

2011-01-01

# Mixed-Signal Distributed Feature Extraction for Classification of Wide-Band Acoustic Signals on Sensor Networks

Humberto Santacruz

University of Texas at El Paso, [hsantacruz@miners.utep.edu](mailto:hsantacruz@miners.utep.edu)

Follow this and additional works at: [https://digitalcommons.utep.edu/open\\_etd](https://digitalcommons.utep.edu/open_etd)



Part of the [Electrical and Electronics Commons](#)

---

## Recommended Citation

Santacruz, Humberto, "Mixed-Signal Distributed Feature Extraction for Classification of Wide-Band Acoustic Signals on Sensor Networks" (2011). *Open Access Theses & Dissertations*. 2389.

[https://digitalcommons.utep.edu/open\\_etd/2389](https://digitalcommons.utep.edu/open_etd/2389)

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact [lweber@utep.edu](mailto:lweber@utep.edu).

MIXED-SIGNAL DISTRIBUTED FEATURE EXTRACTION FOR CLASSIFICATION  
OF WIDE-BAND ACOUSTIC SIGNALS ON SENSOR NETWORKS

HUMBERTO SANTACRUZ

Department of Electrical and Computer Engineering

APPROVED:

---

Jose Gerardo Rosiles, Chair, Ph.D.

---

Bryan Usevitch, Ph.D.

---

Luis R. Contreras, Ph.D.

---

Benjamin Flores, Ph.D.  
Acting Dean of the Graduate School

*Para Mi Querida Familia*

*Mama, Papa, Pedro, y Gilberto*

*los amo.*

MIXED-SIGNAL DISTRIBUTED FEATURE EXTRACTION FOR CLASSIFICATION  
OF WIDE-BAND ACOUSTIC SIGNALS ON SENSOR NETWORKS

by

HUMBERTO SANTACRUZ

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Electrical and Computer Engineering

THE UNIVERSITY OF TEXAS AT EL PASO

August 2011

# Acknowledgements

I would like to give thanks to my parents for all the sacrifices that they have made not only for me but for all of us. Thank you Pedro for given us the example that anything that you set your mind can be accomplished. A very special Thank you to my friend, companion and twin brother Gilberto. Thanks for always making me appreciate the little things in life. Los quiero mucho.

I would like to give my eternal gratitude to Dr. Gerardo Rosiles, without his support and patience I would not be here. Thank you for everything that you have thought me.

To "the gang" it's been a long journey but it's over thanks for all your support through all of this. I know I can count on you guys and as we say "I get by with a little help from my friends".

To my new friends Christopher Romero, Johhny, and Marlene Arrellanes, you guys have shown me what kindness and caring for someone really means. Muchas gracias.

Finally, I will like to thank Iliana Gamboa for being an important part of my life. Thank you for encouraging me to never quit and believing in me even when I didn't.

# Abstract

This thesis proposes a distributed/collaborative scheme for the classification of wide-band acoustic events within the context wireless sensor networks (WSNs). The proposed method is characterized by a mixed-signal processing scheme for the extraction of feature vectors which reduces the processing, memory allocation, and power requirements of the sensor nodes. The mixed-signal scheme assumes that nodes have an analog frontend consisting of a microphone, a bandpass filter, a squaring element and an integrator. This chain of components produces an estimate of the energy over the frequency subband extracted by the filter. The WSN has the ability to form a distributed filter bank where a set of nodes self-organize to capture an acoustic signal (or event) across non-overlapping contiguous frequency subbands. The group of subbands resemble the frequency decomposition of a discrete wavelet packet transform (DWPT). The analog energy estimates are digitized (i.e., sampled) at a very low sampling rate (e.g., one sample per second), reducing the overall power, memory and processing requirements of the WSN. The energy measurements are relayed to a concentrator node where they are grouped to form a feature vector that serves as a signature for the acoustic event. The feature vectors can then be used with a statistical pattern classifier to identify the class to which the event belongs. We used a database of five acoustic classes: birds, explosions, cars, conversation and footsteps. The k-NN classifier, a back-propagation artificial neural network and a radial-basis function support vector machine (RBF-SVM) were evaluated against the dataset. An all-digital reference system was implemented using signals sampled at 44.1 KHz with a DWPT front end. Feature vectors were produced using subband energies and classified with the same pattern classifiers. Both systems were simulated in MATLAB using an eight-band filter bank. Simulations show that the proposed distributed mixed-signal solution performs as well as the all-digital system with 77.7% and 75.7% respectively for the RBF-SVM classifier.

# Contents

	Page
Acknowledgements . . . . .	iv
Abstract . . . . .	v
Table of Contents . . . . .	vi
List of Tables . . . . .	viii
List of Figures . . . . .	ix
<b>Chapter</b>	
1 Introduction . . . . .	1
1.1 Problem Description . . . . .	2
1.2 Proposed Solution . . . . .	3
1.3 Pattern Recognition . . . . .	4
1.4 Wireless Sensor Networks . . . . .	5
1.5 Audio Monitoring Using WSN . . . . .	6
2 An Acoustic Signal Classifier Based On Wavelets Packets . . . . .	8
2.1 Wavelet Packet Transform . . . . .	8
2.2 Support Vector Machines . . . . .	10
2.2.1 Linear Classification . . . . .	11
2.2.2 Non-linear SVM Classifier . . . . .	17
2.2.3 Non-separable case . . . . .	18
2.3 Conclusion . . . . .	19
3 A Classification Algorithm for Acoustic Signals . . . . .	20
3.1 Description of Data Sets . . . . .	20
3.2 Feature Extraction . . . . .	20
3.3 Brief Description of other Non-linear Classifiers . . . . .	24
3.3.1 K-Nearest Neighborhood Classifier . . . . .	24

3.3.2	Back Propagation Neural Network . . . . .	25
3.3.3	Radial Basis Function Support Vector Machine . . . . .	26
3.4	Results for Acoustic Classification . . . . .	26
3.5	Conclusion . . . . .	28
4	A WSN for Mixed-Signal Distributed and Collaborative Sound Classification . .	29
4.1	A WSN model for acoustic signal classification . . . . .	29
4.2	Mixed-Signal Feature Extraction . . . . .	31
4.3	Description of Mixed-Signal Classifier Simulation . . . . .	32
4.4	Results . . . . .	34
4.5	Conclusion . . . . .	35
5	Conclusion and Future Work . . . . .	37
5.1	Conclusions . . . . .	37
5.2	Future Work . . . . .	38
	References . . . . .	40
	Curriculum Vitae . . . . .	43

# List of Tables

3.1	Description and specifications for acoustic classes. . . . .	21
3.2	Correct classification results using 30% of data for training. . . . .	27
3.3	Correct classification results using 50% of data for training. . . . .	27
3.4	Classification confusion matrix for the classifier using the DWPT with RBF-SVM using 50% of data for training. . . . .	28
4.1	Correct classification results for the mixed-signal distributed system using 30% of data for training. . . . .	35
4.2	Correct classification results for the mixed-signal distributed system using 50% of data for training. . . . .	35

# List of Figures

1.1	Example of multi-hop communications over a wireless sensor network . . .	6
2.1	A two-channel maximally decimated filter bank. This structure is the building block of discrete wavelet transforms . . . . .	9
2.2	a) Diagram for a three level 1-D DWT b) Diagram of three-level 1-D DWPT	10
2.3	A linearly-separable two-class classification example . . . . .	11
2.4	Example of a non-linearly separable problem. (a) There is no linear decision surface that can separate the data. (b) Data can be separated with a non-linear decision surface (a parabola). (c) Data projected to a different space using a kernel mapping $\Phi$ , the data is linearly separable in this space. . . .	12
2.5	Illustration of the infinite number of decision boundaries that can be produced for a linearly separable problem. . . . .	13
2.6	Different decision boundaries with different margins for the same dataset. .	14
2.7	Geometrical and vector representation to relate the margin to the decision boundary $\mathbf{w} \cdot \mathbf{x} + b$ . . . . .	15
3.1	Block diagram of typical pattern recognition system. . . . .	21
3.2	Digital signal architecture . . . . .	22
3.3	General Block Diagram . . . . .	22
3.4	Decomposition Filters for DWPT . . . . .	23
4.1	WSN architecture for a distributed/collaborative acoustic signal classifier. .	31
4.2	Architecture for a mote analog front-end for acoustic signal processing. . .	33
4.3	Magnitude responses for the analog filters comprising a distributed filter bank . The total frequency coverage is 16 KHz where each filter has a bandwidth of 2000 Hz. Filter design is described in the text. . . . .	34

# Chapter 1

## Introduction

This chapter introduces the idea of acoustic monitoring within the context of wireless sensor networks (WSNs). We present an overview of the previous and current work that has been done in acoustic classification using Wireless Sensor Networks (WSN). Further, we discuss current developments on computer hearing which is a field in its infancy with several open questions [13].

Although many acoustic monitoring applications have been developed, acoustic monitoring with wireless sensor networks is still new, and very few systems can recognize human voice, discriminate human speech and music, or bird songs [19]. These functionalities or services are important, since they are indispensable for surveillance, environmental monitoring, and smart home environments. For example, an unsupervised WSN can be deployed over a remote region with the goal of identifying and classifying events or entities based on their acoustic signature (i.e., the sound they produce).

Typical approaches for audio monitoring obtain audio signals using powerful sensor nodes (or motes) capable of sampling signals at a high rate and then transmit them to a base station to do the classification of the signals. A recent scheme that has shown good results is EnviroMic [12]. This system takes advantage of a WSN topology to do *cooperative recording*. This is a self-organizing scheme where one mote assumes the leader role and coordinates the remaining motes to store and process acoustics events over non-overlapping time segments. This was accomplished by an innovative algorithm that allowed for lower processing time and power consumption on each device, extending the lifetime of the network.

