to acoustic research was identified as the inability to process all the data being collected. Accordingly, the organizations at the workshop identified the development of, and access to, automated detection and classification software, with both real time and archived data, as their highest priority.

Automated detection and classification software does exist. An open source program named PAMGuard [1] is available, but lacks training on north pacific species. In addition, it is not robust enough to run without manual oversight. Orchive [2], another classification software, has been processing the data from Orca Lab for several years. However, it focuses specifically on northern orca populations. Listening In the Deep Ocean (LIDO) [3], has been processing the ONC hydrophone data for several years. However, it has questionable accuracy, the classifications are not stored, and the algorithms are proprietary. JASCO Applied Sciences Spectroplotter [4] [5] is also proprietary, in that it is only accessible to NGOs as a service provided by the company.

As a result, the North Pacific Underwater Acoustic Research Community remains without a viable detection and classification software solution. Therefore, our underwater audio detection and classification framework called AQUA, has been developed to address this need. We consider AQUA as a fast and reliable underwater audio event detection and classification framework, that generates mammal activity reports to aid in Pacific mammal preservation and provide automated annotation on archived data.

1.1 Dissertation Outline

Each chapter of this dissertation presents our published research, which describes algorithms used by AQUA for detection and classification of underwater audio data.

In Chapter 2, we propose a novel detection method to identify whale activities by using a B-Spline approximation [6] as well as contextual information, to address the problem of generating training and testing datasets for multi-class classification. We argue this method should be able to distinguish recordings that contain mammal activities from those that contain only noise. This will result in the generation of training and testing datasets, that will help to evaluate the performance of different types of classifiers.

In Chapter 3, we evaluate a Hidden Markov Model, utilizing a Gaussian Mix-

ture Model (HMM-GMM) type of classifier with Mel Frequency Cepstral Coefficients (MFCCs) feature extraction method [7] using the generated datasets for humpback whales, sperm whales, and marine vessel sounds. Classifications will be based on the loglikelihood of data distributions in high dimensions. We will start with an HMM-GMM classifier, since they are known to model non-stationary signals. The challenge we face is to determine the HMM-GMM parameters, such as: number of states, number of Gaussians, as well as the parameters for MFCC window size or number of coefficients. We propose a method to adjust the MFCC window size, by observing the classification ratio, while utilizing different number of states and Gaussians in the HMM chain. Additionally, we will apply the b-spline approximation, specifically cubical splines, to Gaussian parameters to find out if it will improve the decision regions. We argue that the classification ratio will improve, based on the finite support nature of the b-spline approximation models. Furthermore, we expand our research to include fin whale calls, and apply our method to recordings that take place over the course of a year for humpback whales, sperm whales, and fin whales.

In Chapter 4, we utilize a multi-classifier fusion technique for the classification ratio [8]. We investigate the performance of entropy-based classifiers with uncharacterized broadband noise (earthquakes, rain) along with other disturbances (power supply noise, pumps, Acoustic Doppler Current Profiler (ADCP) pulses). These noise components have been known to frequently change the ambient noise power spectral density, which can also cause data clusters to become more scattered. This can result the degradation in performance of the classification ratio. Accordingly, we will use the maximum loglikelihood of the Gaussian distributions in an HMM chain, to see if it improves the classification ratio. It is expecting that this new multi-class classifiers framework based on multiple classifiers fusion will be promising over different individual classifiers under diverse conditions (e.g., different data sources, different feature sets).

In chapter 5, we propose a method to identify whale activity regions, and produce a perceptually better sound. Finally, we present a modified filterbank method to extract activities specific to humpback whales, sperm whales, and fin whales. We argue that this will help us to extract candidate calls with accurate time stamps for our classifier.

Finally, chapter 6 presents the AQUA framework and its features. We aim to implement such tools, and demonstrate how they can be used by ONC operators.

1.2 Contributions

Our contributions in this research are as follows:

- AQUA is a real time operating framework, which uses live ONC data, and is intended to process the following NGO archives: ORCALAB, PACIFIC WILD, CETACEA LAB, METLAKATLA, BEAM Reach.
- AQUA shows that maximum loglikelihood based HMM-GMM type classifier works better on non-stationary data with uncharacterized noise.
- The use of maximum loglikelihood based HMM-GMM type classifier made it possible to detect unlearned or rare events.
- AQUA's performance on the humpback whale classification ratio is higher than existing systems.
- AQUA introduces an automated data quality assesment, which is lacking in the ONC passive acoustic data.
- AQUA is able to make call-by-call annotations with accurate time stamps for north pacific whales.

Details of the novel contributions to detection and classification algorithms are given in chapter 7. These contributions have attracted the attention of ONC, and as a result, they plan to employ AQUA on the ONC Marine Mammal Avoidance project with Transport Canada.

Chapter 2

A Novel Approach to Low Frequency Activity Detection in Highly Sampled Hydrophone Data Based on B-Spline Approximation Automatic Activity Detection

We present a novel method for detection of low frequency signals less than 100 Hz in hydrophone data sampled at 96 KHz. The work described in this chapter is described in [6]. The low-frequency activities (e.g., particular whale calls) in the hydrophone data are detected based on B-spline approximations of the hydrophone data. The error pattern of the incoming/detected signal and template signal is derived by calculating the MSEs (mean-square errors) between their B-spline approximations and compared with that of the reference signal and template signal. Here, the incoming signal is a detected (new/non-labeled) hydrophone data, whereas the reference signal is the ensemble of labeled hydrophone data and the template is a target signal that controls the detection. In the decision module, the threshold is selected based on the skewness of the error patterns. The performance of the method is evaluated using real recorded hydrophone data showing promising results. Nowadays, most of the underwater audio recording systems are operating with very high sampling rates creating large volume of hydrophone data, although the special attention is paid in monitoring the activities (such as acoustic events) in low frequency bands. The sound information

captured using a hydrophone plays an important role in monitoring events occurring under ocean. The term acoustic event here represents a short audio segment, which has activities that are rarely occurring and unpredictable in time. Acoustic (rather than visual) monitoring is used primarily for underwater investigations because acoustic waves can travel long distances in the ocean. Visual monitoring is useful only for short range observations up to several tens of meters in depth at most, and is not suitable for monitoring whales or shipping, which may be many kilometers away. The task of detecting acoustic events from hydrophone data is difficult as this type of data is usually noisy and highly correlated. The rare event detection algorithms usually operate in a supervised or semi-supervised manner as they need to be trained with a large set of annotated data (i.e., pre-classified by experts) which makes it quite challenging to create training datasets from the highly sampled noisy hydrophone data. For rare acoustic events detection, the work reported in [9] has proposed a semi-supervised approach based on the adaptive Hidden Markov Model (HMM) by first learning the usual event models from a large set of (commonly available) training data followed by learning the unusual event models through Bayesian adaptation in an unsupervised manner. The rare event detection method in [10] robustly approximates the background for the complex audio scene using the Gaussian mixture model based on the proximity of the distributions determined by entropy. A machine learning and descriptor based supervised rare event detection approach is presented in [11] using support vector novelty detection. Proposed method is used to distinguish 5 min. long recordings that have activities from the ones with noise only.

2.1 The Basic Idea

The proposed low-frequency activity detection scheme is based on the comparison of two error patterns generated by B-spline approximations of the highly sampled real ONC hydrophone data, as depicted in Fig. 2.1. The template has the characteristics of a known signal used as side information, whereas the reference signal is an ensembled average of the observed noisy hydrophone signals. The error pattern is derived based on the mean-square errors (MSEs) between the two B-spline approximated signals at various sampling frequencies within a defined low-frequency band. Finally, the two generated error patterns are used in the decision module to decide whether the new observed data contains the target event or not (see Fig. 4.1). It is worth mentioning that B-splines can be superior to a conventional lowpass filter in terms of

their flexibility by providing a convenient means to adjust the extent of lowpassing and smoothing [12] to offer good quality approximation.

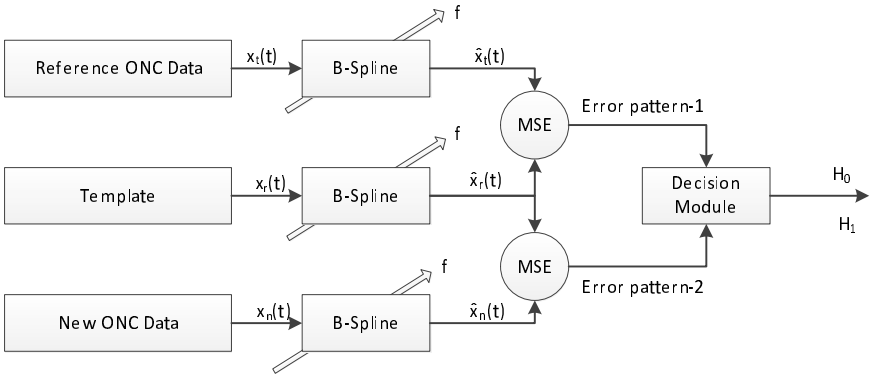


Figure 2.1: The overall block diagram of the proposed detection scheme.

2.2 Method

2.2.1 B-Spline Based Approximation

B-Spline Basis

In this chapter, we use the popular B-spline basis for approximation as it is itself a cubic spline and it has the desirable property of the smallest possible support of any basis for the space of cubic splines. Since B-splines are defined very narrowly, their linear combination is easy to compute and numerically stable. The m-order B-spline with knot sequence $\{k_1, \cdots, k_N\}$ is basically a (m-1)th degree polynomial that is (m-2) times continuously differentiable. The zeroth order discrete B-spline, $B_N^0(k)$, is a rectangular window of width N that is centered with respect to the origin when N is odd. This operator corresponds to a moving average filter of size N that can be implemented recursively[13]. The discrete B-splines of various widths can be constructed from repeated (m+1 times) convolution of simple moving average filters ($B_N^0(k)$) and a correction kernel $B_1^m(k)$ where $'*'$ denotes the convolution[13]. In [13] B-Splines with various widths presented.

$$B_N^m(k) = \begin{cases} \frac{1}{N^m}\left\{B_N^0(k) \underset{m+1}{*} B_N^0(k)\right\} * B_1^m(k), N \text{ is odd} \\ \frac{1}{N^m}\delta_{(m+1)/2} * \left\{B_N^0(k) \underset{m+1}{*} B_N^0(k)\right\} * B_1^m(k), \\ \qquad\qquad m \text{ is odd, } N \text{ is even} \\ \frac{1}{N^m}\delta_{(m+1)/2} * \left\{B_N^0(k) \underset{m+1}{*} B_N^0(k)\right\} \\ *B_1^m(k-0.5), \qquad m \text{ is even, } N \text{ is even} \end{cases} \tag{2.1}$$

B-Spline Model for Approximation

The B-spline approximation model[14] is based on the constraint least-square optimization using the following minimization:

$$\min_\theta \left(\underbrace{(\mathbf{y}-\hat{\mathbf{y}})^T \mathbf{W}(\mathbf{y}-\hat{\mathbf{y}})}_{\text{Original problem}} + \underbrace{\sum_{i=1}^{N} \lambda(i) \left(h''(k_i,\theta)\right)^2}_{\text{Penalty function}} \right) \tag{2.2}$$

where

$$\hat{\mathbf{y}} = \begin{pmatrix} B_1^m(k_1) & \cdots & B_N^m(k_1) \\ \vdots & \ddots & \vdots \\ B_1^m(k_N) & \cdots & B_N^m(k_N) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix}, \tag{2.3}$$

where

$$h(k_i,\theta) = \sum_{j=1}^{N} \theta_j B_j^m(k_i),\ i = 1, \cdots, N \tag{2.4}$$

and $\mathbf{W}$ is a $(N \times N)$ diagonal weighting matrix, $\mathbf{y}$ is the observed data, $\theta = [\theta_1, \cdots, \theta_N]^T$ is the sequence of B-spline coefficients, h'' is the second derivative of h derived as:

$$\begin{aligned} \left(h''(k_i,\theta)\right)^2 &= \left(\sum_{j=1}^{N} \Delta^2\theta_j B_j^m(k_i)\right)^2 \\ &= \sum_{j=1}^{N}\sum_{k=1}^{N} \Delta^2\theta_j \Delta^2\theta_k B_j^m(k_i) B_k^m(k_i) \\ &\approx \sum_{j=1}^{N} (\Delta^2\theta_j)^2 \end{aligned} \tag{2.5}$$

where $\Delta^2\theta_j = \theta_j - 2\theta_{j-1} + \theta_{j-2}$, and $\lambda(i)$ is a smoothing function:

$$\lambda(i) = \beta_0 \ln(i+1) \tag{2.6}$$

where β_0=5000. We found the log-based smoothing function in Eq.(2.6) is the most effective for the B-spline basis. Eq. 2.2 involves the introduction of a variable roughness penalty function to impose additional smoothness in the approximation, where roughness is defined as a departure from local linearity.

Notations

The following notations are used throughout this chapter unless otherwise indicated. Scalars are denoted by small letters, vectors are denoted by small bold letters, and matrices are denoted by capital letter. In addition, the following crucial notations are used:

B_j^m	jth B-spline basis function of order m.
θ_j	jth B-spline coefficient.
k_i	ith elememt of knot-sequence k.
$\hat{\mathbf{y}}$	the approximation of $\mathbf{y}$.
$\mathbf{W}$	weighting matrix.
h''	second derivative of h.
$\lambda(i)$	ith element of λ.
f	normalized sampling frequency.

Parameter Setting

The order of the B-spline function chosen is 4. The frequency range is [10 500] Hz with the lowest and the highest sampling frequencies are 10 and 500 Hz, whereas the increment of the sampling frequency is 50 Hz. (The dataset being used has only activities in between 10 and 500 Hz.)

2.2.2 Reference Background Signal Generation

The reference signal, $\mathbf{r}$ can be taken as the average of $\mathbf{z}_j$ for $J(j{=}1, \cdots, J)$ of the time-synchronized noisy observed signals. It can be obtained by exponential averaging of the time synchronized noisy observed signals [15]. This type of weighted averaging is effective to deal with noisy observed data and obtained recursively as:

$$\begin{aligned} \mathbf{r}_j &= (1-\gamma)\sum_{l=0}^{j}\gamma^{j-l}\mathbf{z}_l \\ &= \mathbf{r}_{j-1} + (1-\gamma)[\mathbf{z}_j - \mathbf{z}_{j-1}] \end{aligned} \tag{2.7}$$

where $\mathbf{z}_j$ is a vector representing the jth observed signal, j is the index of a total of J observed signals, $\gamma(<1)$ is a forgetting factor and $\mathbf{r}_{j-1}$ is the weighted average reference signal at the (j-1)th observed signal input. The reference signal is providing the useful contextual information which indicates the position/location of the events of interest regarding the whale calls in that location.

2.3 Data

Different types of hydrophones are used for different tasks. For example, on the Ocean Networks Canada(ONC)(`http://www.oceannetworks.ca`) observatory, an enhanced version of the Naxys Ethernet Hydrophone 02345 system is used, which includes the hydrophone element (rated to 3000m depth), 40dB pre-amplifier, 16-bit digitizer and Ethernet 100BaseT communication. This particular hydrophone is of high quality, and can be integrated into an existing underwater instrument package. The ONC hydrophone collects data at a constant rate and can generate approximately 5.5 GB of data per day. The sampling frequency of the ONC data used is 96 kHz.

2.4 Results and Evaluation

2.4.1 Reference Signal

The reference signal obtained by exponential averaging with γ=0.9 and J=15, in Eq. (2.7), is plotted in Fig. 2.2.

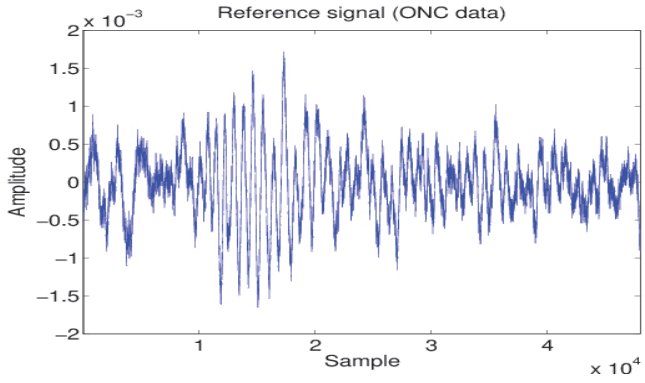


Figure 2.2: The reference signal generated from ONC hydrophone data.

2.4.2 Template Signal

The template signal used in this paper is a particular whale (ORCA whale) call (see Fig. 2.3). The template signal is a preprocessed data referring to a good quality

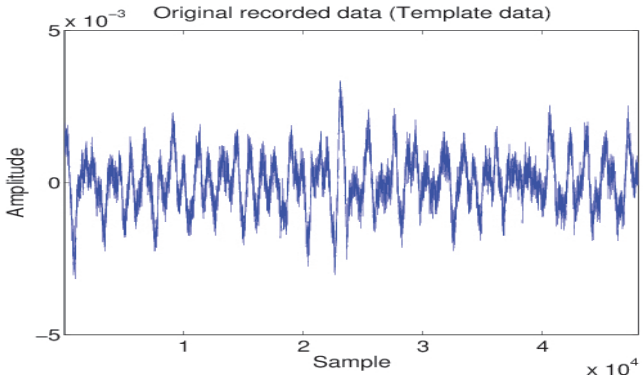


Figure 2.3: The template signal.

whale call, which can have similar characteristics (such as bandwidth, duration) of the observed reference signal.

2.4.3 Error Pattern

The error pattern 1 obtained from the reference signal and the template in terms of MSE versus normalized sampling frequency, is illustrated in Fig. 2.4 with the skewness value of -0.67. Normalized sampling frequency is varied over the observation frequency interval and error patterns are generated between signals in terms of MSE. Fig. 2.4 indicates that we need to use the normalized frequency < 0.005. The error pattern 2 of the ONC noise and the template signal is also illustrated in Fig. 2.4, which has the skewness value of -0.94. The histograms of the skewness of the error patterns for the new ONC signal samples are plotted in Fig. 2.5 for the noisy signals case and the noise-only case. It is noteworthy that the skewness seems the most effective statistical measure when the error pattern departs from more normality (for error pattern 1) to less normality (for error pattern 2) among other possible measures such as mean, median, standard deviation, and kurtosis. Error patterns are extracted by MSE of each bsplined signal as presented in Fig 2.1. The algorithms for the proposed error pattern 1 and error pattern 2 are presented in the Appendix A.

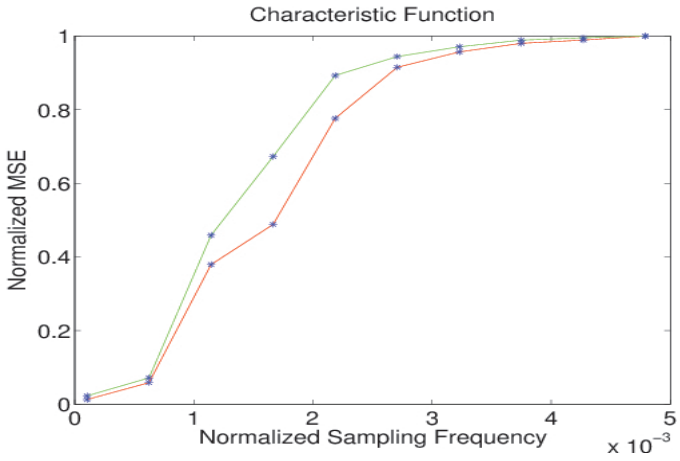


Figure 2.4: The error patterns of the (1) reference signal and the template signal (red) (2) ONC noise and the template signal (green).

2.4.4 Detection Examples

The illustrative plots of the incoming noisy and noise-only ONC data are shown in Fig. 2.6 and Fig. 2.7, whereas the values of the detection index (i.e., skewness of the corresponding error patterns) are presented in Table 2.1. As we see in Table 2.1, the differences of the detection index values between the noisy (signal plus noise) and noise-only data are large for the proposed method, whereas the corresponding entropy values of the ONC data are very small for the acoustic entropy method presented in[16]. Acoustic entropy method calculates the entropy in frequency space.

