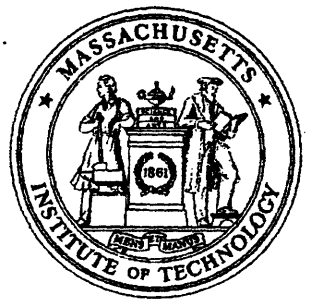

Detection Statistics for Multichannel Data

RLE Technical Report No. 550

December 1989

Tae Hong Joo

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

This work was supported in part by the Defense Advanced Research Projects Agency monitored by the Office of Naval Research under Grant No. N00014-89-J-1489, in part by the National Science Foundation under Grant No. MIP 87-14969, in part by Sanders Associates, Incorporated, and in part by a fellowship from the Amoco Foundation.

Detection Statistics for Multichannel Data

by

Tae Hong Joo

Submitted to the Department of Electrical Engineering and Computer Science
on September 9, 1989, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

A new detection statistic is developed for a class of multichannel detection problems which have the requirement that in the presence of an emitter, a narrowband signal must exist in all channels in addition to wideband noise. When an emitter is absent, the received data may contain narrowband noise components in some, but not all of the channels, as well as wideband noise. A detector which tests each channel separately for the existence of the narrowband component does not perform as well as the detectors which use all channels collectively.

To collectively use the data from different channels, average has been previously used as a detection statistic. However, because the average tests only the total energy, its detection performance noticeably degrades when a narrowband component exists in many channels. As an alternative detection statistic, the semblance, which measures the coherence between the channels, is considered. The receiver operating characteristic curves show that the average performs better than the semblance if more than half of the channels contain only wideband noise when the emitter is absent, while the semblance performs better than the average if more than half of the channels contain narrowband components when the emitter is absent. Therefore, the detection performance of both the average and the semblance can be improved.

An improved detection statistic is developed by combining the average and the semblance. A combining function is determined by satisfying a set of constraints which ensure that the average and the semblance contribute equally to the detection statistic. Before they are combined, the average is transformed to make its probability density function match the probability density function of the semblance. The receiver operating characteristic curves show that the combined statistic performs better than other statistics including the average and the semblance.

This new detection statistic is applied to the gravitational wave signal detection problem. A new algorithm which computes the Fourier transform magnitudes at the exact frequencies using the chirp z-transform is developed. Examples of the gravitational wave signal detection are presented to demonstrate that the new algorithm performs better than the existing algorithm.

Thesis Supervisor: Alan V. Oppenheim
Title: Professor, Department of Electrical Engineering and Computer Science

To Leslie

Acknowledgements

This thesis represents the conclusion of my long and circuitous career as a graduate student. This would not have been possible without the support of my friends and family. I would like to acknowledge their help.

I express my sincere gratitude to my advisor, Prof. Al Oppenheim, who arranged the financial support with unrestricted freedom and guided me through the challenges of the research. I am grateful to Al for his confidence in me and I thank him for all the metaphors. The best part of the doctoral program has been the privilege of working with him.

Prof. Rainer Weiss and Prof. Bruce Musicus served as thesis readers. Bruce suggested the multiple hypothesis testing idea and made many helpful comments. Prof. Weiss provided the gravitational wave data, hence the initial motivation for the multichannel detection problem. Dr. Jeffery Livas introduced me to the gravitational wave signal processing problem and generously answered many questions in the early stage of this research.

Deborah Gage made the Digital Signal Processing Group a nice place to work. Giovanni Aliberti maintained the computers in excellent shape and taught me about them.

Doug Mook encouraged me with his enthusiasm and interest in my work in addition to finding summer jobs for me at Sanders. Peter Doerschuk was always available to discuss many ideas and provide helpful suggestions. Rick Lamb frequently restored my balance by providing bizarre perspectives on life.

Michele Covell and Daniel Cobra, fellow graduate students who are also completing their theses, provided encouragement through their diligence to finish. I also thank the past and the present members of the Digital Signal Processing Group for making it a stimulating place to do research.

During the course of my study, I received generous financial support from the Advanced Research Projects Agency monitored by the Office of Naval Research, the National Science Foundation, Sanders Associates, Inc., and the Amoco Foundation.

My mother, Ja-Ok Kim Joo, and my brother, Tae-Young Joo, have provided unlimited, steadfast support throughout my study. I am grateful to them.

Finally, my wife, Leslie Reade Schenck, deserves my deepest appreciation. Leslie has supported and encouraged me with her unbounded love. Without her, it would not have been as fun or as meaningful. I dedicate this thesis to Leslie with love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

Multichannel signal detection using a receiver array arises in a variety of contexts such as the gravitational wave signal detection problem. Although a gravitational wave is a steady-state signal, it is measured in short bursts because of constraints imposed by the receiver hardware and local noise. The measurement is interpreted as multichannel data by treating each burst of measurement as the output of a channel of a multichannel receiver. When a gravitational wave is measured on earth, the received signal is frequency modulated due to the relative motion between the emitter and the receiver and is contaminated by local disturbances, some of which are narrowband noises. Local narrowband noises exist intermittently or have a constant frequency. If a gravitational wave emitter is present, the gravitational wave signal should exist in all channels of the receiver and the frequency of the signal should be modulated according to the emitter location and frequency. Therefore, the detector must use the existence of the signal with varying frequency in each channel to distinguish the gravitational wave signal from the false alarms caused by local narrowband noise sources.

The frequency variation of the signal over the channels is dependent on the particular detection problem. However, the requirement that the signal exists in all channels in order to decide that an emitter is present is common in many multichannel narrowband signal detection problems. Therefore, the detection problem is formulated as the following hypothesis testing

problem:

$$\begin{cases} H_1 \text{ (emitter is present)} : \underline{R} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \vdots \\ \underline{s}_L + \underline{w}_L \end{pmatrix} \\ H_0 \text{ (emitter is absent)} : \underline{R} \neq \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \vdots \\ \underline{s}_L + \underline{w}_L \end{pmatrix} \end{cases}$$

where $\underline{R}$ denotes the received data, $\underline{s}_l$ is the sinusoidal signal vector $A\cos(\omega_l n + \phi_l)$ for $n = 0, 1, \cdots, N-1$ in the lth channel, and $\underline{w}_l$ is the noise vector in the lth channel for $l = 1, 2, \cdots, L$. The amplitude of the sinusoidal signal, A, is the same for all channels and the frequency and the phase vary over the channels. For the H_0 hypothesis, the received data can also contain the narrowband component in some channels because of intermittent narrowband noise. Unfortunately, the exact number and identity of the channels which contain the narrowband component are unknown, therefore, the conventional likelihood ratio detector cannot be used.

This composite binary hypothesis testing problem can be formulated as a series of single channel binary hypothesis tests. However, because all channels are not considered collectively, the existence of a consistent signal in the channels is not explicitly checked. Thus, one purpose of this thesis is to develop improved algorithms for this detection problem and apply them to the gravitational wave signal detection problem.

1.2 Overview of the Thesis

If the received data under the H_0 hypothesis contain only wideband noise, the likelihood ratio detection statistic for the hypothesis testing problem is the average of the maximum value of the Fourier transform magnitude in each channel. However, when the received data contain narrowband noise, the average performs poorly even though it uses all channels collectively. Therefore, the semblance will be considered as the detection statistic to correct this shortcoming. Semblance is the ratio of the power of the sum of the spectra and the sum of the powers of the individual spectra and is related to the normalized cross-correlation. It is sensitive to the existence of the signal and is, therefore, well matched to the requirement that the signal must exist in all channels for the H_1 hypothesis. Because the semblance measures the

coherence of the signals, it employs fundamentally different criterion than the detectors which employ averaging. Unfortunately, noise alone can produce a high semblance value because the semblance is insensitive to the power of the signal. To take advantage of the sensitivity of the semblance to the correlation between channels and of the sensitivity of the average to the signal power, new detection statistics will be developed by combining the semblance and the average. The combined statistic will be applied to the measured gravitational wave data to demonstrate its effectiveness.

The gravity wave signal detection problem is representative of a large class of multichannel detection problems which have a strict requirement that all channels must contain the signal in order to correctly decide that an emitter is present. While this thesis focuses on the gravitational wave signal detection problem, the results on detection statistics are applicable to other multichannel signal detection problems.

The goal of the gravitational wave signal processing is to accurately and efficiently detect frequency-modulated, periodic gravitational wave signals with unknown emitter location and frequency. Because the emitter parameters are unknown, the gravitational wave detection algorithms first hypothesize a location and frequency, then detect the hypothesized emitter. The algorithm must test for all directions and all frequencies and is, therefore, computationally intensive. Fortunately, however, an algorithm can be developed which tests all possible frequencies simultaneously once the emitter location is selected. Because selecting the emitter location and frequency uniquely specifies the instantaneous frequency of the received signal, the periodic gravitational wave detection algorithm is implemented in two parts. First, the signal component of the hypothesized emitter in each channel is computed where the signal component is taken to be the Fourier transform magnitude at the frequencies predicted by the emitter location. Second, the magnitudes are used to compute the detection statistic which is a combination of the average and the semblance.

Contributions of the Thesis

First, the multichannel detection problem is formulated as a series of single channel binary hypothesis tests. The generalized likelihood ratio detector for each binary hypothesis test is efficiently implemented by the FFT. Because only the arithmetic computational error of the

FFT has been analyzed previously, a deterministic analysis of the FFT coefficient quantization effect on the FFT detector is developed.

Second, the average and the semblance are separately used as a multichannel detection statistic because they use data from all channels collectively. The semblance has been previously used as a detection statistic but its probability density function was unknown. A closed form expression of the probability density function of the semblance is developed for a simple case and a very accurate approximation is developed for the general case.

Third, a new detection statistic with improved detection performance is formed by combining the average and the semblance. A general method of combining two statistics which have different detection characteristics is developed. This method combines the statistics by satisfying a set of functional constraints. Improved detection probability of the combined statistic is demonstrated by comparing it with the average and the semblance.

Fourth, the new statistic, formed by combing the average and the semblance, is applied to the gravitational wave signal detection problem. Before the detection statistic can be applied, the Fourier transform magnitudes must be computed. A more accurate algorithm which uses the chirp z-transform to compute the Fourier transform magnitudes at the exact frequency locations is developed. This new algorithm and the combined statistic are used to process the measured gravitational wave data.

Outline of the Thesis

Chapter 2 begins with a formal statement of the multichannel signal detection problem. By enumerating all possible hypotheses, a detector based on a series of single channel binary hypothesis tests is developed. A solution to the binary hypothesis testing problem is reviewed. It is shown that the FFT provides an efficient implementation of the solution. The effect of quantized FFT coefficients on detection is examined using a deterministic error analysis of the quantized FFT coefficients. Chapter 2 ends with a discussion of the deficiencies of the detection based on a series of binary hypothesis tests.

