ENERGY-EFFICIENT DETECTION SYSTEM IN TIME-VARYING SIGNAL AND NOISE POWER

BY

LONG NGUYEN THANG LE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Professor Douglas L. Jones

ABSTRACT

In many detection applications with battery-powered or energy-harvesting sensors, energy constraints preclude the use of the optimal detector all the time. Optimal energy-performance trade-off is therefore needed in such situations.

In many signal processing applications, the signal and noise power may vary greatly over time, which can be exploited to constrain energy consumption while maintaining the best possible performance.

A detector scheduling algorithm based on the signal and noise power information is developed in this thesis. The resulting algorithm is simple due to its threshold-test structure and can be easily implemented with almost no overhead. A detection system with two detectors using the proposed scheduling scheme is estimated to greatly reduce the energy consumption for a wildlife monitoring application. Hardware implementation also consolidates the empirical evidence for the effectiveness of the proposed method.

To my parents and sweetheart, for their love and support

ACKNOWLEDGMENTS

First and foremost, I would like to express my deep gratitude toward my adviser, Prof. Douglas L. Jones, whose research passion and knowledge continue to be a great inspiration for me. His advising style and deep insights about a research career have helped me to realize the crucial differences between what makes a true scholar and what does not, leading me to become more and more independent in my own early research career.

I could not have realized all of that without the help of my colleague David M. Jun. I am deeply indebted to his patient and helpful responses to my myriad, sometimes silly, inquiries. In addition to that, his eagerness for solving open problems keeps seducing me into the world of the unknown with high excitement.

I would also like to thank my beloved wife, for sympathizing with the road I chose, for understanding all the time that the pursuit of knowledge took me from her, and for giving me valuable life advice whenever I lost my direction.

Finally, I would like to thank my parents for their constant support and love from home. I love you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GCW	Golden-Cheeked Warbler
SNR	Signal-to-Noise Ratio
MAC	Multiply Accumulate
LRT	Likelihood Ratio Test
ML	Maximum Likelihood
NPL	Neyman-Pearson Lemma
AWGN	Additive White Gaussian Noise
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
VAD	Voice Activity Detector
WSS	Wide-Sense Stationary
MCU	Microcontroller Unit
ADC	Analog to Digital Converter
RTC	Real Time Counter
ACMP	Analog Comparator
LWDF	Lattice Wave Digital Filter

CHAPTER 1

INTRODUCTION

"Everything should be made as simple as possible, but no simpler."
– Albert Einstein

In many engineering systems, detection is usually the first operation that needs to be carried out. For energy-constrained systems, the ability to perform energy-efficient detection is therefore very crucial. In particular, a wildlife monitor application is considered to motivate the study of a detection system with an energy constraint.

Specifically, bird scientists are interested in studying (1) the population and (2) the evolution in the call spectrum of an endangered bird species called the Golden-Cheeked Warbler (GCW) [1]. In order to study these birds, vocalization data need to be collected by a monitoring system. The naive approach for monitoring is to have a high sampling rate audio recording system deployed in the field that runs continuously for days before the system's storage depletes. An estimate for a system with 16 kHz sampling at two bytes resolution and that records for three months would require a storage with roughly 230 GB. Such systems are very ineffective, in the sense that they record a large amount of uninteresting data which will eventually be discarded. An alternative approach is to process before writing the data into the storage device. The processing can be done by introducing a microcontroller that can detect, attach a time stamp, and then capture only the calls into memory. According to the preliminary data given to us by our bird scientist collaborators, the GCWs are likely to be present only 10% of the time. When they are present, their calls usually last 1.5 seconds, with the time between calls around 10 seconds. For a period of three months, this means only a storage of 2.5 GB of data is needed to record all bird calls. This effectively condenses 230 GB of data into 2.5 GB of valuable information for bird scientists.

However, there are two problems with this approach. First, these devices are battery powered and have limited energy available. As most birds are seasonal breeders, these battery-powered devices should be able to last at least the entire season without manual battery replacement. The second issue that requires attention is the detection performance, because detection is now a task in the new system. The purpose of this thesis therefore is to investigate and derive a solution for the above problems. The resulting answer turns out to be a detector scheduling algorithm that admits a simple threshold-test structure.

This thesis is organized as follows. Chapter 2 reviews classical detection theory and introduces standard notations that will be useful in later chapters. Chapter 3 discusses major prior work that also looks at the similar problem of detection systems with an energy constraint. The actual formulation and solution is presented in Chapter 4. Then the algorithms in Chapter 5 complement the theory in Chapter 4. Finally, Chapters 6 and 7 give empirical evidences for the effectiveness of the detector scheduling algorithm by simulation and actual hardware implementation, respectively.

CHAPTER 2

BACKGROUND

"Detection is, or ought to be, an exact science, and should be treated in the same cold and unemotional manner."

– Sherlock Holmes, in *The Sign of Four*

Detection theory is briefly reviewed to introduce standard notations, according to [2], that will be useful in later sections. For a more thorough treatment of detection theory, the readers are recommended to see [2].

2.1 Detection Theory Fundamentals

The basic problem of binary hypothesis testing is first introduced. The goal is to decide between two hypotheses H_0 and H_1 based on the observation of a random vector $\mathbf{Y}$. Assume the probability densities, f, on $\mathbf{Y}$ are

$$\begin{aligned} H_0 :& \mathbf{Y} \sim f(\mathbf{y}|H_0) \\ H_1 :& \mathbf{Y} \sim f(\mathbf{y}|H_1) \end{aligned} \tag{2.1}$$

Put in the context of the Golden-Cheeked Warbler (GCW) call detection application, H_0 corresponds to no call event and H_1 corresponds to a GCW call event. In addition, the observation vector $\mathbf{Y}$ is the block of sampled acoustic signals picked up by a microphone.

Once $\mathbf{Y}$ is given, a decision whether H_0 or H_1 is true needs to be made. This is done by the decision function $\delta(\mathbf{Y}) \in \{0, 1\}$ that maps $\mathbf{Y}$ to 1 when H_1 is true and to 0 when H_0 is true. In other words, the observation domain $\mathcal{Y}$ can be segmented into disjoint sets $\mathcal{Y}_0$ and $\mathcal{Y}_1$. Such a decision function is chosen according to certain optimization formulations. There exist two popular formulations which are discussed in the next section: the Bayesian formulation and the Neyman-Pearson formulation.

Table 2.1: Important conditional probabilities in detection theory.

Names	Expressions
Detection	$P[\mathcal{Y}_1\|H_1]$
False Alarm	$P[\mathcal{Y}_1\|H_0]$
Miss	$P[\mathcal{Y}_0\|H_1]$
Correct Rejection	$P[\mathcal{Y}_0\|H_0]$

2.2 Optimal Detection Formulations

2.2.1 Bayesian Formulation

The Bayesian formulation assumes knowledge of priors probability $\pi_0 = P[H_0]$, $\pi_1 = P[H_1]$ and cost function C_{ij} with $0 \leq i, j \leq 1$ deciding H_i when H_j is true. The conditional Bayes risk (BR) under hypothesis H_j is then defined as

$$R(\delta|H_j) = \sum_{i=0}^{1} C_{ij} P[\mathcal{Y}_i|H_j]$$

where

$$P[\mathcal{Y}_i|H_j] = \int_{\mathcal{Y}_i} f(\mathbf{y}|H_j) d\mathbf{y}$$

is the associated conditional probability. Table 2.1 enumerates these terms with their names.

The BR for the decision function δ is given by

$$R(\delta) = \sum_{j=0}^{1} R(\delta|H_j)\pi_j$$

In BR formulation the above quantity is then minimized to find the optimal decision function. It is well known [2] that the solution to BR formulation is a likelihood ratio test of the following form

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & \text{if } T(\mathbf{y}) \geq \tau \\ 0 & \text{if } T(\mathbf{y}) < \tau \end{cases}$$

where

$$T(\mathbf{y}) = \frac{f(\mathbf{y}|H_1)}{f(\mathbf{y}|H_0)} \tag{2.2}$$

is the test statistic and

$$\tau = \frac{(C_{10} - C_{00})\pi_0}{(C_{01} - C_{11})\pi_1} \tag{2.3}$$

is the threshold. Both the test statistic and the threshold take on scalar, real values.

2.2.2 Neyman-Pearson Formulation

An alternative formulation to BR is the Neyman-Pearson (NP) formulation. Unlike BR, NP does not assume knowledge of prior and cost function. Instead, its objective is to maximize the probability of detection $P_D(\delta) = P[\mathcal{Y}_1|H_1]$ while satisfying some probability of false alarm $P_F(\delta) = P[\mathcal{Y}_1|H_0]$ constraint γ.

$$\begin{aligned} \max_{\delta} \quad & P_D(\delta) \\ s.t. \quad & P_F(\delta) \leq \gamma \end{aligned}$$

The solution to the above constrained optimization is given by the Neyman-Pearson lemma [3]

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & \text{if } T(\mathbf{y}) > \tau \\ 1 \text{ or } 0 & \text{if } T(\mathbf{y}) = \tau \\ 0 & \text{if } T(\mathbf{y}) < \tau \end{cases} \tag{2.4}$$

where τ is selected according to the false alarm constraint γ, i.e. $P_F(\tau) = \gamma$.

2.2.3 Alternative Formulations

When the probabilistic criteria discussed in Sections 2.2.1 and 2.2.2 are not tractable, alternative metrics must be used to quantify performance. One such metric is the *deflection* or generalized signal-to-noise ratio (SNR), which is defined as the square of the difference in the mean of $T(\mathbf{Y})$ under H_1 and H_0 over the variance of $T(\mathbf{Y})$ under H_0, i.e.

$$\frac{(\mathrm{E}_{H_1}[T(\mathbf{Y})] - \mathrm{E}_{H_0}[T(\mathbf{Y})])^2}{\mathrm{Var}_{H_0}(T(\mathbf{Y}))} \tag{2.5}$$

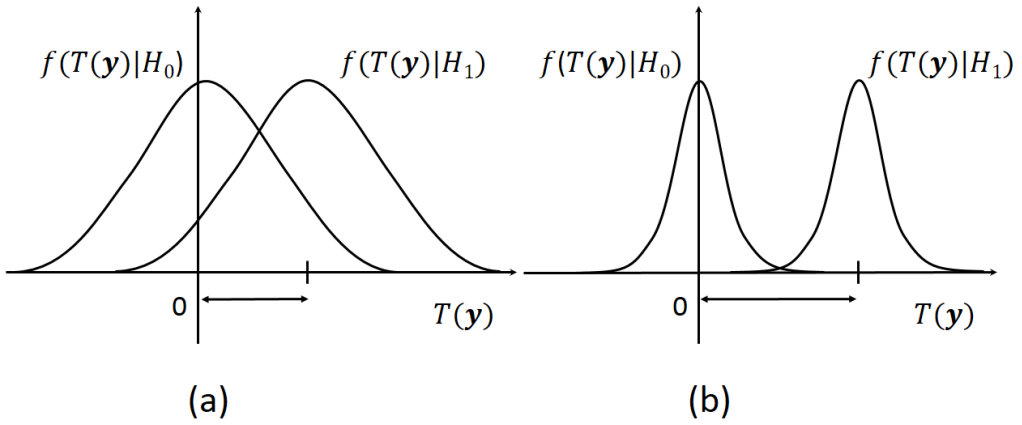

Figure 2.1: Distribution of test statistic under H_0 and H_1 for detection of constant signal in AWGN. (a) has low detection performance while (b) has high detection performance.