## 1.1 Problem Description

Sensor network research for environmental monitoring focuses on low bandwidth sensing such as light, temperature, motion and magnetic fields. In this case, a sample may be acquired every few seconds, minutes or even hours. Notable exceptions include efforts such as structural monitoring [11] and volcano monitoring [22] where vibrations were recorded at a frequency of 100Hz. All of the above services, however, assume the availability of a base station for data uploading and central processing.

Another category of environmental monitoring application focuses on disconnected deployment where data collection occurs only sporadically when researchers drive by the field as in the ZebraNet system [15] used to monitor wild life. Similarly, the SATIRE system [6] monitors targets as they approach the base-station. These applications, however, operate on low-bandwidth sensors that do not produce large data volumes.

On the other hand, acoustic signals are characterized for having spectrums that can consist of a few tones to a large bandwidth spanning the full hearing range (20 Hz to 20 KHz). We refer to the latter as *wideband* signals. Current sound recording hardware can record signals at very high sampling rates, above 44,000 samples per second, which ensure high quality reproduction. However, the acquisition of such high quality signals may not be necessary for monitoring applications. Furthermore, for WSNs, acquisition of wide band acoustic signals poses a difficult or impossible task given the hardware limitations. There are two main issues to consider:

- A mote is a constrained system with limited memory, power and hardware features (e.g., clock speed, memory size, etc.) which severely limit the sampling rate of the ADC. For instance, the *mica2* motes [21] allow a “high speed” sampling of around 4000 Hz which implies a maximum signal bandwidth of 2000 Hz, and even lower for practical applications as some degree of oversampling is always needed in practice.
- Even if high frequency sampling was possible, a WSN provides low data rates and short communication ranges which would not allow the transmission of large amounts

of acoustic data without creating network bottlenecks and high power consumption. As is well known, WSNs try to minimize the number of messages transmitted, as radio modules consume large amount of energy.

Hence, within the WSN context, analysis and classification of acoustic signals poses a challenging problem that is limited by hardware constraints and the inability to transmit wideband digital acoustic signals to a base station (even at moderate sampling rates).

## 1.2 Proposed Solution

In this thesis, we explore the possibility of discriminating between acoustic signals applying the idea of collaborative processing. Collaborative signal processing is one of the main ideas promoted in the area of WSNs [12]. The objective is to implement collaborative WSN schemes which will merge or aggregate their processing and/or communication capabilities to solve a complex problem on a distributed way.

To achieve the above goal, we need to explore and integrate knowledge from multiple disciplines, particularly in audio recognition, machine learning and WSNs. The suggested solution will consists of a set of nodes that perform in-node feature extraction using a distributed acoustic filter bank. The resulting features will be aggregated on a cluster head to form feature vectors which can then be used as inputs to a pattern classifier. To avoid sampling signals at high sampling rates, we will use mixed (analog and digital) signal processing. First we will compute energy-based features in the analog domain and then sample the “energy signal” at a very low rate to continue processing and classification in the digital domain. We present the development of a system that will distinguish between five acoustic classes: human speech, small explosions, bird songs, vehicles, and human steps.

## 1.3 Pattern Recognition

In machine learning, pattern recognition is the assignment of a label to a set of attributes or features that are considered representative characteristics of a pattern. An example of pattern recognition is classification, which attempts to assign each input value to one of a given set of classes [4]. However, pattern recognition is a more general problem that encompasses other types of output as well. Another example is sequence labeling, which assigns a class to each member of a sequence of values. For example, in speech tagging, a part of speech is assigned to each word in an input sentence. Another example is text parsing which assigns a parse tree to an input sentence, describing the syntactic structure of the sentence [24].

Classification has been extensively studied over the last decades where supervised statistical classifiers have taken a significant role in daily life applications [4]. Classical classifiers are based on Bayes decision theory where decisions are based using parameterizations on the feature distributions. Linear discriminant has led into the theory of the perceptron and artificial neural networks which have aimed at recreating the functionality of our brains. More recently, support vector machines (SVMs) have emerged as powerful supervised classifiers which maximize the distance between among the classes represented by the training vectors, typically after taking data to a higher dimensional space using kernel mappings. SVMs address the generalization problem under a strong theoretical framework.

As explained in the next chapters, in this thesis we will explore the use of filter banks (e.g., wavelets) and SVMs for acoustic signal classification. Filter banks, the discrete wavelet transform and related filtering structures have been successfully used for feature extraction in audio and image classification [22]. A simple, yet effective feature set consists of estimated the subband energies which provide a representative signature of the content over the frequency plane.

## 1.4 Wireless Sensor Networks

Wireless Sensor Networks (WSN) research was motivated by military applications and was initially funded by the Defense Advanced Research Projects Agency (DARPA). Recent advances in integrated circuits, *short range* wireless communications, and MEMS, made it feasible to build very small hardware systems with signal processing and wireless communication features which can be integrated to a diverse suite of sensors while achieving low energy consumption. It was envisioned that such devices could be used to form WSNs capable of monitoring and measuring phenomena at unprecedented space-time scales and locations. The combination of sensors and wireless communications has enabled applications like lead to in-network detection and tracking schemes [7], ambient monitoring [22] and smart environments [9].

Perhaps, the main challenge in WSN is to minimize energy consumption as battery replacement is not an option on most cases. A node typically consists of four main sub-systems; microcontroller (MCU), power unit, sensor interface and a transceiver. The main functions of the MCU are to interface with the sensors (typically through an ADC), coordinate wireless communications and to establish collaborations with other motes. A second factor that is important in WSN is the cost of each mote. However, minimization of the WSN mote cost results in limited processing speed and limited memory size of the microcontroller. The sensors installed in the mote must be low cost as well as low power.

An example of a common WSN configuration is shown in Figure 1.4. This network consists of a group of sensor nodes (i.e., motes) connected to gateway which delivers information to a base station. Suppose that one of the motes, located far from the gateway, detects an event of interest. In order to send this information to the base station, a multi-hop communication protocol is necessary since motes have a very short range compared to the geographical regions they cover. Effectively, the WSN organizes to move information from one node to another until the destination is reached. We note that the type of event detection depends on the sensors integrated to the mote and the signal processing

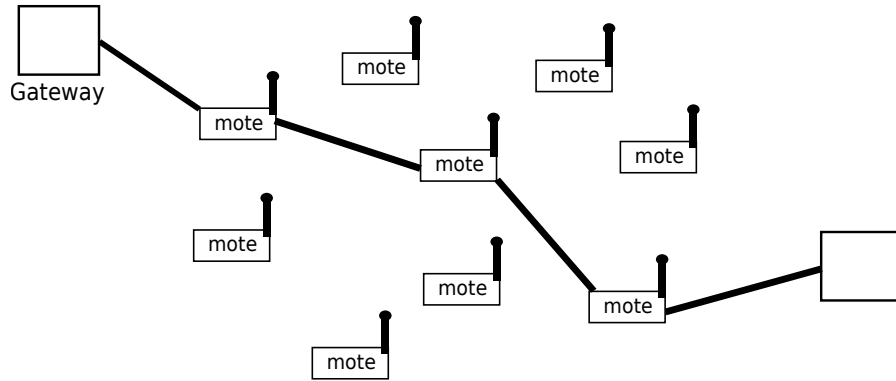


Figure 1.1: Example of multi-hop communications over a wireless sensor network

performed in the mote.

Compared to a single sensor system, wireless sensor networks have many advantages. First, they are very easy to deploy. For instance, sensor motes can be dropped from an airplane into a forest, and then they automatically communicate with each other to form a network suitable for forest fire monitoring. Second, due to the low cost of each individual sensor, sensor nodes can be densely deployed. WSNs seek to provide in-network collaborative schemes which can fuse and aggregate information from several (computationally limited) nodes to produce good parameter estimates, event detection and optimal decisions. Third, when multiple sensors work together the information redundancy among the sensors and the communication channels in the networks enable a better sensing. This allows for a better geographical coverage and provides more reliable information delivery.

## 1.5 Audio Monitoring Using WSN

Acoustic signals contain a great deal of information about their generating sources which can be exploited for detection, classification and recognition. Due to this reason, acoustic monitoring based on wireless sensor networks has received much attention and has been used in a lot of applications. For example, a research group in UCLA built an acoustic

habitat-monitoring sensor network, which recognizes and locates specific animal calls in real time [19]. Simon developed a system based on adhoc wireless sensor networks to detect and locate shooters in urban environment[18]. Phadke designed an embedded speech recognition system, which is capable of recognizing a spoken word from a small vocabulary of about 10-15 words[14]. Lou, et al. [12] created a system know as *EnviroMic* that is being implemented for the study of mating rituals and social behavior of animals in the wild, and for to audio surveillance of military targets.

An emerging field in machine learning is machine hearing. Hearing machines should be able to organize what they hear, they should learn names for recognizable objects, actions, events, places, music styles, etc [13]. These machines should be able to listen and react in real time and take the right actions depending on the event. There are many applications to this newly discovered area, such as, seismic exploration, music recording and compression, speech synthesis and recognition. However, there are much more general applications that are not really related to hearing. For example sonar target detection and classification. As discussed in [13], Google is one of the pioneers on this area. They store information about a lot of sound clips, with an emphasis on video audio tracks. Machine hearing will be an important tool to automate categorization and indexing which is crucial for efficient web-based search. It is expected that the advances on machine learning and machine vision will be usable in machine hearing which should lead to significant developments on this area on the short term.

# Chapter 2

## An Acoustic Signal Classifier Based On Wavelets Packets

In this chapter we provide a introduction to the theory of the Wavelet Packet Transform and of support vector machines. These two algorithms form the basis of the classification algorithm proposed in this thesis. The block diagram for the proposed acoustic signal classifier algorithm is presented on Figure 3.3. It is a classical scheme where data is first decomposed by a multi-resolution transform (e.g., wavelets and filter banks), followed by a set of non-linear operations that estimate the energy content for each subband. These energy measures are grouped as a feature vector that provides a signature for a given sound. It is expected that these feature vectors have discriminative power which allows their use for training a supervised classification algorithm (e.g., neural networks, SVMs, etc.) which can later be used to classify previously unseen patterns.