Chapter 3 includes a discussion of the average and the semblance as detection statistics. The average is the likelihood ratio test solution to the multichannel hypothesis testing problem with the received data containing only wideband noise for the H_0 hypothesis. The receiver operating

characteristic computation is reviewed. Next, the semblance is introduced and its properties are reviewed. The probability density function of the semblance is analytically derived for a simple case and is approximated using the modified beta probability density function for the general case. To evaluate the effectiveness of the average and the semblance as detection statistics, their receiver operating characteristics are estimated using the Monte Carlo detection computation. A comparison of the average and the semblance as the detection statistics of the multichannel hypothesis testing problem is provided. Chapter 3 ends with a new multiple hypothesis testing formulation of the detection problem which suggests a detection statistic that is a combination of the average and the semblance.

Chapter 4 develops different methods of combining the average and the semblance because the detection performance of individual statistics can be improved. The statistics are combined by satisfying a set of intuitive functional constraints. The likelihood ratio combination can be derived for a simple case and is used to evaluate the previous method of combining statistics. For the general case, a combination using the discriminant function method is developed. Chapter 4 ends with a comparison of the detection statistics.

Chapter 5 provides a detailed discussion of gravitational wave signal processing. A description of the received signal and the existing detection algorithm are reviewed. A new algorithm which does not have the shortcomings of the existing algorithm is developed and is applied to the gravitational wave data. Chapter 5 ends with a comparison of the previous algorithm to the new algorithm to show that the new algorithm performs better.

Chapter 6 concludes the thesis by presenting a summary and suggestions for future research.

Chapter 2

Multichannel Detection Problem

In this chapter, a particular type of multichannel detection problem is discussed. Because a simple model of the received data when an emitter is absent is not available for the detection problem, a detector based on a series of single channel binary hypothesis tests is proposed. A single channel binary hypothesis test is efficiently implemented using the FFT. Because of the crucial role played by the FFT, the effect of inaccuracies in the FFT coefficients on the detector performance is analyzed. Lastly, the receiver operating characteristic of the detector based on a series of binary hypothesis tests is determined and the deficiencies of the detector are discussed.

2.1 Problem Statement

Many multichannel detection problems require that *the signal must exist in all channels* to declare that an emitter is present. The H_1 hypothesis denotes that an emitter is present. The gravitational wave signal detection problem is an important problem of this type and will be discussed in Chapter 5. In such hypothesis testing problems, the received data under the H_0 hypothesis, which corresponds to the case in which the emitter is absent, can contain the narrowband component in some of the channels. Therefore, an appropriate hypothesis testing problem to determine the presence or absence of the emitter is formulated with the hypotheses:

$$\begin{cases} H_1: & \text{narrowband component exists in all channels (i.e. emitter is present)} \\ H_0: & \text{narrowband component does not exist in all channels (i.e. emitter is absent).} \end{cases} \tag{2.1}$$

Let L be the total number of channels. Under the H_0 hypothesis, the received data can have the narrowband component in $0, 1, \cdots$, or $L-1$ channels. If the number and the identity of the channels containing the narrowband component are known, the detection algorithm would use only the channels which contain different data under the two hypotheses. Consequently, the problem simplifies to the familiar multichannel, signal-in-noise versus noise-only hypothesis testing problem[72]. Since the number and the identity of the channels which contain the narrowband component are assumed to be unknown, algorithms which make use of all channels must be developed. The next section includes a discussion of a detection procedure based on a series of single channel binary hypothesis tests.

2.2 Detection Based on a Series of Binary Hypothesis Tests

The hypothesis testing problem (2.1) is reformulated as a multiple hypothesis testing problem. In order to simplify the notation but to retain the pertinent properties of this method, the total number of channels, L, is set to 3. For this case, the multiple hypothesis testing formulation of the above problem has 8 hypotheses:

$$H_1: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \underline{s}_2 + \underline{w}_2 \\ \underline{s}_3 + \underline{w}_3 \end{pmatrix}$$

$$H_2: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{w}_1 \\ \underline{s}_2 + \underline{w}_2 \\ \underline{s}_3 + \underline{w}_3 \end{pmatrix}$$

$$H_3: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \underline{w}_2 \\ \underline{s}_3 + \underline{w}_3 \end{pmatrix}$$

$$H_4: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \underline{s}_2 + \underline{w}_2 \\ \underline{w}_3 \end{pmatrix}$$

$$H_5: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{w}_1 \\ \underline{w}_2 \\ \underline{s}_3 + \underline{w}_3 \end{pmatrix}$$

$$H_6: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{w}_1 \\ \underline{s}_2 + \underline{w}_2 \\ \underline{w}_3 \end{pmatrix}$$

$$H_7: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \underline{w}_2 \\ \underline{w}_3 \end{pmatrix}$$

$$H_8: \begin{pmatrix} \underline{r}_1 \\ \underline{r}_2 \\ \underline{r}_3 \end{pmatrix} = \begin{pmatrix} \underline{w}_1 \\ \underline{w}_2 \\ \underline{w}_3 \end{pmatrix}$$

where $\underline{w}_l$ is the lth channel noise vector and $\underline{s}_l$ is the lth channel narrowband component vector with nth sample $s_l(n) = A\cos(\omega_l n + \phi_l)$ for $n = 0, 1, \cdots, N-1$ and $l = 1, 2, 3$. The vector $\underline{r}_l$ denotes N samples $(r_l(0), r_l(1), \cdots, r_l(N-1))^T$ and $r_l(n)$ denotes the nth received sample in the lth channel. The parameters, A, ω_l, and ϕ_l, are unknown and the additive noise, $w_l(n)$, is white Gaussian noise as a function of channel, l, and sample, n, with zero mean and σ^2 variance. The hypotheses $H_2, H_3, \cdots, H_8$ collectively are equivalent to the H_0 hypothesis of (2.1).

The above multiple hypothesis testing problem is solved by minimizing the Bayesian risk $\mathcal{R}$ which is defined as

$$\mathcal{R} = \sum_{i=1}^{8}\sum_{j=1}^{8} P_j c_{ij} \text{ prob}(\textit{decide } H_i | H_j \textit{ is true})$$

where P_j is the *a priori* probability of the jth hypothesis H_j being true and c_{ij} is the cost of deciding H_i given H_j is true. When the risk $\mathcal{R}$ is minimized with the cost assignment $c_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j, \end{cases}$ the multiple hypothesis testing problem is solved by a series of binary tests of the form[72]

$$p_{R|H_j}(R|H_j) \cdot P_j \underset{H_i \text{ or } \gamma_{ij}}{\overset{H_j \text{ or } \gamma_{ij}}{\gtrless}} p_{R|H_i}(R|H_i) \cdot P_i$$

for $i = 1, 2, \cdots, 8$ and $j = i+1, \cdots, 8$ where R denotes the received data such that $R = (\underline{r}_1, \underline{r}_2, \underline{r}_3)^T$ and $p_{R|H_i}(R|H_i)$ is the conditional probability density of the received data R given that H_i is true and $\mathcal{H}_{ij}$ denotes all hypotheses except H_i and H_j. There are 28 binary tests to be performed to determine the minimum risk hypothesis for the $L = 3$ case.

In general, there are 2^L hypotheses and $\binom{2^L}{2} = 2^{L-1}(2^L - 1)$ binary tests which need to be performed for L channels. Even for a modest number of channels, the total number of binary tests which must be implemented becomes prohibitively large. Additionally, this multiple hypothesis testing method also has the following disadvantage. For many problems, the *a priori* probabilities are unknown. Unfortunately, the Neyman-Pearson test[72], which resolves this difficulty in binary hypothesis testing problems, must be extended because the concept of false alarm is not clearly defined for the multiple hypothesis testing problem.

If only the decision H_1 (i.e. narrowband component exists in all channels) versus not H_1 (i.e. narrowband component does not exist in all channels) is desired, then only the tests

$$p_{R|H_j}(R|H_j) \cdot P_j \underset{H_1 \text{ or } \mathcal{H}_{1j}}{\overset{H_j \text{ or } \mathcal{H}_{1j}}{\gtrless}} p_{R|H_1}(R|H_1) \cdot P_1$$

for $j = 2, \cdots, 8$ need to be performed. In general, the required number of binary hypothesis tests is reduced to $2^L - 1$.

In the following, the structures of the above hypotheses are exploited to further reduce the required number of binary hypothesis tests. When the hypothesis H_1 versus the hypothesis H_2 is tested, only the data from the first channel is used because the data in the second and third channels are identical for both H_1 and H_2. Therefore, the hypothesis testing H_1 versus H_2 reduces to testing the hypotheses:

$$\begin{cases} H_1: & A\cos(\omega_1 n + \phi_1) + w_1(n) \\ H_2: & w_1(n). \end{cases}$$

Similarly, only the data from the first channel are used for testing the hypothesis H_3 versus H_5, H_4 versus H_6, and H_7 versus H_8. Therefore, either $\{H_1, H_3, H_4, H_7\}$ or $\{H_2, H_5, H_6, H_8\}$ can be eliminated by testing the data from the first channel.

The above reasoning also applies to testing the data from the second channel and the third channel. The data from the second channel is tested for the existence of the narrowband component, $A\cos(\omega_2 n + \phi_2)$, to eliminate either $\{H_1, H_2, H_4, H_6\}$ or $\{H_3, H_5, H_7, H_8\}$. Similarly,

the data from the third channel is tested for the existence of the narrowband component, $A\cos(\omega_3 n + \phi_3)$, to eliminate either $\{H_1, H_2, H_3, H_5\}$ or $\{H_4, H_6, H_7, H_8\}$. These three tests are sufficient to determine the minimum risk hypothesis because L binary tests can distinguish 2^L possibilities and there are exactly 2^L hypotheses from which to choose. For the general L channel case, each channel is separately tested for the existence of $A\cos(\omega_l n + \phi_l)$, then the H_1 hypothesis is decided only if all channels contain the narrowband component. In Section 2.5, this solution will be shown to be inadequate. Next, the binary hypothesis testing method is discussed because it forms the basis of the multiple hypothesis testing detection solution.

2.3 Binary Hypothesis Testing

By minimizing the Bayesian risk, a solution to the simple binary hypothesis testing problem with the hypotheses H_a and H_b is given by the likelihood ratio test(LRT) $\Lambda(R)$[72,22]

$$\Lambda(R) = \frac{p_{R|H_a}(R|H_a)}{p_{R|H_b}(R|H_b)} \underset{H_b}{\overset{H_a}{\gtrless}} \frac{P_b}{P_a}\left(\frac{c_{ab} - c_{bb}}{c_{ba} - c_{aa}}\right) = \eta$$

where R denotes the received data. Because the costs, c_{aa}, c_{ab}, c_{ba}, and c_{bb}, and the prior probabilities, P_a and P_b, are unknown, the Neyman-Pearson test is used. The Neyman-Pearson test[38,76,72] maximizes the detection probability for a given false alarm probability and its solution is also the LRT. However, the threshold, η, is set by satisfying a specified probability of false alarm.