2.4.5 Detection Performance

The skewness[17] value x of the error pattern of each ONC signal sample is used as the decision variable and the decision rule will be

$$\begin{aligned} H_1 &: x > \gamma \\ H_0 &: x < \gamma \end{aligned} \tag{2.8}$$

where γ is a threshold. The skewness of the incoming signal is utilized for event detection, which assumes the probabilistic character of incoming signal [18], which different for event and non-event cases. Note that here we exploit skewness of the error pattern to differentiate 'noise' and 'signal plus noise' in terms of sparseness. Since the 'noise' only data is more sparse than the 'signal plus noise' data, the former

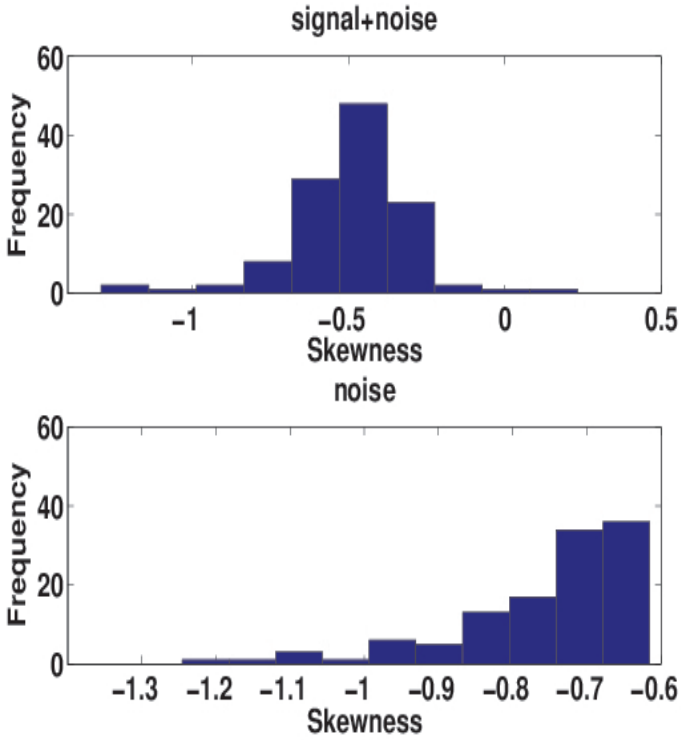


Figure 2.5: The histograms of the skewness of the error patterns for the new ONC samples; (a) signal+noise and (b) noise.

gives the higher negative skewness values. This is caused by the larger mismatch of the error patterns for the noise only case when the noise-only sequence in the low-dimensional (compressive) measurement becomes less coherent, that is more sparse (as the noise has nonzero entries uniformly selected at random)[19, 20], whereas the reference signal can be approximated as a low-rank signal part (more coherent, less sparse) plus a sparse noise part.

The false alarm probability P_F and the probability of detection P_D are used as performance variables. These two probabilities are defined by

$$\begin{aligned} P_F &= \int_\gamma^\infty p(x|H_0)dx \\ P_D &= \int_\gamma^\infty p(x|H_1)dx \end{aligned} \tag{2.9}$$

We have evaluated the performance of the method by using x from ONC data for an approximation of the two conditional probability density functions (pdfs) $p(x|H_0)$ and $p(x|H_1)$.

To obtain the estimates of the two pdfs $p(x|H_0)$ and $p(x|H_1)$, histograms of the

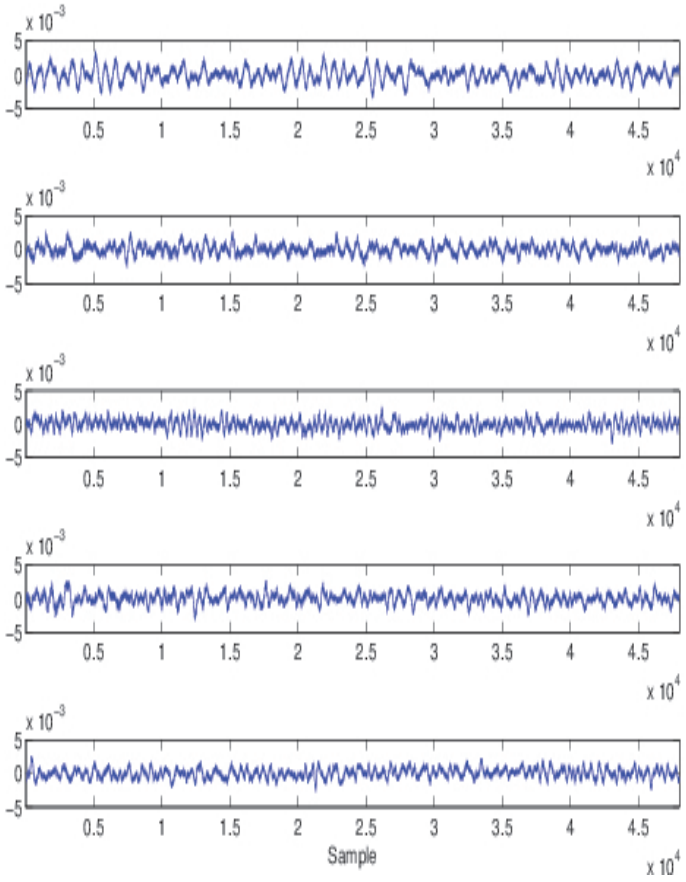


Figure 2.6: Illustrative plots of incoming noisy (signal plus noise) ONC data.

sequences $(x|H_0)$ and $(x|H_1)$ are calculated from the ONC data consists of 265 data samples (115 noisy samples (with events) and 150 noise-only samples). These two histograms are approximated by the normal distributions, given by

$$\begin{aligned} p(x|H_0) &= \exp\{-(x-\mu_0)^2/2\sigma_0^2\}, \quad x \leq 0 \\ p(x|H_1) &= \exp\{-(x-\mu_1)^2/2\sigma_1^2\}, \quad x \leq 0 \end{aligned} \tag{2.10}$$

In Eq. (2.10), μ_0, μ_1 are the means and the σ_0^2, σ_1^2 are the variances for the two distributions, respectively. The parameters calculated in order to fit the histograms in Fig. 2.5 are as follows: $\mu_0 = -0.755$, $\sigma_0 = 0.120$, $\mu_1 = -0.491$, $\sigma_1 = 0.196$, respectively. It can be remarked that the two distributions in Eq. (2.10) are normal (see Fig. 2.8), which is due to the effect of B-spline approximation. To determine the

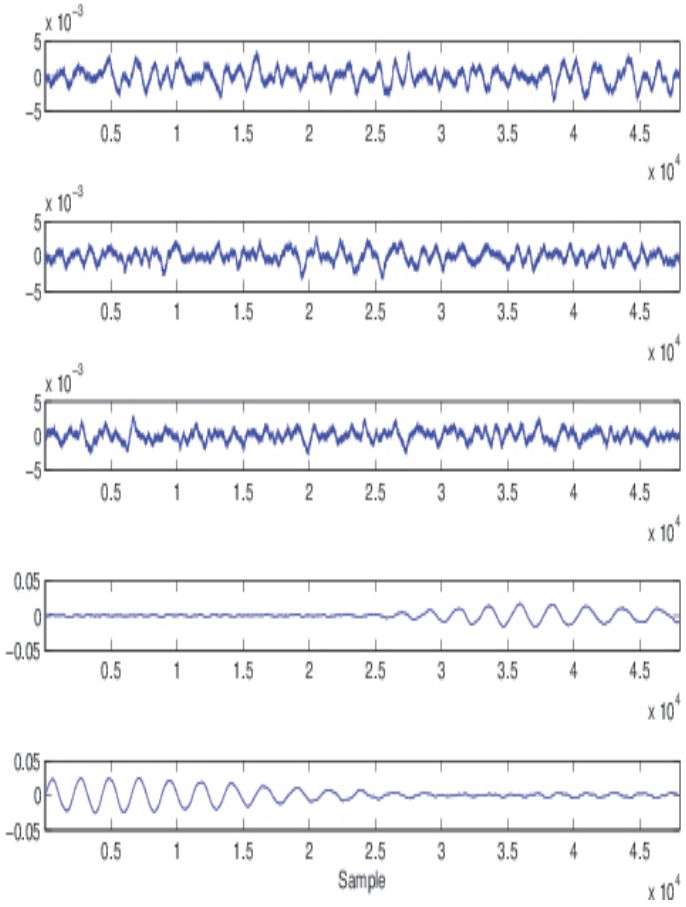


Figure 2.7: Illustrative plots of incoming noise-only ONC data.

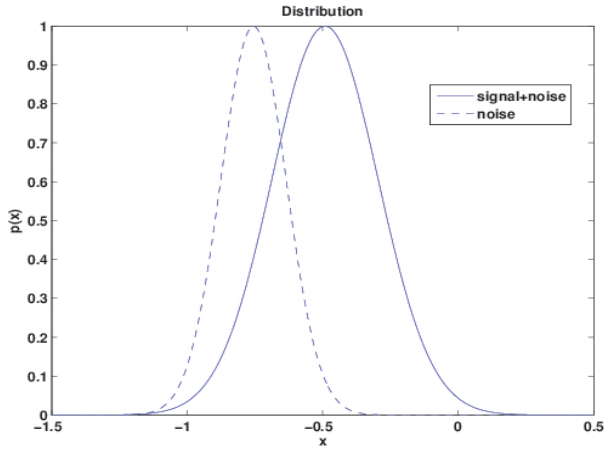


Figure 2.8: The pdfs approximation of the skewness for the detected ONC data.

threshold γ, we set the probability of false alarm $P_F = \alpha$, i.e.,

$$\begin{aligned} P_F &= \int_{\gamma}^{\infty} p(x|H_0)dx \\ &= \int_{\gamma}^{\infty} \exp^{\{-(x-\mu_0)^2/2\sigma_0^2\}} dx = \alpha \end{aligned} \tag{2.11}$$

Table 2.1: The values of detection index for the noisy (signal plus noise) and noise-only ONC data

	Signal + Noise case	
Reference	Proposed	Relative Acoustic Entropy
Fig. 2.6(a)	-0.20	0.85
Fig. 2.6(b)	-0.43	0.87
Fig. 2.6(c)	-0.48	0.90
Fig. 2.6(d)	-0.40	0.86
Fig. 2.6(e)	-0.41	0.88
	Noise only case	
Reference	Proposed	Relative Acoustic Entropy
Fig. 2.7(a)	-0.93	0.86
Fig. 2.7(b)	-0.98	0.85
Fig. 2.7(c)	-0.90	0.85
Fig. 2.7(d)	-1.51	0.84
Fig. 2.7(e)	-1.11	0.87

from which the value of γ can be calculated as:

$$\gamma = \sigma_0 \sqrt{2 \ln(1/\alpha)} + \mu_0 \tag{2.12}$$

The probability of detection is then

$$P_D = P(x|H_1) = \int_{\gamma}^{\infty} \exp^{\{-(x-\mu_1)^2/2\sigma_1^2\}dx} \tag{2.13}$$

After estimating the conditional probabilities in Eq. (2.10), the performance of the method can be obtained by numerical calculations. In Fig. 2.9, the detection performance is presented in terms of receiver operating characteristic (ROC) curves showing the improvement of our detection for the B-spline approximation over the lowpass fil-

tering. The ROC curve for lowpass filtering is obtained by replacing the B-spline with an equiripple linear-phase lowpass FIR filter[21] and varying the cut-off frequency of the lowpass filter within the frequency range [10 500] Hz with a step-size 50 Hz and sampling frequency 96 kHz. The optimal lowpass filter, which has 2400 taps, is obtained by using the Parks-McClellan algorithm[21]. The improvement of the proposed detection method with B-spline can be quantified in terms of AUC (area under the curve), which is higher (0.961) than the AUC with lowpass filter which is 0.880. The skew change for entropy based method is significantly small and cannot be used for detection. As a result region of convergence analysis presented for proposed method and lowpass filter.

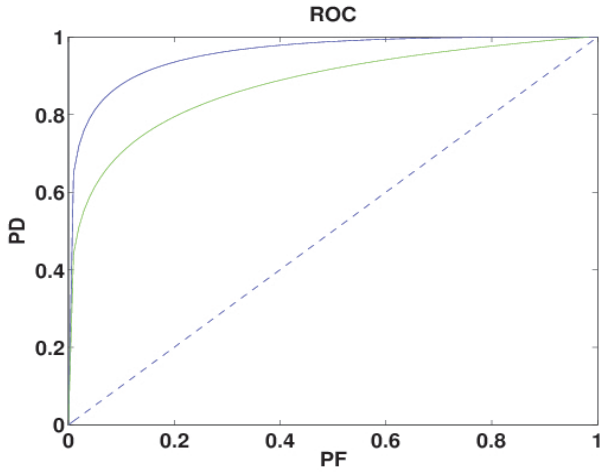


Figure 2.9: The ROC curves of the proposed approach using (1) B-spline approximation (blue), (2) lowpass filtering (green).

2.5 Conclusion

This chapter proposes a novel low-frequency event detection method based on approximations of the hydrophone data. Both the content and contextual information are exploited for our acoustic event detection. The template signal contains the spectral content, whereas the reference signal represents the context which is the circumstance and situation. Moreover, the detection can be performed without knowing the training noise sequence and obtained high detection rate using small sized samples, which is crucial in the case of rare event detection. The proposed method provides a flexible scheme by using a given template,that leads to a reliable detection. It should also

noted that different whale call templates does not change the detection performance. In the future, we would like to evaluate and compare our method with long term datasets and other types of events, for example, dolphin calls and earthquakes.

Chapter 3

Multi-class Acoustic Event Classification of Hydrophone Data Based on Adaptive MFCC Combined with Improved HMM-GMM Topology

The proposed activity detection method in chapter 2, enabled us to prepare 5 min. long training and testing datasets from ONC data. With that dataset, we address the problem of multi-class classification of hydrophone data for acoustic events using low-dimensional features.

A new iterative multi-class classification scheme is proposed based on the combination of adaptive MFCC feature set and an improved HMM-GMM classifier. The adaptive window length for MFCC is important since for acoustic sounds in the ocean, the optimum window length may be different unlike the window length of 16-32 msec., which is optimum for speech signals. Further, in order to increase the classification performance, we perform the B-spline approximation to the generated Gaussians parameters of the multi model HMM-GMM classifier to enhance the separation of the decision region.

Experimental results for the real recorded hydrophone data show that our improved iterative scheme efficiently classifies the acoustic events with high mean accuracy (96%), sensitivity (95%), and specificity (97%). Qualitive definitions for accu-

racy, specifity and sensitivity given in Appendix B.

Bayesian learning based automatic speech recognition systems [22] can be adopted for recognition of acoustic activities in the ocean. Speech is a non-stationary process through the movement of the articulations to be quasi-stationary and it allows the recognition systems extracting the feature vectors over the segments of around 16-32 ms and updating every 8-16 ms. Feature extraction is performed followed by a multi state Hidden Markov Model (HMM) to perform a state transition in every frame (segment), i.e. every 8-16 ms, according to the transition probabilities. The number of states in HMM is defined by the segmentation (frame) size and the HMM stay for a certain time duration in the same state before making a transition. In automatic speech recognition systems, the state durations are usually modelled with certain distributions [23].

The key insight is that while for speech the optimum window length is 16-32 msec, for acoustic sounds in the ocean, the optimum window length may be different. Thus the minimum 'duration constraint' (segmentation/window length) for mammals, boat sounds and other audio related events in the ocean can be adaptively adjusted. This adaptive adjustment can remove the need to use large audio segments for feature extraction, while taking the advantage of having multiple states in HMM-GMM chain to increase recognition/classification performance. Our aim is to adaptively find out an optimum duration constraint for the mel-frequency cepstral coefficients (MFCC) as well as to enhance the HMM-GMM multi-class classification algorithm to achieve better decision spaces.

In our approach, we use MFCC for feature extraction by applying an adaptive window length for the input data. A Bayesian network classifier is used to evaluate the multi-class classification performance. In our approach, we introduce a iterative scheme where our HMM-GMM classifier output gives a feedback to the MFCC feature extraction algorithm to adaptively resize its window length, while incrementing the number of states of the HMM model and tries to maintain a reasonable accuracy rate. We have further improved the traditional HMM-GMM multi-class classifier's performance by applying finite support B-spline approximation to the Gaussian parameters generated in each state. The reason for using B-spline is that it can be superior to conventional lowpass filter in terms of its flexibility by providing a convenient means to adjust the extent of lowpassing and smoothing [24] to offer good quality of approximation. In the multi-class classification of hydrophone data, we have chosen here the HMM-GMM based approach over the other classification approaches for the following

advantages. Firstly, it allows for capturing very long and complex temporal dependencies and secondly it employs a margin maximization paradigm to perform model training, which gives a convex optimization scheme [25]. Moreover, we have chosen the MFCC features over other features since it has a good frequency resolution in the low frequency region, and the robustness to noise is also very good [26, 27]. Moreover, it is simple to calculate and provides Gaussian-like probability density function (pdf), which fits well to the classifier. These make it suitable for the classification of low-frequency events in the noisy hydrophone data. The work described in this chapter is described in [7].

3.1 Methodology

The overall schematic diagram of the proposed scheme is shown in Fig. 3.1. The hydrophone recordings are partitioned first into blocks of defined large segments. Then the 1D data block is converted into a 2D feature map by transforming the data segment from time domain into frequency domain followed by filtering the FFT spectrum through Mel filterbank. The MFCC feature set is then constructed from the feature map by summing its outputs followed by logarithm operation as well as DCT transformation [27] and used as input to the improved HMM-GMM multi-class classifier, which is initialized with 1 state only.

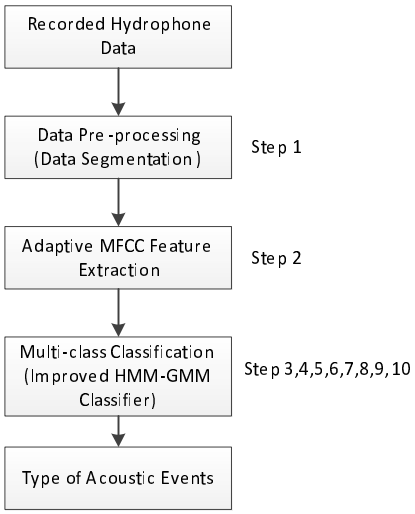


Figure 3.1: The overall schematic diagram of the proposed scheme.

3.1.1 Background Theory

Let us consider a physical observations $\mathbf{X}$ as

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \tag{3.1}$$

where N is the number of $\mathbf{x}_n, 1 \leq n \leq N$ measurements, and the possible target categories or classes $\mathbf{Y}$ as

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_K] \tag{3.2}$$

where the K classes are mutually exclusive. If we denote X as our observation space and Y as our decision space, then the goal of a classifier is to map the observation space to the decision space. Using the fundamental notation of statistical framework, we denote the probability of a measurement vector's $\mathbf{x}_n$ belonging to class $\mathbf{y}_k$ as

$$p(\mathbf{y}_k|\mathbf{x}_n) \quad \begin{array}{l} \text{where } 0 \leq \ p(\mathbf{y}_k|\mathbf{x}_n) \leq 1 \text{ and} \\ \sum_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_n) = 1 \end{array} \tag{3.3}$$

Since this probability can only be estimated after the data has been seen, it is generally referred to as the posterior or a posteriori probability of class $\mathbf{y}_k$. Consequently, we can find the optimum decision that assigns $\mathbf{x}_n$ to class $\mathbf{y}_k$ if

$$p(\mathbf{y}_k|\mathbf{x}_n) > p(\mathbf{y}_j|\mathbf{x}_n) \text{ where } \forall j = 1, 2, ..., K \text{ and } j \neq k \tag{3.4}$$

This optimum strategy which is often called "Bayes" decision rule, assigns the class $\mathbf{y}_k$ that yields the highest posterior probability, given the measurement vector $\mathbf{x}_n$ [28]. In our approach, we use HMM mixed with Gaussian Mixture Models (GMMs) so that in each state of HMM, there are various numbers of Gaussians generated to represent the observation data [29].