Another metric is the divergence between H_1 and H_0, which is defined as the difference in the mean of $T(\mathbf{Y})$ under H_1 and H_0, i.e.

$$\mathrm{E}_{H_1}[T(\mathbf{Y})] - \mathrm{E}_{H_0}[T(\mathbf{Y})] \tag{2.6}$$

The rationale behind these criteria can be understood by means of a simple detection example. Consider the detection of a constant signal in AWGN with unit variance. It is well known [2] that the test statistic $T(\mathbf{Y})$ in this case is Gaussian with non-zero mean under H_1 and zero mean under H_0. It can also be shown analytically [2] or inferred visually from Figure 2.1 that the detection performance will increase as the distance between the means increases and/or the variances of the two hypotheses decrease. This fact gives the motivation for maximizing (2.5) and (2.6) as the means to maximize detection performance in the case where probabilistic measures are intractable [4, 5].

2.3 Detection with Unknown Parameters

In practice, it is often the case that certain parameters $\mathbf{X}$ of the detection model in (2.1) are not known. Common examples include signal amplitude, phase, frequency, time delay, and noise variance [2]. The model that captures this is given as

$$\begin{aligned} H_0 :&\mathbf{Y} \sim f(\mathbf{y}|\mathbf{x}, H_0), \quad \mathbf{x} \in \mathcal{X}_0 \\ H_1 :&\mathbf{Y} \sim f(\mathbf{y}|\mathbf{x}, H_1), \quad \mathbf{x} \in \mathcal{X}_1 \end{aligned} \tag{2.7}$$

The unknown parameters can be viewed either as random, i.e. the Bayesian viewpoint, or nonrandom, i.e. frequentist viewpoint. These two viewpoints lead to two different approaches to address the *composite hypothesis testing* [2] problem in (2.7).

2.3.1 Bayesian Approach

The Bayesian approach assumes prior knowledge of unknown parameters' probability distributions under each hypothesis. This simplifies (2.7) back to (2.1) by marginalization of the unknown parameters. However this falls short in practice as prior knowledge of the distribution of the unknowns is not always available.

2.3.2 GLRT Approach

Another approach where prior statistical knowledge of the unknown parameters is not required is the generalized likelihood ratio test (GLRT). This approach substitutes the unknown parameters in (2.7) with their *maximum likelihood* (ML) estimates, effectively reducing the problem to (2.1). It is worth noting that although being a heuristic-based approach, GLRT possesses the *invariance property* that lends itself to widespread use in signal processing applications [6].

2.4 Gaussian Detection Problem

The classical problem of detecting a zero mean Gaussian signal in additive white Gaussian noise (AWGN), where the signal and noise are assumed to

be independent, is discussed. Two cases are considered, depending on the correlation of the signal to be detected.

2.4.1 Uncorrelated Gaussian Signal

The easier case in which the signal is uncorrelated is first considered. For a Gaussian signal this is also equivalent to having independent samples. Hence the detection problem is given by

$$\begin{aligned} H_0 : \mathbf{Y} &\sim \frac{1}{(\sqrt{2\pi\sigma_0^2})^N} e^{-(\sum_{i=1}^N y_i^2)/2\sigma_0^2} \\ H_1 : \mathbf{Y} &\sim \frac{1}{(\sqrt{2\pi\sigma_1^2})^N} e^{-(\sum_{i=1}^N y_i^2)/2\sigma_1^2} \end{aligned} \tag{2.8}$$

where N is the size of observation vector $\mathbf{Y}$, σ_0^2 is the noise variance, and $\sigma_1^2 = \sigma_0^2 + \sigma_s^2$, with σ_s^2 being the signal variance.

The optimal detector in this case is given by

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & \text{if } \frac{\sum_{i=1}^N y_i^2}{N} \geq 2\frac{\sigma_0^2\sigma_1^2}{\sigma_s^2}\left[\log(\frac{\sigma_1}{\sigma_0}) + \frac{\log(\tau)}{N}\right] \\ 0 & \text{if } \frac{\sum_{i=1}^N y_i^2}{N} < 2\frac{\sigma_0^2\sigma_1^2}{\sigma_s^2}\left[\log(\frac{\sigma_1}{\sigma_0}) + \frac{\log(\tau)}{N}\right] \end{cases} \tag{2.9}$$

Since this detector computes the squared magnitude of the observation vector, it is commonly called the *energy detector* in the literature.

If the variances σ_0^2 and σ_1^2 of the model are unknown, the GLRT approach discussed earlier will substitute them with their ML estimates $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$, where $\hat{\sigma}_0^2 + \hat{\sigma}_1^2 = \frac{\sum_{i=1}^N y_i^2}{N}$, into (2.9).

2.4.2 Correlated Gaussian Signal

Next consider the case when the signal is correlated.

$$\begin{aligned} H_0 : \mathbf{Y} &\sim \frac{1}{(\sqrt{2\pi})^N |\Sigma_0|^{-1/2}} e^{-\frac{\mathbf{y}^T \Sigma_0^{-1} \mathbf{y}}{2}} \\ H_1 : \mathbf{Y} &\sim \frac{1}{(\sqrt{2\pi})^N |\Sigma_1|^{-1/2}} e^{-\frac{\mathbf{y}^T \Sigma_1^{-1} \mathbf{y}}{2}} \end{aligned} \tag{2.10}$$

where $\Sigma_0 = \sigma_0^2 \bar{\Sigma}_0$ is the noise covariance matrix and $\Sigma_1 = \Sigma_0 + \Sigma_s$, with $\Sigma_s = \sigma_s^2 \bar{\Sigma}_s$ being the signal covariance matrix. $\bar{\Sigma}_0$ and $\bar{\Sigma}_s$ denote covariance

matrices with ones on the diagonal.

The optimal detector in this case is given by

$$\delta^*(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y}^T(\Sigma_0^{-1}\Sigma_s\Sigma_1^{-1})\mathbf{y} \geq 2\log(\tau) + \log(|\Sigma_1|) - \log(|\Sigma_0|) \\ 0 & \text{if } \mathbf{y}^T(\Sigma_0^{-1}\Sigma_s\Sigma_1^{-1})\mathbf{y} < 2\log(\tau) + \log(|\Sigma_1|) - \log(|\Sigma_0|) \end{cases} \tag{2.11}$$

where the expression $\Sigma_0^{-1}\Sigma_s\Sigma_1^{-1}$ comes from algebraic manipulation of $\Sigma_0^{-1} - \Sigma_1^{-1}$. Since this detector admits a quadratic form, it is commonly referred to as the *quadratic detector* in the literature [7]. Another way to interpret (2.11) is that the detector has the form of an estimator-correlator [2].

In this case, if the variances σ_0^2 and σ_1^2 of the model are unknown, ML estimation cannot give a closed form expression for $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$. Instead, the expectation-maximization (EM) recursive procedure for finding ML estimates can be employed [2].

This chapter concludes with the remark that traditional detection theory does not take into account computational complexity. This ignorance becomes a problem for detection systems with an energy constraint, because a more complicated detector requires more energy to operate.

CHAPTER 3

PRIOR WORK IN ENERGY CONSTRAINED DETECTION

"If I have seen further than others, it is by standing upon the shoulders of giants."

– Isaac Newton

3.1 Classical Detection with Computational-Complexity Reduction

The interest in reducing the complexity of a detector's structure, and hence energy consumption, can be traced back to [4]. Poor [4] considers the classical detection problem of a Gaussian signal in additive white Gaussian noise (AWGN) with block length N observations. It is well known [4] that if the signal has no correlation, then the optimal detector structure is the simple quadratic (energy) detector. Asymptotic complexity analysis reveals that it only costs $O(n)$ operations. If the signal has correlation, then the optimal detector structure has a quadratic form but requires $O(n^2)$ operations. A reduced-complexity structure for detection is proposed to achieve a compromise between the two structures, namely, capturing the correlation in the signal at the expense of roughly $O(n)$ operations. The proposed quadratic form has the Toeplitz structure imposed on it, hence the name *banded quadratic form* [4], so that it can be implemented by a simple linear filter. The contribution of [4] is to derive the best filter coefficients so that the banded quadratic form achieves the optimal asymptotic deflection.

An alternative approach to simplify the quadratic form in Gaussian detection is studied by Scharf [5]. In [5], a rank reduction technique is used to reduce the rank of the matrix in the quadratic form instead of imposing Toeplitz structure [4]. The criterion for the reduced rank quadratic form

is that it maximizes the divergence between H_1 and H_0. The contribution in [5] is the determination of which eigenvalues should be removed to achieve a reduced rank matrix of the quadratic form.

Sayeed and Jones [8] applied the reduced-rank technique on a more general detection scenario. [8] considered the generic Gaussian detection problem with time and frequency shifts as unknown parameters. Since these parameters do not contribute to the determination of the true hypothesis, they are often called *nuisance parameters* [2]. In [9], the GLRT approach is used to substitute these nuisance parameters with their respective ML estimates, yielding the ML-based GLRT detector. It is also argued in [9] that such an ML-based GLRT detector can be implemented via a bank of spectrograms, which are the squared magnitudes of the short-time Fourier transforms (STFT). The STFT implementation is significant because the STFT is the simplest time-frequency representation (TFR) [8] that has long been widely used in signal processing [10]. However, depending on the complexity of the random signal to be detected, the number of spectrograms in the bank can be large. The contribution in [8] is the derivation of the best, in maximum deflection sense, reduced-size bank of spectrograms.

The idea of incremental refinement is used in [11] to systematically trade off between computational complexity and detection performance. [11] studies the detection problem of a complex exponential with unknown frequency and phase in complex AWGN. It is pointed out that the GLRT detector with ML estimation consists of a bank of correlators and envelope detectors followed by a comparator and a slicer. It is also stated that the bank of correlators is identical to the DFT operation, which can be efficiently implemented by the FFT algorithm, hence the name *FFT detector* [11]. Because FFT is computed in stages, early stage termination can reduce computational complexity. The contribution of [11] is the analysis of detection performance at each stage, assuming radix-2 decimation either in-time or in-frequency. Such information is valuable to operators who are interested in determining when to terminate early, once a certain performance criterion is met.

The previous approaches are optimized for generic situations and do not take into account the existing structures of the problem. For example, in wildlife monitoring, bird call events are rare (roughly 1% of the total time, as discussed in Chapter 1) compared to the non-bird call events. Therefore, it seems that even more aggressive energy saving can be achieved if this rare

event structure can be exploited.

3.2 Detection with Rare Event: Cascade Approach

One of the seminal works in image-based object detection application [12] serves as an epitome of such an approach. [12] studies the face detection problem. Since a face event in an image can be considered as a rare event, a cascade architecture is used to focus the attention and computational power of a face detector to more potential sub-windows while quickly rejecting the others in an image. The cascade architecture arranges simple classifiers with small numbers of features to serve in the early stages, the later in the stage, the more features classifiers have. If the outcome of a classifier in the cascade is H_1, then the next, more intensive classifier in the chain is triggered; if not, the cascade terminates early with an H_0 decision. Only when the last classifier triggers does the system reach an H_1 decision. Each stage in the cascade is trained to select features using a greedy algorithm called AdaBoost [12]. The number of features selected in each stage depends on the desired false alarm and missed rate at each stage. The number of stages in the cascade depends on the final desired false alarm and missed rate. The result is a high-detection-rate face detector that was 15 times faster than any existing system at the time. [12] also enables an extremely rich amount of research effort to further perfect the cascade architecture.