When the probability density functions(PDFs) contain unknown parameters, the above LRT is not applicable. This is, therefore, a composite hypothesis testing problem and two possible solutions exist. First, if the unknown parameter, denoted α, under the hypothesis H_a is random and its PDF, $p_{\alpha|H_a}(\alpha|H_a)$, is known, the LRT is modified to obtain the minimum risk solution as

$$\Lambda(R) = \frac{\int p_{R|H_a,\alpha}(R|H_a,\alpha)p_{\alpha|H_a}(\alpha|H_a)d\alpha}{p_{R|H_b}(R|H_b)} \underset{H_b}{\overset{H_a}{\gtrless}} \eta.$$

If the conditional PDF is unknown then the minimum risk solution is unknown. Second, a reasonable solution is given by the generalized likelihood ratio test(GLRT), however, it is suboptimal because the risk is not minimized. The GLRT solution is obtained by treating the

unknown α as a constant and replacing it by the maximum likelihood(ML) estimate as

$$\Lambda_g(R) = \frac{\max_\alpha p_{R|H_a,\alpha}(R|H_a,\alpha)}{p_{R|H_b}(R|H_b)} \underset{H_b}{\overset{H_a}{\gtrless}} \eta. \tag{2.2}$$

For the multichannel detection problem to be considered here, unknown parameters A , ω_l, and ϕ_l are fixed, hence the GLRT method is used.

2.3.1 Detection Statistic

The binary hypothesis testing problems, resulting from the multiple hypothesis testing formulation described in Section 2.2, have the form

$$\begin{cases} H_a: & r_l(n) = A\cos(\omega_l n + \phi_l) + w_l(n) \\ H_b: & r_l(n) = w_l(n) \end{cases} \tag{2.3}$$

for some fixed l where $r_l(n)$ is the received data. The sample number is $n = 0, 1, \cdots, N-1$ and $A, \omega_l,$ and ϕ_l are unknown parameters. The $w_l(n)$'s are assumed to be white Gaussian noise(WGN) as a function of l and n with zero mean and σ^2 variance. The hypothesis H_a denotes that the received data contain a sinusoid in additive noise and the hypothesis H_b denotes that the received data contain only noise. Because the parameters, A, ω_l, and ϕ_l, are unknown constants, they are replaced by the ML estimates and the detector is given by (2.2).

The ML estimates of A and the ML estimate of ϕ_l are determined next. Because $\omega_l(n)$ is WGN, the probability density function of the received data under the H_0 hypothesis is given by

$$p_{\underline{r}_l|H_a,A,\phi_l,\omega_l}(\underline{r}_l|H_a,A,\phi_l,\omega_l) = (\frac{1}{2\pi\sigma^2})^{\frac{N}{2}} \exp(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}|r_l(n) - A\cos(\omega_l n + \phi_l)|^2)$$

where $\underline{r}_l = (r_l(0), r_l(1), \cdots, r_l(N-1))^T$ is the received data. The ML estimates are computed by assuming that the H_a hypothesis is true and then maximizing the log of the above probability density function with respect to A and ϕ_l as

$$\begin{aligned} \frac{\partial \log p_{\underline{r}_l|H_a,A,\phi_l,\omega_l}(\underline{r}_l|H_a,A,\phi_l,\omega_l)}{\partial A} &= \frac{1}{\sigma^2}(\sum_{n=0}^{N-1} r_l(n)\cos(\omega_l n + \phi_l) - \frac{AN}{2}) \\ \frac{\partial \log p_{\underline{r}_l|H_a,A,\phi_l,\omega_l}(\underline{r}_l|H_a,A,\phi_l,\omega_l)}{\partial \phi_l} &= -\frac{A}{\sigma^2}\sum_{n=0}^{N-1} r_l(n)\sin(\omega_l n + \phi_l). \end{aligned}$$

Setting the above two equations to zero and solving for A and ϕ_l results in the following ML estimates

$$\begin{aligned}\hat{A} &= \frac{2}{N}|\sum_{n=0}^{N-1} r_l(n)e^{-j\omega_l n}| \\ \hat{\phi_l} &= \arg\{\sum_{n=0}^{N-1} r_l(n)e^{-j\omega_l n}\},\end{aligned}$$

and the GLRT becomes

$$\begin{aligned}\Lambda_g(\underline{r_l}) &= \frac{p_{\underline{r_l}|H_a,\hat{A},\hat{\phi_l},\omega_l}(\underline{r_l}|H_a,\hat{A},\hat{\phi_l},\omega_l)}{p_{\underline{r_l}|H_b}(\underline{r_l}|H_b)} \\ &= \exp(\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} 2\hat{A}\cos(\omega_l n+\hat{\phi_l})\underline{r_l}(n) - \hat{A}^2\cos^2(\omega_l n+\hat{\phi_l})).\end{aligned}$$

Using the intermediate result $\hat{A} = \frac{2}{N}\sum_{n=0}^{N-1}\underline{r_l}(n)\ \cos(\omega_l n+\hat{\phi_l})$ and the fact that $\sum_{n=0}^{N-1}\cos^2(\omega_l n+\hat{\phi_l}) = \frac{N}{2}$, the above GLRT simplifies to

$$\Lambda_g(\underline{r_l}) = \exp(\frac{1}{2\sigma^2}\frac{N}{2}\hat{A}^2).$$

Finally, by taking the logarithm of the above expression, the threshold test becomes

$$(\sum_{n=0}^{N-1} r_l(n)\cos\omega_l n)^2 + (\sum_{n=0}^{N-1} r_l(n)\sin\omega_l n)^2 \underset{H_b}{\overset{H_a}{\gtrless}} \eta_1. \tag{2.4}$$

In the above equation, the parameter ω_l is unknown, therefore, it is typically replaced by the ML estimate $\hat{\omega}_l$. The ML estimate of the frequency is determined by matched filtering[70,71] using all possible frequencies which is equivalent to computing the periodogram[50,27] and is efficiently implemented by the FFT. The ML estimate of the unknown frequency is determined by finding the location of the peak of the periodogram[51,57,33]. Therefore, the GLRT solution of the binary hypothesis testing problem (2.3) computes the periodogram and then compares its maximum value against a threshold.

2.3.2 Receiver Operating Characteristic

The PDF of the periodogram of the signal $r_l(n)$ is determined by computing the PDF of the magnitude of the discrete Fourier transform(DFT) and then transforming the PDF to

accommodate the squaring operation. If $r_l(n)$ contains WGN only, the DFT of the received data is

$$X_l(k) = \sum_{n=0}^{N-1} r_l(n) e^{-j\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} w_l(n) e^{-j\frac{2\pi}{N}kn}$$

and its mean is determined as

$$\mathbf{E}(X_l(k)) = \sum_{n=0}^{N-1} \mathbf{E}(w_l(n)) e^{-j\frac{2\pi}{N}kn} = 0.$$

The covariance is computed as

$$\mathbf{E}(X_l(k) X_l^*(m)) = \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \mathbf{E}(w_l(n_1) w_l^*(n_2)) e^{-j\frac{2\pi}{N}kn_1} e^{j\frac{2\pi}{N}mn_2} = N\sigma^2\delta(k-m)$$

where $\delta(.)$ denotes the delta function. The DFT of WGN with zero mean and σ^2 variance is WGN with zero mean and $N\sigma^2$ variance and its magnitude is Rayleigh distributed[52].

For the sinusoidal signal in additive WGN case, the mean of the DFT is determined as

$$\mathbf{E}(X_l(k)) = \sum_{n=0}^{N-1} \mathbf{E}(A\cos\omega_l n + w_l(n)) e^{-j\frac{2\pi}{N}kn} = \frac{AN}{2}\delta(k-k_0)$$

where $\omega_l = \frac{2\pi}{N}k_0$. The covariance is computed as

$$\begin{aligned}
&\mathbf{E}((X_l(k) - \mathbf{E}(X_l(k)))(X_l(m) - \mathbf{E}(X_l(m)))^*) \\
&= \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \mathbf{E}((A\cos\omega_l n_1 + w_l(n_1))(A\cos\omega_l n_2 + w_l(n_2))^*) e^{-j\frac{2\pi}{N}kn_1} e^{j\frac{2\pi}{N}mn_2} \\
&\quad -\mathbf{E}(X_l(k))\mathbf{E}(X_l(m)) \\
&= N\sigma^2\delta(k-m).
\end{aligned}$$

The DFT has zero mean for all DFT bins except at the k_0th bin where the mean is $\frac{AN}{2}$, and the variance is $N\sigma^2$. The PDF of the magnitude of the k_0th bin is Rician[56]. Therefore, if the DFT of $r_l(n) = A\cos(\omega_l n + \phi_l) + w_l(n)$ evaluated at k_0, where $w_l(n)$ is WGN with zero mean and σ^2 variance and $\omega_l = \frac{2\pi}{N}k_0$, is $X_l(k_0) = \frac{AN}{2} + X_R - jX_I$, then $\mathbf{E}(X_R) = \mathbf{E}(X_I) = 0$, $\mathbf{E}(|X_R|^2) = \mathbf{E}(|X_I|^2) = \frac{N\sigma^2}{2}$, and $\mathbf{E}(X_R\, X_I) = 0$. The random variables, X_R and X_I, are uncorrelated Gaussian densities with zero mean and $\frac{N\sigma^2}{2}$ variance.

$$p_{X_R,X_I}(X_R, X_I) = \frac{1}{\pi N\sigma^2} e^{-\frac{1}{N\sigma^2}((X_R - \frac{AN}{2})^2 + X_I^2)}$$

and let $y_l = |X_l(k)|$, then its PDF is

$$p_{y_l}(y_l) = \begin{cases} \frac{2y_l}{N\sigma^2} e^{-\frac{y_l^2}{N\sigma^2}} u(y_l) & \text{if } k \neq k_0 \text{ (Rayleigh)} \\ \frac{2y_l}{N\sigma^2} I_0(y_l \frac{A}{\sigma^2}) e^{-\frac{(\frac{AN}{2})^2 + y_l^2}{N\sigma^2}} u(y_l) & \text{if } k = k_0 \text{ (Rician)} \end{cases}$$

where $I_0(.)$ is the modified Bessel function of first kind, zeroth order and $u(.)$ is the unit step function. The periodogram, which is denoted by v_l, is the square of the magnitude, hence its PDF is given by $p_{v_l}(v_l) = \frac{p_{y_l}(\sqrt{v_l})}{2\sqrt{v_l}}$. The PDF of the periodogram at the k_0th bin is

$$p_{v_l}^{k_0}(v_l) = \frac{1}{N\sigma^2} I_0(\frac{A}{\sigma^2}\sqrt{v_l}) e^{-\frac{(\frac{AN}{2})^2 + v_l}{N\sigma^2}} u(v_l) \tag{2.5}$$

and the PDF for bins other than the k_0th bin is

$$p_{v_l}^{o}(v_l) = \frac{1}{N\sigma^2} e^{-\frac{v_l}{N\sigma^2}} u(v_l) \tag{2.6}$$

which is an exponential PDF.