### 2.1 Wavelet Packet Transform

Wavelets and filter banks have emerged as important signal analysis tools that allow their decomposition in the form of a multi-resolution analysis. In particular, the discrete wavelet transform (DWT) has been extensively studied. Algorithmically, the building block of the 1-D DWT is implemented with the two-channel filter bank shown in Figure 2.1. An input signal  $x[n]$  is filtered with is low and high passed filtered by  $h_0[n]$  and  $h_1[n]$ , respectively, and followed by a down-sampling by a factor of 2 [25]. Mathematically this is expressed as

$$a[n] = \sum x[k]h_0[2n - k]$$

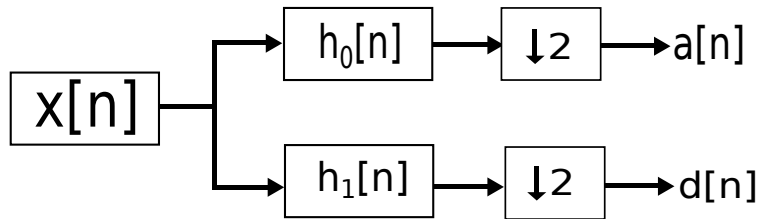


Figure 2.1: A two-channel maximally decimated filter bank. This structure is the building block of discrete wavelet transforms

and

$$d[n] = \sum x[k]h_1[2n - k].$$

The downsampling operations ensure that there is no data expansion. The resulting signals  $a[n]$  and  $d[n]$  are typically called subbands and represent, respectively, approximation and detail transform coefficients. They are also known as lowpass and highpass frequency channels. When the DWT is applied to a low frequency signal, most of the energy will be compacted in the  $a[n]$  subband resulting in small coefficients for  $d[n]$ . This property is useful for signal compression of many types of signals found in nature.

The DWT is implemented with the filtering structure shown in Figure 2.2-(a). The DWT is obtained by iterating the two-channel filter bank on  $a[n]$ . The intuition behind this structure is that it applies to signals where energy is concentrated at low frequencies and repeated decomposition allows separation of signal components at lower and lower resolutions.

A related decomposition is the discrete wavelet packet transform (DWPT). The algorithmic implementation is shown in Figure 2.2-(b) where the two-channel filter bank is uniformly applied to  $a[n]$  and  $d[n]$  across all levels of decomposition. In this case, the frequency plane is partitioned into subbands with uniform bandwidth. A uniform filter bank is part of the MP3 standard for coding and compression of audio. In the case of acoustic signals, the DWPT seems to be a more appropriate decomposition. We should note that

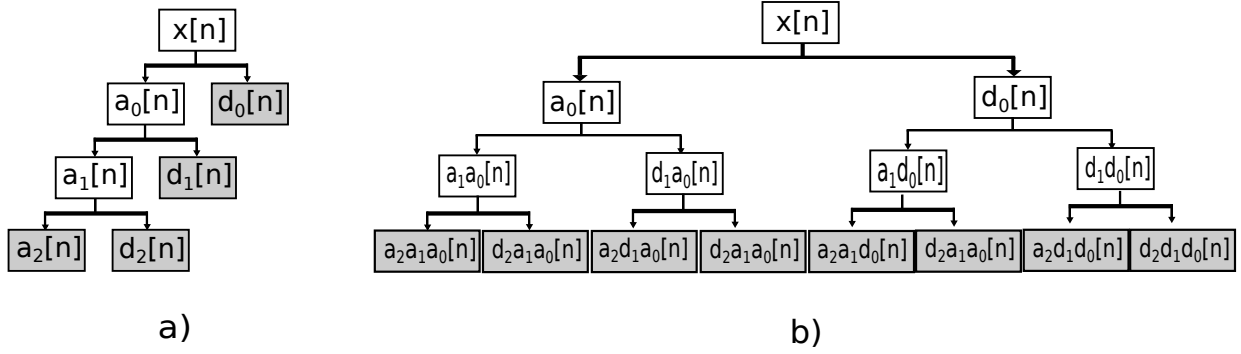


Figure 2.2: a) Diagram for a three level 1-D DWT b) Diagram of three-level 1-D DWPT

for a three-level DWPT, the overall downsampling ratio is eight.

## 2.2 Support Vector Machines

Support vector machines (SVMs) are a type of supervised learning methods that are used for classification and regression analysis. The original SVM algorithm was invented by Vladimir Vapnik. SVMs are used in different areas of machine learning applications. Some of these applications include image processing, face and hand-written character recognition, speaker identification/verification, and speech recognition.

The standard SVM is a supervised binary classifier where each input pattern is assigned one of two possible classes, for example a value from the set  $\{-1, +1\}$ . SVMs are characterized by finding an optimal linear decision boundary between the two classes based on the geometrical properties of the data over the N-dimensional feature space. They are also characterized by formulating the problem without the need to approximate a posterior probability as is done by many classifiers based on Bayes decision theory. The SVM training algorithm used known, labeled, examples to find the optimal parameters of the linear model that defines a decision hyperplane [9]. As will be discussed later, the optimality criteria is based on minimizing the separation margin within classes.

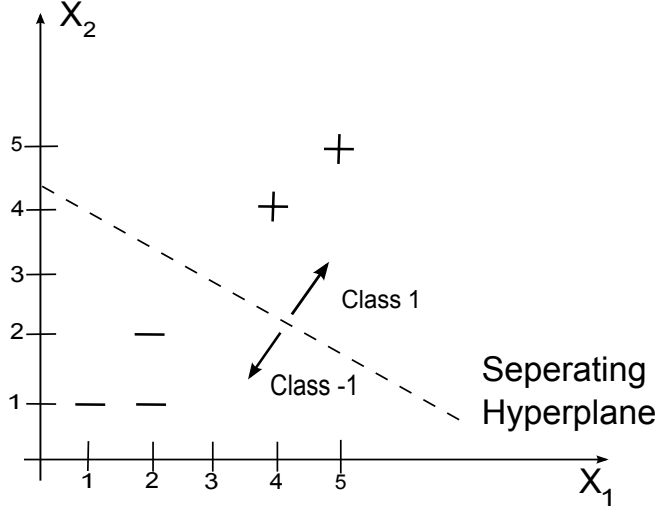


Figure 2.3: A linearly-separable two-class classification example

### 2.2.1 Linear Classification

The purpose of classification is to find decision boundaries that satisfy certain conditions and separate input data into different classes. Figure 2.3 illustrates an example of linear classifier boundary that divides class  $+$  and  $-$ . These boundaries are called separating hyperplanes.

To further appreciate the concept of a linear classifier, we need to introduce the idea of class separability and non-linear decision boundaries. Suppose we have a set of labeled examples or training feature vectors  $\mathcal{T} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$  where  $\mathbf{x}_i \in \mathbb{R}^n$  and  $y_i \in \{-1, 1\}$ . For our discussion we refer to Figure 2.3 which shows a an example of a two class problem with  $n = 2$ . Figure 2.4-(a) shows an example where, clearly, the training data cannot be separated by a line; consequently, we say that the data is non-separable. Further inspection shows that there is a parabola-shaped decision boundary as depicted in Figure 2.4-(b). In this case, a *non-linear* classifier could be used. Non-linear classifiers, like artificial neural networks (ANNs), have been developed over the years to deal with data set that exhibit complex decision boundaries. Finally, Figure 2.4-(c) shows the application of the so-called “kernel trick” which applies a mapping operation  $\Phi$  that maps the data to an alternate

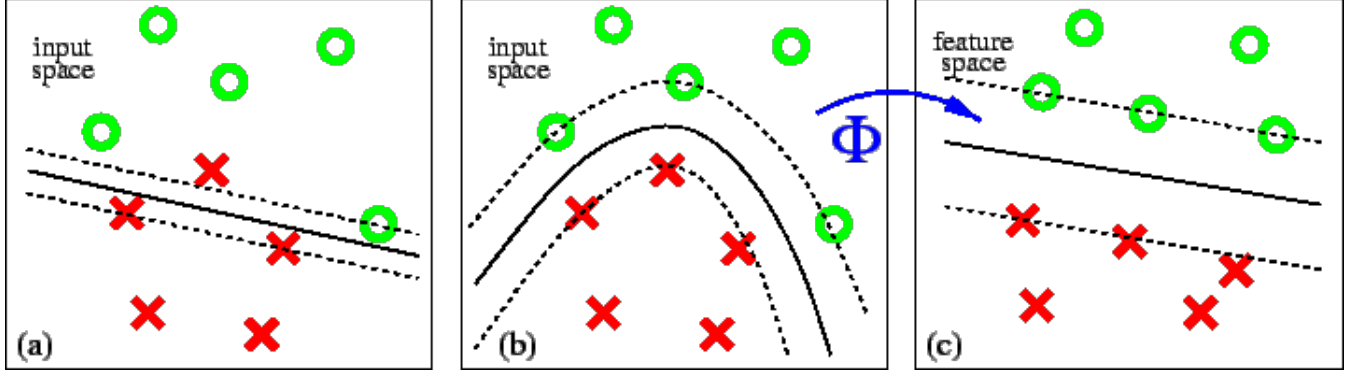


Figure 2.4: Example of a non-linearly separable problem. (a) There is no linear decision surface that can separate the data. (b) Data can be separated with a non-linear decision surface (a parabola). (c) Data projected to a different space using a kernel mapping  $\Phi$ , the data is linearly separable in this space.

space where data becomes linearly separable. The use of kernels is an important component of SVMs as will be discussed later.