Inspired by [12], Jun [13] studies a system with multiple detectors in cascade. Again, this strategy can achieve energy efficiency if H_1 is rare because the more energy-intensive detectors are only activated by simpler detectors for a small fraction of time. Energy consideration is explicitly taken into account by introducing the energy constraint into the classical BR and NP criteria, hence formulating new optimization problems, the energy-aware Bayes risk (EABR) and energy-aware Neyman Pearson (EANP) [13], respectively. This constraint allows for a systematic and energy efficient way to set the false alarm and missed rates in the cascade, which was missing in [12]. We note that this optimization, if carried out over all the detectors' decision functions δ, is too difficult; hence, the problem is relaxed to the optimization over the detectors' thresholds τ with fixed test statistics $T(\mathbf{Y})$. Such relaxation does not make the problem trivial since without proper threshold

management, the energy saving characteristic of the cascade will degrade badly. For example, with a system of two detectors, if the first detector's threshold is so low that it triggers the second detector all the time, then the energy saving of the cascade is lost. A case study for an optimized cascade of two detectors, energy and FFT detector [11], is shown to outperform the FFT detector with incremental refinement described in Section 3.1.

As mentioned in the previous paragraph, the cascade architecture can be used as a way to save energy in detection systems with an energy constraint. However, it is determined in this thesis that there are scenarios where the use of the cascade is not very energy-efficient. To illustrate this point, consider again the wildlife monitoring application. A detection system operating over the course of time will obviously experience time-varying signal and noise power. If the target bird is nearby and background noise is low, the decision about H_1 made by the simple energy detector could be just as accurate as any sophisticated detector. However, if the cascade is used, such sophisticated detectors still need to be triggered before the system decision can be made. This is where energy inefficiency manifests itself. A similar story goes for the case when the target bird is far away and background noise is high; then running the sophisticated detector is inevitable for good performance as a simple energy detector might falsely reject the weak signal. Thus directly skipping the simple detector might bring about additional performance and energy saving. This observation suggests that exploiting the time-varying signal and noise power in the problem might be beneficial in saving energy, and one of the related areas that studies detection in time-varying environment is the design of voice activity detector (VAD).

3.3 Detection in the Time-Varying Environment

VAD design is a research area in speech processing that focuses on detecting the existence of noisy speech between silent periods. It has important application in digital cellular systems due to the fact that a party is typically only active 35% of the time in a conversation [14], i.e. the voice activity factor is 0.35. Hence, bandwidth can be saved if transmission of the encoded speech is triggered only when speech is actually present. In this detection problem, the time-varying nature of the environmental noise was subjected

to extensive investigation, because real-world noise exhibits both stationary and non-stationary characteristics, and the worst-case SNR can even be less than 0 dB [15].

Srinivasan and Gersho [14] extend the basic VAD used in the GSM standard [16] for two typically encountered noise types: vehicular noise, which is stationary, and babble noise, which is non-stationary. The basic VAD used in the GSM standard is made of an adaptive FIR noise suppression filter with LPC coefficients [14] followed by an energy detector with an adaptive threshold. The *mobile VAD* proposed by [14] adds the ability to perform energy thresholding for four individual sub-bands. It also measures spectral flatness at the output of the filter for better indication of speech and uses an adaptive hangover scheme to reduce clipping at the end of speech. The mobile VAD is experimentally shown to drastically reduce clipping in high noise levels and is even slightly better than the standard GSM VAD in low noise levels. The *babble VAD* is also proposed to address babble noise. It consists of two VADs, the primary one that makes the decision about speech existence and a secondary one that detects the noise-only frame to update the thresholds used by the primary one. It is shown that this VAD indeed outperforms the mobile VAD for babble noise background. To combine the benefits of both VADs, a heuristic scheme is proposed to switch between mobile VAD and babble VAD depending on the stationarity of the input signal. Namely mobile VAD is used for stationary input and babble VAD is used for non-stationary input.

Instead of switching between VADs, a fusion method studied in [15] combines the outputs of the periodicity-based VAD [17] and energy-threshold-based VAD [18], both of which are used to address non-stationary noise. The periodicity-based VAD can operate at SNRs lower than 0 dB but suffers from false alarms due to periodic noise or interference signals. The energy-threshold-based VAD, on the other hand, does not suffer from periodicity of noise and interference signal, but usually requires a lower bound on SNR to operate without a high missed rate. The fusion is done by a weighted sum of the binary outputs of the two VADs, thresholding on 0.5 for the final decision. Hence the weighting on the two VADs can be used to trade off between miss rate and false-alarm rate. It is shown experimentally that this fusion results in a VAD that outperforms all previously known VAD methods.

From VAD design in time-varying environments, it seems that for consis-

tently high performance, a scheduling or fusion method needs to be employed to combine the benefits of different VADs that were designed specifically for different environments. The formulation in the next chapter will be carried out in light of this. However, unlike previous approaches which are heuristic-based, it is attempted to rigorously optimize the combination to maximize the total system performance.

CHAPTER 4

FORMULATION

"For since the fabric of the universe is most perfect and the work of a most wise Creator, nothing at all takes place in the universe in which some rule of maximum or minimum does not appear."

– Leonhard Euler

In order to address the limitations that were identified in previous work, a new system block diagram is proposed (see Figure 4.1). The purpose of this block diagram is to avoid the issues with the cascade discussed in Section 3.2, i.e. it allows the selection of an appropriate detector according to the SNR information. In addition, the blocks are subjected to optimization in order to overcome the sub-optimality problem with the heuristic approaches in 3.3.

The system consists of an estimator block that estimates the signal and noise power. For now it is assumed that the estimator is perfect and the signal and noise power estimates are true ones. This signal and noise power information is then fed into a scheduler block which selects an appropriate detector for the situation. The selected detector also uses the signal and noise power information as model parameters. The design problem is then to find the best scheduling scheme μ and the detectors' thresholds τ_U that maximize detection performance while satisfying the energy constraint. This chapter therefore presents the framework for solving such design problems.

Let the input information to the scheduler be the 2-tuple random process $X_n = (P_n, Q_n)$ with P_n being the signal power and Q_n being the noise power. The output decision $U \in \{1, 2\}$ is the detector to use, assuming for simplicity that there are only two available detectors in the system. The scheduler itself is then a (possibly randomized) policy μ that maps X_n to 0 if the first detector is used and 1 if the second detector is used. From detection

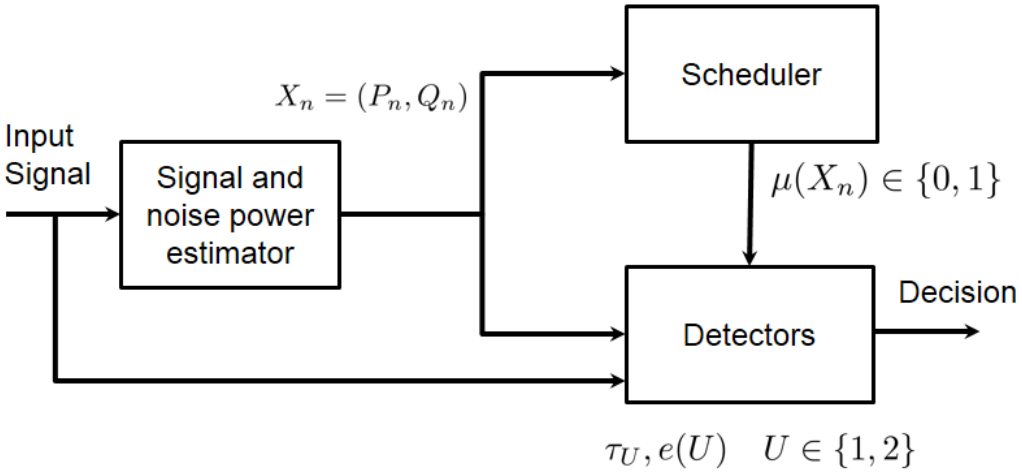

Figure 4.1: Proposed system block diagram.

theory [2], a detector can be modeled by the equation

$$T_U(\mathbf{Y}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau_U$$

where H_0, H_1 are the two standard hypotheses: noise and signal plus noise, respectively, and $\mathbf{Y}$ is the noisy observation vector from one of the two hypotheses. The function T_U maps the given observation to a test statistic which is then thresholded by τ_U to decide which hypothesis was true. The test statistic for each detector is fixed so that the energy cost to compute them can be quantified and denoted by $e(U)$. Hence, at a particular time instance n, given X_n, the scheduling policy μ, and the threshold τ_U, the system risk and energy consumption (EC) can be defined as

$$\begin{aligned} R^{sys}(X_n, \mu, \tau_U) &\triangleq \mu(X_n)R(X_n, \tau_1) + (1 - \mu(X_n))R(X_n, \tau_2) \\ EC(X_n, \mu) &\triangleq \mu(X_n)e(1) + (1 - \mu(X_n))e(2) \end{aligned}$$

where $R(X_n, \tau_U)$ is the respective detector risk that depends on the threshold and the signal and noise power. The goal then is to find the scheduling policy μ and thresholds τ_U that minimize the average system risk subject to an average energy consumption. Therefore, the optimization problem to be

solved is

$$\begin{aligned} \min_{\mu,\tau_U} \quad & \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} R^{sys}(X_n, \mu, \tau_U) \quad \text{m.s.} \\ \text{s.t.} \quad & \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} EC(X_n, \mu) \leq \beta \quad \text{m.s.} \end{aligned} \tag{4.1}$$

where β is the average energy constraint and the limit is taken in the mean square (m.s.) sense.

In this work, it is assumed that the process X_n is wide-sense stationary (WSS) and ergodic. Therefore, the risk and energy consumption of the system are also WSS and ergodic, because the policy μ cannot affect the natural process X_n. This assumption allows us to convert the time average in (4.1) to the ensemble average with respect to the joint long-term statistics $p(\cdot)$ of X.

$$\begin{aligned} \min_{\mu,\tau_U} \quad & \mathbb{E}[R^{sys}(X, \mu, \tau_U)] \\ \text{s.t.} \quad & \mathbb{E}[EC(X, \mu)] \leq \beta \end{aligned} \tag{4.2}$$

Expanding (4.2) yields

$$\begin{aligned} \min_{\mu,\tau_1,\tau_2} \quad & \int dx\, p(x) \Big\{ \mu(x)[R(x,\tau_1) - R(x,\tau_2)] + R(x,\tau_2) \Big\} \\ s.t. \quad & \int dx\, p(x) \Big\{ \mu(x)[e(1) - e(2)] + e(2) \Big\} \leq \beta \end{aligned} \tag{4.3}$$

Observe in (4.3) that the thresholds only appear inside the detector risk so that the minimization over thresholds can be moved inside. This fact decouples (4.3) into two subproblems: optimization over thresholds and optimization over policy.

4.1 Detector Threshold Optimization

The first subproblem is given by

$$\min_{\tau_U} R(x, \tau_U), \quad U = 1, 2$$

with

$$\begin{aligned} R(x, \tau_U) =& \pi_1 \int dy\, f_1(y) I(T_U(\mathbf{y}|x) < \tau_U) \\ &+ \pi_0 \int dy\, f_0(y) I(T_U(\mathbf{y}|x) \geq \tau_U) \end{aligned}$$

where $\pi_1 = \Pr(H_1)$, $\pi_0 = \Pr(H_0)$, and $f_0(\mathbf{y})$ and $f_1(\mathbf{y})$ are the observation's densities under the two hypotheses H_0 and H_1, respectively. $I(\cdot)$ denotes the indicator function.

Except for the special case where the ratio between observation densities $f_1(\mathbf{y})/f_0(\mathbf{y})$ is the same as the test statistic $T_U(\mathbf{y})$, the optimal threshold is not the simple Bayesian threshold $\tau_U^* = \pi_0/\pi_1$ [2]. In general, τ_U^* might need to be determined empirically, especially if a reliable observation model is not available. A technique for adaptively tracking the thresholds is discussed in Section 5.1.