3.1.2 Proposed Framework

Our proposed framework consists of the following steps (see also Fig 3.1):
Repeat until the convergence is achieved in terms of maximum classification accuracy.
{

1. **Set the MFCC window length** - We slice the input data into fixed length segments with 50% overlapping. Our recordings are 15 sec long, so our initial

MFCC window length is 15 sec duration.

2. **MFCC Feature Extraction** - We extract 15 MFCC features.

3. **Data Generation for Training and Testing** - We use 2/3 of the features data for training and 1/3 for testing.

4. **Initialize the HMM and Gaussians** - We use 5 Gaussians for each state. Initial number of states is 1 which is increased by 1 in each iteration. We started with 5 Gaussians, due to restricted processing resources.

5. **Train the HMM-GMM** - Generate multiple models (each model represents a variation of call belongs to a particular class) and build *a posteriori* as well as transition probability matrices and Gaussian parameters of μ and σ that represent for the training data the best in terms of maximum likelihood.

6. **Apply B-spline Approx to Gaussian Parameters** - After generating the μ (mean) and σ (co-variance) values for GMM mixtures, we apply the 4th-order B-spline function for their approximation.

7. **Test the HMM-GMM** - Extract the maximum likelihood values for each observation of the 15 sec test sequence.

8. **Create the Confusion Matrix** - Calculate the classification performance.

9. **Adaptive Adjustment** - Increase the number of states and decrease the MFCC window length. There are 5 different parameters to adjust in an HMM chain utilizing GMMs and MFCC feature extraction algorithm. Two of the most important parameters has been selected to observe classification ratio.

10. **Go to step 2.**

}

It is worth to mention that the single model HHM-GMM generates only 1 model in step 5 as well as omits step 6, and the multi model HMM-GMM omits step 6 whereas our improved multi model HMM-GMM includes all steps. It should also be noted that, in testing mode, decisions are made in according to mahalonobis distance metric between the learnt models and input test data [30].

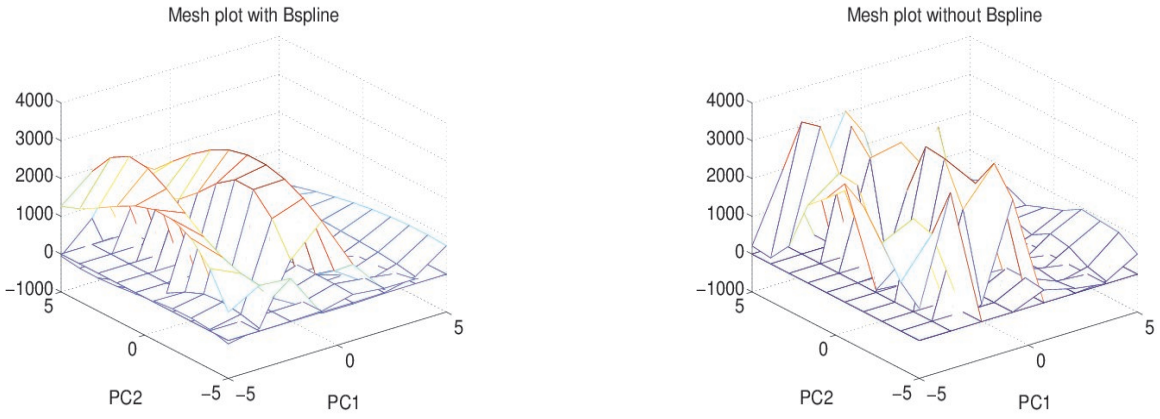
Discussion

In our framework we have used 4 different types of activities; Sperm whale call, Humpback whale call, boat sound and noise as our audio sources. We first slice the input data into fixed length segments with 50% overlapping by moving a sliding window. The choice of the initial sliding window length corresponds to the whole length of the data but reduced by HMM framework in each iteration while calculating the classification rate in terms of confusion matrix. In step 2, 15 MFCC features are extracted (we also tested with 20 MFCC features, however the changes in terms of classification performance are not noticeable). In step 3, we randomly separate the features into training dataset and test dataset (2/3 of the input features are used for training and 1/3 for testing). In step 4, we placed the Gaussians randomly in space, by using the K-means algorithm to generate the initial Gaussian parameters (μ and σ values), while priori probabilities and the transition matrices are randomly generated. In step 5, HMM-GMM model is trained to generate the Gaussians so that the input dataset can be represented the best while iteratively maximizing the likelihood of the observation sequences. With GMMs the observation space is modelled using Gaussian multivariate densities, which are consequently weighted and added to compute the emission likelihoods of each of the states or the state output probability. The Gaussian components are state specific and parameterized by the mean vector (representing the mean of the component as a d-dimensional vector) and by the covariance matrix (describing the metric of the space spanned by d-dimension) [31]. The HMM state output probability $P(x_t\,|q_j\,)$ is calculated from the state pdf $P(X\,|Q\,)$ by $p(x_t\,|q_j) = P(X = x_t\,|Q = q_j\,)$ with the given model as follow:

$$\theta^* = \arg\max_{\theta} \prod_{j=1}^{K} P(X_j|\theta) \tag{3.5}$$

where estimation of the GMM parameters $\theta = \{\mu, \sigma\}$ is done by the standard Expectation-Maximization (EM) algorithm [31] and θ consists of a set of parameters (i.e means and covariances) for M number of Gaussians, whereas θ^* is an optimum value of θ that maximizes the $\prod_{j=1}^{K} P(X_j|\theta)$. In step 6, Gaussian parameters are smoothed with a finite support B-spline interpolator. B-spline approximation algorithm runs on the μ and σ values where the order of B-spline is chosen as 4 [24]. In order to show the effectiveness of the proposed modification, PCA (principal compo-

nent analysis) is applied to the multivariate Gaussians estimated by the EM algorithm and the differences between non-interpolated and interpolated Gaussians are shown Fig. 3.2 for the sake of visualization. As we see in Fig. 3.2(a), the bivariate distribution of estimated GMMs is very smooth with B-spline compared to the distribution without B-spline (see Fig. 3.2(b)). Since the interference between the generated Gaussian surfaces are smoothed out with the B-spline, it gives more precise decision boundaries for the improved HMM-GMM classifier.



(a) (b)

Figure 3.2: Estimated GMMs (a) with B-spline Approximation and (b) without B-spline Approximation.

3.2 Experimental Results and Analysis

3.2.1 Data

We have collected the data from the Ocean Networks Canada (ONC)(`http://www.oceannetworks.ca`) observatory where an enhanced version of the Naxys Ethernet Hydrophone 02345 system is used, which includes the hydrophone element (rated to 3000m depth), 40dB pre-amplifier, 16-bit digitizer and Ethernet 100BaseT communication. This particular hydrophone is of high quality, and can be integrated into an existing underwater instrument package. The ONC hydrophone collects the data at a constant rate and can generate approximately 5.5 GB of data per day where the sampling frequency of the ONC data is 96 kHz. The ONC dataset used consists of 46 whale-call, 23 boat-sound, and 23 noise recordings. Each recording is 5 minute long and among the 46 whale-call recordings, 23 recordings contain Sperm whale calls

while the remaining 23 recordings have Humback whale calls. Each 5 minute recording is then sliced into 20 segments of 15 sec durations for processing. Therefore, the dataset consists of 1840 15-sec segments. It should be noted that splitting of training and testing datasets done at the recording level not at the segment level.

3.2.2 Results and Performances

The proposed scheme is evaluated in terms of classification results for real hydrophone data with and without B-spline approximation. Results are obtained over 100 different runs in which the feature sets are split randomly by recording where 2/3 of the data are used for training and 1/3 of the data are retained for testing. In each case, the feature set is normalized to have zero mean and unit standard deviation. In Figs. 3.3 and 3.4, the classification results of our improved multi model HMM-GMM (improved with B-spline applied to Gaussian parameters) are presented with respect to various number of Gaussians (M), number of states (Q), MFCC window length, and number of MFCC features. It can be seen that classification accuracy of the improved mult-model HMM-GMM is quite high around (97–98)%. Moreover, the results are quite consistent with respect to the number of Gaussians when the number of states is higher (e.g. 4) and the window length is smaller (e.g. 25% of the data length of the full length window). Here, we did not consider the delta-MFCC and delta delta-MFCC due to high classification performance achieved by the MFCC features. Also, in Fig. 3.4(c), it is seen that for a window of 3.75 sec, the accuracy does not remain constant, but increases slightly for a number of Gaussians greater than 3. Because, at lower number of states, it requires a fairly large number (e.g. 4) Gaussians for modeling the feature vectors to any required level of accuracy. Note that in Figs. 3.3 and 3.4, when the window length decreases, the performance goes better, while the performance drops when the window length is less than 3.75 sec (e.g. the classification accuracies are dropped by (6–7)% and (10–13)% when the window length is decreased to 1.5 sec and 1 sec, respectively). In this case, the cluster points of the source features are closer, or the respective transitional regions are overlapped (the corresponding scatter plot is not shown here). It makes the decision boundaries specially between the whale calls and boat sounds are getting closer and deteriorates the performance.

We compare our results with the decision tree [32] and multi-class SVM (MSVM) [33] classifiers, since they are useful tool for multi-class classification. In decision tree, the leaf node represents the complete classification at a given instance of the attribute

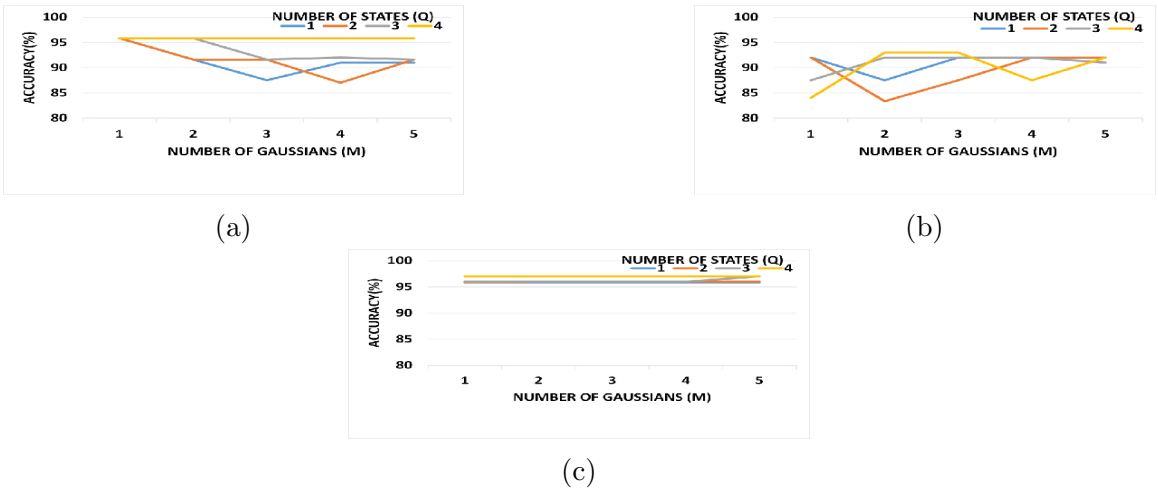
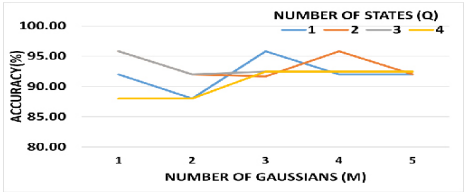


Figure 3.3: Performances (classification accuracy) of the improved multi model HMM-GMM with 15 MFCC coefficients and window length (a) 15 sec (b) 7.5 sec (c) 3.75 sec.

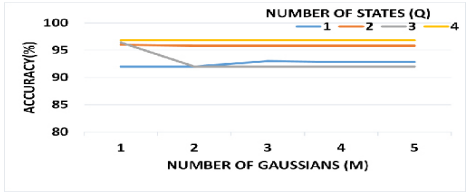
and the decision node specifies the test that is carried out to produce the leaf node. Thus with a decision tree, the sub tree that is created after any node is necessarily the outcome of the test that is conducted.

The overall performances and the comparison results are presented using single model HMM-GMM, multi model HMM-GMM, improved multi model HMM-GMM, decision tree, and MSVM classifiers in Table 3.1, where it indicates that similar performances can be achieved for the number of MFCC coefficients increased from 15 to 20. In Table 3.1, the low performance of the decision tree classification is due to the fact that the attribute of cross-entropy in the decision tree could not able to differentiate the whale calls from boat sounds. Table 3.1 shows the lowest performance for the MSVM classifier which can be due to its less resilient with the noisy features. For Table 3.1, the number of states and the number of Gaussians are 4 and 5, respectively. It should also noted that, proposed scheme enable to adjust window length adaptively.

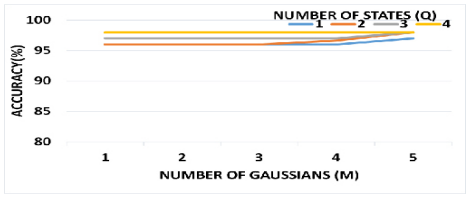
For illustration, the confusion matrix for the HMM-GMM and the improved HMM-GMM classifications are presented in Table 3.2, where the parameters used are as follows: MFCC window length=3.75 sec, number of Gaussians=5, number of states=4, number of features=15. As we can see in Table 3.2, significantly improved sensitivity and accuracy have been achieved by our improved multi model HMM-GMM classification. As explained earlier, our method looks for the maximum likelihood of the estimated Gaussian parameters with test data. Even though having



(a)



(b)



(c)

Figure 3.4: Performances (classification accuracy) of the improved multi model HMM-GMM with 20 MFCC coefficients and window length (a) 15 sec (b) 7.5 sec (c) 3.75 sec.

Table 3.1: Average results of classification accuracy (%) for different classifiers

	Number of Features					
	15			20		
	MFCC Window Length			MFCC Window Length		
	15 sec	7.5 sec	3.75 sec	15 sec	7.5 sec	3.75 sec
Single Model HMM-GMM	20	23	53	22	25	53
Multi Model HMM-GMM	91	89	98	91	92	98
Improved Multi Model HMM-GMM	93	91	99	93	95	99
Decision Tree	64	59	64	58	62.5	60
MSVM	41	41	41	48	45	48

multiple states helps to increase the classification accuracy, using only one Gaussian is not sufficient since we are dealing with real noisy hydrophone data.

Table 3.2: Confusion matrix when multi model HMM-GMM (**A**) and improved multi model HMM-GMM (**B**) are used, the classification accuracy as indicated in the right bottom corner (bold face) is calculated from the confusion matrix as $\left(\frac{\text{Sum of diagonal elements}}{\text{Sum of all elements}}\right)$

	Whale	*Boat*	*Noise*	*Specificity*
	A/B	**A/B**	**A/B**	**A/B**
Whale	1080/1320	240/120	120/0	0.75/0.91
Boat	0/0	720/720	0/0	1/1
Noise	0/0	0/0	720/720	1/1
Sensitivity	1/1	0.75/0.86	0.86/1	**0.87/0.96**

3.3 Conclusion

This paper proposes an iterative multi-class classification scheme by combining the adaptive MFCC feature extraction method with a multi model HMM-GMM classifier with B-spline for classification of acoustic events for real recorded hydrophone data. This new scheme introduces the adaptive MFCC features by adjusting the window length as well as B-spline approximation of the generated Gaussian parameters. As a result, the HMM-GMM classifier achieves high performance using low-sized features. In our approach, separation of decision regions is enhanced by applying B-spline approximation to the generated Gaussian parameters, while approximation of the parameters with B-spline also reduces the variance of classification performance. Moreover, proposed method allows us to automatically annotate the new test data and use them as a part of training data for further classification. A key finding in this work is that the optimum window length for our ocean acoustic data is almost three orders of magnitude longer than for speech. This fits well with the whale calls with average duration of 2-3 sec. Therefore, in our future work, we would like to expand this work to a two-stage classification scheme where various types of whales calls (e.g. Sperm, Humback and Fin whales) as well as other events such as earthquakes in the ocean, will be classified further using larger data sets. Furthermore, we extended our to include 48 days of data recorded in 2014. Results can be seen in Table 3.3:

Table 3.3: Confusion matrix for long-term data

Results	**humpback whale**	**sperm whale**	**fin whale**	**unknown**
humpback whale	97%	1%	0%	2%
sperm whale	2%	88%	0%	10%
fin whale	0%	0%	76%	24%

Chapter 4

Multiple Classifiers Fusion to Classify Acoustic Events in ONC Hydrophone Data

In this chapter, we present a new framework of multiple classifiers fusion to classify acoustic events in ONC (Ocean Network Canada) hydrophone data. The outputs of three different classifiers are fused based on aggregation of a generated decision matrix. An ensemble class label is thereby obtained for the classification of acoustic events into multiple classes of whale calls, boat sounds and noise. The classification performances are evaluated using real recorded hydrophone data showing an overall improvement of the classification accuracy by 10% then the proposed method in chapter 3.

Hydrophone data classification problems have become more complex due to stream of large amount of data. Not only the highly sampled data but also noise, making the classification problem even more complicated. As a consequence, the classifiers are ending up dealing with difficult situations such as, spectral overlapping, very low SNR (signal to noise ratio), highly correlated spectral feature sets, etc. In order to handle such situations while maintaining a classification accuracy, a fused classification framework has been purposed here which utilizes a Bayesian based stochastic classifier (HMM-GMM), a decision tree (DT) classifier and an artificial neural network (ANN) classifier. Currently multiple-classifiers fusion is receiving increasing attention[34]. Although multiple more sophisticated approaches proposed in[34], designing a real time operating system stays as a challenge in terms of processing power. The work

published so far demonstrates that success of the ensemble approach to classification in a variety of application domains[35]. Research on classifier ensembles permeate many strands in machine learning including streaming data[36], biometrices[37], concept drift and incremental learning[38].

Bayesian based stochastic classifiers have advantages such as they can classify high dimensional features of an observed data. One example for this type of classifier is an HMM (Hidden Markov Model) chain utilizes multiple GMMs (Gaussian Mixture Model). HMM-GMM type of classifiers are more sensitive to temporal changes. However, HMM-GMM approach has certain limitation: the input feature vectors are assumed to be statistically independent, or the corresponding HMM-GMM model assumes that there is no correlation between the consecutive frames. DT classifiers use predictive models which are fast in terms of processing speed and relatively robust to noise since they tend to over fit the noisy data. However, DT classifiers are not efficient for online learning (when data streams continuously coming and model has to be continuously updated as well) since any data can include exceptional situation (randomness) may force the decision tree to be fall apart and need to be constructed again. ANN classifiers are capable of reflecting the information of new instance on a model efficiently by just changing the weight values. However, ANN models come with some disadvantages such as, difficulty in adjustment of parameters (learning rate, regularizer coefficient, number of hidden layers, selection of activation function) and need long time to be trained compared to other methods like decision tree or HMM-GMM. The work described in this chapter is described in [8].

4.1 Basic Idea

The basic idea is that several classifiers are employed to make a classification decision about the object submitted at the input, and the individual decisions are subsequently aggregated. The output of the ensemble is a class label for the object. By combining classifiers, we are aiming at an accurate classification decision which is not achievable using simple trainable classifiers. Instead of looking for the best set of features and the best classifier, here we look for the set of classifiers and their combination method. Early work on this idea is presented in [34]. In general, the classifier fusion should work due to the following reasons: Statistical reasons - The empirical estimate of the classification performance is a random value depending on a given data and the training algorithm. So, there is always uncertainty associated with the performance

estimate. Instead of pricing just one classifier, a safer option would be to use them all and average their outputs. Computational reasons - Imperfect training algorithm: Suppose that the quality of the estimate of the classification performance depends entirely upon the training algorithm. A combination of the outputs of several diverse suboptimal classifiers may lead to a better overall classification. Representation reasons: A complex classification boundary (of any shape) can be approximated with a desired precision by simple boundaries. Classifiers fusion of different classifiers can approximate a highly nonlinear classification boundary.