The density functions, (2.5) and (2.6), become increasingly dissimilar as N increases and the probability of detection improves. The probability of false alarm is

$$P_F = \int_{\eta}^{\infty} \frac{1}{N\sigma^2} e^{-\frac{v_l}{N\sigma^2}} dv_l = e^{-\frac{\eta}{N\sigma^2}}. \tag{2.7}$$

For a constant false alarm rate, the threshold value is set to $\eta = -2N\sigma^2 \ln(P_F)$. The probability of detection is

$$P_D = \int_{\eta}^{\infty} \frac{1}{N\sigma^2} I_0(\frac{A}{\sigma^2}\sqrt{v_l}) e^{-\frac{v_l + \frac{A^2N^2}{4}}{N\sigma^2}} dv_l. \tag{2.8}$$

The above expression can be represented using Marcum's Q-function[40,72][1] as

$$P_D = \frac{1}{2} \int_{\frac{2\eta}{N\sigma^2}}^{\infty} e^{-\frac{t + \frac{A^2N}{2\sigma^2}}{2}} I_0(A\sqrt{\frac{N}{2\sigma^2}}\sqrt{t}) dt = Q_M(A\sqrt{\frac{N}{2\sigma^2}}, \sqrt{-2\ln(P_F)}) \tag{2.9}$$

Since the DFT effectively operates as a bank of matched filters in which the data length corresponds to the integration time, the probability of detection of the sinusoid increases with the transform length. In other words, when P_F is fixed, P_D increases as N increases. The FFT is used to efficiently compute the DFT. When the FFT is implemented with finite precision arithmetic, the above detection probability may degrade. In the next section, the effect of the finite precision FFT coefficients on the probability of detection is examined.

[1]Marcum's Q-function is defined as $Q_M(a,b) = \int_b^{\infty} te^{-\frac{t^2+a^2}{2}} I_0(at)dt = \frac{1}{2}\int_{b^2}^{\infty} e^{-\frac{t+a^2}{2}} I_0(a\sqrt{t})dt$.

2.4 FFT Coefficient Quantization Effects on Detection

In computing the FFT, errors due to arithmetic roundoff and coefficient quantization increase with increasing FFT length and degrade the detection performance. The effect of arithmetic roundoff on the FFT output has been analyzed and is well documented[74,50]. In the sinusoidal signal detection problem, the arithmetic roundoff can be modeled as additive white noise in the FFT output. On the other hand, the effect of the coefficient quantization on detection is less straightforward to analyze. A deterministic bound for the error of the FFT computed with the quantized coefficients is presented here. This bound is used in conjunction with the experimental measurements to obtain an empirical formula for the FFT output error. Finally, the degradation in the probability of detection due to the coefficient quantization is computed using the empirical formula[31].

2.4.1 Coefficient Quantization

The effects of coefficient quantization on the FFT output have been analyzed previously[37, 26,69,23,24], however, a close, analytical bound of the FFT output error has not been derived previously. In the following, the error bound of the decimation-in-time, radix-2 FFT algorithm is considered. The required FFT coefficients, $e^{-\frac{2\pi}{N}m}$, for $m = 0, 1, \cdots, \frac{N}{2} - 1$, can be either precomputed and stored in a table, or recursively computed at each stage of the FFT computation. In the recursive implementation, only $\log_2 N$ complex values must be stored to be used as the initial values of the recursion. For large N, this results in significant savings in storage. However, as will be shown, using a table of $\frac{N}{2}$ precomputed coefficients is more accurate than using the recursively computed coefficients.

In computing the FFT, the coefficients W_N^{kn} of the DFT

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} x(n) W_N^{kn} \tag{2.10}$$

are realized through combinations of coefficients associated with shorter FFT length. Specifically, in the decimation-in-time FFT algorithm, (2.10) is effectively replaced by

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kb_{0,n}} W_N^{kb_{1,n}2} \cdots W_N^{kb_{M-1,n}2^{M-1}} \tag{2.11}$$

where $M = \log_2 N$, $n = b_{0,n} + b_{1,n}2 + \cdots + b_{M-1,n}2^{M-1}$ and $b_{i,n} = 0$ or 1 for $i = 0, 1, \cdots, M-1$. Thus, $(b_{M-1,n}b_{M-2,n}\cdots b_{1,n}b_{0,n})$ is the binary representation of n . When the coefficients are quantized, (2.11) becomes

$$\hat{X}(k) = \sum_{n=0}^{N-1} x(n)(W_N^{kb_{0,n}} + \epsilon_{0,k}) \cdots (W_N^{kb_{M-1,n}2^{M-1}} + \epsilon_{M-1,k}) \tag{2.12}$$

where the difference between the true and the quantized coefficients is denoted by $\epsilon_{i,k}$. Because there is no quantization error in representing 1 or -1, $\epsilon_{M-1,k} = 0$

In analyzing the error, it is convenient to use matrix notation. The transformation (2.10) is expressed as $\underline{X} = F\underline{x}$ where $\underline{x} = (x(0), x(1), \cdots, x(N-1))^T$, $\underline{X} = (X(0), X(1), \cdots, X(N-1))^T$ and F is the $N \times N$ matrix of coefficients with knth element $F_{kn} = W_N^{kn}$. Correspondingly, (2.12) is written as $\underline{\hat{X}} = \hat{F}\underline{x}$ where $\hat{F}$ is formed by the quantized coefficients and the error vector is defined as $\underline{e} = \underline{\hat{X}} - \underline{X} = (\hat{F} - F)\underline{x}$. The maximum FFT output error over all frequency bins is used in determining the degradation in the probability of detection of a sinusoid. Therefore, the error measure is the infinity norm of $\underline{e}$

$$\|\underline{e}\|_\infty = \max_k |e(k)| \tag{2.13}$$

where $e(k) = \hat{X}(k) - X(k)$. To derive an error measure which is independent of the input $\underline{x}$, the following inequality[19,49] is used.

$$\|A\underline{v}\|_\infty \leq \|A\|_\infty \|\underline{v}\|_\infty$$

where A is a matrix, $\underline{v}$ is a vector, and the matrix norm is defined as $\|A\|_\infty = \max_k \sum_{n=0}^{N-1} |a_{kn}|$. Using the inequality,

$$\max_k |e(k)| \leq \|\hat{F} - F\|_\infty \|\underline{x}\|_\infty$$

with the (k,n)th element of the difference matrix $(\hat{F} - F)_{kn}$ as

$$(\hat{F} - F)_{kn} = \tag{2.14}$$
$$(W_N^{kb_{0,n}} + \epsilon_{0,k}) \cdots (W_N^{kb_{M-1,n}2^{M-1}} + \epsilon_{M-1,k}) - W_N^{kb_{0,n}} W_N^{kb_{1,n}2} \cdots W_N^{kb_{M-1,n}2^{M-1}}$$

Two different methods of coefficient quantization are considered next.

Precomputed Coefficient

An error bound of the FFT output where the coefficients are computed using a table of $\frac{N}{2}$ precomputed coefficients is derived. Each coefficient W_N^p is quantized such that $|\epsilon_{i,k}| \leq \sqrt{2}\Delta$ where Δ is $\frac{1}{2}$ of the quantization step size. Assume that $|\epsilon_{i,k}|$ is small enough such that second and higher order error terms in (2.14) can be ignored. The error in the coefficient matrix (2.14) becomes

$$\begin{aligned}(\hat{F} - F)_{kn} \approx\ & b_{0,n}\epsilon_{0,k}(W_N^{kb_{1,n}2}W_N^{kb_{2,n}4}\cdots W_N^{kb_{M-1,n}2^{M-1}}) \\ & + b_{1,n}\epsilon_{1,k}(W_N^{kb_{0,n}}W_N^{kb_{2,n}4}\cdots W_N^{kb_{M-1,n}2^{M-1}}) + \cdots \\ & + b_{M-1,n}\epsilon_{M-1,k}(W_N^{kb_{0,n}}\cdots W_N^{kb_{M-2,n}2^{M-2}}) \end{aligned} \tag{2.15}$$

Applying the triangle inequality to the above equation and using the fact that $|\epsilon_{i,k}| \leq \sqrt{2}\Delta$ and $\epsilon_{M-1,k} = 0$, an upper bound is obtained as

$$|(\hat{F} - F)_{kn}| \leq \sqrt{2}\Delta(b_{0,n} + b_{1,n} + \cdots + b_{M-2,n}).$$

Consequently, the matrix norm is

$$\|\hat{F} - F\|_\infty = \max_k \sum_{n=0}^{N-1} |(\hat{F} - F)_{kn}| \leq \sqrt{2}\Delta \sum_{n=0}^{N-1} b_{0,n} + \cdots + b_{M-2,n}.$$

As n ranges from 0 to $N-1$, $b_{i,n}$ for each i will be one for one half of the terms and zero for the remaining half of the terms. Consequently, since $M = \log_2 N$,

$$\|\hat{F} - F\|_\infty \leq \sqrt{2}\Delta\frac{N}{2}(\log_2 N - 1).$$

Therefore, the maximum magnitude output error (2.13) of the quantized-coefficient FFT using a table of $\frac{N}{2}$ values is bounded by

$$\max_k |e(k)| \leq \sqrt{2}\Delta\frac{N}{2}(\log_2 N - 1)\|\underline{x}\|_\infty. \tag{2.16}$$

Recursively Computed Coefficient

An error bound of the FFT output where the coefficients are computed recursively using $\log_2 N$ stored initial values is derived. At the ith stage, the initial value $W_N^{2^{M-i}}$ is used to generate all the required coefficients for that stage. However, because $W_N^{2^{M-i}}$ is quantized,

recursive computation results in the FFT output error. Assume that the quantization error for the initial value $W_N^{2^{M-i}}$ is $|\epsilon_{i,1}| \leq \sqrt{2}\Delta$. As the recursion is used to compute the next coefficient, the coefficient quantization error increases linearly. For example, assuming that $|\epsilon_p^2| \ll 1$ and using the triangular inequality, the error in computing W_N^{2p} is

$$|(W_N^p + \epsilon_p)(W_N^p + \epsilon_p) - W_N^p W_N^p| \approx |2\epsilon_p W_N^p| \leq 2\sqrt{2}\Delta.$$