4.2 Scheduler Policy Optimization

The second subproblem is given by

$$\begin{aligned} \min_{\mu} \quad & \int dx\, p(x) \Big\{ \mu(x)[R(x, \tau_1^*) - R(x, \tau_2^*)] + R(x, \tau_2^*) \Big\} \\ s.t. \quad & \int dx\, p(x) \Big\{ \mu(x)[e(1) - e(2)] + e(2) \Big\} \leq \beta \end{aligned} \tag{4.4}$$

Even though the above problem is a scheduling problem, it shares exactly the same structure as the well-known detection problem in the Neymann-Pearson lemma [3]. Applying the same machinery (see Appendix B) yields

$$\mu^*(x) = \begin{cases} 0 & & \text{if } M(x) > \lambda \\ 1 & \text{w.p.} \rho & \text{if } M(x) = \lambda \\ 1 & & \text{if } M(x) < \lambda \end{cases} \tag{4.5}$$

where $M(x) = \frac{R(x,\tau_1^*) - R(x,\tau_2^*)}{[e(2)-e(1)]}$ denotes the scaled relative Bayesian risk between two detectors, and $\lambda \in [0, \infty)$, $\rho \in [0, 1]$ are artificial variables [19]. The exact values of ρ and λ are determined from the energy constraint, as will be shown in Section 5.2.

The fact that the optimal policy given in (4.5) is a simple threshold test

is significant, because it implies that the addition of the scheduling module to the system adds virtually no extra overhead.

4.3 Robust Scheduling Policy

So far it is assumed that the signal and noise power are estimated perfectly by the estimator block (see Figure 4.1). However, in practice, the errors in signal and noise power estimates are unavoidable. Therefore a natural question to ask is how the estimation errors might affect the scheduling policy obtained in (4.5). In fact, it can be shown that the optimal policy structure in (4.5) is robust to the estimated signal and noise power. On the other hand, the policy threshold needs to be modified to ensure that the energy constraint is not violated even in the worst case (see Appendix C).

Encouraged by the robustness of the policy structure, a heuristic implementation of the estimator block is given in Section 5.3.

CHAPTER 5

ALGORITHMS

"An algorithm must be seen to be believed."

– Donald Knuth

This chapter complements the discussion in Chapter 4 by describing the algorithm used to (1) adapt the detectors' thresholds, (2) find the optimal policy threshold, and (3) estimate signal and noise power.

5.1 Detector Threshold Adaptation

As discussed in Section 4.1, the threshold τ_U of detector U might need to be determined empirically when a reliable observation model for $\mathbf{Y}$ is not available. In these situations, one approach is to refine the threshold over time, a process called threshold adaptation, based on the stochastic approximation theory [20]. The goal of this approach is to find the threshold τ_U such that the resulting false alarm rate satisfies the desired level γ regardless of the observation model. The motivation for this goal is based on the optimality condition of the Neyman-Pearson formulation in Section 2.2.2, in which the optimal threshold is the one that satisfies the desired level γ. For the rare-event case, it can be shown (see Appendix A) that the stochastic approximation equation for threshold $\tau_{U,n}$ at time n is given by

$$\tau_{U,n+1} = \alpha_{\tau_U}\tau_{U,n} + (1 - \alpha_{\tau_U})(I(T_U(\mathbf{y}_n) \geq \tau_{U,n}) - \gamma) \tag{5.1}$$

where $\alpha_{\tau_U} \in [0, 1]$ is the threshold smoothing constant. Hence the remaining problem is to choose the appropriate α_{τ_U}. It can also be shown that (see Appendix A), due to the characteristic of (5.1), a necessary condition for the resulting false alarm to converge to γ is the existence of an appropriate scaling constant C_U for the test statistic. Hence the complete update equations are

given by

$$\begin{aligned} T_U(\mathbf{y}_n) &= C_U T_U(\mathbf{y}_n) \\ \tau_{U,n+1} &= \alpha_{\tau_U}\tau_{U,n} + (1-\alpha_{\tau_U})(I(T_U(\mathbf{y}_n)) - \gamma) \end{aligned} \tag{5.2}$$

5.2 Optimal Policy Threshold

Knowledge of $\mathbb{E}[EC(X, \mu^*(\lambda))]$ is needed in order to find the optimal policy threshold. One way to do that is through analysis. Without loss of generality, assume that $e(2) > e(1)$. Substituting the optimal policy mapping μ^* into the energy constraint yields

$$\begin{aligned} \mathbb{E}[EC(X, \mu^*(\lambda))] = &\int_{-\infty}^{\lambda} dt\, m(t)e(1) + \int_{\lambda}^{\infty} dt\, m(t)e(2) \\ &+ m(\lambda)\big[(1-\rho)e(1) + \rho e(2)\big] \end{aligned}$$

where $m(t) = \int dx p(x) I(M(x) = t)$. Another way to obtain $\mathbb{E}[EC(X, \mu^*(\lambda))]$, useful when the analytical method becomes intractable, is to apply machine learning techniques on training data.

Using either method, a sketch of $\mathbb{E}[EC(X, \mu^*(\lambda))]$ is given in Figure 5.1. From this figure, the solution for λ and ρ can be obtained using the following algorithm. Depending on the energy constraint β,

- If $\beta > e(2)$ then the average energy constraint is redundant. This corresponds to the case $\beta = \beta_1$, hence $\lambda = \rho = 0$.
- If $e(1) \leq \beta \leq e(2)$ then the energy constraint can be satisfied with equality. If there is no point mass, i.e. $\beta = \beta_2$, then $\lambda = \lambda_2^*$ and $\rho = 0$. If there is point mass, i.e. if $\beta = \beta_3$, then $\lambda = \lambda_3^*$ and $\rho = \rho^*$. The λ^* and ρ^* can be found using any root-finding method, i.e. if the function is monotonic, then bisection search can be used.
- If $\beta < e(1)$ then the energy constraint is so stringent that it cannot be satisfied.

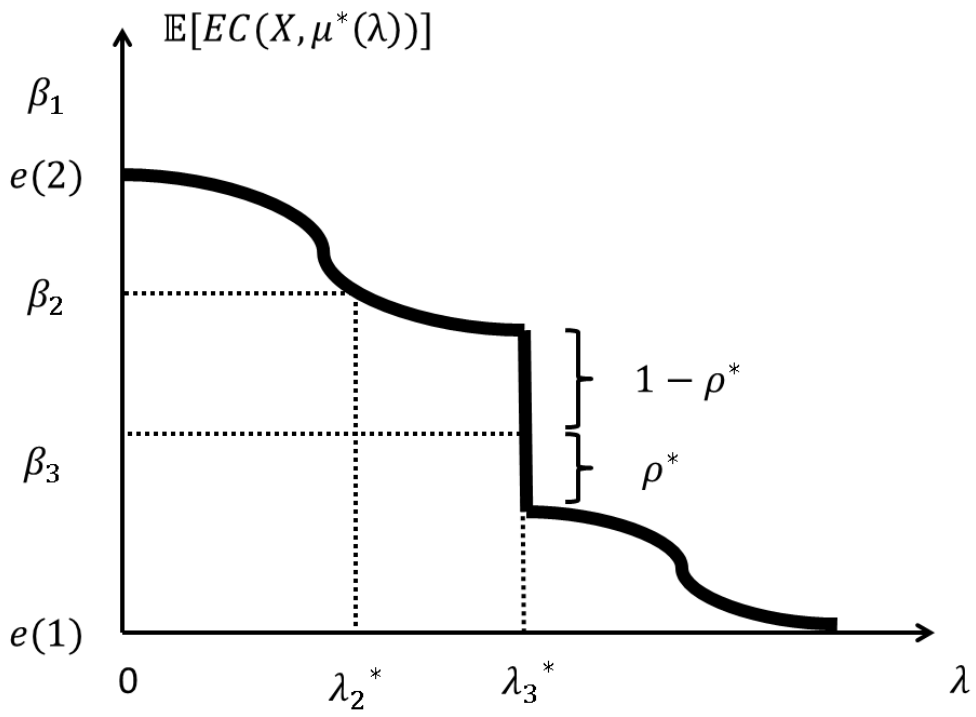

Figure 5.1: Solution of λ^* and ρ.

5.3 Signal and Noise Power Estimation

Signal power P_n and noise power Q_n are estimated using recursive averaging [21]. By modeling the signal and noise power as variances, i.e. $P = \sigma_s^2$ and $Q = \sigma_0^2$, the ML estimates for signal and noise variances discussed in Section 2.4.1 can be used to estimate signal and noise power. Thus the desired update equations for noise power with smoothing coefficient α_Q can be given by

$$\begin{aligned} H_0^n &: Q_{n+1} = \alpha_Q Q_n + (1-\alpha_Q)\|Y_n\|^2 \\ H_1^n &: Q_{n+1} = Q_n \end{aligned} \tag{5.3}$$

and the equations for the signal power with smoothing coefficient α_P by

$$\begin{aligned} H_0^n &: P_{n+1} = P_n \\ H_1^n &: P_{n+1} = \alpha_P P_n + (1-\alpha_P)\Big[\|Y_n\|^2 - Q_n\Big] \end{aligned} \tag{5.4}$$

where H_0^n and H_1^n are the hypotheses decided by the system and thus different from the true hypotheses H_0 and H_1.

Let $\pi_1^n = \Pr(H_1^n)$, then the update equation for π_1^n with smoothing coefficient α_π is

$$\pi_1^n = \alpha_\pi \pi_1^{n-1} + (1 - \alpha_\pi) I(H_1^n)$$

Hence the two equations in (5.3) can be combined into

$$Q_{n+1} = \widetilde{\alpha}_Q^n Q_n + (1 - \widetilde{\alpha}_Q^n) \|Y_n\|^2$$

where $\widetilde{\alpha}_Q^n = \alpha_Q + (1 - \alpha_Q)\pi_1^n$. Similarly for (5.4)

$$P_{n+1} = \widetilde{\alpha}_P^n P_n + (1 - \widetilde{\alpha}_P^n) \Big[\|Y_n\|^2 - Q_n \Big]$$

where $\widetilde{\alpha}_P^n = 1 - (1 - \alpha_P)\pi_1^n$.

In the case when the signal-to-noise-ratio (SNR) estimate at time n is desired, it can be given by the ratio between P_n and Q_n.

CHAPTER 6

DETECTION OF GOLDEN-CHEEKED WARBLER: SIMULATION

"Anyone who attempts to generate random numbers by deterministic means is, of course, living in a state of sin."

– John von Neumann

In this chapter we present the experimental performance of a detection system using the scheduling algorithm developed in Section 4.2. The target of detection is the call of an endangered bird species named the Golden-Cheeked Warbler (GCW) [1]. The observations are collected in blocks with the presumed size N. The system employs two detectors. The first detector is the energy detector described in Section 2.4.1. It has relatively low complexity, namely $O(N)$ multiply-accumulate (MAC) operations [7]. The second detector is the *quadratic detector* described in Section 2.4.2. Its implementation requires $O(N^2)$ MAC operations [7]. Therefore their respective costs $e(1)$ and $e(2)$ can be assigned to be N and N^2.