4.2 Method

The block diagram of the proposed classification method is presented in Fig. 4.1, where the outputs of the L classifiers are fed into the multiple-classifiers fusion, which gives the output of the class label.

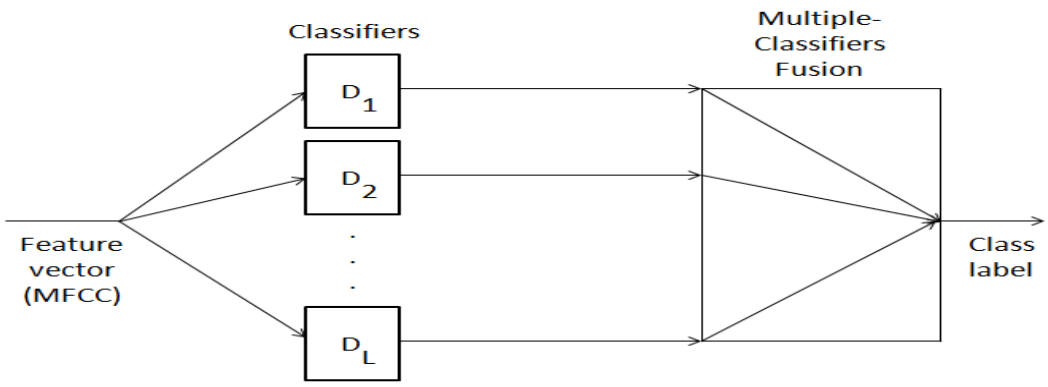


Figure 4.1: The block diagram of the proposed scheme.

4.2.1 Feature Sets Generation

The input hydrophone recordings are partitioned first into blocks of defined large segments of 15 seconds since it is large enough to contain the target events under consideration. Then the 1D data block is converted into a feature map by transforming the data segment from time domain into frequency domain followed by filtering the FFT spectrum through Mel filter bank. The MFCC feature set is then

constructed from the feature map by summing its outputs followed by logarithm operation as well as DCT transformation[39] and used as input to a classifier. The MFCC feature set is considered here since it is one of the most effective and general purpose features[40]. Moreover, it has also strong low-frequency sound capabilities, and weaker high-frequency sound perception which is suitable for our low-frequency events classification.

4.2.2 Multiple Classifiers Fusion

Let us consider that we have L classifiers and K objects, i.e. classes. We have the following decision profile matrix $DP(x)$ for each input x based on the outputs of L classifiers $C_i, i = 1, 2, \cdots, L$ as

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \cdots & d_{1,k}(x) & \cdots & d_{1,K}(x) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i,1}(x) & \cdots & d_{i,2}(x) & \cdots & d_{2,K}(x) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{L,1}(x) & \cdots & d_{L,k}(x) & \cdots & d_{L,K}(x) \end{bmatrix} \tag{4.1}$$

Then the support of the kth class $P(k)$ from the classifiers $C_1, \cdots, C_L$ is:

$$P(k) = \left(\frac{1}{L} \sum_{i=1}^{L} d_{i,k}^{\alpha}(x) \right)^{1/\alpha} \tag{4.2}$$

where $\alpha = 1/L$ and $k = 1, \cdots, K$. (Note that matrix elements are 1s and 0s)
Finally, we assign the ensemble label k^* to the object where

$$k^* = \arg \max_{k=1}^{K} \quad P(k) \tag{4.3}$$

Here is the algorithm for each new object:

1. Classify the new object x to find its decision profile $DP(x)$.

2. Calculate support for each class by Eq. (4.2).

3. Assign the ensemble label k^* to the object using Eq. (4.3).

For example, in our case, we have three classes/objects, i.e. whale calls, boat sounds and noise and three classifiers, i.e. Modified HMM-GMM, Decision Tree (DT), and Artificial Neural Network (ANN).

For instance, we could have the following illustrative scenario for an input x: The

Table 4.1: An Illustrative Scenario

Classifier	Decision		
	Whale	Boat	Noise
Modified HMM-GMM	1	0	0
DT	1	0	0
ANN	0	0	1

flow graph of the multiple classifiers fusion is presented in Fig. 4.2.

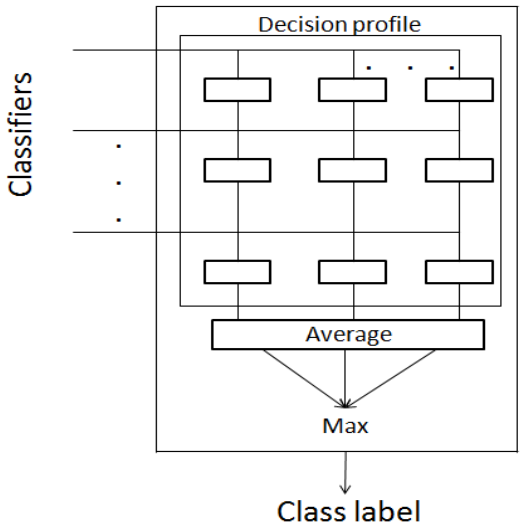


Figure 4.2: The flow graph of the multiple classifiers fusion.

4.2.3 Classifiers

The description of the different types of classifiers used are presented below.

Modified HMM-GMM Classifier

The modified HMM-GMM classifier is initialised by setting the number of states to 5 and the number of Gaussian components per state to be 3. It is worthwhile to highlight that 5 states with 3 Gaussians gives the best choice. Increasing the number of states reduces the duration of each state while it lowers the transition probabilities and does not change the classification performance. Increasing the number of Gaussians does not show also the increment in the performance since the data points represented by 3 Gaussians seems to be good enough for our classification. The HMM-GMM classifier then generates *a posteriori* and transition probability matrices as well as the Gaussian parameters of μ and σ that represent for the training data the best in terms of maximum likelihood. After training has been done and models has been generated, a train data cross check algorithm runs on the trained models and cross compares them in terms of maximum likelihood. This step increases the trained model's accuracy further. For example, if there is a whale call followed by noise then the latter one is placed to the noise dataset instead of leaving it in the whale dataset. After reorganizing the training dataset, a B-spline approximation is performed to the Gaussian parameters using a 4^{th}-order B-spline function. For testing, the maximum likelihood values are extracted for each test sample to find the respective class label. Here, for the modified HMM-GMM classifier, 2/3 of the dataset is used for training and 1/3 for testing. It is worth to mention that the modification has been made by adopting the cross check algorithm and B-spline approximation to improve the training data and the decision boundaries.

Decision Tree (DT) Classifier

Here, we have considered the decision tree classifier, since it is a useful tool for multi-class classification[41]. In decision tree, the leaf node represents the complete classification at a given instance of the attribute and the decision node specifies the test that is carried out to produce the leaf node. Thus with a decision tree, the sub tree that is created after any node is necessarily the outcome of the test that is conducted. Decision tree training leads itself to a recursive tree-growing algorithm [41]. Starting at the root, decide whether the data is pure enough to warrent termination of the

training. If yes, make a leaf and add it to the tree. If not, find the feature with the maximum discrimination ability. Split the data into left and right children nodes according to the best threshold for that feature. Repeat the procedure for the left and then for the right child, taking forward the respective portions of the data reaching that node. For the stopping criterion, we can use a measurable objective function (e.g. entropy based measure of impurity of the distribution of the class labels) to decide when to stop splitting. To label a new data x, start at the root of the tree and follow the path according to the feature values of x. Then we assign to x, the label of the leaf if it finally arrives at.

Artificial Neural Network (ANN) Classifier

ANNs are widely used to approximate complex systems and one of the most common application is classification. In this study, the ANN classification algorithm is used the most widely used back-propagation neural network, which is trained to perform the classification. Back-propagation attempts to minimize error through gradient descent by adjusting each value of a network proportional to the derivative of error with respect to that value. In back-propagation learning, the actual outputs are compared with the target values to derive the error signals, which are propagated backward layer by layer for updating of the synaptic weights in all lower layers[42]. The input variables consist of 20 MFCC coefficients and the ANN structure has an input layer including 20 neurons, a hidden layer including 3 neurons, and an output layer including 4 neurons, which are found adequate for our classification using 70% data for training and the remaining data for testing.

4.3 Data

The experimental data set is collected by Ocean Networks Canada(ONC)(`http://www.oceannetworks.ca`) observatory, where an enhanced version of the Naxys Ethernet Hydrophone 02345 system is used, which includes the hydrophone element (rated to 3000m depth), 40dB pre-amplifier, 16-bit digitizer and Ethernet 100BaseT communication. This particular hydrophone is of high quality, and can be integrated into an existing underwater instrument package. The ONC hydrophone collects data at a constant rate and can generate approximately 5.5 GB of data per day. The sampling frequency of the ONC data used is 96 kHz. Same number of dataset used

as in previous chapter.

4.4 Results

We have considered three classes of whale calls, boat sounds and noise to show the efficiency of the proposed method. The corresponding classification results are shown in Figs.4.3-4.5. In Fig. 4.3, the number of the true classification (class label 1) and misclassifications (class labels 2 and 3) of the whale calls are presented. Similarly, the number of the true classification (class label 2) and misclassifications (class labels 1 and 3) of the boat sounds are shown in Fig. 4.4, whereas Fig. 4.5 shows the the number of the true classification (class label 3) and misclassifications (class labels 1 and 2) of the noise. Based on the results of histograms in Figs.4.3-4.5, we have shown the performances in Table 4.2, where we can see that the proposed classification method (D) outperforms the three classifiers (A,B,C) in terms of higher sensitivity, specificity and accuracy. Qualitive definitions for accuracy, specifity and sensitivity given in Appendix B.

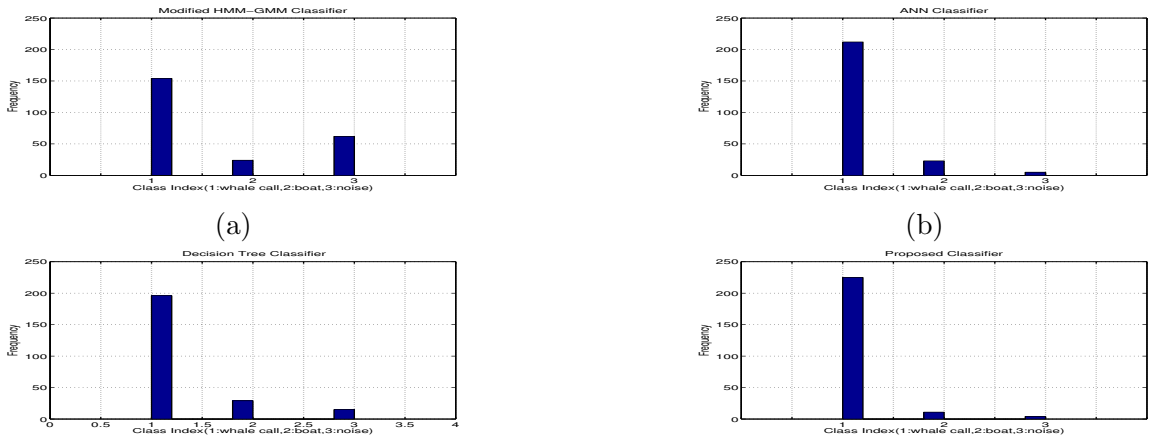


Figure 4.3: The results of the histograms for classification-misclassification of whale calls based on (a) Modified HMM-GMM, (b) ANN, (c) Decision Tree, and (d) Proposed classifications, respectively.

Table 4.2: Confusion matrix when Modified HMM-GMM (**A**), ANN (**B**), DT(**C**), and Proposed method (**D**) are used, where the classification accuracy of each classifier is shown in the right bottom corner (bold face) which is calculated as $\left(\frac{\text{Sum of diagonal elements}}{\text{Sum of all elements}}\right)$

	Whale	*Boat*	*Noise*	*Specificity*
	A/B/C/D	**A/B/C/D**	**A/B/C/D**	**A/B/C/D**
Whale	154/212/196/225	24/23/29/11	62/5/15/4	0.64/0.88/0.81/0.94
Boat	28/7/27/10	89/105/92/109	15/8/1/1	0.74/0.87/0.76/0.91
Noise	1/1/1/1	1/4/1/1	118/115/118/118	0.98/0.96/0.98/0.98
Sensitivity	0.84/0.96/0.87/0.95	0.78/0.79/0.75/0.90	0.60/0.89/0.88/0.96	**0.75/0.90/0.84/0.94**

.

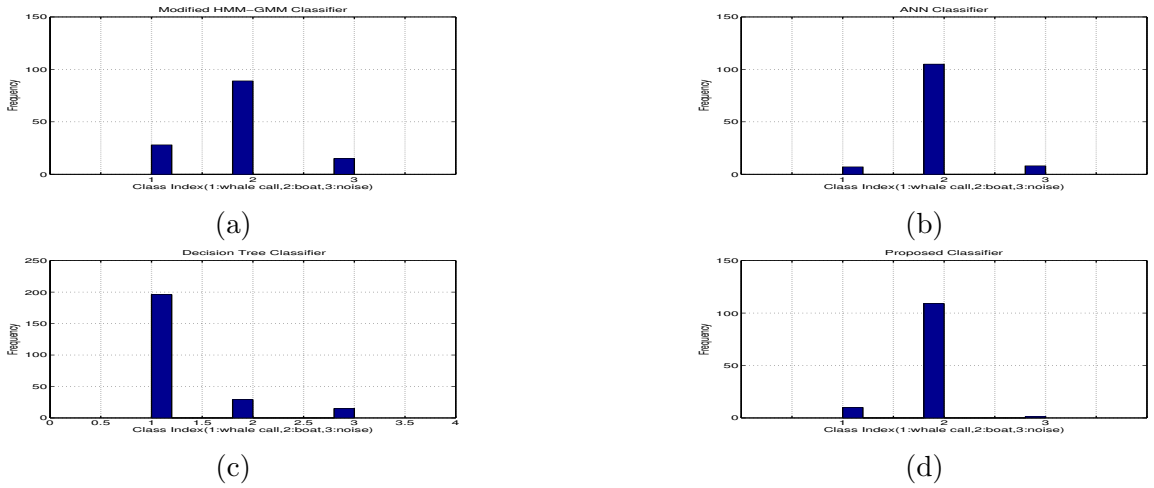


Figure 4.4: The results of the histograms for classification-misclassification of boat sounds based on (a) Modified HMM-GMM, (b) ANN, (c) Decision Tree, and (d) Proposed classifications, respectively.

4.5 Conclusion

This paper proposes an Modified framework for acoustic event classification of real-recorded hydrophone data based on multiple-classifiers fusion. The proposed classifier which is based on multiple-classifiers fusion, is providing promising results with the highest classification accuracy than the Modified HMM-GMM, ANN and DT classifiers. In the future, we would consider data fusion with this proposed scheme. Sometimes data come from different sources, and the features may be of different nature (distinct pattern representation). Instead of pooling all features, it might be better to build separate classifiers on the different groups of features and combine the classifier outputs. In another way, it might be more effective if different classifiers are trained on the different modalities.

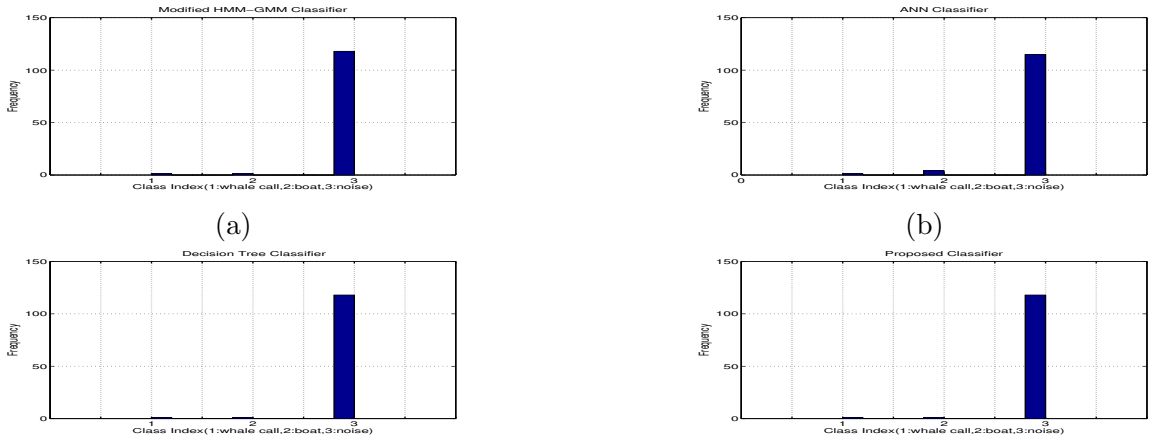


Figure 4.5: The results of the histograms for classification-misclassification of noise based on (a) Modified HMM-GMM, (b) ANN, (c) Decision Tree, and (d) Proposed classifications, respectively.

Chapter 5

A New Automatic Whale Calls Detection Algorithm Using a Modified Filter Bank

5.1 Introduction

In this chapter, we introduce a novel approach for detecting whale calls in noisy hydrophone data. The goal is to provide call-by-call annotated data to our classifier. The problem is challenging due to both parameter selection and uncharacterized broadband noise. Our method is based on a modified perceptual filter bank to distinguish broadband noise from data which enables reliable detection of whale activities in low frequencies (below 1 kHz for humpback and fin whale calls) as well as higher frequencies (between 4 kHz and 9 kHz for sperm whale calls). Our method was tested with 14370 minutes of data spanning a single year, and using a human operator's annotations as ground truth. The proposed method achieved over 90% success in finding activity segments with their time stamps. Long-term monitoring and interpretation of the sound field in the ocean can help answer important questions associated with marine mammal conservation. When monitoring events that occur under the ocean, sound information captured using hydrophones plays an important role because acoustic waves can travel long distances in the ocean.

In this work, we use hydrophone data from Ocean Networks Canada (ONC) [43]. The task of detecting whale calls from this data is challenging as the ONC recordings mostly contains highly-correlated non-Gaussian noise as well as other distur-

bances such as acoustic doppler current profiler (ADCP) interference and boat noise. However, call-by-call detection and classification is important for real-time long-term whale activity monitoring.

Methods for automated event detection in ocean acoustic data are reviewed here. Bowhead whale detection has been studied In [44] for bowhead whales, the method flags any transient event that occurs in the acoustic data using a version of a cell averaging/clutter map constant false alarm rate (CFAR) detector. However, this method would not work for humpback whale and fin whale calls due to highly uncategorized (non-Gaussian) low frequency noise presented in the ONC hydrophone recordings [43] when the harmonics of both whale calls are not strong enough to be detected by a simple energy detector. In [45], an automatic detection method with a fixed threshold has been developed in high SNR without interfering sounds to detect blue whale calls in Southwestern Indian Ocean. In [46], detection and recognition of North Atlantic right whale calls in the presence of ambient noise is described based on multistage hypothesis testing including a generalized likelihood ratio test (GLRT) detector, spectrogram testing and feature vector testing algorithms. The algorithm of [46] has high probability of recognition on synthetic signals however when tried on real calls, the ratio dropped down due to low signal to noise ratio. Furthermore, authors mentioned that investigation of the influence of the SNR on the recognition probability was not a purpose of that work, test results demonstrate that the recognition probability decreases as the SNR goes down. The rest of the work done in literature is presented in [47], [48] and [49].

Our proposed technique improves the signal to noise ratio of sperm whale, humpback whale and fin whale calls in the presence of wide band uncharacterized background noise, and thus enabled us to find accurate start and end time detections for each call. Our method uses the Gammatone perceptual filter bank combined with a peak detector that generates promising results. Unlike the other studies mentioned above, our work has addressed the following practical issues,: 1) call-by-call detection for different types of whales in Pacific Ocean (humpback, sperm, and fin whales), 2) call-by-call detection for low SNR signals with interfering noise, 3) detection using the ONC hydrophone data taken in the Pacific Ocean, 4) detection using long-term data sets.