If L terms of quantized coefficients are multiplied, the error is bounded approximately by $L\sqrt{2}\Delta$ because the quantization step size is assumed to be small enough to ignore non-linear error terms. Again, applying the triangular inequality to (2.15),

$$\|\hat{F} - F\|_\infty \leq \max_k \sum_{n=0}^{N-1} b_{0,n}|\epsilon_{0,k}| + \cdots + b_{M-1,n}|\epsilon_{M-1,k}|.$$

The error $|\epsilon_{i,k}|$ reaches its maximum when $k = \frac{N}{2} - 1$ such that $|\epsilon_{0,\frac{N}{2}-1}| \leq \sqrt{2}\Delta(\frac{N}{2} - 1), |\epsilon_{1,\frac{N}{2}-1}| \leq \sqrt{2}\Delta(\frac{N}{4} - 1)$, etc. Because $b_{i,n} = 1$ for only $\frac{N}{2}$ terms and $\log_2 N - 1$ terms are summed,

$$\|\hat{F} - F\|_\infty \leq \sqrt{2}\Delta\frac{N}{2}((\frac{N}{2} + \frac{N}{4} + \cdots + 1) - M).$$

Therefore, the maximum magnitude output error (2.13) of the quantized-coefficient FFT using a recursion is bounded by

$$\max_k |e(k)| \leq \sqrt{2}\Delta\frac{N}{2}(N - 1 - \log_2 N)\|\underline{x}\|_\infty. \tag{2.17}$$

When the table of coefficients is used, the error of the FFT output is proportional to $N \log_2 N$, as shown by (2.16). However, when the recursion is used, the error of the FFT output is proportional to N^2, as shown by (2.17). Therefore, although more storage is required, using the precomputed table of coefficients proves to be more accurate.

To verify and measure the closeness of the above bounds to the exact FFT output error, (2.16) is checked by computing the error $\underline{e}$ explicitly for a sinusoidal input

$$\underline{x} = (1, e^{j\omega_0}, \cdots, e^{j\omega_0(N-1)})^T.$$

The results of the simulations employing an 8 bit uniform quantizer is shown in Table 2.1. The values in the *measured* column are obtained by explicitly searching for the maximum error. The upper bounds predicted by (2.16) are listed in the *bound* column. These values indicate

$\log_2 N$	*measured*	*bound*	*predicted*
1	.0	.0	.086
2	.011	.011	.091
3	.022	.044	.107
4	.071	.132	.147
5	.201	.353	.249
6	.536	.883	.494
7	1.228	2.122	1.064
8	2.603	4.949	2.368
9	5.567	11.313	5.301
10	11.360	25.455	11.818
11	26.284	56.568	26.158
12	58.228	124.450	57.443

Table 2.1: The maximum magnitude error of the FFT output which is computed using a table of precomputed coefficients. The coefficients are quantized by the 8 bit uniform quantizer. The measured error is listed under the *measured* column. The analytical bound is listed under the *bound* column. The error predicted by the empirical formula is listed under the *predicted* column.

that the calculated bounds are approximately twice the measured values. This suggests that the bound can be scaled to predict the FFT output error. The functional form of the bound and the measured errors are used to derive an empirical formula for the FFT output error. A linear model of the form $\alpha\,\Delta\frac{N}{2}(\log_2 N - 1) + \beta$ is fit to measured data by choosing α and β to minimize the squared error. Thus

$$\alpha\,\Delta\frac{N}{2}(\log_2 N - 1) + \beta \approx \text{measured } \|\underline{e}\|_\infty \tag{2.18}$$

with least squares solution $\alpha = 0.65$ and $\beta = 0.08$. The values in the *predicted* column on Table 2.1 are computed using the above equation. There is an excellent agreement between the predicted and the measured values. To check the bound for very large N, the error for a 2^{16} length FFT is computed. Because the FFT length is rather large, the search for ω_0 was limited to a small frequency range. The search for the maximum resulted in $\|\underline{e}\|_\infty = 1356.65$. The upper bound given by (2.16) is 2715.3 and $\|\underline{e}\|_\infty$ is predicted to be 1251.45 by (2.18). Figure 2.1 plots the measured $\|\underline{e}\|_\infty$, the predicted (2.18), and the bound (2.16) for an 8 bit uniform quantizer. The figure demonstrates that the bound is reasonable and the empirical formula accurately predicts the error.

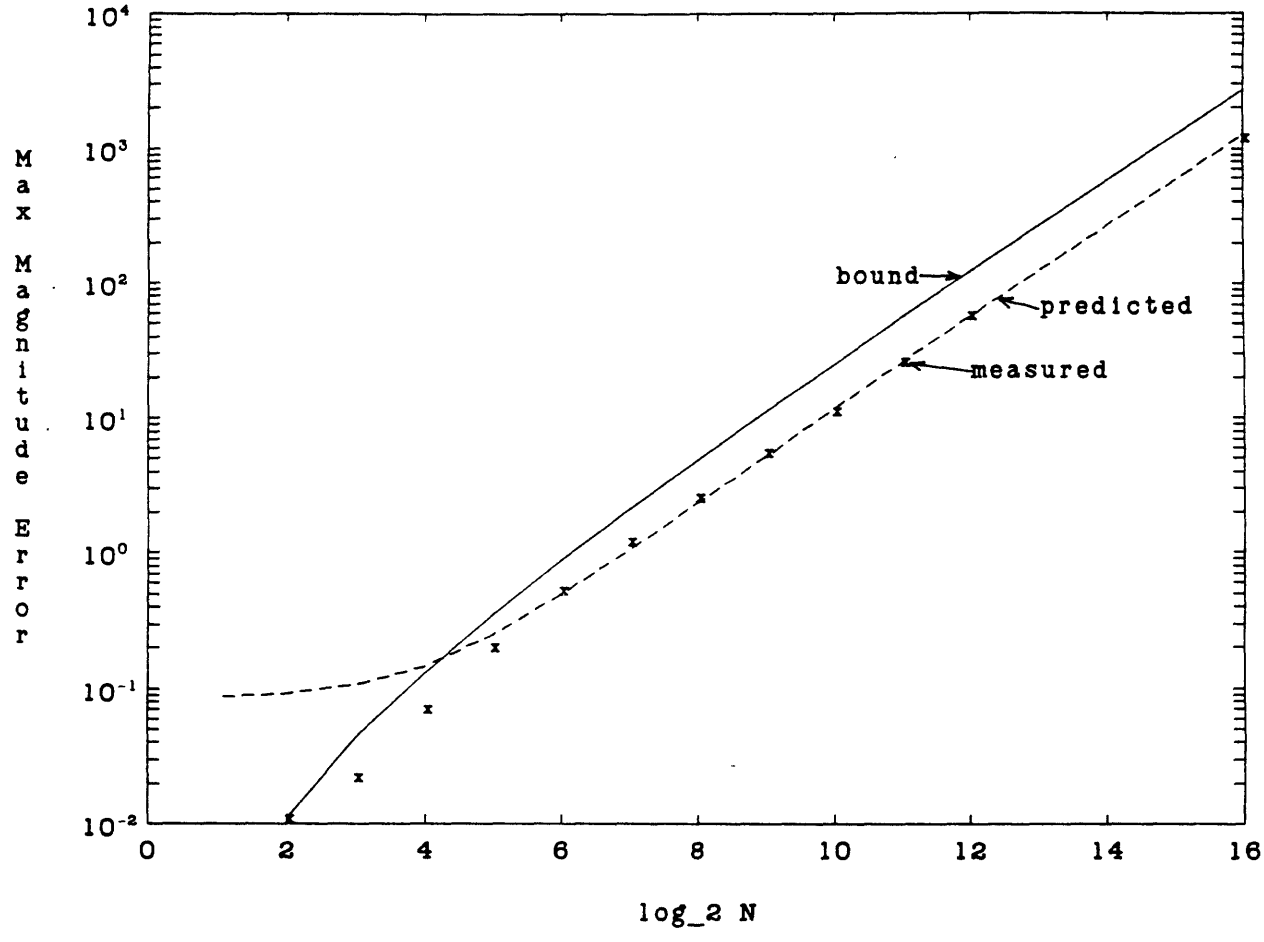

Figure 2.1: The maximum magnitude error of the FFT output which is computed using a table of precomputed coefficients. The coefficients are quantized by the 8 bit uniform quantizer. The measured error, denoted by 'x', the analytical bound, and the empirical formula are plotted.

2.4.2 Probability of Detection using Quantized Coefficients

The FFT output error due to coefficient quantization degrades the probability of detection. The probability of detection at the correct FFT bin is given by (2.9). Because the FFT implements a bank of matched filters, the definitions of both the probability of detection, denoted $\tilde{P}_D$, and the probability of false alarm, denoted $\tilde{P}_F$, are modified. The detection over all FFT bins is defined as a decision that a signal exists at the correct bin, $k = k_0$. The false alarm over all FFT bins is defined as a decision that a signal exists at an incorrect bin, $k \neq k_0$, where the frequency of the signal is $\omega_0 = \frac{2\pi}{N} k_0$.