Figure 6.1 illustrates the system's operation on five hours of real GCW data, recorded by Professor Rama Ratnam from the Biology Department at the University of Texas at San Antonio [1]. The first window shows the data re-sampled at 16 kHz, since all of the bird call spectrum is below 8 kHz, and processed in frames of size $N = 128$. It is then manually labeled as shown in the second window from the top of Figure 6.1. It is worth noting that because the detectors make decisions on individual frames while labels are given in multiple frames for each call, post-processing is required to evaluate bird call detection. Namely, in our experiment, if more than 25 % of the frame in a call is labeled as detected, then the whole call is considered detected. The third window shows the tracked SNR by using the algorithm in Section 5.3 and label information; that is, we assume perfect SNR estimates. The two detectors with Bayesian thresholds are managed by the scheduling algorithm described in Section 4.2. As can be seen from the fourth window, the second

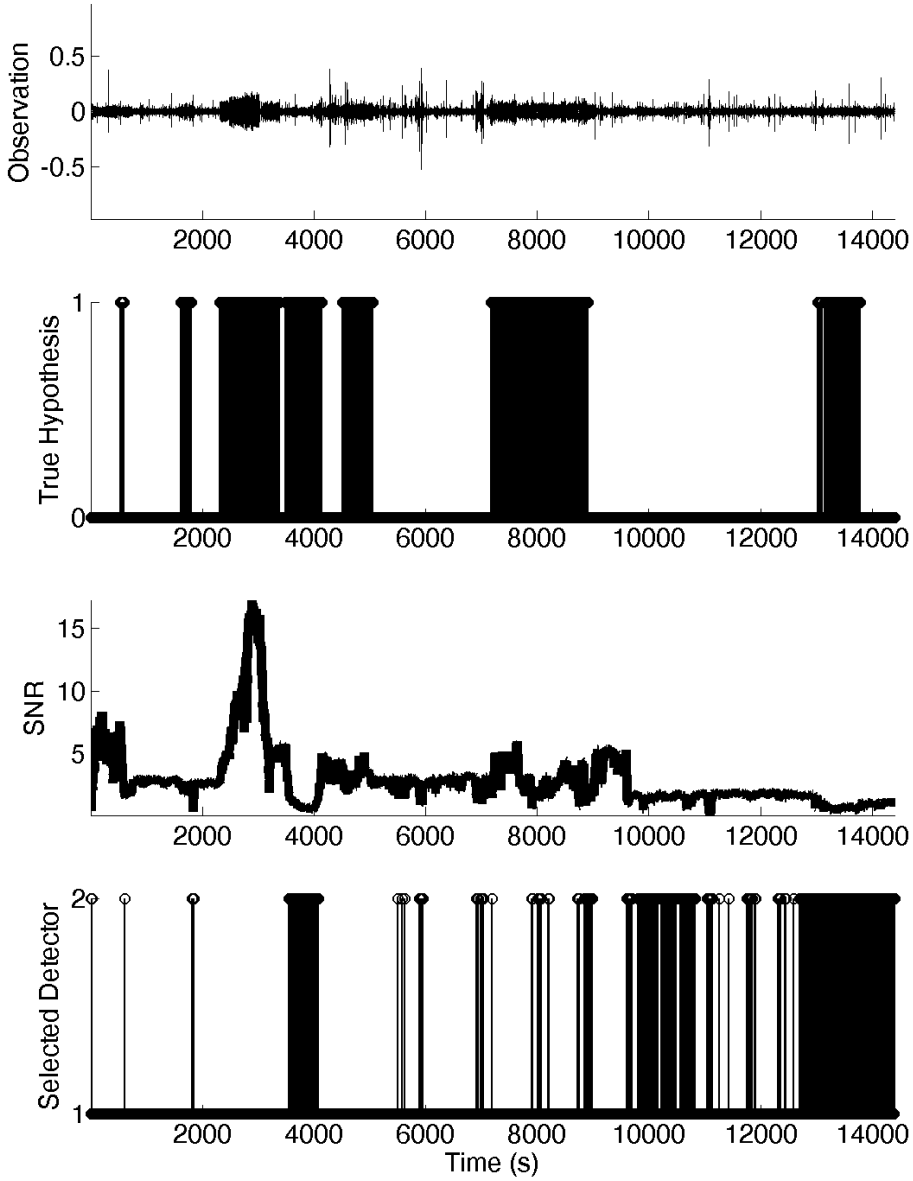

Figure 6.1: Optimal detector scheduling on sample GCW data.

detector is run only when SNR is low while the first detector runs in the remaining time. The fraction of time between running the second detector and the first detector is determined by the energy budget. The operating point used in this case is labeled in Figure 6.2.

The optimal energy-performance curve is shown in Figure 6.2 as the solid

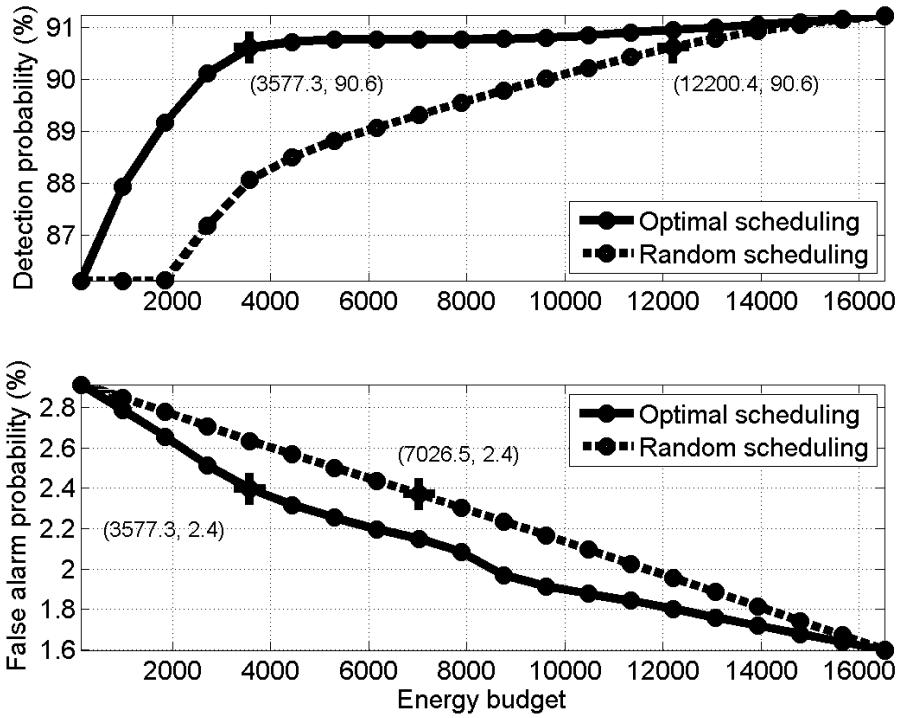

Figure 6.2: Comparison between optimal scheduling and random scheduling over various energy budgets.

line. The energy-performance curve of a random scheduling system, which is a straightforward but naive approach for this application, is shown in the dashed line. In this figure, the labeled points together illustrate the gap in the energy budget required between the optimal scheduling system and the random scheduling system for the same level of desired performance. Recall that for these simulations, the energy cost is assumed to be proportional to the number of MAC operations. Namely, the optimal scheme is 3.5x more energy efficient than the random scheme at the same level of detection probability. The number is 2x for false alarm probability. Furthermore, the optimal scheduling system's performance scales gracefully over an order of magnitude of the energy constraint.

Using the same setup, but with adaptive thresholds (see Section 5.1) instead of Bayesian thresholds for the detectors, the energy-performance curves of the optimal scheduling scheme and the cascade scheme, discussed in Section 3.2, are compared in Figure 6.3. The cascade is implemented using the state-of-the-art algorithm in [22], with the energy detector serving as the first

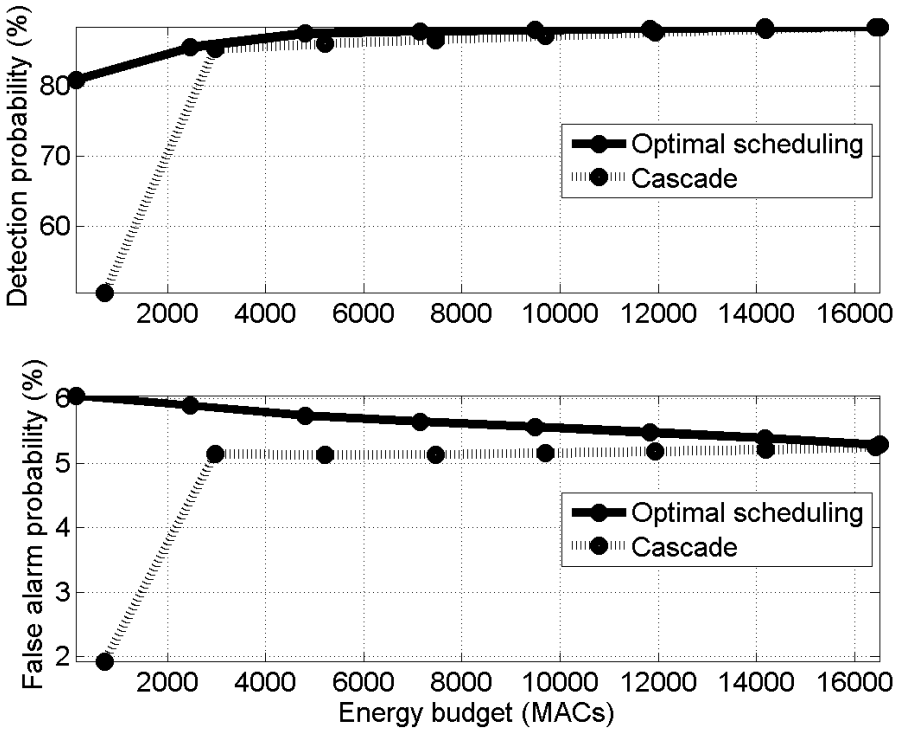

Figure 6.3: Comparison between optimal scheduling and cascading over various energy budgets.

stage to trigger the quadratic detector. The target false alarm is set to be equal to the prior probability of H_1.

As illustrated by the labeled points in Figure 6.3, the probability of detection using the optimal scheduling scheme is significantly better than the one using the cascade scheme at a very low energy budget. The reason is because in the cascade, the H_1 decision can only be made after executing the quadratic detector, which is not possible under a stringent energy budget. On the other hand, the optimal scheduling scheme has higher false-alarm rate. In fact, the false-alarm rate of any scheduling scheme cannot be lower than that of the cascade because the quadratic detector is uniformly better [22] than the energy detector. However, this might not be a problem for rare event detection, in which high detection probability is often more desirable at the expense of a reasonable false alarm rate. For example, in studying the evolution in the call spectrum of the GCW, it is important that the call data are collected intact. Missing too many calls during the collection process might lead to incorrect inference about the evolution in the call spectrum

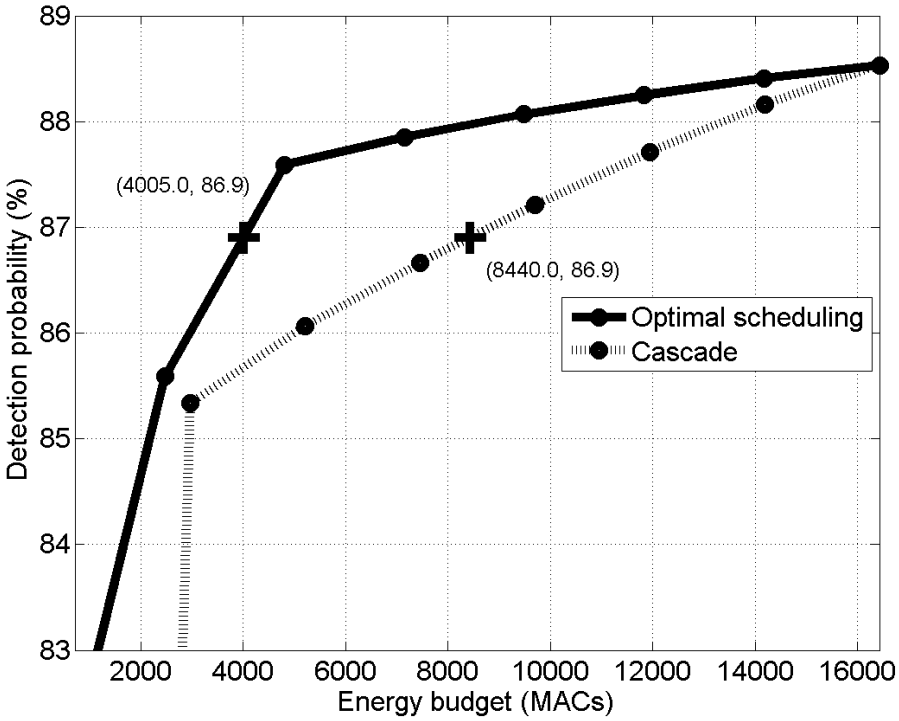

Figure 6.4: Zoom-in on probability of detection between optimal scheduling and cascading.

over time.