In our study, we use call-by-call annotations created by ONC staff available for the first four days of each month for the year 2014 as the ground truth. We have chosen three species of whale (humpback, sperm and fin) since these calls can cover different

frequency bands including low-frequency band (10-50 Hz for fin whale activities), mid-frequency band (100-800 Hz for humpback whale activities) and high-frequency band (4-9 kHz for sperm whale activities). The data consists of 14370 minutes of recordings segmented into 2546 single call events by ONC. 2281 single calls (90%) with time stamps close to the annotated data were detected by our method.

5.2 The Proposed Algorithm

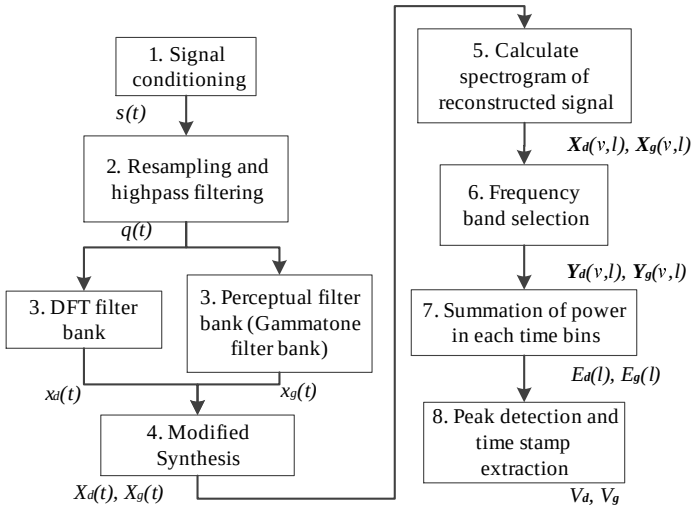


Figure 5.1: Flow diagram for the proposed method.

The steps of the proposed algorithm are explained below, the overall flow diagram of the complete proposed algorithm is presented in Fig. 5.1.and the relevant parameters are in Table 5.1.

Step 1 - Signal conditioning: This is performed by removing the mean-value of the input signal and normalizing the amplitude to yield the conditioned signal $s(t)$.

Step 2 - Resampling based on configuration: The conditioned signal $s(t)$ is resampled and highpass filtered according to the values shown in Table 5.1 to yield $q(t)$. The resampling improve the resolution in the desired frequency range listed in Table 5.1.

Step 3 - Perceptual filter bank and Discrete Fourier Transform (DFT) filter bank: Resampled and highpass filtered data $q(t)$ is applied to a perceptual (i.e. gammatone) filter bank as well as to a DFT filter bank [50] for SNR comparison. Gammatone filter bank consists of multiple gammatone filters with different center frequencies.

Table 5.1: Configuration for Different Whale Types

Configuration	**Symbols**	**humpback whale**	**sperm whale**	**fin whale**
Original sampling rate (Hz)		64000	64000	64000
Original duration of the recordings (min.)		5	5	5
Resampling rate (Hz)	F_s	4000	18000	100
Highpass filter cut-off freq. (Hz)		100	4000	10
Window size (samples)	W	12000	8192	700
Hop size (samples)	H	6000	4096	350
FFT size	F	8192	16384	512
Min. freq. of observation interval (Hz)	f_1	100	4000	10
Max. freq. of observation interval (Hz)	f_2	800	9000	50
Number of channels for gammatone filter bank	K	32	4	16

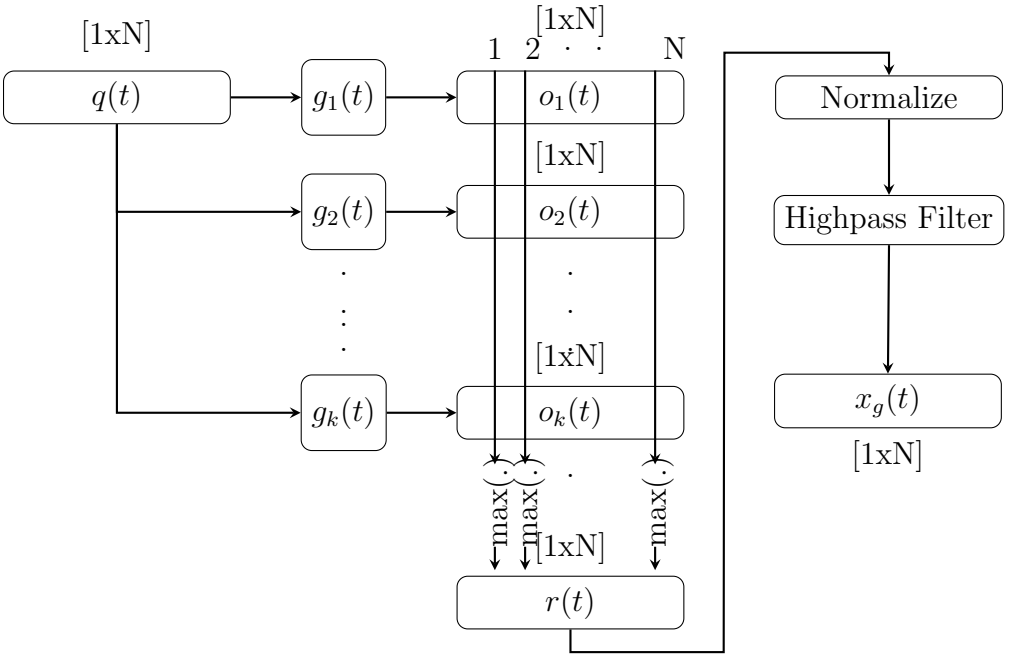


Figure 5.2: The analysis and synthesis of the gammatone filter bank. The same figure applies for the DFT filter bank using $d_k(t)$ instead of, $g_k(t)$.

The gammatone filter bank may be represented in the form of impulse response in the time domain:

$$g_k(t) = at^{n-1}e^{(2\pi b_k t)}\cos(2\pi w_k t + \phi); \quad k = 1, \cdots, K \tag{5.1}$$

In Eq. 5.1, w_k is the center frequency of each filter in Hz, ϕ is the phase (usually set to 0), a controls the gain, n is the order of the filter (usually set to be less than or equal to 4) and b_k is the decay factor for each filter, which is related to w_k given by:

$$b_k = 1.019(24.7 + 0.108w_k) \tag{5.2}$$

where 1.019 is the proportionality constant for a 4th order gammatone filter and is given in [51]. The values of w_k for a 4th-order gammatone filter bank are derived in the appendix. DFT filter bank impulse response is given in Eq. 5.3.

$$\begin{aligned} d_k(t) &= h(t)W_N^{-tk}, \quad k = 1, 2, ..., N \\ W_N &= e^{-j2\pi/N} \end{aligned} \tag{5.3}$$

Table 5.2: Configuration for Window Size

Configuration	**humpback whale**	**sperm whale**	**fin whale**
Max. call duration (sec.)	3	0.5	7
Selected window size (sec.)	3	0.45	7
Selected window size (samples)	12000	8192	700

Step 4 - Modified filter bank synthesis: The signal $q(t)$ is input to each channel $g_k(t)$ and $r(t)$ is the reconstructed signal.

$$o_k(t) = q(t) * g_k(t), \quad k = 1, 2, \cdots, K \text{ and } \quad t = 1, 2, \cdots, N \quad \text{(gammatone filter bank)} \tag{5.4}$$

$$o_k(t) = q(t) * d_k(t), \quad k = 1, 2, \cdots, K \text{ and } \quad t = 1, 2, \cdots, N \quad \text{(DFT filter bank)} \tag{5.5}$$

$$r(t) = \max([o_1(t), o_2(t), \cdots, o_K(t)]), \quad t = 1, 2, \cdots, N \tag{5.6}$$

Reconstructed signal $r(t)$ is obtained by taking the maximum energy samples at each filter's output, $o_k(t)$. A highpass filter is then applied to remove the dc offset that appears on $r(t)$ while taking the magnitude of the maximum energy samples. The cutoff frequencies used for the highpass filter are listed in Table 5.1 for both DFT and gammatone filter bank.

Step 5 - Calculate spectrogram of the reconstructed signal: For the output of the modified gammatone and DFT filter banks $X_g(t)$, $X_d(t)$ we have calculated the spectrogram $\boldsymbol{X_g}(\nu, l)$ and $\boldsymbol{X_d}(\nu, l)$ with the window size that is used to obtain optimum time resolution. The sizes of the windows are selected based on our observations for call durations, which are listed in Table 5.2.

Step 6 - Each row of $\boldsymbol{X_g}(\nu, l)$, $\boldsymbol{X_d}(\nu, l) \in [F/2 \times T]$ is indexed by frequency(ν) and each column is indexed by time(l). F is the size of FFT and T is the total number of time bins in the spectrogram given as (provided that overlap is 50% ($H = W/2$)):

$$T = \frac{N}{H} \tag{5.7}$$

where N is the length of input signal and H is the hop size. We select the rows of$\boldsymbol{X_g}(\nu, l)$ and $\boldsymbol{X_d}(\nu, l)$ in the frequency range f_1, f_2. Thus $\boldsymbol{Y}_g, \boldsymbol{Y}_d \in (n_o \times T)$ is the

observation matrix obtained from the spectrogram $\boldsymbol{X}_g(\nu, l)$, $\boldsymbol{X}_d(\nu, l)$ as:

$$\begin{aligned} \mathbf{Y}_g &= \boldsymbol{X}_g(\nu, l), \quad \nu \in [m, m+1, \cdots, n], \quad l = 1, 2, \cdots, T \\ \mathbf{Y}_d &= \boldsymbol{X}_d(\nu, l), \quad \nu \in [m, m+1, \cdots, n], \quad l = 1, 2, \cdots, T \end{aligned} \tag{5.8}$$

where the start and end frequency bins m and n, are calculated for sampling frequency F_s:

$$m = \frac{f_1 F}{F_s}, \quad n = \frac{f_2 F}{F_s} \tag{5.9}$$

Step 7 - Summation of power in each time bin: The total power in each time bin of Y_g and Y_d is calculated to find the energy differences according to Eq. (5.10);

$$E(l) = 20 \log_{10}(\textstyle\sum_{i=1}^{n_o}(|\mathbf{Y}(i, l)|)), \quad l = 1, 2, \cdots, T \tag{5.10}$$

The resultant time sequence $E(l)$ then has the energy peaks of the activity regions.

Step 8 - Perform peak detection: We have utilized MATLAB's peak detector [52] on E to find the peak locations (local maxima) within the selected frequency bins. Peak detection is done by using default values. Detected peaks at the output of the synthesized gammatone filter bank can be seen in Fig. 5.3. Peak detector runs on $E(l)$ and finds the center time location of each activity and holds the results in V. Then +/- 1.5 secs. margin for humpback whale, +/- 0.25 secs. margin for sperm whale, and +/- 3 secs. margin for fin whale calls is added to define the time limits of bounding boxes in Fig. 5.3. The frequency limits are f_1, f_2. Those margin values are added in according to the maximum duration of the individual calls presented in Table 5.2. We then compared the relative time stamp errors between detected regions and ONC annotations. Furthermore we also measure the SNR by calculating the RMS magnitude difference between the activity regions and non activity regions.

5.3 Results and Evaluation

The proposed method is developed to detect begin and end time stamps of whale activities to provide call-by-call annotations to our classification framework while generating a perceptually better call sounds. This will help human operators to perceptually confirm their annotations. We compared performance of the proposed detector/de-noiser with standard DFT filter bank. For this comparison the bandwidths of filters in DFT filter bank are adjusted to match the number of filters in

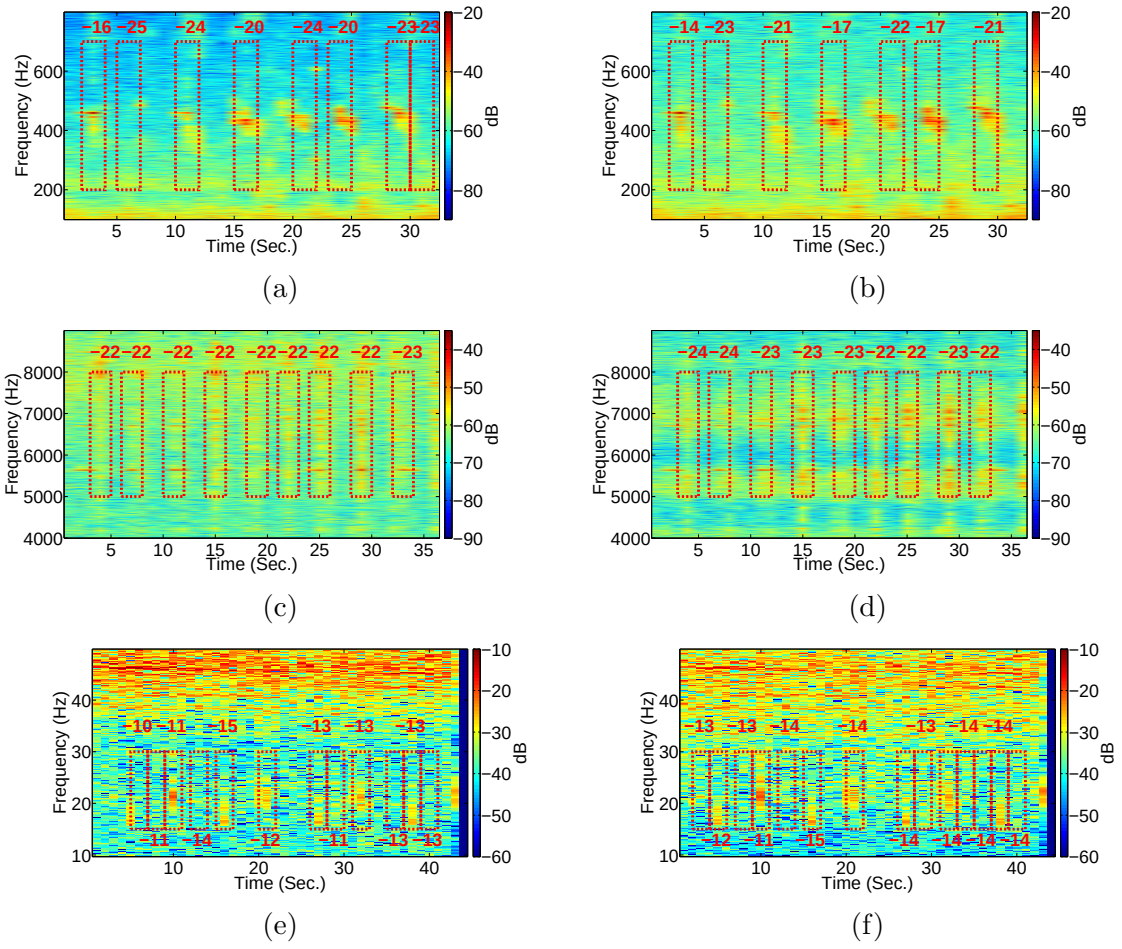


Figure 5.3: Illustrative spectrograms of proposed method. The dB scale is relative to the power of a full scale sinewave. The numbers above or below of each bounding box represent the value of $E(l)$ from Eq. 5.10 (a) humpback whale calls by DFT filter bank, (b) humpback whale calls by gammatone filter bank, (c) sperm whale calls by DFT filter bank (the periodic signal between 5000-6000 Hz is a result of the ADCP pulses), (d) sperm whale calls by gammatone filter bank, (e) fin whale calls by DFT filter bank, (f) fin whale calls by gammatone filter bank.

Table 5.3: Statistical values for measured SNR (dB)

Signal	**humpback whale**			**sperm whale**			**fin whale**		
	mean	***std***	***skew***	***mean***	***std***	***skew***	***mean***	***std***	***skew***
$\boldsymbol{q}(t)$	2.26	0.99	0.76	3.87	4.55	2.00	13.62	6.16	1.25
$\boldsymbol{x}_d(t)$	6.89	3.22	0.57	4.45	4.95	1.87	13.87	7.2	1.17
$\boldsymbol{x}_g(t)$	5.97	2.61	0.34	4.69	5.00	1.89	13.06	6.22	0.63

gammatone filter bank. The results for three example recordings for different mammals types can be seen in Fig. 5.3. In Fig. 5.3 the red bounding boxes show the detected activity regions for the reconstructed signal and the associated values represent the sum of the energy in each region $(E_g(l), E_d(l))$. In Table 5.3, the statistics of SNR are shown for different types of whale calls SNR is calculated as the ratio between the average power in the detected activity region and the average power in no activity region. Note that the statistics of SNR are obtained by considering all the annotated calls, see Table 5.4.

The detection results in terms of detection ratio are presented in Table 5.4. The proposed method with gammatone filter bank always provides higher detection ratio than the DFT filter bank based method. This is explained when a histogram of $E(l)$ is considered for the entire dataset. With the uncharacterized broadband noise, the total energy difference between activity and non-activity regions are smaller with DFT filter bank. Furthermore, we observed that the gammatone filter bank spreads the distribution of call energy such that, even in low SNR values we observe energy peaks. We also observed gammatone filter bank yielded a better perceived sound quality which helped ONC operator to confirm the results by listening.

Table 5.4: Detection Ratio (%)

	Call-by-call annotations done by ONC	**Correct detection with DFT filter bank**	**Correct detection with Gammatone filter bank**
Humpback whale	1032	812 (78.6%)	1018 (98.64%)
Sperm whale	263	182 (69.2%)	251 (95.4%)
Fin whale	1251	542 (43.3%)	1112 (88.9%)
Total	2546	1536 (60.3%)	2381 (93.5%)

For further evaluation we used time stamps from call-by-call annotations provided by ONC to verify our results. We, therefore, calculated the relative error of the

extracted time stamps:

$$\mathrm{RE}(\%) = \left(\frac{\frac{|a_s - e_s| + |a_e - e_e|}{2}}{a_e - a_s} \right) \times 100 \tag{5.11}$$

where, a_s is actual start time, e_s is extracted start time, a_e is actual end time and e_e is extracted end time. In Table 5.5, the corresponding results are shown. The

Table 5.5: Relative Error of the Extracted Time Stamps

	DFT filter bank	**Gammatone filter bank**
Humpback whale	2.57%	0.43%
Sperm whale	1.65%	0.29%
Fin whale	3.26%	0.86%

proposed perceptual filter bank based method not only provide improved detection but also enabled us to make call-by-call annotations with accurate time stamps for our back end processing, such as classification.

5.4 Conclusion

We have developed a method for extracting activity segments from noisy hydrophone data (containing various whale calls) to be classified by our classifier, with the goal of achieving call-by-call annotations and classifications with accurate time stamps. The data used for testing was from Ocean Networks Canada (ONC) and contained a significant level of broadband and correlated noise. To find the peak energy regions, we developed a method that could increase the spectral strength of the signal of interest while attenuating the background noise. We used a perceptual filter bank using gammatone filter bank and standard DFT filter bank with maximum sample synthesisi method to increase the SNR, and we were able to achieve a detection ration of 90%. In addition to this work, we also developed an automated framework for publishing our results on the ONC servers. To date, 38324 calls have been detected by the proposed detector, and 38119 of them were classified successfully. We would like to extend our detector to cover different mammals in other frequency bands (e.g. Pacific white sided dolphins), as well as other disturbances (ADCP pulses, small/big sized boats).

Chapter 6

AQUA Framework and Tools

6.1 Abstract

An underwater audio detection, identification, and classification framework (AQUA) was generated for this dissertation. AQUA is a real time system, running on Westgrid servers, that automizes the annotation process on both live and archived data. It retrieves data from the ONC data management system (DMAS), in which it perfoms detections and classifications.

6.2 Introduction

The overall block diagram of AQUA framework and its modules can be seen in Fig. 6.1.

AQUA runs on multiple Matlab threads controlled by Python scripts. AQUA has the capability to be trained and tested with 5-minute long recordings (default ONC file size), as well as single calls. However, file sizes, memory, and process requirements makes it difficult to achieve this goal with Matlab. With the help of scipy and numpy modules of python, we generated a real time system, which can manage downloads, synchronize mutliple matlab threads with different configurations, and generate annotation reports. In addition to detection, identification, and classification, AQUA provides tools to help human operators on manual annotation tasks. These tools are listed as follows:

- Download manager: manages downloading sound files from ONC data management server with specified dates.