Hence, for linear classification the separating hyperplane is defined by a weight vector noted as  $\mathbf{w}$  and its bias factor  $b$ . Figure 2.3 shows the separating hyperplane in a two-dimensional space. We can generally express the decision boundary as

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (2.1)$$

where  $\mathbf{w} = [w_1 \ w_2]^T$  and  $\mathbf{x} = [x_1 \ x_2]^T$ . and superscript  $T$  denotes the transpose of a vector. As denoted in Figure 2.3 class 1 is represented by a plus-sign and class  $-1$  is represented by a minus-sign. Using Equation 2.1 we can now determine the class of any vector  $x_i$  as

$$y_i = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b) = 1 \quad (2.2)$$

for class  $+$  and

$$y_i = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b) = -1 \quad (2.3)$$

for class  $-$  where  $\text{sgn}()$  is the signum operation. As shown in Figure 2.5, there is an infinite

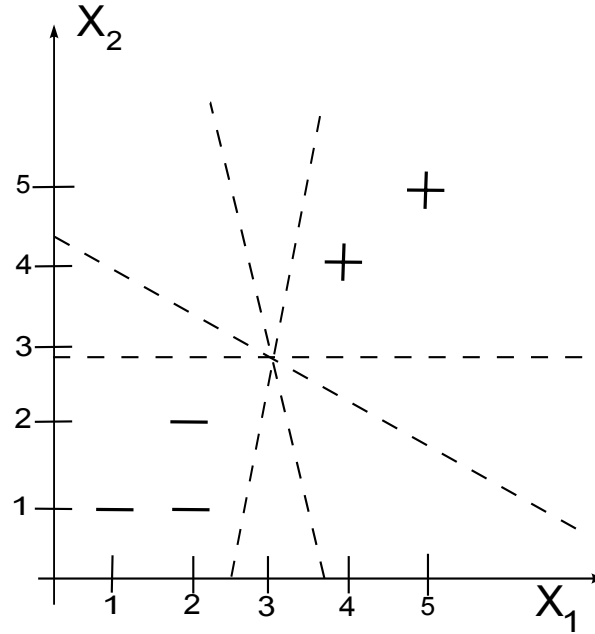


Figure 2.5: Illustration of the infinite number of decision boundaries that can be produced for a linearly separable problem.

number of lines, hence values of  $w$  and  $b$  that will generate a valid decision boundary. Is there a decision boundary that is better than all the others?

To motivate the development of the support vector machines (SVM), we start by considering the two plots presented on Figure 2.6. Besides the decision boundaries, each plot depicts an additional pair of parallel lines that define a band-shaped area. The width of this band is known as the margin, and is delimited by the imaginary lines that intersect those feature vectors that are closest to  $\mathbf{w} \cdot \mathbf{x} + b = 0$ . The two examples in Figure 2.6 show a wide margin and a narrow margin example. Intuitively, a wider gap will be preferred since it provides more wiggle room to account for unseen feature vectors and minimizes the probability of misclassification.

Hence, it makes sense to look for the values of  $\mathbf{w}$  and  $b$  that maximized the margin. This idea is known as classifier generalization and warrants that we the classifier will not suffer from data over fit to the training data. Support vector machines (SVMs) addresses these

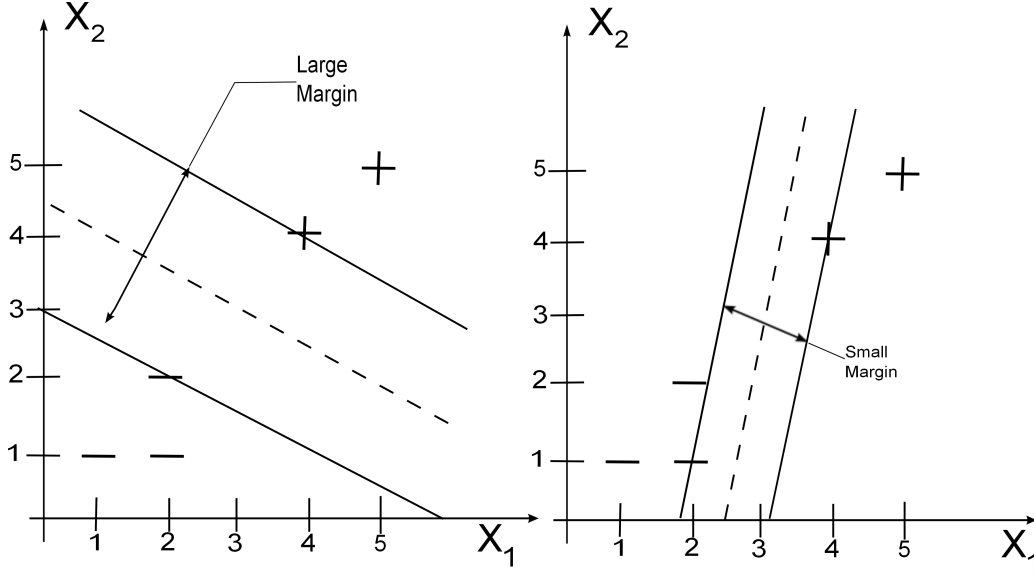


Figure 2.6: Different decision boundaries with different margins for the same dataset.

issue by providing a method that finds line (i.e., hyperplane for higher dimensions) that maximizes the *margin* between classes. The name of the SVM algorithm comes from its direct use of the training data to find the line that maximizes the margin. Checking Figure 2.6, we define the boundary lines as

$$\mathbf{w} \cdot \mathbf{x} + b = 1 \quad (2.4)$$

and

$$\mathbf{w} \cdot \mathbf{x} + b = -1 \quad (2.5)$$

where the right sides are normalized to 1 and  $-1$  respectively. Then, it is straightforward to see that  $\mathbf{w} \cdot \mathbf{x} + b \geq 1$  for all training vectors belonging to class  $+$ , and  $\mathbf{w} \cdot \mathbf{x} + b \leq -1$  for all vectors belonging to class  $-$ . We see that for the wide margin case, there are two vectors that intersect the lines (one for class  $+$  and one for class  $-$ ). For the narrow margin case, there are two  $-$  vectors and one  $+$  vector intersecting their corresponding lines. The vectors where equality holds are called *support vectors*. More generally, the set of support

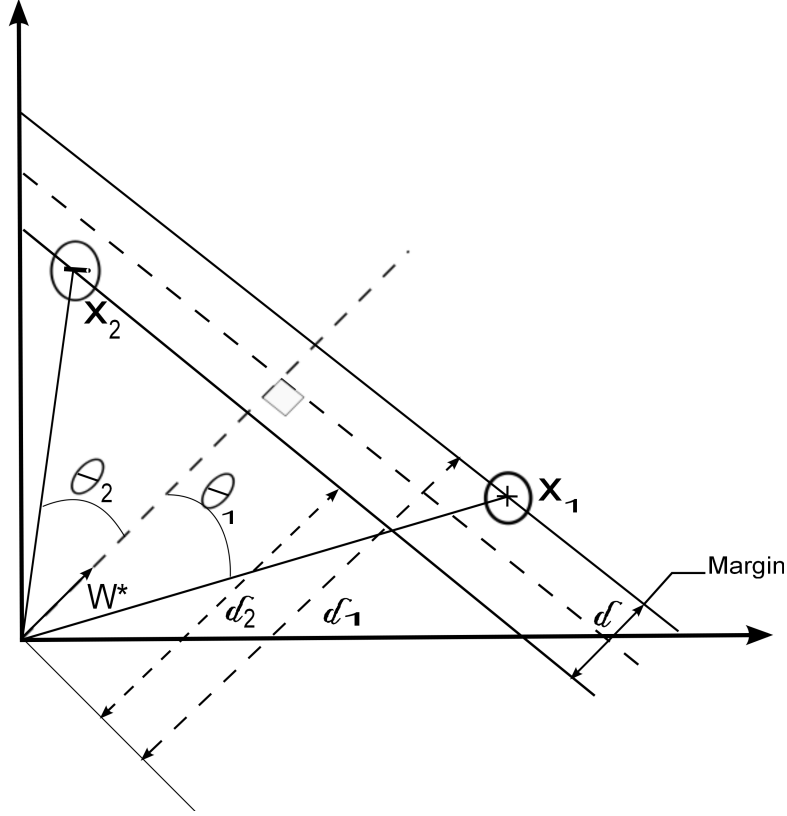


Figure 2.7: Geometrical and vector representation to relate the margin to the decision boundary  $\mathbf{w} \cdot \mathbf{x} + b$ .

vectors  $SV \subset \mathcal{T}$  is defined as

$$SV = \{\mathbf{x} \mid \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b) = 1\}.$$

The name implies that these are the vectors that are used to help (i.e., *support*) the process of finding the maximum margin through the solution of an optimization program. In summary, the optimal separating hyperplane is that achieves class separation and maximizes their margin in order to achieve a maximum gap between classes.

To establish a relationship between the margin and the parameters  $(\mathbf{w}, b)$ , Figure 2.7 shows a geometrical perspective. Consider the two points on the margin boundaries,  $x_1$  and  $x_2$ , corresponding to the classes  $\{1, -1\}$  respectively. The vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , consider

form angles  $\theta_1$  and  $\theta_2$  with vector  $\mathbf{w}$ . Note that  $\mathbf{w}$  is orthogonal to the decision line. Using basic trigonometry, we can obtain expressions for the distances  $d_1$  and  $d_2$  as

$$d_1 = \|\mathbf{x}_1\| \cos \theta_1 \quad (2.6)$$

and

$$d_2 = \|\mathbf{x}_2\| \cos \theta_2 \quad (2.7)$$

respectively. Next, the margin can be obtained as distance between the margins is given by

$$d = d_1 - d_2. \quad (2.8)$$

Substituting 2.6 and 2.7 into 2.8 we obtain,

$$d = \|\mathbf{x}_1\| \cos \theta_1 - \|\mathbf{x}_2\| \cos \theta_2. \quad (2.9)$$

Using the definition of the dot product for vectors, we have

$$\cos \theta_1 = \frac{\mathbf{w} \cdot \mathbf{x}_1}{\|\mathbf{w}\| \|\mathbf{x}_1\|}, \quad (2.10)$$

with a similar expression for  $\cos \theta_2$ . Then, the expression for the margin can be rewritten as

$$d = \frac{\mathbf{w} \cdot \mathbf{x}_1 - \mathbf{w} \cdot \mathbf{x}_2}{\|\mathbf{w}\|}. \quad (2.11)$$

Using the margin boundary equations (2.4) and (2.5) with  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , we finally arrive to

$$d = \frac{1 - b - (-1 - b)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}. \quad (2.12)$$

Clearly, there is an inverse relationship between the margin and the norm of  $\mathbf{w}$  which implies that we need to minimize  $\|\mathbf{w}\|$  to achieve maximum margin while satisfying the inequalities formed by the training vectors.