A closer look at the detection probability when the energy budget is not so tight, as illustrated by the labeled points in Figure 6.4, reveals that the optimal scheduling strategy can achieve the same performance as the cascade strategy at 2x reduction in energy budget.

This chapter concludes with the remark that the optimal scheduling is always better than a random scheduling. Furthermore, the optimal scheduling strategy can outperform the cascade strategy in the detection rate, especially at an extremely stringent energy budget, but not in the false alarm rate. Depending on the application, one might prefer to use one over the other. For rare-event detection, in which high detection rate is important, optimal scheduling might be preferable.

CHAPTER 7

DETECTION OF GOLDEN-CHEEKED WARBLER: EFM32TG IMPLEMENTATION

"Machines take me by surprise with great frequency."

– Alan Turing

Early empirical evidence for the validity of the method developed in Chapter 4 was obtained through computer simulation in Chapter 6. However, simulation is usually not enough to fully characterize all the tradeoffs arising in a real system implementation. Hardware implementation, on the other hand, will be able to affirm the practical validity of the proposed method. Furthermore, system realization also provides realistic power numbers that will be useful for estimating the battery life of the system once deployed.

7.1 STK 3300 and EFM32TG840F32

The platform of choice for implementation is the Starter Kit (STK) 3300 from Energy Micro [23], which comes with the EFM32TG840F32 *Tiny Gecko* series microcontroller unit (MCU). The advantages of using this MCU over other off-the-shelf competitors in the market such as TI MSP430 is that it has (1) very low energy consumption thanks to specialized peripherals [24] and (2) convenient, software-based energy debugging support [25]. The specialized peripherals will be extremely useful for implementing the ultra low-power energy detector, as can be seen in Section 7.2. The software-based energy debugging tool allows a quick and accurate [25] interface for current measurement of the EFM32TG840F32. Since supply voltage is known and fixed at 3.3 V, power consumption is simply current times voltage.

Additional features can also be added on the board. Due to limited sensors available on the STK 3300, an additional microphone and preamp circuit board for acoustic sensing is added through the expansion headers. The circuit is designed by Texas Instruments and is used in the MSP-EXP430F5438

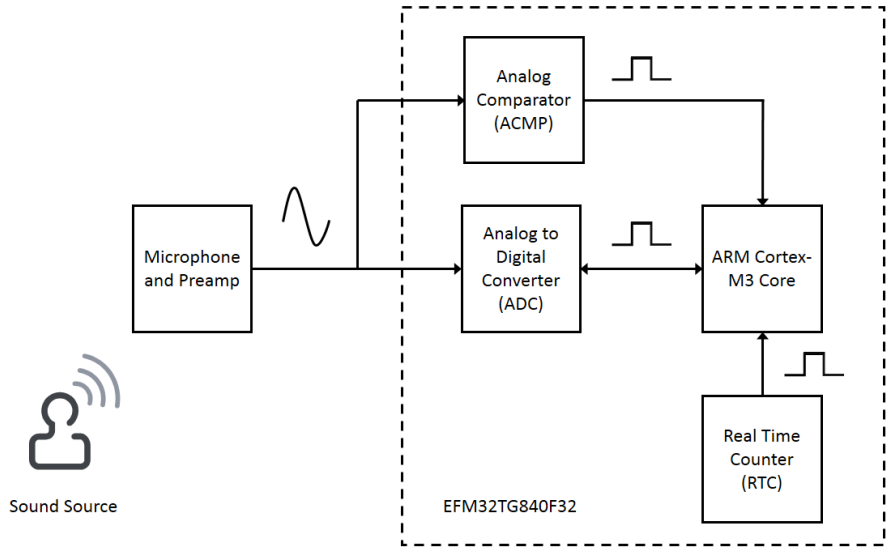

Figure 7.1: Low-power peripherals setup for data acquisition. The sine wave represents the analog signal while the square wave represents the digital signal and/or trigger.

experimenter board [26]. Another desirable feature is the voltage-scaling capability instead of a fixed 3.3 V supply voltage. This feature is recently demonstrated in [22] as having significant impact on energy saving. However, this is more appropriate for the next step in the development process when a dedicated PCB is fabricated because the STK 3300 does not allow the supply wire to be exposed by any headers. Hence dynamic voltage-scaling remains future work to be done.

The EFM32TG840F32 MCU is based on an ARM Cortex M-3 core with 32 KB of Flash and 4 KB of RAM. It also possesses standard MCU peripherals such as the analog-to-digital converter (ADC), real time counter (RTC), direct memory access (DMA), etc., plus specialized peripherals from Energy Micro such as the analog comparator (ACMP) and the low-energy sensor interface (LESENSE), etc. Since the MCU is targeted for low-power application, it does not have a floating point unit and hence all signal processing is done in fixed point arithmetics.

The MCU is configured for audio signal processing. The block diagram is illustrated in Figure 7.1 and includes the following components:

- The RTC is configured to run on a low-frequency crystal clock source of 32.768 kHz and has a variable trigger rate. It triggers at the rate of

16.384 kHz when the ADC is used, i.e. sampling rate of 16.384 kHz, and 128 Hz when ACMP is used. The reason for this is discussed in more detail in Section 7.2.

- The ACMP is configured as part of the implementation of the energy detector, which is discussed in detail in Section 7.2.

- The ADC is configured to provide the quadratic detector with 12-bit-resolution samples. Whenever triggered by the MCU, the ADC converts the input analog signal and return with a digitized value. This is done instead of employing the service of a DMA because on this platform, for a slow sampling rate, i.e. less than 20 kHz, running DMA actually consumes more power [27].

- The ARM Cortex-M3 core is configured to be in deep-sleep mode [28] most of the time. When awakened, it runs on the high-frequency crystal clock source of 32.768 MHz. The MCU executes either the energy detector or the quadratic detector for a time period of 1/128 seconds and accumulates the test statistic. Detailed implementation of the energy detector and the quadratic detector are discussed in Section 7.2 and 7.3, respectively. Each detector's threshold is adapted using the algorithm in Section 5.1. Then the decision about the presence or absence of a bird call is made at the end of each period.

  The MCU also decides which detector to schedule for the next period by comparing *directly* the SNR estimate, which is updated using the algorithm in Section 5.3, with a policy threshold λ that was determined experimentally. The justification for this can be based on the discussion in Section 2.2.3, in which a probabilistic criterion such as risk (see Section 4.2) can be substituted by a more amenable SNR criterion. It is also worth noting that the overhead for the scheduler in this case is merely an *"if $\cdots$ else"* statement which incurred zero overhead. Figure 7.2 summarizes the discussion by a software flowchart.

Due to the limited memory capacity of the starter kit STK 3300, the preliminary performance evaluation of the system is done using an eight-minute sample of the recorded data mentioned in Chapter 6. This sample data, shown in Figure 7.3, is a concatenation of representative sections across

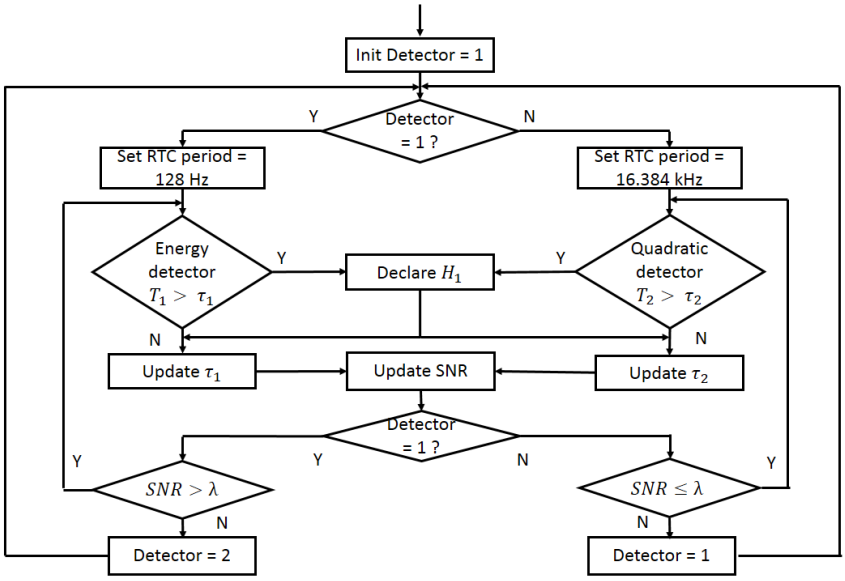

Figure 7.2: The flowchart of the software implemented on the MCU.

the five-hour data. Table 7.1 shows the performance of two configurations for the quadratic and the energy detector, respectively.

Table 7.1: Performance evaluation of the two detectors using the sample data in Figure 7.3, which contains a total of 42 GCW calls. Recall that there are 128 decisions made per second, and a call is detected once a significant number of correct detection decisions are made.

MCU configuration	Detection rate (%)	False alarm rate (%)
RTC + ACMP + Core(estimator, energy detector)	85.71	20.97
RTC + ADC + Core(estimator, quadratic detector)	100	4.7

Table 7.2 shows the average current consumption of the above setup, whose gap between two detector configurations is significantly larger than the standard setup in Table 7.3. The standard setup is the one in which both energy detector and quadratic detector require samples from the ADC, as illustrated in Figure 7.4. In the standard setup, the gap is small because, as current consumption analysis of individual components in Table 7.4[1, 2] reveals, the

[1]RTC current measurement is configuration-independent, and is taken from the datasheet.

[2]Table 7.3 and Table 7.4 are consistent up to 1 μA of measurement error.

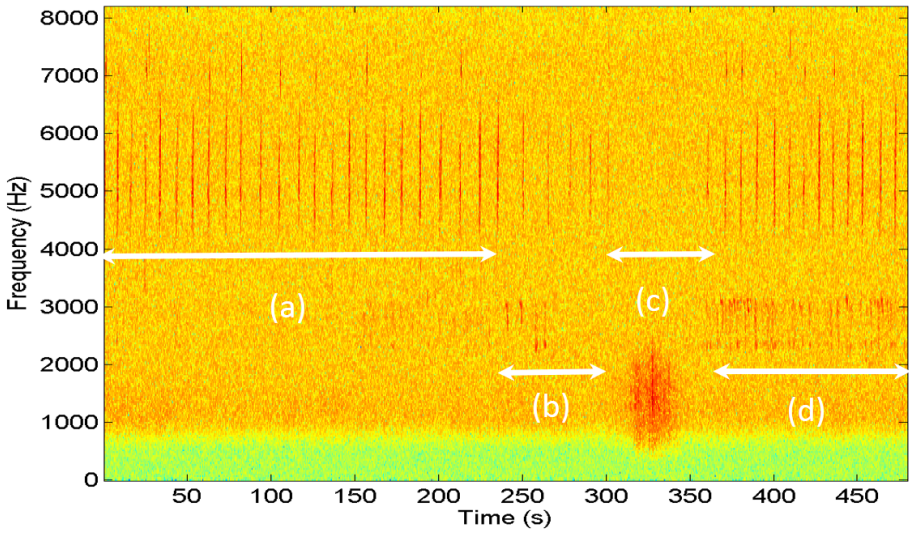

Figure 7.3: Spectrogram of the sample data with 512-sample window size, 50% overlap and 16.384 kHz sampling rate. Energy spikes in the range of 4.5 kHz to 7.5 kHz indicate the GCW calls. This sample data is concatenated from four representative sections. (a) has high SNR, (b) has low SNR, (c) has high noise power, and (d) has high SNR with interference from other bird calls.