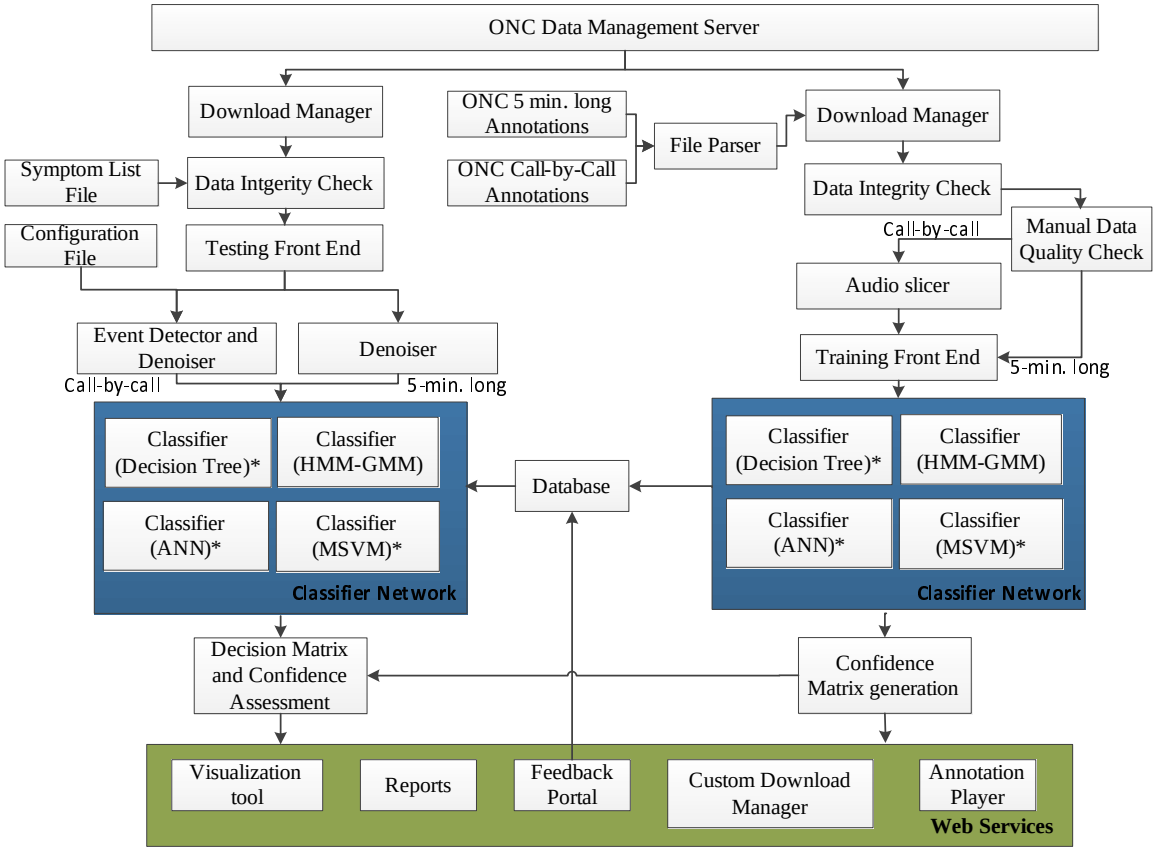


Figure 6.1: Overall block diagram of AQUA framework.

- Data integrity checker: checks the integrity of each downloaded file (sample rate, duration, broken compressed files, etc.).
- File parser: parses the annotation files delivered by ONC and generates download strings for the download manager.
- Audio slicer: prepares slices of audio with specific time stamps, in accordance to the ONC annotation file.
- Manual data quality checker: an application for the operator to be able to virtually check the quality of annotations before a training session.
- Visualization: plots spectrograms with bounding boxes around detected events, and publishes the results on the AQUA web page.
- Reports: prepares annotation reports and uploads them to ONC data management server.
- Annotation player: enables operators to listen to annotations done with a given sampling rate over a web browser.
- AQUA data request service: enables operators to download files in custom duration and sampling rates.

6.3 File Parser and Downloader

The first step of the training of AQUA is to parse the annotation file delivered by ONC. An example of an ONC annotation file for both 5 min. long and a call-by-call basis can be seen in Fig. 6.2.

In training mode, the file parser reads the annotation file and parses the given annotations in accordance to the chosen whale type, to generate a download string for the download manager. The file downloader is a python class tool to enable access to actual audio files (.wav) in DMAS. The file downloader gets the download string from the file parser module, where it schedules the downloads. In case the desired mode of classification is a call-by-call basis, the download manager runs the audio slicer tool. This process is presented in Fig. 6.3.

144	Resource Name	Species	Start Date (UTC)	End Date (UTC)
145	Ocean Sonics icListen LF Hydrophone 224	Fin whale	08-Oct-2014 00:01:04	08-Oct-2014 17:58:30
146	Ocean Sonics icListen LF Hydrophone 224	Fin whale	13-Oct-2014 22:46:22	13-Oct-2014 23:56:22
147	Ocean Sonics icListen LF Hydrophone 224	Fin whale	12-Oct-2014 23:51:21	12-Oct-2014 23:56:21
148	Ocean Sonics icListen LF Hydrophone 224	Fin whale	01-Oct-2014 05:35:17	01-Oct-2014 23:51:17
149	Ocean Sonics icListen LF Hydrophone 224	Fin whale	04-Oct-2014 01:22:46	04-Oct-2014 23:56:32
150	Ocean Sonics icListen LF Hydrophone 224	Fin whale	12-Oct-2014 10:26:21	12-Oct-2014 12:56:21
151	Ocean Sonics icListen LF Hydrophone 224	Fin whale	14-Oct-2014 00:51:35	14-Oct-2014 09:36:35
152	Ocean Sonics icListen LF Hydrophone 224	Fin whale	25-Oct-2014 12:17:56	25-Oct-2014 20:45:55
153	Ocean Sonics icListen LF Hydrophone 224	Fin whale	09-Oct-2014 00:23:30	09-Oct-2014 23:57:37
154	Ocean Sonics icListen LF Hydrophone 224	Fin whale	24-Oct-2014 00:02:06	24-Oct-2014 07:20:36
155	Ocean Sonics icListen LF Hydrophone 224	Fin whale	09-Mar-2015 01:03:37	09-Mar-2015 23:58:38
156	Ocean Sonics icListen LF Hydrophone 224	Fin whale	11-Oct-2014 17:10:57	11-Oct-2014 23:56:10
157	Ocean Sonics icListen LF Hydrophone 224	Fin whale	01-Oct-2014 00:00:41	01-Oct-2014 05:25:55
158	Ocean Sonics icListen LF Hydrophone 224	Fin whale	05-Oct-2014 00:26:32	05-Oct-2014 01:11:32
159	Ocean Sonics icListen LF Hydrophone 224	Fin whale	13-Oct-2014 00:11:21	13-Oct-2014 02:21:21
160	Ocean Sonics icListen LF Hydrophone 224	Fin whale	13-Oct-2014 11:21:22	13-Oct-2014 11:51:22
161	Ocean Sonics icListen LF Hydrophone 224	Fin whale	26-Oct-2014 18:26:00	26-Oct-2014 21:44:43
162	Ocean Sonics icListen LF Hydrophone 224	Fin whale	13-Oct-2014 18:36:22	13-Oct-2014 18:41:22
163	Ocean Sonics icListen LF Hydrophone 224	Fin whale	19-Oct-2014 01:43:37	19-Oct-2014 23:58:37
164	Ocean Sonics icListen LF Hydrophone 224	Fin whale	19-Oct-2014 01:43:37	19-Oct-2014 11:38:37
165	Ocean Sonics icListen LF Hydrophone 224	Fin whale	13-Oct-2014 19:26:22	13-Oct-2014 22:06:22
166	Ocean Sonics icListen LF Hydrophone 224	Fin whale	10-Mar-2015 00:33:38	10-Mar-2015 00:53:38

(a)

1	Resource Name	Start time (sec)	End time (sec)	Species
2	ICLISTENHF1251_20140101T000159.729Z.wav	4.72203	6.08841	HB
3	ICLISTENHF1251_20140101T000159.729Z.wav	68.28692	69.27152	HB
4	ICLISTENHF1251_20140101T000159.729Z.wav	119.1085	124.1085	NN
5	ICLISTENHF1251_20140101T000159.729Z.wav	271.5444	271.9061	FW
6	ICLISTENHF1251_20140101T001159.731Z.wav	243.0684	244.2137	HB
7	ICLISTENHF1251_20140101T001159.731Z.wav	244.5754	244.9773	FW
8	ICLISTENHF1251_20140101T002159.733Z.wav	293.192	294.4177	FW
9	ICLISTENHF1251_20140101T003159.734Z.wav	1.36637	2.81313	HB
10	ICLISTENHF1251_20140101T003159.734Z.wav	221.4586	222.182	FW
11	ICLISTENHF1251_20140101T004159.736Z.wav	177.6536	182.6569	NN
12	ICLISTENHF1251_20140101T004159.736Z.wav	185.1104	186.095	FW
13	ICLISTENHF1251_20140101T005159.738Z.wav	240.3793	240.8816	FW
14	ICLISTENHF1251_20140101T010159.740Z.wav	11.25409	12.13823	FW
15	ICLISTENHF1251_20140101T010159.740Z.wav	138.7057	140.2328	HB
16	ICLISTENHF1251_20140101T011159.741Z.wav	17.62923	18.85495	HB
17	ICLISTENHF1251_20140101T011159.741Z.wav	43.67083	44.09279	FW
18	ICLISTENHF1251_20140101T012159.743Z.wav	47.02649	47.62931	FW
19	ICLISTENHF1251_20140101T012159.743Z.wav	112.5327	113.4168	HB
20	ICLISTENHF1251_20140101T013159.745Z.wav	182.1616	182.6238	FW
21	ICLISTENHF1251_20140101T014159.747Z.wav	107.7743	108.4776	FW
22	ICLISTENHF1251_20140101T014159.747Z.wav	196.5763	197.601	FW

(b)

Figure 6.2: An example taken from two ONC annotation files. Screenshot includes 5 min. long annotations (a), and call-by-call annotations (b).

6.4 Data Integrity Checker

The AQUA Data Integrity Checker (ADIC) is a file checking tool for downloaded files. A text file, called the symptom list file, is fed to ADIC to provide potential errors to identify (i.e., broken compressed files, unexpected sample rates, unexpected durations, etc.). The identified files are logged and reported to ONC operators over a web service. An example screenshot of the web service can be seen in Fig. 6.4.

6.5 Audio Slicer

The call-by-call audio slicer is a tool for extracting audio segments from the annotations provided by ONC.

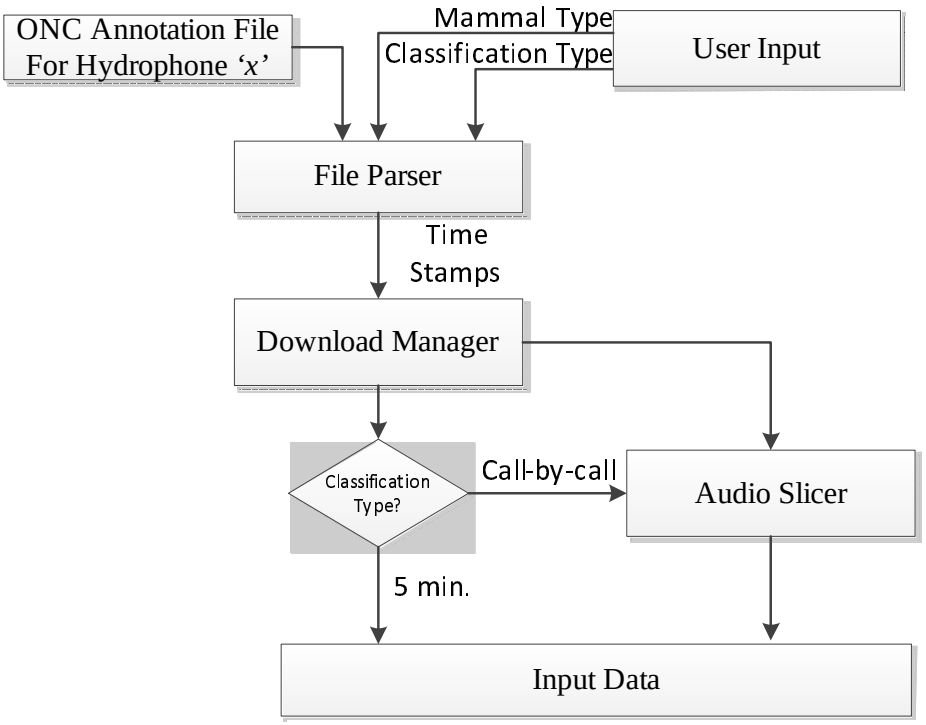


Figure 6.3: The file parse and download process.

6.6 AQUA Web Service

The AQUA web service, which is implemented with python, provides easy access to the generated classification results. Additionally, the AQUA web service has an interactive user interface, that allows operators to visualize or listen to the annotations. Example results with live data can be seen in Fig. 6.5

6.7 AQUA Data Request Service

The AQUA data request service is a web service for requesting ONC data. The files provided by ONC are 5 minutes long. This web-based service enabled us to request arbirtary duration files with custom sampling rate.

6.8 Confusion Matrix Generator

In evaluation mode, AQUA generates a text file, which has both the classification results by ONC and the classification results by AQUA. This function prepares the

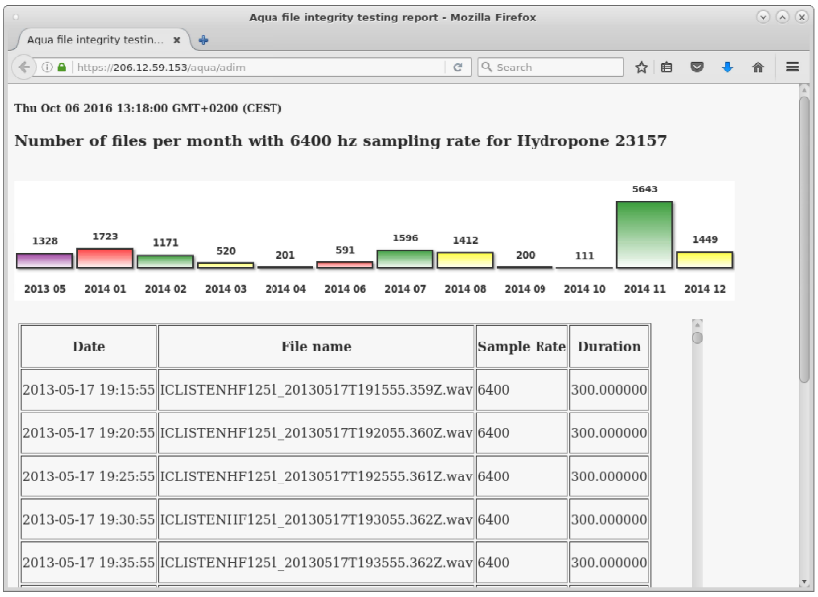


Figure 6.4: ADIC data visualization.

confusion matrices for the given file, which helps evaluate the classification performance of AQUA.

6.9 Annotation Overlap Finder

The annotation overlap finder is to calculate the relative time stamp error between the annotations done by AQUA and those by ONC.

6.10 Manual Annotation Quality Checker

The manual annotation quality checker is a tool which helps the human operator to check the quality of training datasets. This tool enables the operator to change the brightness and contrast of the spectrogram, as well as the sampling rate, and the duration shown. An example screen shot of this tool can be seen in Fig. 6.9 and 6.8.

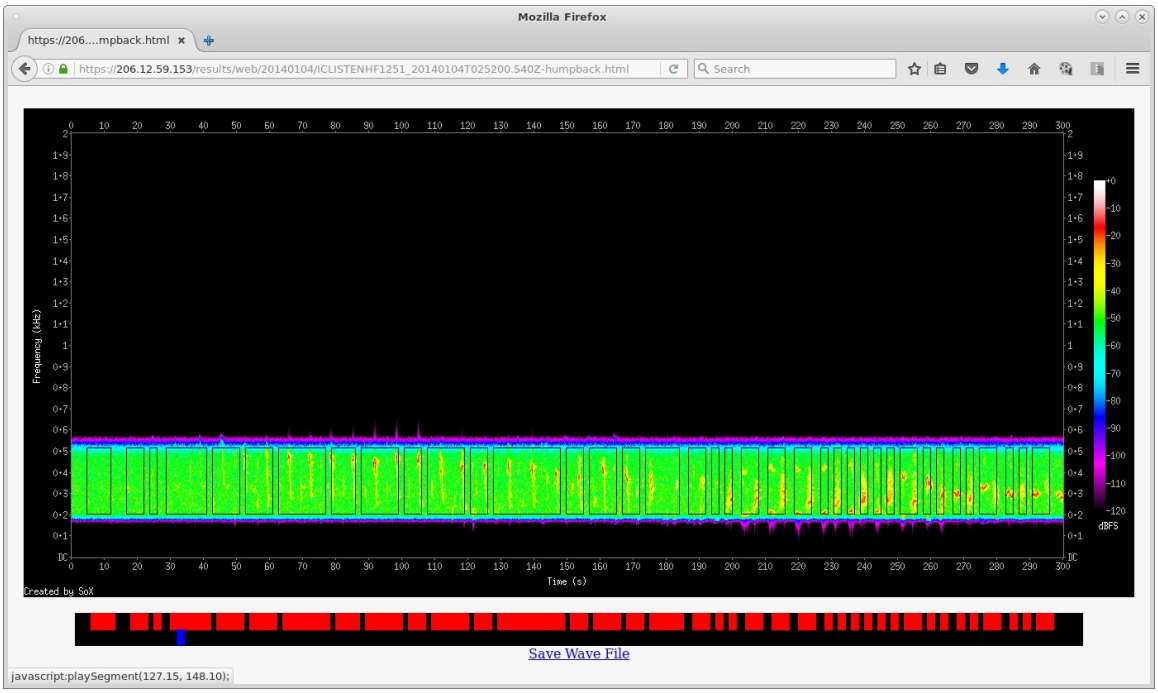
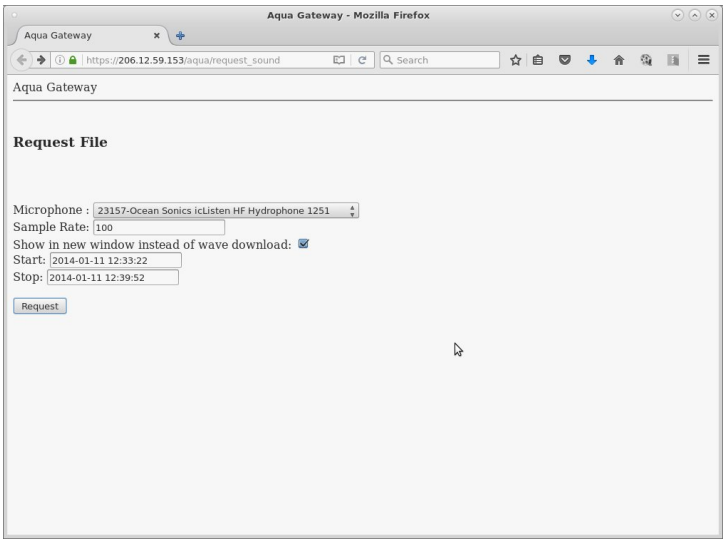
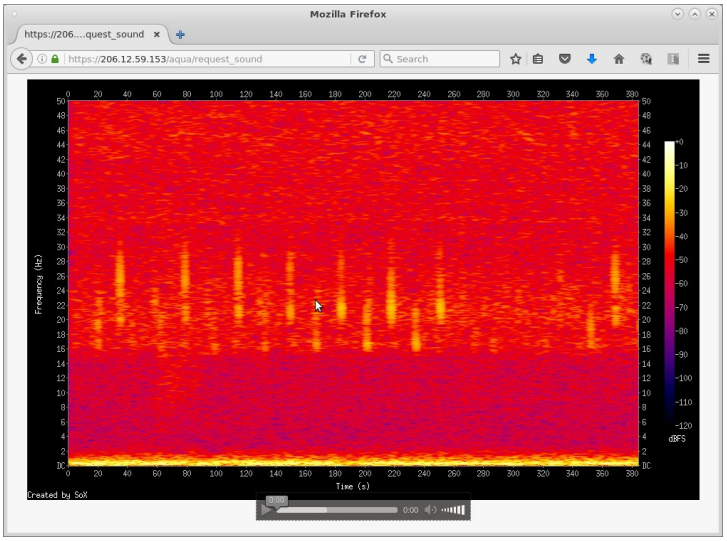


Figure 6.5: AQUA interactive web service.