More formally, the minimization problem is defined as

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$

subject to

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$$

for  $i = 1, 2, \dots, N$ . The square of the norm is used to avoid the complexity of dealing with a square root, and the  $1/2$  term is used for mathematical convenience. This is a quadratic optimization problem that can be solved through well established techniques. One approach is to use to form the Lagrangian

$$L(\mathbf{w}, b) = \left\{ \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \right\} \quad (2.13)$$

where  $\alpha_i$  are the Lagrange multipliers. Taking the derivative of  $L(\mathbf{w}, b)$  with respect to  $\mathbf{w}$  and  $b$  and equating to zero, it is possible to show that

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (2.14)$$

and

$$\sum_{i=1}^N \alpha_i y_i = 0. \quad (2.15)$$

Substituting (2.14) into (2.13), an equivalent optimization program is found as a function of only the  $\alpha_i$ 's. This program is known as the dual and can be expressed as

$$\max_{\alpha_i} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.16)$$

for  $i = 1, 2, \dots, N$ , and subject to  $\alpha_i \geq 0$ , and equation (2.15). This last problem is also a quadratic problem and is easier to solve in practice using different algorithms. It is relevant to note that expressions 2.14 and 2.16 rely solely on the computation of dot products.

## 2.2.2 Non-linear SVM Classifier

Most problems will present data with non-linear decision boundaries. As it turns out, it is straightforward to extend the concept of linear SVMs to the non-linear case. This extension is achieved through the use of a non-linear kernel mapping  $\Phi : \mathbb{R}^n \mapsto \mathbb{R}^m$  where  $m > n$ .

Hence, training vectors  $\mathbf{x}_i$  are mapped to vectors  $\mathbf{z}_i = \Phi(\mathbf{x}_i)$  on a higher dimensional space in such a way that the data becomes linearly separable. Such a mapping is illustrated in Figure 2.4. Once the data is transformed into this m-dimensional space, the linear SVM algorithm can be applied.

To solve the problem, we need to substitute the dot products  $\mathbf{x}_i \cdot \mathbf{x}_j$  on equations (2.16) and (2.14) with  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ . However, a big disadvantage is that the computational complexity can be substantially increased by the kernel mapping operation (and the dot product operation). Fortunately, to apply the SVM algorithm, we only need the numerical result of the dot product operation which can be obtained through the so-called “kernel trick.” In many cases of interest, is possible to find a function  $K(\mathbf{x}_1, \mathbf{x}_2)$  that computes  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$  directly in  $\mathbb{R}^n$ . Some standard choices for kernel functions are the polynomial function and the radial basis functions [23].

Hence, the new optimization problem becomes

$$\max_{\alpha_i} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

for  $i = 1, 2, \dots, N$ , and subject to  $\alpha \geq 0$ , and equation (2.15). Similarly, the decision function becomes

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x})\right)$$

### 2.2.3 Non-separable case

When data remains non-separable due to the presence of noisy samples, a soft-threshold model is adopted such that we look for the solution to

$$\min_{\mathbf{w}, \xi, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right\}$$

subject to

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i,$$

with  $\xi_i \geq 0$ . The  $\xi_i$  terms measure the distance error between a miss-located vector (i.e., on the wrong side of the decision boundary) and the boundary line for the correct class

(eqns. (2.4) and (2.5)). The parameter  $C$  is a degree of freedom that sets the penalty of these errors, and is obtained by trial and error. A detailed discussion of this scheme is found in [3].

## 2.3 Conclusion

In this chapter, we reviewed the theory of the discrete wavelet packet transform (DWPT) and of support vector machines (SVMs). These are the key components for the acoustic signal classifier developed in the next chapters. The DWPT will be used as the front end processing element for an energy-base feature extraction scheme. The SVM will be the classifier of choice which will be compared against other well known schemes.

# Chapter 3

## A Classification Algorithm for Acoustic Signals

The following section gives the results of the proposed classification algorithm using Support Vector Machines. The results obtained are compared with other classifiers including the nearest neighbor classifier and the backpropagation neural network. The algorithm was tested using different parameters in order to obtain the best results. The parameters are the percentage of data points from the data set used to train the classifier, and the bandwidth of the filter.

### 3.1 Description of Data Sets

This thesis focuses on classifying five classes. We identify class 0 as birds, class 1 as bombs, class 2 as car speeding/revving engine, class 3 as people having a conversation, and finally class 4 is people walking on foot. A data set of acoustic signals was obtained from *www.freesound.org* which offers an extensive database of high quality audio segments. Table 3.1 includes a break down of the number of audio clips. All of the audio clips had a sample rate of 44,100 Hz.

### 3.2 Feature Extraction

This section will explain the procedure of how the feature vectors were obtained. The section will describe the architecture of the theoretical block diagram of the system. Figure 3.1

Table 3.1: Description and specifications for acoustic classes.

Class ID	Class	# Audio Clips	Total Seconds	Sample Rate
0	Birds	6	355	44100
1	Bombs	7	74	44100
2	Car	6	199	44100
3	Conversation	6	359	44100
4	Foot	6	202	44100

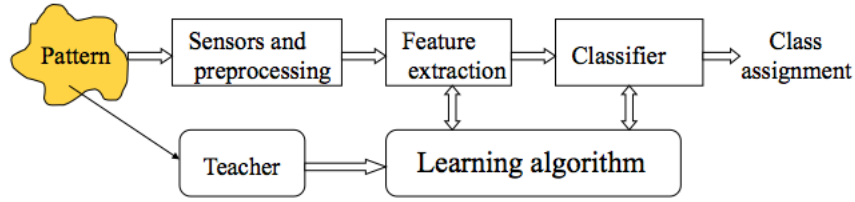


Figure 3.1: Block diagram of typical pattern recognition system.

illustrates a typical block diagram of pattern recognition system.

Feature extraction is often an important pre-processing step in classifier design, in order to overcome the problems associated with having a large input space. A common way of doing this is to use Principle Component Analysis (PCA) to find the most important features. However, it has been recognized that this may not produce an optimal set of features in some problems since the method relies on the second order statistics (covariance structure) of the data. Figure 3.2 shows the proposed architecture for the the digital feature extraction system used in this thesis. The system computes a three-level DWPT of a signal and then estimated the energy for each subband. The energy estimates are the input feature to the classifier [20].

The feature vectors were obtained by first breaking each signal from Table 3.1 into one second segments (i.e., 44,100 samples) which then where processed as in Figure 3.3