The quantized-coefficient FFT for input $Ae^{j\omega_0 n} + w(n)$ is given by

$$\hat{X}(k) = \sum_{n=0}^{N-1} Ae^{j\omega_0 n} \widehat{W_N^{kn}} + \sum_{n=0}^{N-1} w(n) \widehat{W_N^{kn}}$$

where $\widehat{W_N^{kn}}$ denotes the quantized coefficients shown in (2.12). If the coefficients have no quantization error then $\sum_{n=0}^{N-1} Ae^{j\omega_0 n} W_N^{kn} = AN\delta(k - k_0)$. For $k \neq k_0$,

$$|\sum_{n=0}^{N-1} Ae^{\omega_0 n} \widehat{W_N^{kn}}| \leq A\|\underline{e}\|_\infty.$$

The equality is assumed for a conservative $\tilde{P}_D$ estimation. Therefore, the probability density function of $\hat{z}$, where $\hat{z} = |\hat{X}(k)|$, becomes

$$p(\hat{z}) = \frac{\hat{z}}{N\sigma^2} I_0(\frac{\hat{z}}{N\sigma^2} A\|\underline{e}\|_\infty) \exp(-\frac{\hat{z}^2 + A^2\|\underline{e}\|_\infty^2}{2N\sigma^2}) u(\hat{z})$$

hence the probability of false alarm over all FFT bins is

$$\tilde{P}_F = \int_\eta^\infty \frac{\hat{z}}{N\sigma^2} I_0(\frac{\hat{z}}{N\sigma^2} A\|\underline{e}\|_\infty) \exp(-\frac{\hat{z}^2 + A^2\|\underline{e}\|_\infty^2}{2N\sigma^2}) d\hat{z} = Q(\frac{A\|\underline{e}\|_\infty}{\sqrt{N}\sigma}, \frac{\eta}{\sqrt{N}\sigma}). \tag{2.19}$$

For $k = k_0$,

$$|\sum_{n=0}^{N-1} Ae^{j\omega_0 n} \widehat{W_N^{kn}}| \geq A(N - \|\underline{e}\|_\infty).$$

Again the equality is assumed for a conservative estimate of $\tilde{P}_D$. Therefore, the probability density function of $\hat{z}$ becomes

$$p(\hat{z}) = \frac{\hat{z}}{N\sigma^2} I_0(\frac{\hat{z}}{N\sigma^2} A(N - \|\underline{e}\|_\infty)) \exp(-\frac{\hat{z}^2 + A^2(N - \|\underline{e}\|_\infty)^2}{2N\sigma^2}) u(\hat{z})$$

hence the probability of detection over all FFT bins is

$$\begin{aligned}\tilde{P}_D &= \int_{\eta}^{\infty} \frac{\hat{z}}{N\sigma^2} I_0\left(\frac{\hat{z}}{N\sigma^2} A(N - \|\underline{e}\|_\infty)\right) \exp\left(-\frac{\hat{z}^2 + A^2(N - \|\underline{e}\|_\infty)^2}{2N\sigma^2}\right) d\hat{z} \\ &= Q\left(\frac{A(N - \|\underline{e}\|_\infty)}{\sqrt{N}\sigma}, \frac{\eta}{\sqrt{N}\sigma}\right). \end{aligned} \tag{2.20}$$

The empirical formula (2.18) is used for $\|\underline{e}\|_\infty$ in (2.19) and (2.20). Figure 2.2 shows $\tilde{P}_D$ as a function of FFT length. It is generated assuming that the amplitude of the sinusoid is 0.1 and the variance of the WGN is 1. The threshold is determined by satisfying the constant false alarm rate of $\tilde{P}_F = 0.01$. Even though the error, $\|\underline{e}\|_\infty$, increases as the data length increases, $\tilde{P}_D$ also improves. The simulation shows that 8 bit quantization only slightly degrades the probability of detection while 4 bit quantization noticeably degrades the probability of detection. This analysis indicates that it is important to quantize the coefficients using more than 4 bits. In this thesis, the FFTs are computed with coefficients that are quantized using 8 bits or more, so that the degradation in the probability of detection due to the inaccurate coefficients can be ignored.

2.5 Deficiency of Detection Based on a Series of Binary Hypothesis Tests

The receiver operating characteristic of the detector based on a series of single channel binary hypothesis tests as proposed in Section 2.2 is discussed in this section. Assume that the received data have the narrowband component in K of L channels. In this case, the correct output of the detector is the decision that the emitter is absent or, equivalently, that the narrowband component does not exist in all L channels. If the detector decides that the narrowband component exists in all channels, it is a false alarm. Because the proposed detector tests L channels separately, the probability of false alarm is determined by the product of K probabilities of correctly deciding that the channel contains the narrowband component when it does and $L - K$ probabilities of incorrectly deciding that the channel contains the narrowband component when it does not. Therefore, the probability of false alarm, denoted $P_{F,m}$, is given by

$$P_{F,m} = P_F^{L-K} \times P_D^K$$

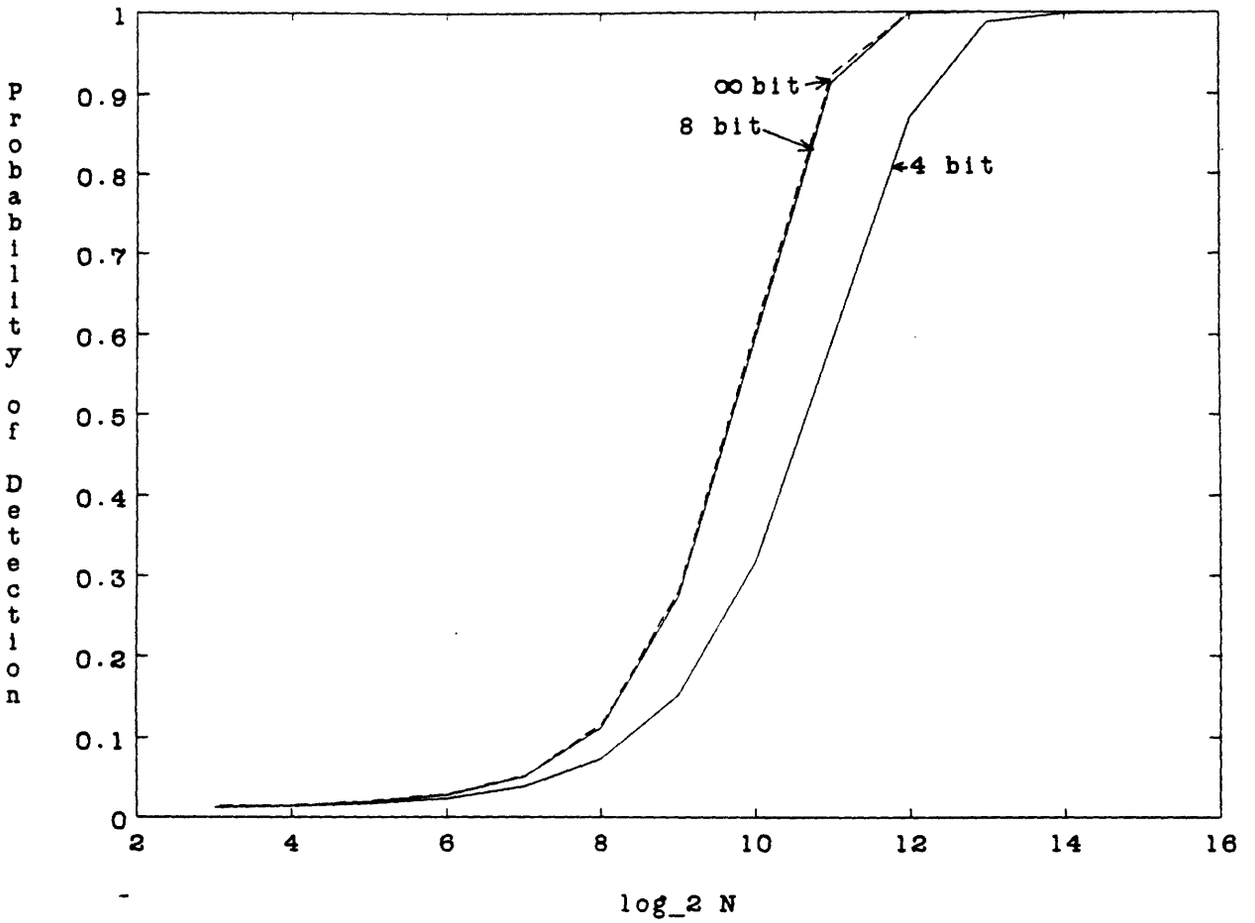

Figure 2.2: The probability of detection($\tilde{P}_D$) of the FFT which is computed with the precomputed table of quantized coefficients. The sinusoidal amplitude(A) is 0.1 and the variance of the WGN(σ^2) is 1. The false alarm is set to 0.01. $\tilde{P}_D$ for the ideal FFT coefficients, 8 bit uniformly-quantized coefficients, and 4 bit uniformly-quantized coefficients are plotted as a function of the data length($\log_2 N$).

where P_F and P_D are the probability of false alarm (2.7) and the probability of detection (2.8) of the single channel detection, respectively.

The probability of detection, denoted $P_{D,m}$, is given by

$$P_{D,m} = P_D^L.$$

Because $P_F < P_D$, $P_{F,m}$ increases as K increases and the detection performance degrades because $P_{D,m}$ is constant for all K. For the example, let $L = 20$, $P_D = 0.95$ and $P_F = 0.5$, then the probability of detection $P_{D,m}$ is 0.35 and the probability of false alarm is 1.8×10^{-6} for $K = 1$, 5.8×10^{-4} for $K = 10$, and 0.19 for $K = 19$. This confirms that there is a significant loss in the probability of detection as K increases.

The amplitude of the narrowband component can vary substantially without affecting the detection performance of the detector based on a series of binary hypothesis tests. Thus, this method is insensitive to the amplitude variation of the narrowband component and it cannot enforce the requirement that the amplitude A of $A\cos(\omega_l n + \phi_l)$ for $l = 1, 2, \cdots, L$ must be the same. Additionally, this detector requires many tests and the simple Neyman-Pearson threshold cannot be used. To remedy these problems, all channels are used simultaneously to derive detection statistics in the next chapter.

2.6 Summary

In this chapter, the multichannel detection problem was formulated as a series of single channel binary hypothesis testing problems. Each single channel binary hypothesis testing problem is solved by threshold testing the Fourier transform magnitude or, equivalently, the periodogram of the data. The periodograms are computed using the FFT and it was shown that the quantization of the FFT coefficients has a minimal effect on detection if 8 bits or more are used for the quantization. The receiver operating characteristic of the single channel binary hypothesis test was derived and was used to determine the performance of the detector based on a series of binary hypothesis tests. It was shown that the performance of the detector degrades noticeably when many channels contain the narrowband component under the H_0 hypothesis.

Chapter 3

Multichannel Detection Statistics

The performance of the detector based on a series of single channel binary hypothesis tests discussed in Chapter 2 can be improved upon because it does not use data from all channels collectively. In this chapter, both *average* and *semblance*, which use data from all channels collectively, are discussed as detection statistics for the multichannel detection problem. As is well known, the average is the generalized likelihood ratio test(GLRT) detection statistic for a particular multichannel detection problem. It will be shown that the average performs well if the received data fit the model of the detection problem. Deviations from the model result in degradation in detection performance. To accommodate the received data for which the average performs poorly, the semblance is introduced and its detection performance is analyzed. It will be shown that the semblance does not perform as well as the average in some important cases. Therefore, to maintain good performance over a wide range of received data, the average and the semblance can be combined to produce a new detection statistic. Combining the average and the semblance is discussed in Chapter 4. This chapter separately analyzes the average and the semblance.

3.1 Average

A multichannel hypothesis testing problem for which the average is the GLRT detection statistic is discussed in this section. Testing the average is identical to testing the sum because any monotonic transformation of a test statistic does not affect the ROC. For convenience, the

probability density function of the average is discussed using the sum in this section and the detection performance of the average will be examined in Section 3.4.