Table 7.2: Current drawn by the MCU using the specialized peripherals.

MCU configuration	Average current drawn
RTC + ACMP + Core(estimator, energy detector)	33.93 μA
RTC + ADC + Core(estimator, quadratic detector)	2.915 mA

current consumption of the RTC, ADC, and Core with estimator and energy detector in Figure 7.4 is dominated by the current consumption of the ADC. This small gap dwarfs the need for scheduling in the standard setup. Hence an implementation of the energy detector without ADC, which is the topic of Section 7.2, is important both to reduce the total system current consumption, and to motivate the need for scheduling.

Table 7.3: Current drawn by the MCU with standard data acquisition setup.

MCU configuration	Average current drawn
RTC + ADC + Core(estimator, energy detector)	673.215 μA
RTC + ADC + Core(estimator, quadratic detector)	2.915 mA

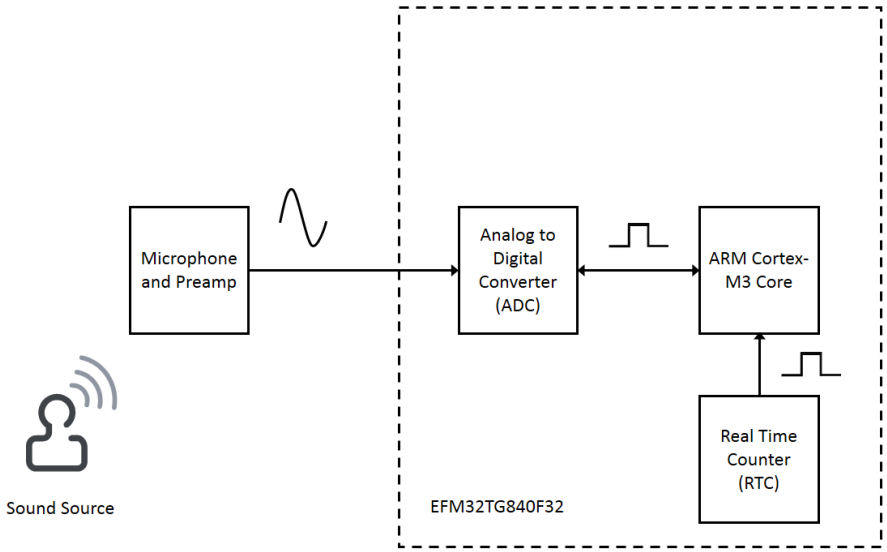

Figure 7.4: Standard peripheral setup for data acquisition. The sine wave represents the analog signal while the square wave represents digital signal and/or trigger.

7.2 Low-Power Implementation of the Energy Detector

The analog comparator (ACMP) is a specialized peripheral by Energy Micro [29] that can be used to replace the ADC in the implementation of the energy detector, using the setup in Figure 7.1.

The ACMP is configured with two reference voltages, a positive and a negative one. Every time the input voltage crosses the boundary specified by the two reference voltages, the ACMP triggers an interrupt. By reading the difference in the RTC's counter values at different ACMP triggers, the amount of time that the input signal overshoots or undershoots can be calculated. If the two reference voltages are both fixed at the same level, then the energy of the input signal during the overshoot and undershoot time can be approximated as the square of that voltage level times the overshoot or

Table 7.4: Current drawn by individual components in the MCU.

Components	Average current drawn
RTC	100 nA
ADC	513.82 μA
Core(estimator, energy detector)	160.075 μA
Core(estimator, quadratic detector)	2401.18 μA

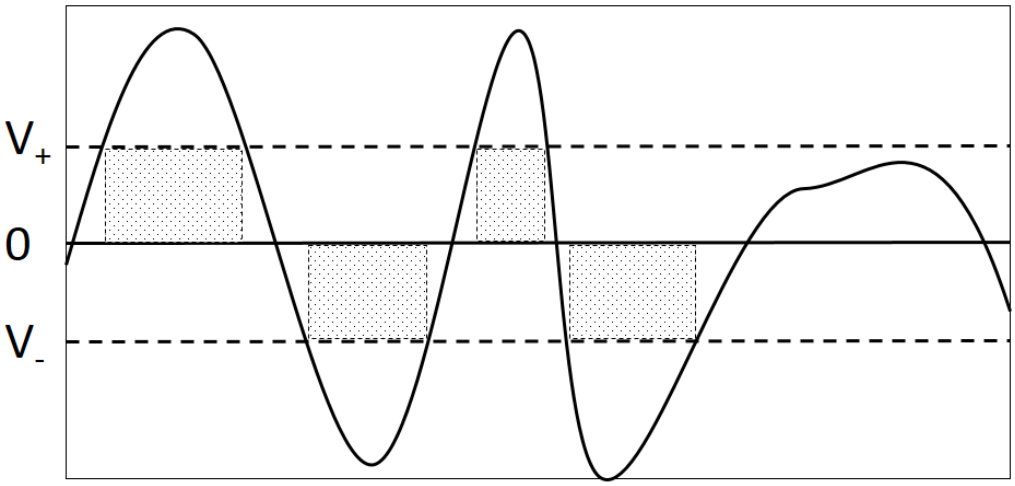

Figure 7.5: Approximation of the energy detector.

undershoot time. This approximation is illustrated in Figure 7.5, where the shaded areas represent the integrated time. The partial energy is accumulated during the elapsed time between RTC triggers. The final energy value is the test statistic. It is worth mentioning that the RTC trigger period needs to be 128 Hz for this test statistic to be consistent with the decision rate of 128 Hz of the system.

As can be seen from Table 7.2, this new configuration drastically reduces the current drawn by the MCU. The ratio in the current drawn, and hence the power consumption, between the two configuration is 86x, thus validating the motivation for doing scheduling.

7.3 Lattice Wave Digital Filter Implementation of the Quadratic Detector

The quadratic detector can be approximately implemented as a filter followed by an energy detector [4]. Since the GCW calls only occupy the frequency spectrum from 4.5 kHz to 7.5 kHz, a bandpass filter with the above passband can be used to filter out noise in other bands.

As all arithmetic operations are carried out in fixed-point representation, it is desirable to implement a lattice wave-digital filter (LWDF) [30] for numerical stability. This is because all the multiplications inside this filter operate on coefficients which are always positive and less than 0.5 [30]. Furthermore, the LWDF is a class of IIR filter, hence it has relatively a smaller order compared to a stable FIR filter with the same performance [30].

The implementation of a LWDF is assisted by two intermediate programs. The LWDF filter-design program that comes with [30] takes in the above passband specification, along with the sampling rate of 16.384 kHz, and generates a filter coefficients file. This file is then fed into an automated code generator that outputs the assembly code that realizes the specified filter. [3] Then this assembly code can then be called from the main routine, using the calling convention of the compiler, to process blocks of $N = 128$ samples before making a detection decision.

[3]The source code, written in C, for this LWDF code generator is open source, although at the time of this thesis it only supports MSP430 and ARM Cortex-M3 assembly languages.

CHAPTER 8

CONCLUSION

In detection problems with an energy constraint, information about the time-varying signal and noise power can be used to optimally elect between the high-performance but sophisticated detector and a simple but inexpensive one. The machinery to exploit this information by an optimal scheduler is simple and hence almost no overhead cost is added. The result is a system that consumes much less energy while preserving adequate performance.

On the other hand, the technique offered in this thesis also has several drawbacks. Firstly, it will not be able to offer much energy-saving in constant signal and noise power environment. Secondly, it will not work reliably in non-stationary signal and noise power environment. One way to address this problem is to extend the model of the signal and noise power random processes from WSS to piecewise-stationary using the Hidden Markov Model. For example, in the wildlife monitoring application, the hidden states can be the seasons in a year and different seasons specify different stationary distributions of the observable signal and noise power random processes. Thirdly, the simple constant energy consumption model of a detector in this thesis might be too restrictive to be true in practice. Therefore a straightforward extension of this work can be about more general energy consumption models of a detector. Finally, as mentioned in Chapter 6, scheduling architectures, which interconnect detectors in parallel, usually suffer from higher false alarm rates than the cascade architectures, which interconnect detectors in serial. An interesting extension of this work is to consider a generalized interconnection of detectors that includes both parallel and serial interconnection to achieve the best performance in energy-limited systems.

APPENDIX A

DERIVATION OF THE ADAPTIVE DETECTOR THRESHOLD EQUATION

Equation (5.1) is derived using stochastic approximation theory and the rare event assumption. Under the rare event assumption, the probability of event H_1 occurring is much lower compared to the probability of event H_0 occurring. Hence most of the system triggers can be attributed to false alarm, i.e. $P[T_U(\mathbf{Y}) \geq \tau_U | H_0] \approx P[T_U(\mathbf{Y}) \geq \tau_U]$. Given a false-alarm level γ, the threshold τ_U that results in such false alarm can be found by finding the root of the equation

$$P[T_U(\mathbf{Y}) \geq \tau_U] - \gamma = \mathbb{E}[I(T_U(\mathbf{Y}) \geq \tau_U) - \gamma] = 0 \tag{A.1}$$

If the knowledge about the density of $\mathbf{Y}$ was available, then (A.1) could be solved using the following recursion.

$$\tau_{U,n+1} = \alpha_{\tau_U} \tau_{U,n} + (1 - \alpha_{\tau_U})(\mathbb{E}[I(T_U(\mathbf{Y}) \geq \tau_{U,n})] - \gamma)$$

Stochastic approximation theory, which is widely used in adaptive filtering [10], proposes a recursive method to solve (A.1) even without the knowledge about the density of $\mathbf{Y}$. Instead, a sample path $\mathbf{y}_n$ of $\mathbf{Y}$ is needed for the recursion. The central idea is to replace the terms inside the expectation with their sample path. Namely,

$$\tau_{U,n+1} = \alpha_{\tau_U} \tau_{U,n} + (1 - \alpha_{\tau_U})(I(T_U(\mathbf{y}_n) \geq \tau_{U,n}) - \gamma)$$

which is simply a restatement of (5.1).

Observe that if

- The target γ is strictly less than 1 or greater than 0, and
- $T_U(\mathbf{y}_n)$ is *always* greater or less than γ

then the false-alarm rate resulting from the recursion in (5.1) might not

converge to γ. The goal of the remainder of this section is a derivation of a necessary condition to guarantee the convergence of the false-alarm rate using (5.1). The analysis of the recursion in (5.1) reveals an important characteristic that is useful to ensure such convergence. First consider the following lemma.

Lemma A.0.1 $\tau_{U,n} \in [-\gamma, 1-\gamma], \quad \forall n$

Proof Consider the case when H_1 is *always* decided. Assuming the convergence of the threshold, i.e. $\tau_{U,n} \xrightarrow{n\to\infty} \tau_U$, the asymptotic recursion in this case becomes

$$\begin{aligned} & \tau_U = \alpha_{\tau_U}\tau_U + (1-\alpha_{\tau_U})(1-\gamma) \\ \Leftrightarrow\ & \tau_U = (1-\gamma) \end{aligned}$$

This is the upper bound on the value of τ_U. Similar analysis in the case H_0 is *always* decided yields the lower bound on the value of τ_U.

$$\begin{aligned} & \tau_U = \alpha_{\tau_U}\tau_U + (1-\alpha_{\tau_U})(0-\gamma) \\ \Leftrightarrow\ & \tau_U = -\gamma \end{aligned}$$

Therefore $\tau_{U,n} \in [-\gamma, 1-\gamma], \quad \forall n$. ■

Because τ lies in a bounded interval, $T_U(\mathbf{Y})$ must also be bounded in an appropriate sense, as discussed in the following theorem, for the resulting false alarm to converge to γ.