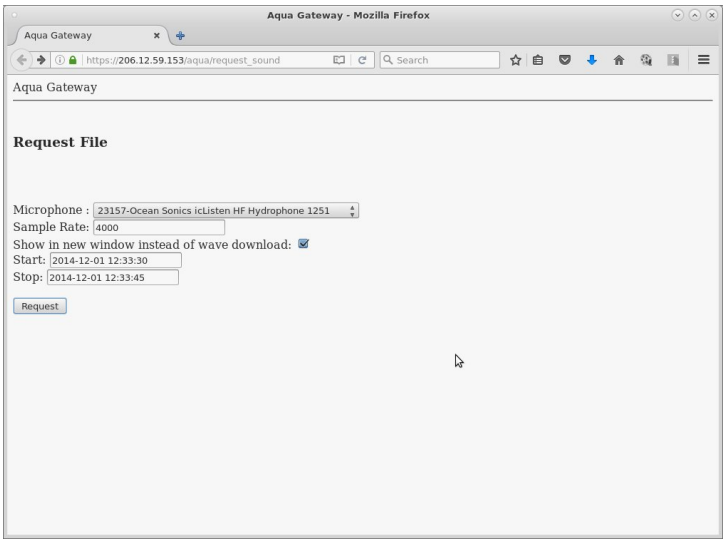


(a)

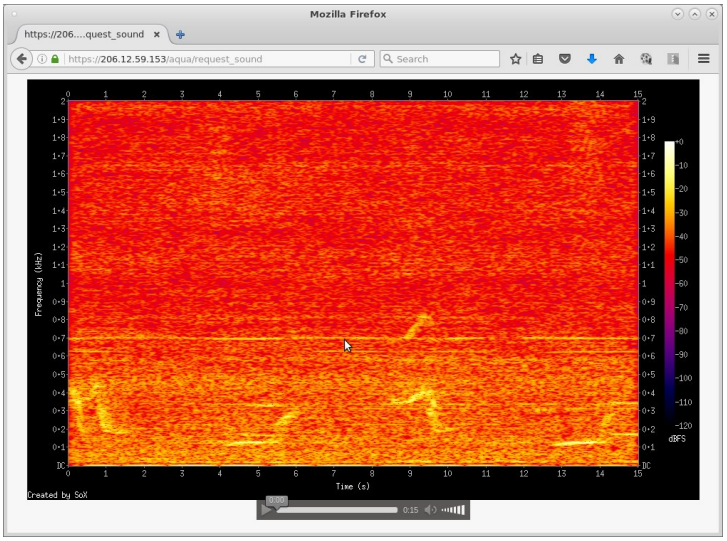


(b)

Figure 6.6: (a), parameters for request of hydrophone 1251 with 100 hz sampling rate and time interval of 2014-01-11 12:33:22 and 12:39:52 (390 secs), (b) response of the request.

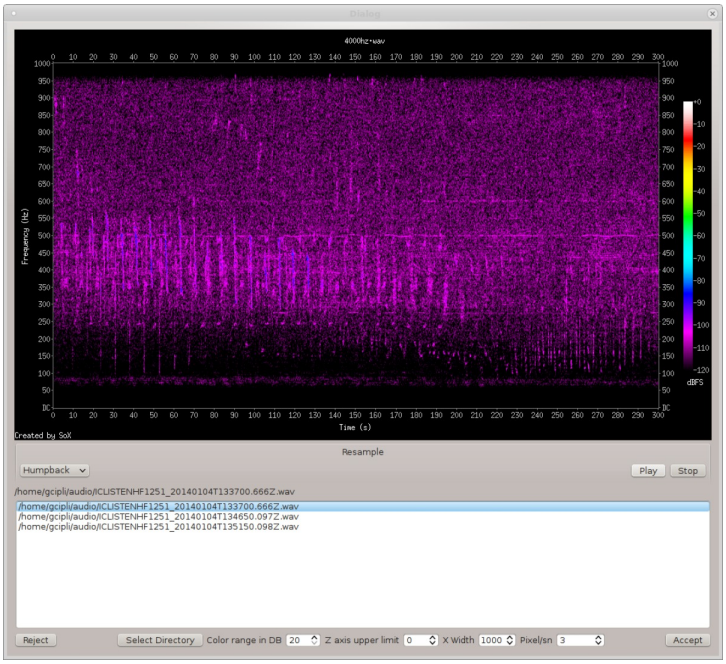


(a)

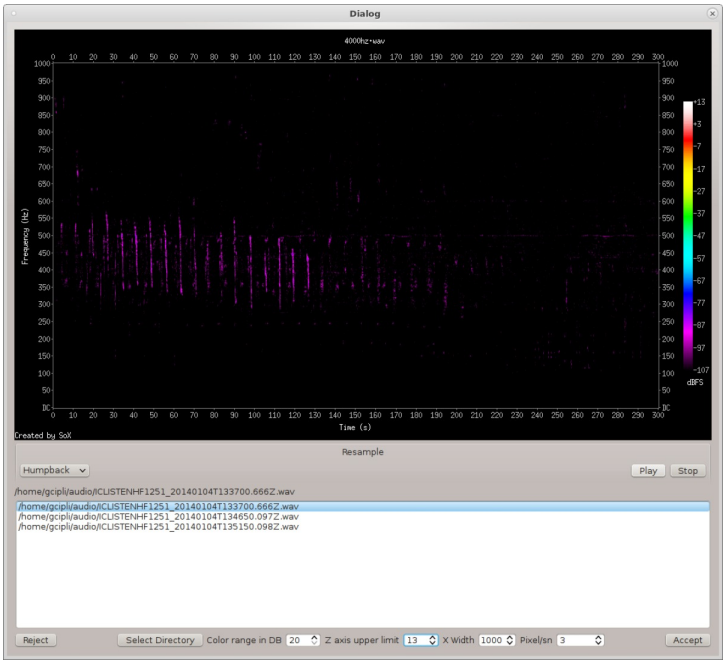


(b)

Figure 6.7: (a) parameters for request of hydrophone 1251 with 4000 hz sampling rate and time interval of 2014-12-01 12:33:30 and 12:33:45 (15 secs), (b) response of the request.

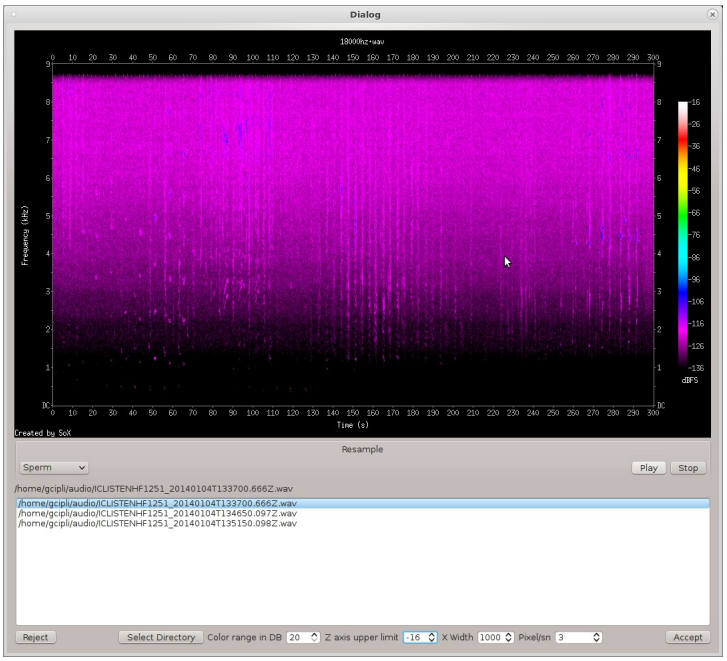


(a)

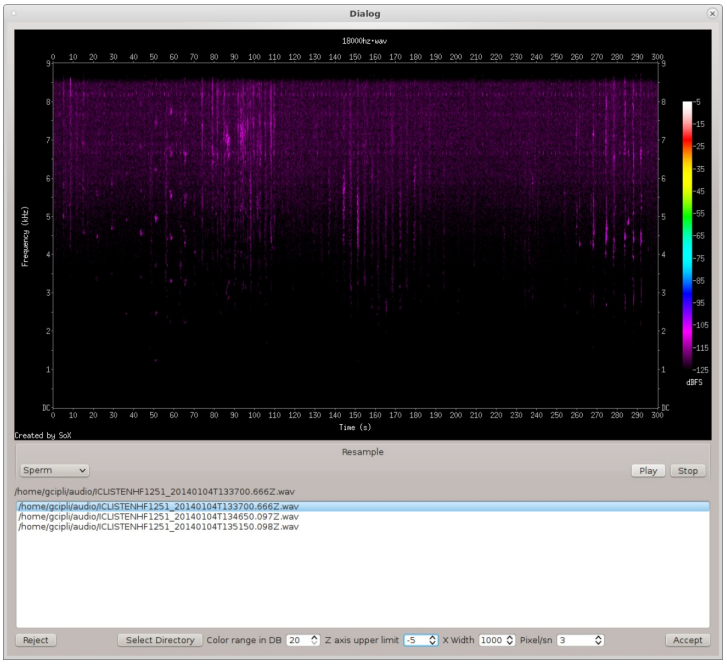


(b)

Figure 6.8: Spectrogram of humpback whale call. (a) untouched spectrograms. (b) is given Z axis upper limit as 13 to make whale calls more visible.



(a)



(b)

Figure 6.9: Spectrogram of sperm whale call. (a), spectrogram with Z axis upper limit -16. (b), spectrogram with Z axis upper limit -5.

Chapter 7

Conclusion

In this research, we developed an underwater real time detection and classification framework named AQUA, for three types of whales living in the Pacific Ocean.

In chapter 2, the proposed algorithm successfully identified the recordings with whale activities. This will allow ONC operators to identify the recordings that contain whale activities, without listening to every recording. This drastically reduced the amount of time required for doing annotations, as well as provided training and testing datasets for our classifier.

In chapter 3, we utilized a maximum log likelihood based HMM-GMM classifier to make classifications specifically for humpback whales, sperm whales, and marine vessel sounds. Classifications were performed by using the loglikelihood of the input data versus trained models. We developed a method to adaptively determine the number of states, and Gaussians, as well as the parameters for MFCC window size and number of coefficients. Furthermore, cubical splines applied to generated Gaussian parameters improved the classification ratio.

In chapter 4, we combined different types of classifiers, such as decision tree, Multi-class Support Vector Machine (MSVM), and artificial neural networks. Next, we fused them with the maximum loglikelihood based HMM-GMM classifier to improve the classification ratio. The results showed that uncharacterized broadband noise, along with other disturbances, reduced the performance of spectral entropy based classifiers. Furthermore, the maximum loglikelihood based HMM-GMM classifier enabled us to make decisions based on the confidency level of each trained model. This resulted in the addition of a new category, which accounts for unknown (i.e., rare) events. In contrast, the other classifiers were not capable of making classifications outside of their learned models.

In chapter 5, we presented a method that utilized modified filterbanks to detect the activity regions. Using the automatic whale call detection algorithm, we were able to accurately determine the time stamps of each specific whale activity. In addition, it successfully produced a perceptually better sound, which helped human operators to confirm their annotations. To date, AQUA detected and classified 48 days of data recorded in 2014 for humpback, sperm and fin whales. The method had a 90% detection ratio in the presence of uncatagorized broadband noise or other external distribunces.

Finally, chapter 6 presented the overall AQUA framework, as well as the utilization of its tools.

Chapter 8

Discussion and Ongoing Work

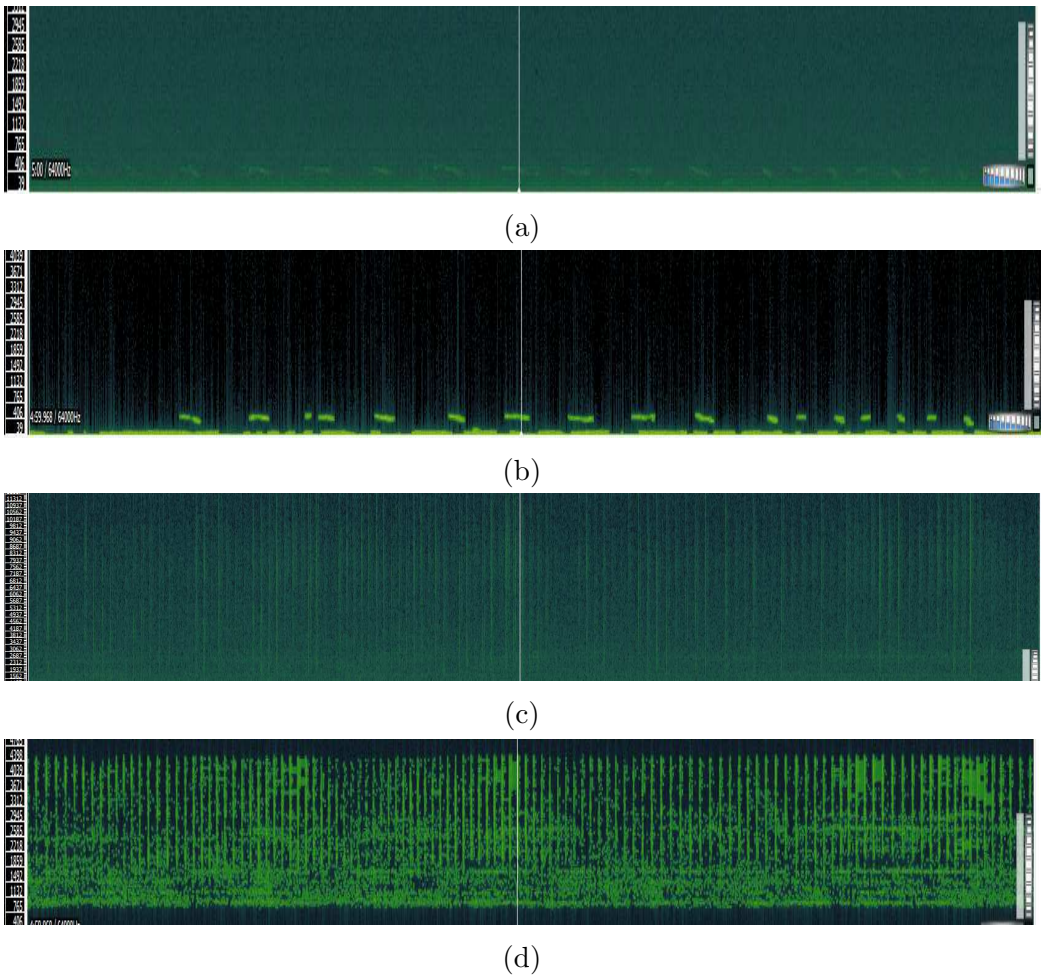
The use of the underwater audio event detection, identification, and classification framework, AQUA, along with its findings, have widespread implications in various fields of study. For example, the application of the AQUA framework to the large amount of archived data will provide NGOs with annotated data for humpback whales, sperm whales, and fin whales in the North Pacific Ocean. This will provide comprehensive reports on the activities of these types of whales, which was previosuly unavailable. Additionally, the AQUA framework has the capacity to eliminate the need for manual annotations, which will drastically reduce data processing time, therefore, increasing efficiency.

The application of AQUA also has the potential to have a direct impact on local whale conservation efforts, in that it can help to reduce whale deaths caused by ships. Accordingly, AQUA would provide information on the time, and location for a specific type of whale to the Canadian Coast Guard. This information would then be used to notify ships of the need for a course change.

Topics for future research with preliminary results to further enhance the performance of AQUA include:

8.1 Morphological image processing based de-noiser:

We are developing a morphological feature based de-noiser. Preliminary results for a humpback whale and sperm whale can be seen in Fig. 8.1.



(a)

(b)

(c)

(d)

Figure 8.1: Preliminary results for morphological feature base denoiser (a) Original humpback whale recording, (b) De-noised humpback whale recording, (c) Original sperm whale recording, and (d) De-noised spermk whale recording, respectively.

8.2 Evaluation of multitaper spectrogram for weak call detection:

Example sperm whale's multiptaper spectrogram can be seen in Fig. 8.2

8.3 Image processing based activity detection and identification.

We are also working on an image pattern recognition based activity detection and identification method based on call-by-call images from spectrogram of each type of

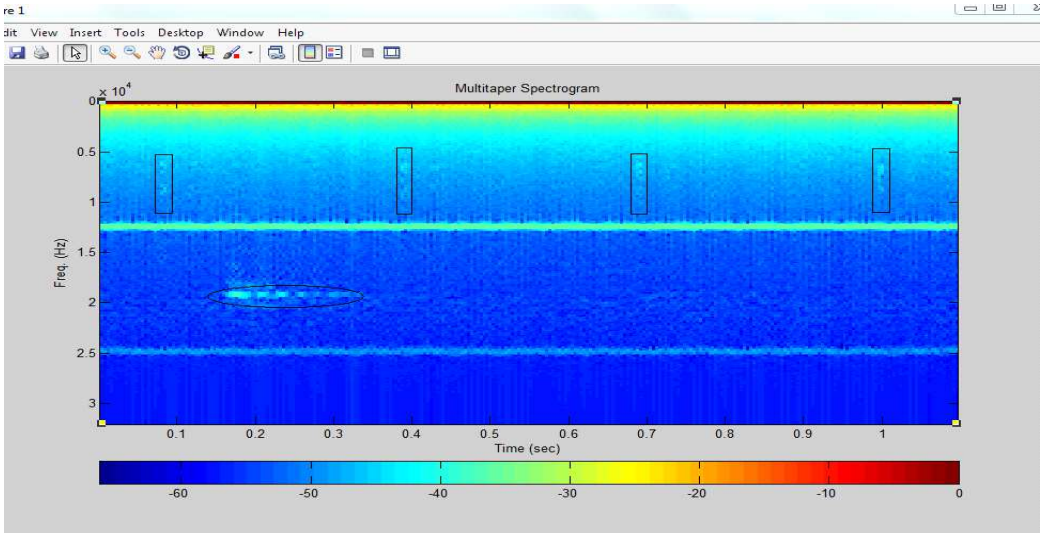


Figure 8.2: Multitaper spectrogram of a sperm whale call recording with detected divert calls.

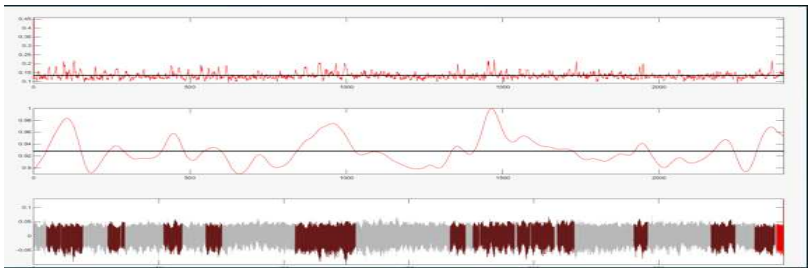
a class. Preliminary results for humpback whales can be seen in Fig. 8.3.