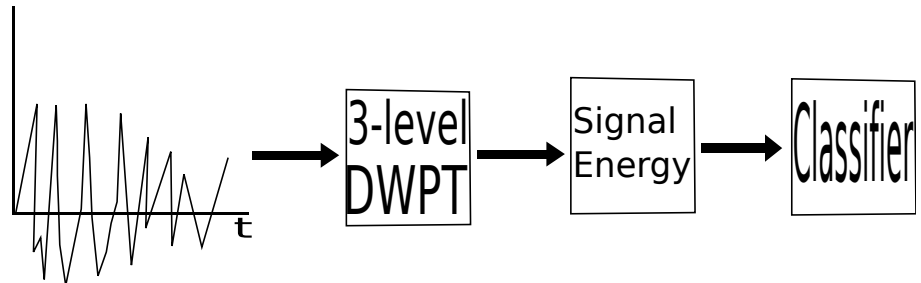


Figure 3.2: Digital signal architecture

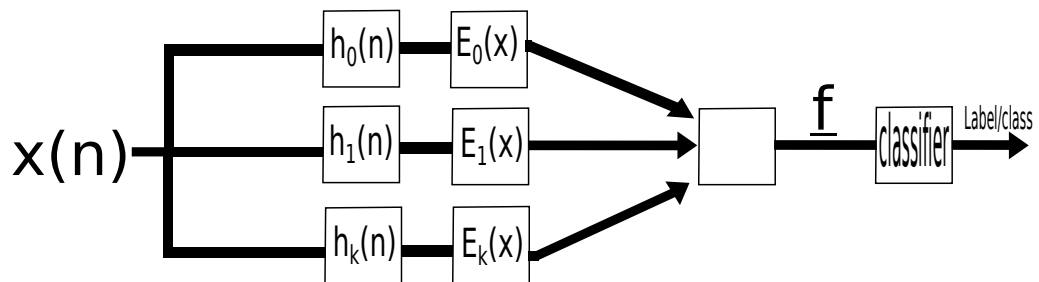


Figure 3.3: General Block Diagram

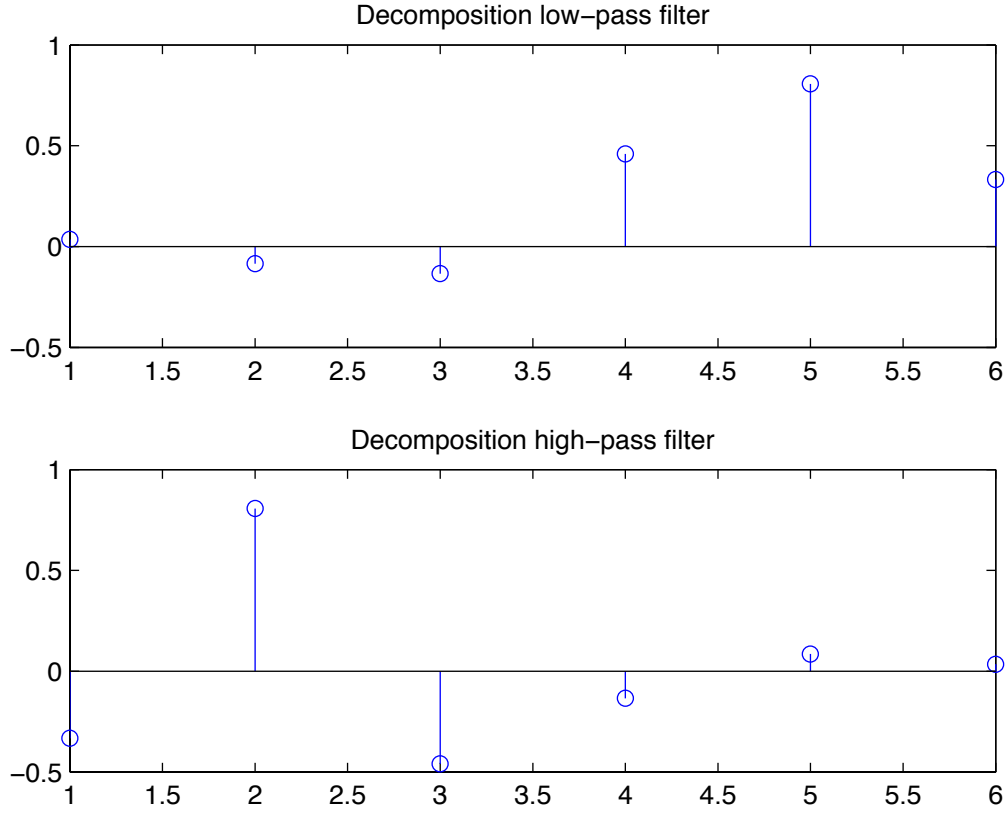


Figure 3.4: Decomposition Filters for DWPT

using an wavelet packet transform. The three-level DWPT refers to the Discrete Wavelet Packet Transform described in Section 2.1 and in Figure 2.2. For this thesis we used the Daubechies wavelet filters with the filter responses shown in Figure 3.4. Recall that the DWPT decomposes the input signal into the low and high frequency band (i.e  $a[n]$  and  $d[n]$ ) at every level of the decomposition. Thus, giving us a total of eight output signals at the 3-level DWPT.

Once the subband signals are obtained, feature extraction is performed by computing the subband energy over one second periods for each audio sample. In this thesis, we use

the squared  $\ell_2$  norm given by

$$||x||_2 = \sum_{k=0}^{K-1} (|x[k]|^2). \quad (3.1)$$

as an energy measure. (Note that for a sampling rate of 44100 samples per second,  $K = 44100/8$  given that the overall downsampling ratio of a three-level DWPT is eight.) Hence, a feature vector is formed by concatenating the subband energies  $e_i$  as

$$\mathbf{f} = [e_1 \ e_2 \ e_2 \ \dots \ e_8]^T \quad (3.2)$$

Totalling the fourth column on Table 3.1, we have 1189 seconds which corresponds to the same number of feature vectors for all classes.

### 3.3 Brief Description of other Non-linear Classifiers

In this section we briefly describe three well known classification algorithms that we will use in this thesis. The three classifiers are non-linear, implying that they will find non-linear decision surfaces. First we introduce the k-Nearest Neighbor classifier, followed by artificial neural networks (ANNs), and finally by support vector machines (SVMs). We note that the SVM theory was presented in Section 2.2. For a detailed treatment of ANN and k-NN classifiers we refer the reader to [4].

#### 3.3.1 K-Nearest Neighborhood Classifier

As its name implies, the k-NN algorithm classifies a test pattern based on its proximity to the training examples in the feature space. Proximity is determined according to some predefined distance function. The k-NN classifier approximates a classification rule locally without actual training; decisions are fully determined at classification time using the complete training set. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the test pattern being assigned to the most common class within its k nearest neighbors.

If  $k = 1$ , then the object is simply assigned to the class of its nearest neighbor [10]. For instance suppose that  $k = 3$ , and that the  $k$  nearest vectors to the test feature vector correspond to classes A, A, and B, then the test point will be classified to class A because it has the majority vote.

The value of  $k$  is a user-defined constant which has to be determined through experimentation. If the statistical properties of the data are stationary, the best value of  $k$  found experimentally, will provide the best results over unseen patterns.

### 3.3.2 Back Propagation Neural Network

An Artificial Neural Networks (ANN) are a family of classifiers inspired by the way biological nervous systems, such as the brain, process information. In particular, the multi-layer perceptron (MLP) is a feed-forward connected network that has been extensively applied to classification problems. The MLP is a structure composed of a large number of highly interconnected processing elements (neurons). The neurons are organized in layers identified as input, output and hidden layers. An input layer is used to assign a neuron to each feature in the feature vector. An output layer assigns a neuron to each of the data classes. There can be as many hidden layers as desired, but for classification purposes, a single hidden layer is enough to design a classifier that fits any set of non-linear decision boundaries. Each neuron is composed of many inputs, a single output and a non-linear activation function that “fires” when the sum of the inputs reaches some threshold. An important issue for ANNs is to determine the correct architecture, implying the optimal number of layers and number of neurons per layer for a given application. These parameters have to be determined by extensive experimentation [11].

ANN’s are supervised classifiers. There are different learning processes (i.e., training methods) reported in the literature. Error back-propagation is a commonly used technique that minimizes an error function iteratively by propagating the classification error of training data from output nodes to input nodes. The error is minimized through finding a set of weights used to inputs to all neurons.

### 3.3.3 Radial Basis Function Support Vector Machine

As described in Section 2.2.2 SVM can be converted to non-linear classifier through the use of a kernel mapping  $K(\mathbf{u}, \mathbf{v})$ . The radial basis function (RBF)

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \|\mathbf{u} - \mathbf{v}\|^2). \quad (3.3)$$

is one of the most popular kernel functions. The parameter  $\gamma$  determines the spread of the function. The value of  $\gamma$  needs to be determined experimentally. For the rest of this thesis we will use the RBF SVM in our simulations.

## 3.4 Results for Acoustic Classification

Classification experiments were performed using the DWPT for feature extraction and Radial Basis Function SVM (RBF-SVM) and Back Propagation Neural Network (BP-NN) and K-nearest neighborhood (k-NN) classifiers. All of these algorithms are non-linear classifiers. A training set was derived using the acoustic signals from Section 3.3 and the feature extraction method from section 3.2. The experiments were performed in MATLAB using the wavelet toolbox to implement the wavelet packet transform and the PRTools [5] classification toolbox to implement the classifiers. For the k-NN classifier a value of  $k = 5$  was obtained using leave-one-out cross-validation over the training set. A back-propagation neural network was designed with a single hidden layer consisting of 10 neurons and the threshold function. The number of hidden nodes was determined through experimentation. The back-propagation algorithm was run for 100 epochs. Finally, the  $\gamma$  parameter and the regularization parameter  $C$  for the RBF-SVM were obtained through an internal optimization of the classifier using cross-validation. To deal with multiple classes the *one-against-all* SVM scheme was used.

The algorithms were trained using the hold-out method. The training and testing set were obtained using two training/testing data splits: 30/70 and 50/50 percent. The classification results are shown in Tables 2.2 and 2.3. As expected, the second split provides

Table 3.2: Correct classification results using 30% of data for training.

	RBF SVM	BP NN	k-NN
Percentage Accuracy	57.8	56	25.8

Table 3.3: Correct classification results using 50% of data for training.

	RBF SVM	BP NN	k-NN
Percentage Accuracy	77.7	77.3	73.1

as better results with the SVM and ANN having a similar performance of 77%. It is also relevant to note that the improvement from 30/70 to 50/50 is very noticeable, in particular for the k-NN classifier.

Table 3.4 shows the confusion matrix. A confusion matrix is a tool typically used in machine learning. Each column of the matrix represents the instances in a predicted class, while each row represents the misclassification in an actual class. One benefit of a confusion matrix is that it is easy to see if the system is commonly mislabeling one as another. For example, for class 0, there are a total of 355 samples, which out of those, 323 were classified correctly. However, the classification algorithm misclassified class 0 five times to class 1, zero times for class 2, five times to class 3, and twenty two times for class 4. The classifier had the most problem between class 3 and class 2. If we look at table 3.4 most of class 2 was misclassified as class 3. A hypothesis of this error is that the audio clips contained similar background noise. Both of the clips used for this thesis didn't isolate the sound we were trying to classify. The conversation audio clips for class 3 were taken in an open cafe where you can hear on the background as cars pass in the background or you can hear the moving of other guests in the cafe.

Table 3.4: Classification confusion matrix for the classifier using the DWPT with RBF-SVM using 50% of data for training.

	Estimated Labels					
True Labels	0	1	2	3	4	Totals
0	323	5	0	5	22	355
1	2	71	0	1	0	74
2	13	5	52	111	18	199
3	50	2	6	266	35	359
4	44	0	0	39	119	202
Totals	432	83	58	422	194	1189

## 3.5 Conclusion

In this chapter we have presented three classifiers for acoustical signals using the DWPT for feature extraction. We note that the Car class exhibited the largest individual classification error while Birds, Bombs and Conversation had a good performance. The best classification error was found using support vector machines with artificial neural networks following closely. Our results show that the proposed classification system is appropriate for acoustics signals. There is a lot of room for improvement. In addition, the small size of the sample (i.e., number of patterns for training and testing) should be addressed through the acquisition of more acoustic signals and/or the development of schemes that account for small sample sets. This should be the subject of future work.

# Chapter 4

## A WSN for Mixed-Signal Distributed and Collaborative Sound Classification

### 4.1 A WSN model for acoustic signal classification

In this chapter we introduce an adaptation of an acoustic signal classifier to the WSN context. We expand the ideas from Chapter 3 for implementation into WSNs. As mentioned before, WSNs have many constraints with respect to power, computation, storage and communication data rate. This thesis proposes an approach to deal with these limitations in order to achieve acoustic signal recognition without the need to acquire, compress, and transmit large amounts of data through the network. The proposed solution has the following characteristics:

- Feature extraction is performed on a distributed and collaborative way. A set of motes self-organizes to form a filter bank.
- Each mote extracts one frequency subband from the signal and performs feature extraction.
- Each node transmits the extracted features to a concentrator node that forms the feature vector. The concentrator node can implement the classifier or send it to a base station for classification. All this is done at a very low data rate as opposed to transmitting full acoustic signals through the network [1].