Theorem A.0.2 *A necessary condition for the convergence of the false-alarm rate* $P[(T_U(\mathbf{Y}) \geq \tau_U)]$ *to* $\gamma \in (0,1)$ *is* $\mathbb{E}[T_U(\mathbf{Y})] \in (-\gamma, 1-\gamma)$.

Proof From Lemma A.0.1 it is known that $\tau_U \leq 1-\gamma$. Therefore

$$P[(T_U(\mathbf{Y}) \geq 1-\gamma)] \leq P[(T_U(\mathbf{Y}) \geq \tau_U)]$$

A necessary condition for convergence requires that

$$P[(T_U(\mathbf{Y}) \geq \tau_U)] = \gamma < 1$$

Considering the case when the test statistic $T_U(\mathbf{Y})$ is nonnegative and applying Markov's inequality yields

$$P[(T_U(\mathbf{Y}) \geq 1-\gamma)] \leq \frac{\mathbb{E}[T_U(\mathbf{Y})]}{1-\gamma}$$

Therefore, the upper bound on the expected value of the test statistic is given by $\mathbb{E}[T_U(\mathbf{Y})] < 1-\gamma$. Similarly for the case when $T_U(\mathbf{Y})$ is non-positive. Namely,

$$P[(T_U(\mathbf{Y}) \geq -\gamma)] \geq P[(T_U(\mathbf{Y}) \geq \tau_U)] > 0$$

and

$$\begin{aligned} P[(T_U(\mathbf{Y}) \geq -\gamma)] &= P[(-T_U(\mathbf{Y}) \leq \gamma)] \\ &= 1 - P[(-T_U(\mathbf{Y}) \geq \gamma)] \\ &\geq 1 - \frac{\mathbb{E}[-T_U(\mathbf{Y})]}{\gamma} \end{aligned}$$

implies $\mathbb{E}[T_U(\mathbf{Y})] > -\gamma$. ■

Theorem A.0.2 suggests the use of a test statistic scaling constant C_U described in Section 5.1. Namely, C_U should be selected so that $\mathbb{E}[C_U T_U(\mathbf{Y})] \in (-\gamma, 1-\gamma)$. For example, C_U can be 1 for the detector in Section 2.4.1, and $\|\Sigma_0^{-1}\Sigma_s\Sigma_1^{-1}\|_2$ [1] for the one in Section 2.4.2.

[1] $\|\cdot\|_2$ denotes the operator induced or spectral norm.

APPENDIX B

DERIVATION OF THE OPTIMAL SCHEDULING POLICY

The optimization problem over policy in (4.4) is a continuous constrained optimization problem. It is well known from Lagrangian theory [19] that to solve for the optimal policy μ, we can solve a related unconstrained optimization problem that includes an additional, artificial Lagrange multiplier λ. The Lagrangian is given by

$$L(\mu,\lambda) = \int dx p(x)\mu(x)\Big\{[R(x,\tau_1^*) - R(x,\tau_2^*)] + \lambda[e(1) - e(2)]\Big\} + const \quad \text{(B.1)}$$

Applying the Karush-Kuhn-Tucker (KKT) conditions, which establish the first-order necessary conditions for the optimality of constrained optimization problems, yields

$$\begin{cases} \mu^* = \arg\min\limits_{\mu} L(\mu,\lambda^*) \\ \lambda^*(\mathbb{E}[EC(X,\mu^*)] - \beta) = 0 \end{cases} \quad \text{(B.2)}$$

As the set of all randomized policies is trivially convex, these are also sufficient conditions. Using (B.1) and the first equation in (B.2), $\min\limits_{\mu} L(\mu,\lambda^*)$ is equivalent to

$$\mu^*(x) = \begin{cases} 0 & & \text{if } M(x) > \lambda^* \\ 1 & \text{w.p. } \rho & \text{if } M(x) = \lambda^* \\ 1 & & \text{if } M(x) < \lambda^* \end{cases}$$

where $M(x) = \frac{R(x,\tau_1^*) - R(x,\tau_2^*)}{[e(2)-e(1)]}$ denotes the relative risk between two detectors and ρ is yet another artificial variable that lies in $[0,1]$. The complementary slackness condition, i.e. the second equation in (B.2), plays an important role in determining the exact value of ρ^* and λ^*. Solving this condition is equivalent to the algorithm described in Section 5.2.

It is worth pointing out that the analysis in this section is analogous to

the Neymann-Pearson lemma [3] with the optimal policy replaced by the randomized threshold test.

APPENDIX C

SOLUTION TO A ROBUST FORMULATION

Assume that the estimated signal and noise power has a joint long-term statistic $g(\cdot)$ that is different from the true $p(\cdot)$. For accuracy, $g(\cdot)$ needs to be modeled in (4.3), and a contaminated model [2] is proposed. For a given tolerance $\epsilon > 0$, this model assumes that the estimated signal and noise power are correct (i.e. drawn from the true density $p(\cdot)$) with probability $1-\epsilon$ and incorrect (i.e. drawn from some bad density $h(\cdot)$) with probability ϵ [31]. Namely,

$$g(x) = (1-\epsilon)p(x) + \epsilon h(x)$$

It is desirable to design a robust scheduling policy μ^R that can safeguard against the least favorable density g^L, i.e.

$$\begin{aligned} \min_{\mu}\max_{g} \quad & \int dx g(x)\Big\{\mu(x)[R(x,\tau_1^*) - R(x,\tau_2^*)] + R(x,\tau_2^*)\Big\} \\ s.t. \quad & \int dx g(x)\Big\{\mu(x)[e(1) - e(2)] + e(2)\Big\} \leq \beta \end{aligned} \tag{C.1}$$

Before finding the solution to (C.1) we need to verify that it exists.

- The policy μ lies in the convex set that is compact with respect to the infinity norm, i.e. $\|\mu\|_\infty = \sup_{x\in\mathbb{X}} \mu(x) \leq 1$.
- The density $g(\cdot)$ lies in a convex, compact set by construction.
- The objective function of (C.1) is linear in both μ and $g(\cdot)$, and hence is convex in μ and concave in $g(\cdot)$.

Therefore according to von Neumann's minimax theorem [2], there exists a saddle-point (μ^R, $g^L(\cdot)$) that will solve (C.1). Now observe that for *any* density g, the robust policy structure is still given by (4.5), i.e. $\mu^R = \mu^*$. In other words, the optimal policy in (4.5) is robust to estimated signal and noise power. The robust policy threshold λ^R, however, depends on the least

favorable density that is yet to be found. Finding the least favorable g^L is the same as finding the least favorable h^L. This is equivalent to solving

$$\max_h \int dx h(x) \Big\{ \mu(x)[R(x, \tau_1^*) - R(x, \tau_2^*)] + R(x, \tau_2^*) \Big\} \tag{C.2}$$

Define all the terms inside the curly bracket in (C.2) to be $f(x)$. Applying Schwarz's inequality to the integral in (C.2) reveals that the least favorable $h^L(x)$ is proportional to $f(x)$. Since h^L is a density, it is given by $h^L(x) = f(x)/\int f(t)dt$. Using this least favorable density, a robust policy threshold can be found using the approach in Section 5.2. The purpose of the robust policy threshold is to ensure that the energy constraint is not violated even in the least favorable case.

REFERENCES

[1] W. Leonard, J. Neal, and R. Ratnam, "Variation of type B song in the endangered Golden-Cheeked Warbler (Dendroica chrysoparia)," *The Wilson Journal of Ornithology*, vol. 122, no. 4, pp. 777–780, 2010.

[2] B. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer Verlag, 2008.

[3] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. New York, NY: Prentice Hall, 2000, vol. 1.

[4] H. Poor and C. Chang, "A reduced-complexity quadratic structure for the detection of stochastic signals," *The Journal of the Acoustical Society of America*, vol. 78, p. 1652, 1985.

[5] L. Scharf and B. Van Veen, "Low rank detectors for Gaussian random vectors," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 11, pp. 1579–1582, 1987.

[6] S. M. Kay and J. R. Gabriel, "An invariance property of the generalized likelihood ratio test," *Signal Processing Letters, IEEE*, vol. 10, no. 12, pp. 352–355, 2003.

[7] Y. Sung, L. Tong, and H. V. Poor, "Optimal and suboptimal detection of Gaussian signals in noise: Asymptotic relative efficiency," *CoRR*, vol. abs/cs/0507018, 2005.

[8] A. Sayeed and D. Jones, "Optimal reduced-rank time-frequency/time-scale quadratic detectors," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1996, pp. 209–212.

[9] A. Sayeed and D. Jones, "Optimal quadratic detection using bilinear time-frequency and time-scale representations," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1994, pp. 365–368.

[10] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed. New York, NY: Prentice Hall, 2007.

[11] J. Winograd, S. Nawab, and A. Oppenheim, "FFT-based incremental refinement of suboptimal detection," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, Conference Proceedings*, vol. 5, 1996, pp. 2479–2482 vol. 5.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. I–511–I–518.

[13] D. Jun, "An energy-aware framework for cascaded detection algorithms," M.S. thesis, University of Illinois at Urbana-Champaign, December 2010. [Online]. Available: http://www.musyc.org/pubs/219.html

[14] K. Srinivasan and A. Gersho, "Voice activity detection for cellular networks," in *Proceedings IEEE Workshop on Speech Coding for Telecommunications*, 1993, pp. 85–86.

[15] S. Tanyer and H. Ozer, "Voice activity detection in nonstationary noise," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 4, pp. 478–482, 2000.

[16] D. K. Freeman, G. Cosier, C. Southcott, and I. Boyd, "The voice activity detector for the pan-european digital cellular mobile telephone service," in *International Conference on Acoustics, Speech, and Signal Processing*, 1989, pp. 369–372 vol.1.

[17] R. Tucker, "Voice activity detection using a periodicity measure," in *IEEE Proceedings I in Communications, Speech and Vision*, vol. 139, no. 4. IET, 1992, pp. 377–380.

[18] H. Ozer and G. Tanyer, "A geometric algorithm for voice activity detection in nonstationary Gaussian noise," *Algorithms*, vol. 3, p. 11, 1998.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[20] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400–407, 1951.

[21] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *Signal Processing Letters, IEEE*, vol. 9, no. 1, pp. 12–15, 2002.

[22] D. Jun and D. Jones, "Cascading signal-model complexity for energy-aware detection," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 1, pp. 65–74, 2013.

[23] *User Manual Starter Kit EFM32TG-STK3300*, Energy Micro, 2011.

[24] "EFM32 introduction white paper," White Paper, Energy Micro, Sep. 2009.

[25] "EFM32 energy debugging white paper," White Paper, Energy Micro, Oct. 2009.

[26] *MSP-EXP430F5438 Experimenter Board User's Guide*, Texas Instruments, 2011.

[27] *Analog to Digital Converter Application Note*, Energy Micro, 2012.

[28] *EFM32TG Reference Manual Tiny Gecko Series*, Energy Micro, 2011.

[29] *Analog Comparator Application Note*, Energy Micro, 2012.

[30] *Wave Digital Filtering Using the MSP430 Application Report*, Texas Instruments, 2006.

[31] P. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965.