(a) (b)

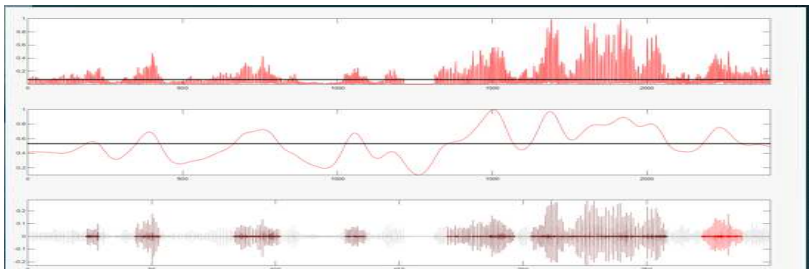
Figure 8.3: Image processing based activity detection and identification on spectrogram (a) detection and identification results for humpback whale recording, (b) training images used

8.4 Event detection with gammatonegram utilizing B-Spline approximation:

Event detection with gammatonegram utilizing B-Spline approximation method is also an ongoing research. The aim is to extract very weak call segments. Preliminary results can be seen in Fig. 8.4



(a)



(b)

Figure 8.4: Event detection with gammatonegram utilizing B-Spline approximation (a) humpback whale recording, (b) sperm whale recording

Chapter 9

Future Research

We also would like to add the following features to AQUA:

- Modeling low frequency noise.
- Creating models for earthquake and rain sounds, as well as automatically recognizing and removing other types of disturbances.
- Low frequency activity detection (earthquake, rain etc...)
- Multi-sensor processing and profiling
- ADCP pulse estimator and cleaner
- Optimization of EM (Expectation maximization) algorithm,
- Optimization of HMM-GMM parameters,
- Optimization of the parameters of feature extraction algorithm,
- Subclass classification,

Appendix A

B-Spline event detector algorithm

Algorithm for Error Pattern 1

Require: $\mathbf{x}_t \in \mathbb{R}^{L\times 1}$ {template signal}
Require: $\mathbf{x}_r \in \mathbb{R}^{L\times 1}$ {reference signal}
Require: f_s {sampling frequency}
Require: f_l {lowest frequency}
Require: f_h {highest frequency}
Require: Δ_f {frequency increment}
Require: $f = f_l$ {initial frequency}
1. **while** $f > f_h$ (i.e. stopping criterion) not satisfied **do**
2. Approximate the $\mathbf{x}_t$ signal as $\hat{\mathbf{x}}_t$
3. Approximate the $\mathbf{x}_r$ signal as $\hat{\mathbf{x}}_r$
4. Find MSE (mean-square error) between $\mathbf{x}_t$ and $\mathbf{x}_r$ signals
5. store MSE and f
6. update the frequency f as $f = f + \Delta_f$
7. **end while**
Output: Error sequence, f sequence

Algorithm for Error Pattern 2

Require: $\mathbf{x}_t \in \mathbb{R}^{L\times 1}$ {template signal}
Require: $\mathbf{x}_o \in \mathbb{R}^{L\times 1}$ {observed signal}
Require: f_s {sampling frequency}

Require: f_l {lowest frequency}
Require: f_h {highest frequency}
Require: Δ_f {frequency increment}
Require: $f = f_l$ {initial frequency}
1. **while** $f > f_h$ (i.e. stopping criterion) not satisfied **do**
2. Approximate the $\mathbf{x}_t$ signal as $\hat{\mathbf{x}}_t$
3. Approximate the $\mathbf{x}_o$ signal as $\hat{\mathbf{x}}_o$
4. Find MSE (mean-square error) between $\mathbf{x}_t$ and $\mathbf{x}_r$ signals
5. store MSE and f
6. update the frequency f as $f = f + \Delta_f$
7. **end while**
Output: Error sequence, f sequence

Appendix B

Definitions for Qualitative Measures

Qualitative definitions for accuracy, sensitivity and selectivity is calculated from confusion matrix as follows:

Table B.1: Confusion Matrix

	Predicted: NO (N)	**Predicted: YES(Y)**
Actual: NO(N)	True Negative (TN)	False Positive (FP)
Actual: YES (Y)	False Negative (FN)	True Positive (TP)

Accuracy: Represents how often is the classifier is correct and calculated as:

$$\frac{TP + TN}{total} \tag{B.1}$$

Sensitivity: Represents true positive rate and calculated as:

$$\frac{TP}{TP + FN} \tag{B.2}$$

Specifity: Represents how often is the classifier predict a no, when it is actually a no. It is calculated as:

$$\frac{TN}{N} \tag{B.3}$$

Appendix C

Derivation of Gammatone Filter Center Frequencies $w_k, k \in [1, 2, \cdots, K]$

We have implemented a 4th-order gammatone filter bank using the process shown in Fig. 5.2. We generated w which has K center frequencies between f_1 and f_2 that are placed in ERB-scale. Vector $w = [w_1, \cdots, w_K]$ holds the center frequencies of each filter in gammatone filterbank.

First of all, f_1 and f_2 are converted to the equivalent ERB numbers in according to [53]:

$$h_1 = 21.4 \log_{10} 0.00437 f_1 + 1, \quad h_2 = 21.4 \log_{10} 0.00437 f_2 + 1 \tag{C.1}$$

Secondly, K center frequencies in ERB scale are generated between h_1 and h_2:

$$z_k = h_1 + \left(\frac{h_2 - h_1}{K - 1} \right) g, \quad g = 0, \cdots, K - 1 \tag{C.2}$$

Thirdly,generated ERB numbers converted back to Hz values (Eq. (C.3)):

$$w_k = \frac{10^{(z_k/21.4)-1}}{0.00437}, \quad k = 1, \cdots, K \tag{C.3}$$

Appendix D

Whale Call Detection Algorithm Further Analysis with Synthetic Data

In appendix D, we further analyzed whale call detection algorithm proposed in chapter 6 with synthehtic data. We generated 32 filters for each filter bank in the frequency range of 100 Hz-800 Hz (corresponding to observation interval for humpback whales). First of all, we evaluated the filter banks with a noisy logarithmic chirp signal sweeping from 250 hz to 350 Hz. Magnitude of generated chirp signal is 1V p-p. Results can be seen in Fig D.1.

We also evaluated a single humpback whale call with proposed method and results can be seen in Fig. D.2,

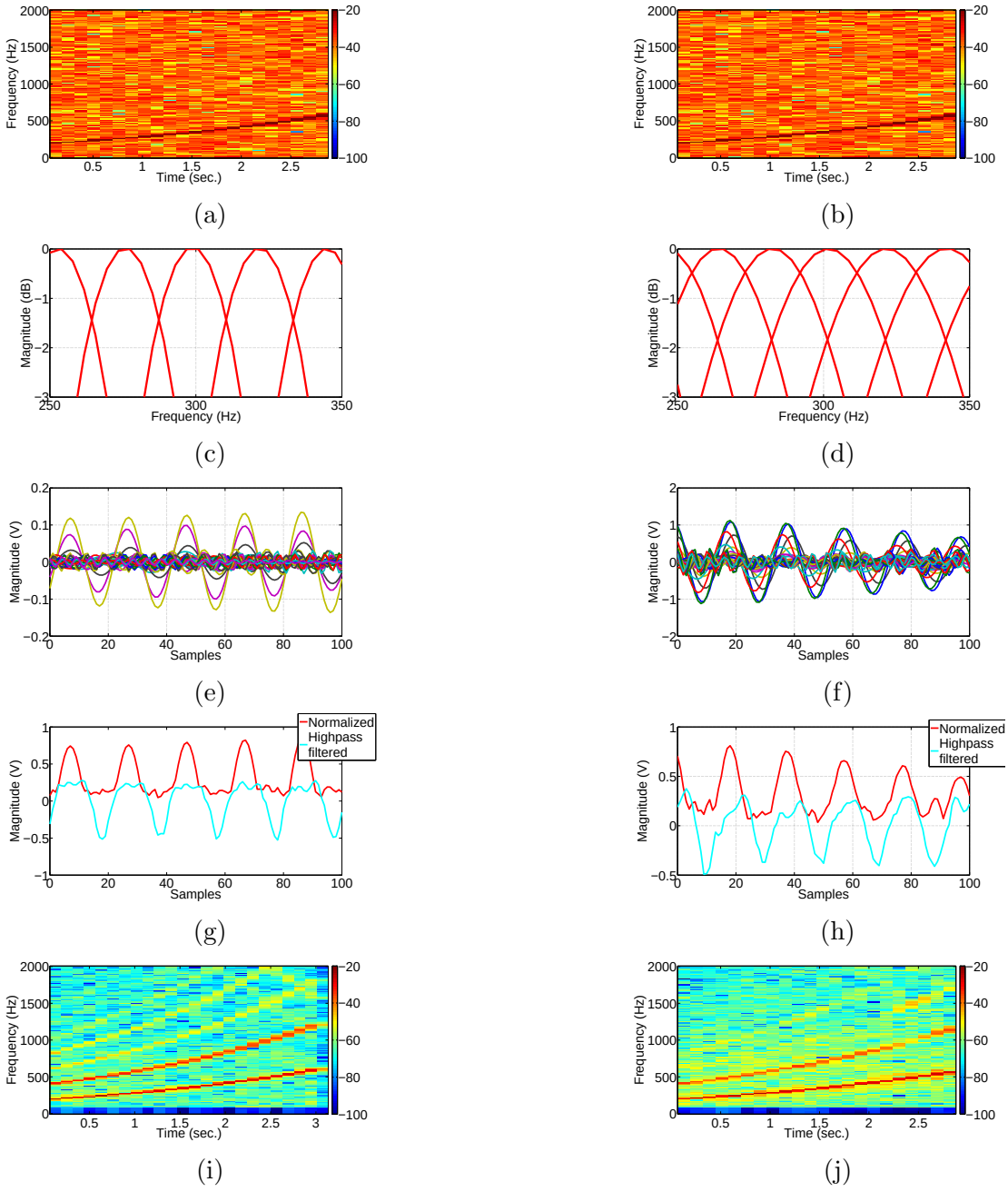
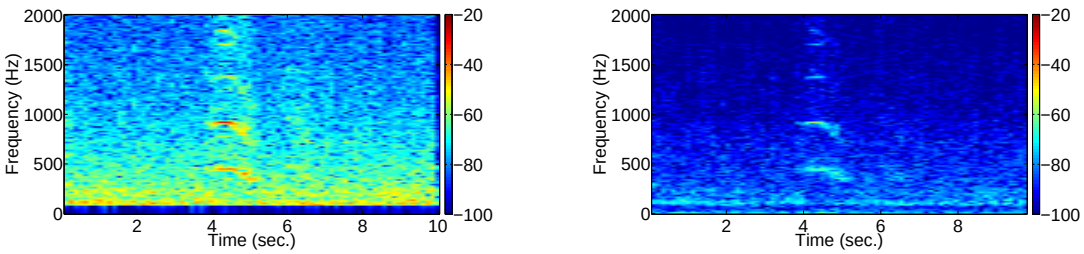


Figure D.1: Detected maximum energy samples in filter banks (a) Spectrogram of noisy input chirp signal, (b) Spectrogram of noisy input chirp signal. (c) Frequency response of DFT filter bank (d) Frequency response of gammatone filter bank (e) Output of each channel in DFT filter bank (f) Output of each channel in gammatone filter bank (g) Output of DFT filter bank (h) Output of gammatone filter bank (i) Resultant spectrogram after DFT filter bank (j) Resultant spectrogram after Gammatone filter bank.



(a) (b)

Figure D.2: Single humpback whale call (a) After DFT filter bank, (b) After gammatone filter bank.

Bibliography

[1] M. Oswald1, J. N. Oswald1, M. O. Lammers, S. Rankin, and W. W. L. Au, *Integration of realtime odontocete call classification algorithm into PAMGUARD signal processing software.* Acoustical Society of America, 2011.

[2] S. Ness, H. Symonds, P. Spong, and G. Tzanetakis., *The Orchive : Data mining a massive bioacoustic archive.* CoRR-Computing Research Repository, 2013.

[3] M. Andre, M. V. D. Schaar, A. Mas, A. Roma, J. Castell, M. Morell, M. Sol, J. Rolin, and R. Person, *Realtime acoustic monitoring of the deep ocean environment.* The Journal of the Acoustical Society of America, 2008.

[4] R. Burham, R. Palm, D. Duffus, and A. Riera, *The combined use of visual and acoustic data collection techniques for winter killer whale (Orcinus orca) observations.* Global Ecology and Conservation Conf., 2016.

[5] B. Martin, K. Kowarski, X. Mouy, and H. Moors-Murphy, *Recording and identification of marine mammal vocalizations on the scotian shelf and slope.* Oceans, 2015.

[6] G. Cipli, F. Sattar, and P. F. Driessen, *A Novel Approach to Low Frequency Activity Detection in Highly Sampled Hydrophone Data Based on B-Spline Approximation.* IEEE Pacific Rim Conf. Comm. Comp. and Signal Proc. (PACRIM), 2015.

[7] G. Cipli, F. Sattar, and P. F. Driessen, *Multi-class Acoustic Event Classification of Hydrophone Data Based on Adaptive MFCC Combined with Improved HMM-GMM Topology.* IEEE Pacific Rim Conf. Comm. Comp. and Signal Proc. (PACRIM), 2015.

[8] G. Cipli, F. Sattar, and P. F. Driessen, *Multiple Classifiers Fusion to Classify Acoustic Events in ONC Hydrophone Data*. IEEE Pacific Rim Conf. Comm. Comp. and Signal Proc. (PACRIM), 2015.

[9] V. Hodge and J. Austin, *A survey of outlier detection methodologies*, vol. 22. Artificial Intelligence Review, Springer, 2004.

[10] D. Z. Daniel, G.-P. S. Bengio, and I. Mc-Cowan, *Semi-supervised adapted HMMs for unusual event detection*. IEEE Confenerence on Computer Vision and Pattern Recognition(CVPR05), 2005.

[11] M. Davy, F. Desobry, A. Gretton, , and C. Doncarli, *An online support vector machine for abnormal events detection*, vol. 86. Signal Processing, 2006.

[12] S. Li and K. Mueller, *Spline-based gradient filters for high-quality refraction computations in discrete datasets*. IEEE Conference on Visualization, 2005.

[13] M. Unser, A. Aldroubi, and M. Eden, *B-Spline signal processing: part I-Theory*, vol. 41. IEEE Transactions on Signal Proccessing., 1993.

[14] P. H. C. Eilers and B. D. Marx, *Flexible smothing with B-Splines and penalties*, vol. 11. Statistical Science, 1996.

[15] T. Peeters and O. Ciftcioglu, *Statistics on exponential averaging of periodograms*. IEEE Transactions on Signal Processing, 2002.

[16] C. E. Stilp, M. Kiefte, and K. R. Kluender, *Signal detection as a function of relative acoustic entropy*, vol. 127. Int J. Acoustic. Soc. Am., 2010.

[17] H. P. (Ed.), *Handbook of Engineering Statistics*. Springer, 2006.

[18] H. O. S. G. Tanyer, *Voice activity detection in nonstationary noise*, vol. 8. IEEE Transactions on Speech and Audio Processing, July 2000.

[19] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank, *Sensitivity to basis mismatch in compressed sensing*, vol. 59. IEEE Transactions on Signal Processing, 2011.

[20] R. Zahedii, A. Pezeshki, , and E. K. P. Chong, *Measurement design for detecting sparse signals*, vol. 5. Physical Communication, 2012.

[21] A. V. Oppenheim and R. W. Schafer, *Discrete time Signal Processing*. Prentice Hall, 1989.

[22] L. R. Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, vol. 77. Proceedings of the IEEE, 1989.

[23] X. Wang, L. F. M. Bosch, and L. C. W. Pols, *Integration of Context-Dependent Durational Knowledge into HMM Based Speech Recognition*. Proc ICSLP, 1996.

[24] M. Unser, A. Aldroubi, and M. Eden, *B-spline Signal Processing: Part I-Theory*, vol. 41. IEEE Transaction on Signal Proccessing, 1993.

[25] S. P. Chatzis, *Margin-Maximizing Classification of Sequential Data with Infinitely-Long Temporal Dependencies*, vol. 40. Expert Systems with Applications, 2013.

[26] Y. Han, G. Wang, and Y. Yang, *Speech Emotion Recognition Based on MFCC*, vol. 20. ChongQing University of Posts and Telecommunications, 2008.

[27] S. Davis and P. Mermelstein, *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*, vol. 28. IEEE Transactions on Acoustics, Speech and Signal Process., 1980.

[28] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press Ltd., 1990.

[29] D. Geoff, *Pattern Recognition and Classification: An Introduction*. Springer, 2013.

[30] C. Z. Shiming Xiang, Feiping Nie, *Learning a Mahalanobis distance metric for data clustering and classification*, vol. 41. ELSEVIER, December 2008.

[31] S. Sakti, K. Markov, S. Nakamura, and W. Minker, *Incorporating Knowledge Sources into Statistical Speech Recognition*, vol. 42. Lecture Notes in Electrical Engineering, Springer, 2009.

[32] B. de Ville, *Decision trees*. Wiley, October 2013.

[33] H. Altun and G. Polat, *Boosting Selection of Speech Related Features to Improve Performance of Multi-class SVMs in Emotion Detection*, vol. 36. Expert Systems with Applications, 2009.

[34] L. Kuncheva, *Combining Pattern Classifiers - Methods and Algorithms*, vol. Second Ed. Wiley, 2014.

[35] N. C. Oza and K. Tuner, *Classifier Ensembles: Select Real-World Applications*, vol. 9(1). Information Fusion, 2008.

[36] V. Grossi and F. Turini, *Stream Mining: A Novel Architecture for Ensemble-based Classification*, vol. 30. Knowledge and Information Systems, 2012.

[37] J. Read, A. Bifet, G. Holmes, and B. Pfahringer, *Scalable and Efficient Multi-Lable Classification for Evolving Data Streams*, vol. 88(1–2). Machine Learning, 2012.

[38] R. Elwell and R. Polikar, *Incremental Learning of Concept Drift in Nonstationary Environments*, vol. 22(10). IEEE Transactions on Neural Networks, 2011.

[39] M. R. Devi and T. Ravichandran, *A Novel Approach for Speech Feature Extraction by Cubic-Log Compression in MFCC*. IEEE Conf. on Pattern Recognition, Informatics and Mobile Eng., 2013.

[40] S. Davis and P. Mermelstein, *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*, vol. 28. IEEE Transactions on Acoustics, Speech and Signal Process, 1980.

[41] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Wadsworth Int, 1984.

[42] B. Bas, O. Ozgonenel, B. Ozden, B. Bekcioglu, E. Bulut, and M. Kurt, *Use of Artificial Neural Network in Differentiation of Subgroups of Temporomandibular Internal Derangements: A Pre-liminary Study*. J. Oral Maxillofac Surg., 2012.

[43] *http://www.oceannetworks.ca/*.

[44] A. M. Thodea, K. H. Kim, S. B. Blackwell, C. R. Greene, S. Christopher, and T. L. McDonald, *Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys*. Acoustical Society of America, 2012.

[45] F. Samarana, O. Adamb, and C. Guineta, *Classification and Localization of Marine Mammals Using Passive Acoustics*, vol. 71. Int. Workshop on Detection, 2010.

[46] I. R. Urazghildiiev, C. W. Clark, T. P. Krein, , and S. E. Parks, *Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise*, vol. 34. IEEE Journal of Oceanic Engineering, 2009.

[47] M. Bittle and A. Duncan, *A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring.* Proceedings of Acoustics 2013, November 2013.

[48] O. A. Flore Samarana and C. Guineta, *Detection range modeling of blue whale calls in Southwestern Indian Ocean*, vol. 71. Applied Acoustics, November 2010.

[49] S. Frank and A. Ferris, *Analysis and localization of blue whale vocalizations in the Solomon Sea using waveform amplitude data.* Journal of the Acoustical Society of America, 2011.

[50] U. Zolzer, *DAFX: Digital Audio Effects.* No. 0-471-49078-4, John Wiley and Sons, 2002.

[51] J. H. I. N.-Smith, R. Patterson, and P. Rice, *Implementing a gammatone filter bank.* Cambridge Electronic Design, John Holdsworth(Ed), 1988.

[52] *http://www.mathworks.com/help/signal/ref/findpeaks.html.*

[53] R. D. Patterson and B. C. J. Moore, *A revision of Zwicker's loudness model*, vol. 82. Acta Acustica, 1996.