3.1.1 Multichannel Binary Hypothesis Testing Problem

In Section 2.2 the multichannel detection problem was analyzed as a multiple hypothesis testing problem. In this section, a special case of the multichannel detection problem for which the average is the likelihood ratio detection statistic is examined. The multichannel detection problem (2.1) is analyzed by assuming that all channels contain only wideband noise under the H_0 hypothesis and all channels contain the narrowband component in additive noise under the H_1 hypothesis. The received data model for the composite hypothesis testing, then, becomes

$$\begin{cases} H_1: \ \underline{R} = \begin{pmatrix} \underline{r}_1 \\ \vdots \\ \underline{r}_L \end{pmatrix} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \vdots \\ \underline{s}_L + \underline{w}_L \end{pmatrix} \\ H_0: \ \underline{R} = \begin{pmatrix} \underline{r}_1 \\ \vdots \\ \underline{r}_L \end{pmatrix} = \begin{pmatrix} \underline{w}_1 \\ \vdots \\ \underline{w}_L \end{pmatrix} \end{cases} \tag{3.1}$$

where $\underline{R}$ denotes the received data, L is the number of channels, and N is the number of samples. As in Section 2.2, the nth sample of the lth channel narrowband component vector, $\underline{s}_l$, is $s_l(n) = A\cos(\omega_l n + \phi_l)$ where $A, \omega_l,$ and ϕ_l are unknown parameters. The nth sample of the lth channel noise vector, $\underline{w}_l$, is $w_l(n)$ and the $w_l(n)$'s are assumed to be white Gaussian noise as a function of l and n with zero mean and σ^2 variance. The above received data model for the H_0 hypothesis does not contain the narrowband component, therefore, it is simpler than the model employed in Section 2.2.

The multichannel detection problem (3.1) with unknown parameters is solved using the GLRT by replacing the unknowns with the ML estimates. The GLRT solution for this detection problem is given by a vector form of (2.2) as

$$\Lambda_g(\underline{R}) = \frac{\max_{A,\underline{\phi},\underline{\omega}} p_{\underline{R}|H_1,A,\underline{\phi},\underline{\omega}}(\underline{R}|H_1, A, \underline{\phi}, \underline{\omega})}{p_{\underline{R}|H_0}(\underline{R}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \tag{3.2}$$

where the received data is $\underline{R} = (\underline{r}_1, \underline{r}_2, \cdots, \underline{r}_L)^T$ with $\underline{r}_l = (r_l(0), r_l(1), \cdots, r_l(N-1))^T$, $\underline{\omega} = (\omega_1, \omega_2, \cdots, \omega_L)^T$, and $\underline{\phi} = (\phi_1, \phi_2, \cdots, \phi_L)^T$. The threshold value η_0 is set to match a specified

constant false alarm rate. The ML estimates of the unknowns are computed by assuming that H_1 is correct and by maximizing the likelihood over the desired parameters.

The probability density function of the received data under the H_1 hypothesis is

$$p_{\underline{R}|H_1,A,\underline{\phi},\underline{\omega}}(\underline{R}|H_1,A,\underline{\phi},\underline{\omega}) = (\frac{1}{2\pi\sigma^2})^{\frac{NL}{2}} \exp(-\frac{1}{2\sigma^2}\sum_{l=1}^{L}\sum_{n=0}^{N-1}|r_l(n)-s_l(n)|^2).$$

To maximize $\log p_{\underline{R}|H_1,A,\underline{\phi},\underline{\omega}}(\underline{R}|H_1,A,\underline{\phi},\underline{\omega})$ with respect to A and $\underline{\phi}$, the following equations are solved for $\hat{A}$ and $\hat{\phi}_l$.

$$\begin{aligned}\hat{A} &= \frac{2}{LN}\sum_{l=1}^{L}\sum_{n=0}^{N-1} r_l(n)\cos(\omega_l n+\hat{\phi}_l)\\ 0 &= \sum_{l=1}^{L}\sum_{n=0}^{N-1} r_l(n)\sin(\omega_l n+\hat{\phi}_l)\end{aligned}$$

The ML estimates are

$$\begin{aligned}\hat{A} &= \frac{2}{LN}\sum_{l=1}^{L}|\sum_{n=0}^{N-1} r_l(n)e^{-j\omega_l n}|\\ \hat{\phi}_l &= \arg\{\sum_{n=0}^{N-1} r_l(n)e^{-j\omega_l n}\},\end{aligned}$$

and the GLRT (3.2) becomes

$$\begin{aligned}\Lambda_g(\underline{R}) &= \exp(\frac{1}{2\sigma^2}\sum_{l=1}^{L}\sum_{n=0}^{N-1} 2\hat{A}\cos(\omega_l n+\hat{\phi}_l)r_l(n)-\hat{A}^2\cos^2(\omega_l n+\hat{\phi}_l))\\ &= \exp(\frac{1}{2\sigma^2}\frac{LN}{2}\hat{A}^2)\end{aligned}$$

because $\hat{A}=\frac{2}{LN}\sum_{l=1}^{L}\sum_{n=0}^{N-1} r_l(n)\ \cos(\omega_l n+\hat{\phi}_l)$ and $\sum_{n=0}^{N-1}\cos^2(\omega_l n+\hat{\phi}_l)=\frac{N}{2}$. Finally, the threshold test, (3.2), simplifies to

$$\sum_{l=1}^{L}|\sum_{n=0}^{N-1} r_l(n)e^{-j\omega_l n}| \underset{H_0}{\overset{H_1}{\gtrless}} \eta_1$$

for some constant η_1 where the squaring operation is removed without affecting the detection performance. The above test is a generalization of the single channel test (2.4). In the above equation, as in Section 2.3.1, the ω_l's are typically replaced by the ML estimates determined by the location of the peak of the periodograms[51,57,33]. Therefore, the GLRT solution of the multichannel detection problem is a summation of the peak values of the Fourier transform magnitudes which are compared against a threshold. Of course, the average can be used instead of the sum because they are related by a monotonic transformation.

3.1.2 Approximation of Probability Density Functions

In the previous section, it was shown that when the received data under the H_0 hypothesis contain only wideband noise, the GLRT statistic is the average of the peak values of the Fourier transform magnitudes. In this section, the performance of the average as a detection statistic is discussed. The PDF of the magnitude of the Fourier transform, denoted y_l, was derived in Section 2.3.2 and is repeated here for convenience:

$$p_{y_l}(y_l) = \begin{cases} \frac{2y_l}{N\sigma^2} e^{-\frac{y_l^2}{N\sigma^2}} u(y_l) & \text{if WGN only } (H_0) \\ \frac{2y_l}{N\sigma^2} I_0(y_l \frac{A}{\sigma^2}) e^{-\frac{(\frac{AN}{2})^2 + y_l^2}{N\sigma^2}} u(y_l) & \text{if narrowband component+WGN } (H_1). \end{cases}$$

The PDF of the sum $M = \sum_{l=1}^{L} y_l$ is derived. The PDF of the average, denoted $p_a(a)$, is simply related to the PDF of the sum, denoted $p_M(M)$, by $p_a(a) = L\, p_M(La)$.

The sum of the Rayleigh densities, which corresponds to the WGN only case, is discussed first. When $L = 2$, the sum is $M = y_1 + y_2$ and because y_1 and y_2 are independent Rayleigh densities,

$$\begin{aligned} p_M(M|H_0) &= \int_0^M p_{y_1}(t) p_{y_2}(M-t)dt \\ &= e^{-\frac{M^2}{N\sigma^2}} \left(\frac{M}{N\sigma^2} + \sqrt{\frac{\pi}{2N\sigma^2}} \left(\frac{M^2}{N\sigma^2} - 1\right) \text{erf}\left(M\sqrt{\frac{2}{N\sigma^2}}\right)\right) \end{aligned}$$

where $\text{erf}(x) = \frac{1}{2\pi} \int_0^x e^{-\frac{t^2}{2}} dt$. For $L > 2$ case, the PDF of M involves the convolution of the above expression with the Rayleigh density, hence a closed form expression is not known.

Alternatively, a closed form expression of the PDF of M can be attempted using the characteristic function of y_l. Because the y_l's are statistically independent, the characteristic function of M is determined by the product of the individual characteristic functions. However, the characteristic function of y_l also involves erf(.) function[40] because the characteristic function of the normalized Rayleigh PDF $p(x) = x \exp(-\frac{x^2}{2}) u(x)$ is given by $1 - j\sqrt{\frac{\pi}{2}}\omega \exp(-\frac{\omega^2}{2})\ \text{erf}(\frac{j\omega}{\sqrt{2}})$. Therefore, a closed form is unobtainable and the PDF must be approximated.

For the H_1 hypothesis, the PDF computation of M involves the Rician densities and $p_{M|H_1}(M|H_1)$ also cannot be expressed in a closed form. An approximation method based on the Gram-Charlier series is presented next.

The Gram-Charlier series uses the Gaussian density to expand the desired PDF $p(x)$ as[40,

22,10]

$$p(x) = \frac{1}{\sigma}\sum_{k=0}^{\infty} c_k \psi_k(\frac{x-\bar{x}}{\sigma})$$

where $\bar{x}$ is the mean and σ^2 is the variance of the random variable x. The basis functions are $\psi_k(x) = \frac{d^k\psi(x)}{dx^k}$ with $\psi_0(x) = \psi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Therefore,

$$\psi_k(x) = \frac{(-1)^k}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} H_k(x)$$

where $H_k(x)$ is the Hermite function[15]. For example, the first few terms are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and $H_3(x) = x^3 - 3x$.