Before we continue with our discussion, we note that the scope of this thesis is to introduce a method for classification of acoustic signals in WSNs. However, there are several more components/services that are needed to successfully implement and deploy such system. These components/services are not developed here, but we assume they will be available. At a minimum, a complete acoustic signal classification system will require:

- Geo-localization and self-organizing algorithms and protocols that will allow the nodes to form a distributed filter bank.
- Routing protocols that will allow the movement of feature data and classification decisions across the network.
- Algorithms to detect the presence of sounds different from the background. This allows the classification system to be only active when an event occurs in order to reduce power consumption.
- Algorithms to compensate for time of arrival in order to synchronize feature extraction across nodes. This is related to the geo-localization system.
- DSP algorithms to compensate for signal attenuation and for noise reduction.

In Figure 4.2 we depict a realization of the proposed network. As a sound source approaches the WSN, each node detects its presence and starts the self organizing process. We assume in our case that we need a sub-network of eight nodes to attend the acoustic source. As part of the self-organizing process, each node picks a frequency range (i.e., subband) over which it will process the signal. Next, acquisition, processing and feature extraction is done in the analog domain. The analog “feature” signal is digitized as a very low rate (one sample per second). Finally the features are transmitted to a concentrator node for classification. This process could be performed continuously as long as an acoustic source is present in the WSN field.

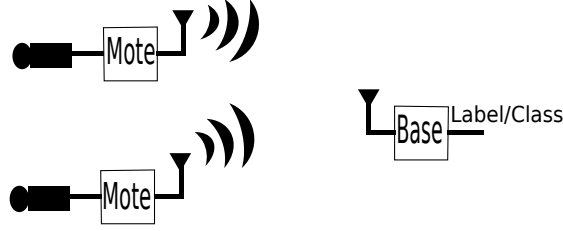


Figure 4.1: WSN architecture for a distributed/collaborative acoustic signal classifier.

## 4.2 Mixed-Signal Feature Extraction

In Chapter 3 feature vectors were created by using an all digital/discrete systems. The process was done by processing signals sampled at 44.1 KHz with the DWPT. In this section we introduce a scheme for mixed signal feature extraction. The main idea is to use a distributed bank of analog filters to process a signal and compute features with analog components before digitization. As we will discuss, this dramatically reduces the analog-to-digital converter requirements, power consumption, and memory storage requirements.

In signal processing, a filter bank is an array of band-pass filters that decomposes the input signal into multiple subbands. Each subband captures the signal content over a specific range of frequencies. The process of decomposition performed by the filter bank is called analysis[16]. In this thesis we consider an analog filter bank with eight contiguous channels. For testing purposes the filters were created using three different bandwidths. The bandwidths for the filters were of 2KHz, 500Hz, and of 250Hz. This means that in the WSN each mote will have a different filter bandwidth to implement an analog distributed filter bank. The full frequency range for the 2kHz filter is 0 to 16KHz, the range for 500 Hz is from 0 to 8Khz, and for 250Hz 1 is from 0 to 2 KHz.

Our acoustic classification system requires the design of a customized mote with an analog front-end that allows the acquisition of an acoustic signal and the extraction of features in the analog domain. Hence, each mote has the functionality of one channel of the DWPT. A depiction of the front end is shown if Figure 4.1; we note that this

front end is fully realizable with analog components. Further work on an actual hardware implementation will follow in the future. The analog part for the  $i$ th mote (or channel) starts capturing the event  $x(t)$  using a microphone and signal conditioning sub-system. The conditioned signal is then processed with a analog filter such that the subband signal  $y_i(t)$  can be expressed as

$$y_i(t) = x(t) * h_i(t) = \int x(\tau)h(t - \tau)d\tau.$$

Next, we need to estimate the energy of  $y_i(t)$  using a squaring circuit and an integrator such that

$$e_i(t) = \int_{t-1}^t y_i^2(\tau)d\tau.$$

The signal  $e_i(t)$  represents the energy of  $y_i(t)$  over the previous second. We use signal energy as our discriminating features as in Chapter 3. In this case, we limit the integration period to one second in order to account for the non-stationarity of acoustic signals (i.e., the signal statistics change over time). Finally, we can obtain features by sampling  $e_i(t)$  at regular sampling intervals of one second, such that  $e_i[n] = e_i(nT)$  with  $T = 1$  second.

Hence, we have virtually eliminated any high sampling requirements of the system by allowing most of the feature extraction process to occur in analog domain. Once the signal is digitized, the feature values can be (digitally) transmitted to a classifier mote (or base station) where the classifier is implemented in software (or digital hardware). We assume the classifier has been previously trained or has an in-line training scheme that allows learning as new data arrives.

### 4.3 Description of Mixed-Signal Classifier Simulation

In this thesis, we have implemented a simulation of the proposed distributed algorithm. An implementation over a real network is left for future work. An initial attempt was made to simulate the system using the SimuLink tool from MATLAB. We found that there are

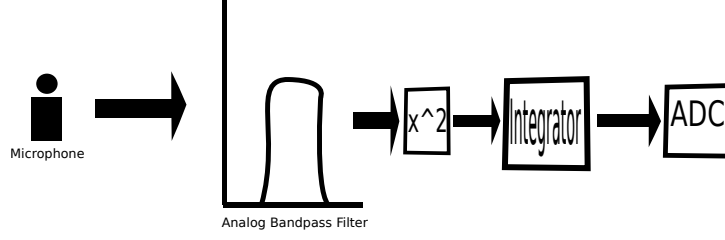


Figure 4.2: Architecture for a mote analog front-end for acoustic signal processing.

severe computation limitations to simulate analog filtering operations in SimuLink which make it impractical with current computer resources.

A SimuLink simulation focusses on interpolating a discrete signal (oversampling followed by filtering) to generate a continuous-like signal that can be processed with the analog block set from the software. In our case, we note that the analog feature acquisition sub-system for the mote (described in the previous section) has a very low sampling rate of one sample per second. Hence, given that the signal set was acquired at 44100 samples per second (see Table 3.1), for practical purposes we can use them in lieu of analog signals and implement our simulation of analog processing with a highly oversampled digital signal.

We implemented the simulation in MATLAB with the following components. First, we used the filter design tool to design eight Butterworth filters. The filters have the same bandwidth and replicate the spectral partitioning of a three-level wavelet packet transform. The filters were designed using a passband ripple of 1dB, and a stopband attenuations of 60dB (left) and 80dB (right). The filter orders were automatically determined by the design tool. Three filter banks were designed with individual filter bandwidths of 2000 Hz, 500Hz, and 250 Hz. Figure 4.3 shows the spectrum partition from 0 to 16 KHz using the 2000 Hz filters. The squaring element was simple to implement with basic MATLAB operations. Finally, the integration operation was performed as a sliding window using the `trapz()` function. This function uses the trapezoidal method to approximate integrals numerically. Finally, the output of the integration step is sampled at 1Hz to produce the subband features which are grouped with the other subbands to compose a feature vector

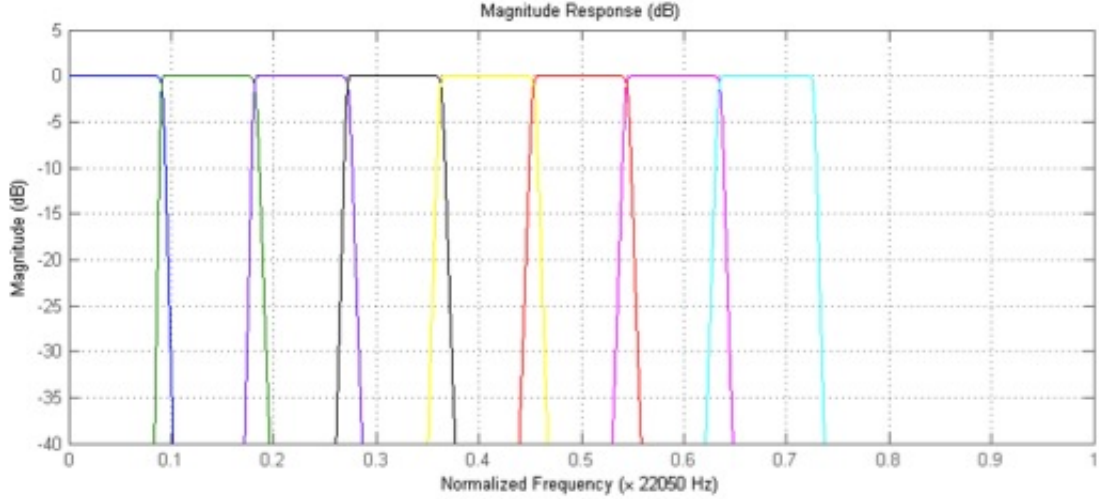


Figure 4.3: Magnitude responses for the analog filters comprising a distributed filter bank . The total frequency coverage is 16 KHz where each filter has a bandwidth of 2000 Hz. Filter design is described in the text.

for each second of audio data.

## 4.4 Results

We performed similar simulations to those in Section 3.2. A sample set of 1189 feature vectors was produced which was used to train the three classifiers described before. The classifiers were trained using the hold out method with a 30/70 and 50/50 split.

The results of the mixed-signal system described in the previous section are presented in Tables 4.1 and 4.2. If we look at the results obtained in section 3.5 The results are similar. There are three rows on each table, each of these rows shows the correct classification rates obtained for each classifier under specific filter bandwidths. The best results were obtained using the RBF-SVM and using 250 Hz filters with 75.7% correct classification. The overall bandwidth for this system is 2000 Hz, implying that there is significant less information retained by this system when compared to the others. This implies that this frequency

Table 4.1: Correct classification results for the mixed-signal distributed system using 30% of data for training.

	RBF SVM	BP NN	k-NN
2KHz BW	63.7	52.8	58.8
500Hz BW	49	65.2	61.7
250Hz BW	62.7	55.3	60.1

Table 4.2: Correct classification results for the mixed-signal distributed system using 50% of data for training.

Percentage Error	RBF SVM	BP NN	k-NN
2KHz BW	66.2	64.2	64.9
500Hz BW	72.6	70	72.1
250Hz BW	75.7	73.2	75.1

range provides enough discriminative information to perform classification tasks at better or similar levels. The advantage of having filters with smaller bandwidth is that they can be implemented at a lower cost in hardware.

## 4.5 Conclusion

In this chapter we have presented a distributed/collaborative scheme for classification of acoustic signals over WSNs. We took a holistic approach where we consider the memory, storage and power constraints of a WSN to distribute the feature extraction task across many nodes. We used a classification scheme based on filter banks as these adapt naturally to distributed processing. Each node extracts a pre-determined frequency subband and computes an energy estimate that is used as a discriminative feature for the signal under study. We used a novel technique to extract features by keeping all the processing in the analog domain up to the point that an estimate of the feature needs to be digitized. This

scheme reduces dramatically the sampling requirements (i.e., analog-to-digital conversion hardware), memory storage and transmission data rate. The system is collaborative in the sense that all nodes transmit the computed features to a concentrator node in order to form a feature vector which can be used as the input to a statistical pattern classifier. Simulations show that the proposed scheme has a similar classification performance compared to the all-digital classifier presented on Chapter 3.

# Chapter 5

## Conclusion and Future Work

### 5.1 Conclusions

The contributions of this thesis can be framed within the context of *machine hearing* which is a nascent area of research [13]. Our objective was to be able to classify five different types of acoustic signals (birds, explosions, car, conversation, and steps) using statistical pattern classifiers. The specific contributions of this method are:

- The acoustic classification problem was posed within the collaborative/distributed signal processing paradigm which has been one of the promises of WSNs.
- We adopted a distributed filter bank (i.e., wavelet) approach for the feature extraction stage. Each channel from the filter bank is assigned to an individual node where a signal feature is computed as the subband energy. The individual features are collected by a concentrator node that performs classification or relies the information to a base station.
- An important component of this work is the introduction of a mixed-signal feature extraction scheme, where motes use an analog front end consisting of an analog filter and a square-integrator device that computes the energy subband. This approach has the important benefit of significantly reducing the power, computing, storage and transmission requirements for the classification task.
- We demonstrated classification of acoustic signals using three classifiers: the k-NN, the back-propagation neural network and the radial-basis function support vector

machine (RBF-SVM). We compared the proposed approach with an all-digital baseline system, and observed similar performance for both systems. In particular, the RBF-SVM provided slighter better performance than the other two classifiers.

## 5.2 Future Work

The developments presented in this thesis has been at the theory and simulation level. Moreover, it has focussed on the classification task on an isolated manner. The final objective of this research line would be to achieve integration and deployment on a real WSN platform. The following list of research and development topics are envisioned to achieve this goal:

- Integrate the system to a complete simulation suite that allows the simulation of communications, networking, and acoustic events simultaneously for further algorithm validation.
- Design, develop, build the analog frontend for acoustic signal acquisition described in this thesis. Simultaneously integrate the frontend to commercial-off-the-shelf wireless motes.
- Research and develop the algorithms and protocols for filter bank self organization and collaborative classification.
- Deploy and test the proposed classification system over a real WSN scenario.

Another important aspect for future research is the creation of a database of acoustic events that is representative and the universe of interest for the context under which the system will be used. The field of *computer hearing* is on its infancy and there is a lot of research to be done (beyond the scope of this thesis) on signal processing, feature extraction and machine learning directed to acoustic signals. Finally, as discussed in Chapter 4, the proposed distributed classifier will require support and interaction with other WSN services.

These include geo-localization, data routing, sound detection and delay compensation, and signal enhancement. The required level of system integration to merge all these operations in an open problem in WSNs that opens many lines of research to explore in the future.

# References

- [1] A. M. Ali, K. Yao, T. C. Collier, "An Empirical Study of Collaborative Acoustic Source Localization," *IPSN '07*, April 25-27, 2007. Cambridge, Massachusetts.
- [2] W. Bingchen, O. Sigeru, and A. Toshiro, "Identification of the Defective Transmission Devices Using the Wavelet Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June 2005, Vol. 27 Issue 6. pp. 919-928.
- [3] C. Bishop, "Pattern Recognition and Machine Learning," 2006. Berlin, Springer.
- [4] R. O. Duda, P. E. Hart, and D. G. Stork. (2001) *Pattern Classification*. Wiley.
- [5] R.P.W. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D.M.J. Tax, S. Verzaikov PRTTools4.1, A Matlab Toolbox for Pattern Recognition, Delft University of Technology, 2007.
- [6] R. K. Ganti, P. Jayachandran, T. F. Abdelzaher, and J. A. Stankovic. "Satire: a software architecture for smart attire," *In MobiSys* , 2006.
- [7] L. Girod, V. Bychkovskiy, J. Elson, and D. Estrin, "Locating Tiny Sensors in Time and Space: A Case Study," *IEEE International Conference on Computer Design*, 2002.
- [8] L. Girod, D. Estrin, "Robust Range Estimation Using Acoustic and Multimodal Sensing," *Intelligent Robots and Systems*, 2001. Oct. 29, 2001. Maui, Hawaii.
- [9] L. Gu, D. Jia, P. Vicaire, "Lightweight Detection and Classification for Wireless Sensor Networks in Realistic Environments," *SenSys' 05*, Nov. 2-4, 2005, SAndiego, California pp. 205 -217.
- [10] A. Klausner, A. Teng, and B. Rinner, "Vehicle Classification on Multi-Sensor Smart

- Cameras Using Feature and Decision-Fusion,” *Distributed Smart Cameras*, Sept. 2007 pp. 67-74.
- [11] M. J. Larkin, ”Sensor Fusion and Classification of Acoustic Signals Using Bayesian Networks,” *IEEE Explorer*, 1998. Nov. 4, 1998. pp. 1359-1362 Pacific Grove, California.
  - [12] L. Luo, ”Enviromic:Towards Cooperative Storage and Retrieval in Audio Sensor Networks,” *In proceedings of the 27th International Conference on Distributed Computing Systems*, Washington, DC, 2007, pp. 34.
  - [13] R. F. Lyon. ”Machine Hearing: An Emerging Field,” *IEEE Signal Processing Mag.*, September 2010 pp. 131-139.
  - [14] S. Phadke, R. Limaye, S. Verma, and K. Subramanian, On design and implementation of an embedded automatic speech recognition system, in *In proceedings of the 17th International conference on VLSI design*, 2004, pp. 127132.
  - [15] Philo Juang, Hidekazu Oki, YOUNG Wang, Margaret Martonosi, Lishiuan, Peh, and Daniel Rubenstein, ”Energy-Efficient Computing for Wildlife Tracking: Design Trade-offs and Early Experiences with ZebraNet. In Proceedings of the Tenth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-X). October, 2002. San Jose, CA.
  - [16] V. C. Ravindra, Y. Bar-Shalom, and T. Damarla, ”Feature-Aided Localization of ground Vehicles Using Passive Acoustic Sensor Arrays” *12th International Conference on Information Fusion*, july 6-9, 2009.
  - [17] J. G. Rosiles, M. J. T. Smith, ”Texture Classification with a Biorthogonal Directional Filter Bank”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2001. Volume 3 pp 1549-1552.
  - [18] G. Simon, A. Ledezdzi, and M. Maroti, Sensor network-based countersniper system, itSenSys, Nov. 2004.

- [19] V. Trifa, L. Girod, T. Collier, D. Blumstein, and C. Taylor. "Automated wildlife monitoring using self- configuring sensor networks deployed in natural habitats", *Proceedings of the International Symposium on Artificial Life and Robotics*, AROB 12th 2007 (Beppu, Japan).
- [20] V. M. Trifa, A. N. G. Kirschel, and C. E. Taylor, "Automated Species Recognition of Antbirds in a Mexican Rainforest Using Hidden Markov Models," *J. Acoust. Soc. Am.*, 2008. Volume 123 Issue 4 pp. 2424-2431.
- [21] Tinyos Hardware. 2006. TinyOS Alliance. 15 May 2009  
[www.tinyos.net/hardware/hardware.html](http://www.tinyos.net/hardware/hardware.html)
- [22] Tzanetakis, George, Essl, Georg, and Perry Cook "Audio Analysis using the Discrete Wavelet Transform " In. Proc. WSES Int. Conf. Acoustics and Music: Theory and Applications (AMTA 2001) Skiathos, Greece, 2001
- [23] A. Ukil, "Intelligent Systems and Signal Processing in Power Engineering" Springer-Verlag 2007.
- [24] M. Unser, "Texture Classification and Segmentation Using Wavelet Frames" *IEEE Transactions on Image Processing*, Nov. 1995. Volume 4 Issue 11 pp. 1549-1560.
- [25] W. Bingchen, O. Sigeru, and A. Toshiro. "Identification of the Defective Transmission Devices Using the Wavelet Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27 No.6 pp. 919-928.

# Curriculum Vitae

Humberto Santacruz was born in El Paso, Texas. The second son of Jesus Enrique Santacruz and Eustacia Santacruz Casas, he graduated from Mountain View High School, El Paso Texas, in the spring of 2004. In the fall he entered the University of Texas at El Paso. He obtained his Bachelor's Degree in Electrical Engineering and was involved in different organizations including the Institute of Electrical and Electronics Engineers (IEEE), Tau Beta Pi, and Eta Kappa Nu. In fall 2008 he was accepted to graduate school pursuing a Masters degree in Electrical Engineering at The University of Texas at El Paso with concentration in the Telecom/DSP area under the guidance of Dr. Gerardo Rosiles. In the summer of 2011 Humberto will begin working for Sandia National Laboratories.

Permanent address: 3799 Judy Marie

El Paso, Texas 79938

This thesis was typed by the author.