The coefficients are computed as

$$c_k = \frac{(-1)^k}{k!}\int_{-\infty}^{\infty} p(x) H_k(\frac{x-\bar{x}}{\sigma})dx$$

and the first few terms are $c_0 = 1$, $c_1 = 0$, $c_2 = 0$, and $c_3 = -\frac{1}{3!}m_3$ with the moment $m_k = \frac{1}{\sigma^k}\int_{-\infty}^{\infty}(x-\bar{x})^k p(x)dx$. A rough approximation of the PDF is obtained by the following truncated Gram-Charlier series.

$$p(x) \approx \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\bar{x})^2}{2\sigma^2})(1 + \frac{m_3}{6}((x-\bar{x})^3 - 3(x-\bar{x}))).$$

The Gram-Charlier series of the sum of the Rayleigh densities is obtained using[52]

$$\begin{aligned}
\overline{M}_0 &= L\,\mathbf{E}(y_l|H_0) = \frac{L}{2}\sqrt{\pi N\sigma^2} \\
\sigma_0^2 &= L\,\mathbf{E}((y_l - \mathbf{E}(y_l))^2|H_0) = L(2 - \frac{\pi}{2})\frac{N\sigma^2}{2}
\end{aligned}$$

because $M = \sum_{l=1}^{L} y_l$ and the y_l's are independent. The PDF is approximated using the first K terms as

$$p_{M|H_0}(M|H_0) = \frac{1}{\sigma_0}\sum_{k=0}^{K-1} c_{0,k}\psi_k(\frac{M-\overline{M}_0}{\sigma_0})$$

and the probability of false alarm is approximated as

$$P_F = \int_{\eta}^{\infty} p_{M|H_0}(M|H_0)dM = \frac{1}{\sigma_0}\sum_{k=0}^{K-1} c_{0,k}\int_{\eta}^{\infty}\psi_k(\frac{M-\overline{M}_0}{\sigma_0})dM.$$

The Gram-Charlier series of the sum of the Rician densities is obtained using[56]

$$\begin{aligned}
\overline{M}_1 &= L\,\mathbf{E}(y_l|H_1) = \frac{L}{2}\sqrt{\pi N\sigma^2}\exp(-\frac{A^2N}{8\sigma^2})((1 + \frac{A^2N}{4\sigma^2})I_0(\frac{A^2N}{8\sigma^2}) + \frac{A^2N}{8\sigma^2}I_1(\frac{A^2N}{8\sigma^2})) \\
\sigma_1^2 &= L\,\mathbf{E}((y_l - \mathbf{E}(y_l))^2|H_1) = L(\frac{A^2N^2}{4} + N\sigma^2 - (\frac{\overline{M}_1}{L})^2).
\end{aligned}$$

The PDF is approximated using the first K terms as

$$p_{M|H_1}(M|H_1) = \frac{1}{\sigma_1} \sum_{k=0}^{K-1} c_{1,k} \psi_k(\frac{M - \overline{M}_1}{\sigma_1})$$

and the probability of detection is approximated as

$$P_D = \int_\eta^\infty p_{M|H_1}(M|H_1) dM = \frac{1}{\sigma_1} \sum_{k=0}^{K-1} c_{1,k} \int_\eta^\infty \psi_k(\frac{M - \overline{M}_1}{\sigma_1}) dM.$$

To obtain an accurate approximation, a large number of terms are often required. Therefore, in Section 3.3, the ROCs are computed using the more general method of Monte Carlo simulations. However, before the ROC of the average is examined, the semblance as an alternative detection statistic is discussed.

3.2 Semblance

The approach which leads to the average assumes that conditional PDFs under each hypothesis, $p_{\underline{R}|H_1}(\underline{R}|H_1)$ and $p_{\underline{R}|H_0}(\underline{R}|H_0)$, are available. However, for many problems, neither the exact PDFs nor functional forms of the PDFs are known. In such cases, the unknown densities can be approximated by proposing a functional form and then estimating its parameters using the measurements. Unfortunately, the detection statistic obtained using approximated PDFs often performs poorly.

Nonparametric or distribution free statistics are derived without assuming a specific PDF for the received data. These statistics are insensitive to the errors caused by the inaccuracies of the approximated PDFs or the model used for the received data[5,35,18]. Of course, the cost of the nonparametric statistics' relatively good performance over a broad range of data is that the parametric statistics perform better than the nonparametric statistics when the data model is accurate. Any function of the measurements can be used as a detection statistic and any detection statistic devised without assuming the PDFs can be viewed as a nonparametric statistic. A particular nonparametric statistic must be selected by comparing the ROCs of all candidate statistics, however, the selection is frequently made based upon the ease of the ROC computation.

In this section, the semblance is introduced. Instead of computing the average of the Fourier transform magnitudes, the semblance of the Fourier transform magnitudes is computed as the

detection statistic. The semblance can be viewed as a nonparametric statistic for the detection problem.

3.2.1 Background

In reflection seismology, multiple seismic traces are averaged, or "stacked" in geophysical parlance, to improve the signal to noise ratio[43,47]. Because the traces are made with varying source-receiver configurations, they must be aligned before performing the averaging operation. The alignment, called the moveout correction, is dependent on the earth's structure and propagation characteristics. A simple and frequently used model is the horizontally layered medium model which is completely specified by the thickness of the layers and the wave propagation velocity of the layers. For a given source-receiver separation distance and arrival time of the seismic reflections, the alignment function is described by a hyperbolic equation called the moveout equation[59]. The parameters of the hyperbolic equation are the normal incident time and the velocity, called the stacking velocity. The goal is to estimate the velocity for each incident time which is then used to align the traces for averaging. For a selected incident time, the stacking velocity is estimated by computing some measure of coherence between the seismic traces for a range of possible values and then selecting the velocity value which has the maximum coherence. The semblance was introduced to be used as a coherence measure in the stacking velocity estimation problem[48,65]. It has been used in other geophysical signal processing problems[36] because it is a coherence measure with useful properties.

3.2.2 Properties

Semblance, which is denoted by s, is defined as[48,65]

$$s = \frac{\sum_{n=-M}^{M} \left(\sum_{l=1}^{L} x_l(n) \right)^2}{L \sum_{n=-M}^{M} \sum_{l=1}^{L} x_l(n)^2} \tag{3.3}$$

where L is the number of channels, $2M+1$ is the semblance gate length and $x_l(n)$ denotes the nth sample of lth channel data. Important characteristics of the semblance are its insensitivity to the overall gain and its sensitivity to the amplitude variation between the channels. This sensitivity to the amplitude variation suggests that the semblance might be an effective statistic for the multichannel detection problem since the narrowband component with the same amplitude must

exist in all channels in order to decide that an emitter is present. Before using the semblance for the detection problem (2.1), pertinent properties of the semblance are listed in this section. The proofs of these properties are presented in Appendix 3.A.

Property 3.2.1 $0 \leq s \leq 1$.

Property 3.2.2 *If* $x_l(n) \geq 0$ *for* $l = 1, \cdots, L$ *and* $n = -M, \cdots, M$ *then* $s \geq \frac{1}{L}$.

Property 3.2.3 *If the semblance of* $\{x_l(n) : l = 1, \cdots, L; n = -M, \cdots, M\}$ *is* s_x *then the semblance of* $\{cx_l(n) : l = 1, \cdots, L; n = -M, \cdots, M\}$ *for some constant* c *is also* s_x.

Property 3.2.4 *If* $x_l(n) = x(n)$ *for* $l = 1, \cdots, L$ *then* $s = 1$.

Property 3.2.5 *Let* $x_l(n) = c_l y(n)$ *for* $l = 1, \cdots, L$ *and* $n = -M, \cdots, M$ *where* c_l*'s are constants. The semblance is maximized and has the value of* 1 *if and only if* $c_l = c$ *for all* l.

Property 3.2.6 *If* $x_l(n) = s(n) + w_l(n)$ *where* $s(n)$ *is the signal and* $w_l(n)$ *is zero mean noise, then as the number of channels increases* $s \approx$ *signal power/(signal power + noise power).*

Property 3.2.5 states that the semblance is sensitive to the channel to channel amplitude variation and it is, therefore, utilized for the detection problem.

3.2.3 Derivation of Probability Density Functions

A closed form expression of the PDF of the semblance for arbitrary L and M is not known. The PDF has been approximated by the F-statistic[11] by assuming that the individual $x_l(n)$s have a Gaussian distribution. The $x_l(n)$'s considered in this thesis violate this assumption. They have either Rayleigh or Rician distribution because $x_l(n)$ is the magnitude of the Fourier transform. In this section, the PDF of the semblance for a simple case is derived. The simple case to be considered uses $L = 2$ and $M = 0$, hence the semblance (3.3) simplifies to

$$s = \frac{(x_1(0) + x_2(0))^2}{2(x_1(0)^2 + x_2(0)^2)} \tag{3.4}$$

To simplify the notation, $x_1(0)$ and $x_2(0)$ are denoted by x_1 and x_2, respectively. The PDF of s can be derived by first computing the cumulative distribution function, which is determined by

integrating the joint PDF $p_{x_1,x_2}(x_1,x_2)$ over appropriate regions, then differentiating it. This method is conceptually straightforward but implementationally tedious. A simpler derivation uses the polar form of x_1 and x_2 so that the semblance, (3.4), becomes

$$s = \frac{1}{2} + \frac{1}{2}\sin(2\theta) \tag{3.5}$$

where $\theta = \tan^{-1}(\frac{x_2}{x_1})$ because $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$. To find the PDF of θ, which is denoted $p_\theta(\theta)$, the joint PDF, $p_{r,\theta}(r,\theta)$, is computed first, then the unwanted variable, r, is integrated out. Because the Jacobian of the polar transformation is $\frac{1}{r}$, the joint PDF is

$$p_{r,\theta}(r,\theta) = r\, p_{x_1,x_2}(r\cos\theta, r\sin\theta).$$

The random variables, x_1 and x_2, are independent and identically distributed. Because the PDF of x_1 is Rayleigh density $p_{x_1}(x_1) = \frac{x_1}{a^2}\exp(-\frac{x_1^2}{2a^2})$ for some constant a and $x_1 \geq 0$,

$$p_{r,\theta}(r,\theta) = \frac{r^3}{2a^4}\sin(2\theta)e^{-\frac{r^2}{2a^2}}$$

for $r \geq 0$ and $0 \leq \theta \leq \frac{\pi}{2}$ because $r\cos\theta \geq 0$ and $r\sin\theta \geq 0$. The PDF of θ is obtained by computing the marginal density

$$p_\theta(\theta) = \int_{r=0}^{\infty} p_{r,\theta}(r,\theta)dr = \sin(2\theta).$$

To compute the PDF of s, given by (3.5), the PDF of $\phi = 2\theta$ is computed first by

$$p_\phi(\phi) = \frac{1}{2}\, p_\theta(\frac{\phi}{2}) = \frac{\sin\phi}{2}$$

for $0 \leq \phi \leq \pi$. The PDF of $\tilde{s} = \sin\phi$ is determined as [52, p.99]

$$p_{\tilde{s}}(\tilde{s}) = \frac{1}{\sqrt{1-\tilde{s}^2}}(p_\phi(\sin^{-1}(\tilde{s})) + p_\phi(\pi - \sin^{-1}(\tilde{s}))) = \frac{\tilde{s}}{\sqrt{1-\tilde{s}^2}}$$

for $0 \leq \tilde{s} \leq 1$. Finally, the PDF of $s = \frac{1}{2} + \frac{1}{2}\tilde{s}$ is

$$p_s(s) = 2p_{\tilde{s}}(2s-1) = \frac{2s-1}{\sqrt{s(1-s)}} \tag{3.6}$$

for $\frac{1}{2} \leq s \leq 1$. Figure 3.1 shows the analytically derived PDF (3.6) compared with the histogram which is known to provide an unbiased and consistent estimate of the PDF[66,53,44,17,12]. The histogram is made using 1000 samples and 20 bins. The length of the data is 1024 samples and the variance of WGN is 2. As expected, the analytically derived PDF fits the histogram closely. The figure also shows the modified beta PDF approximation, which will be described in the next section.

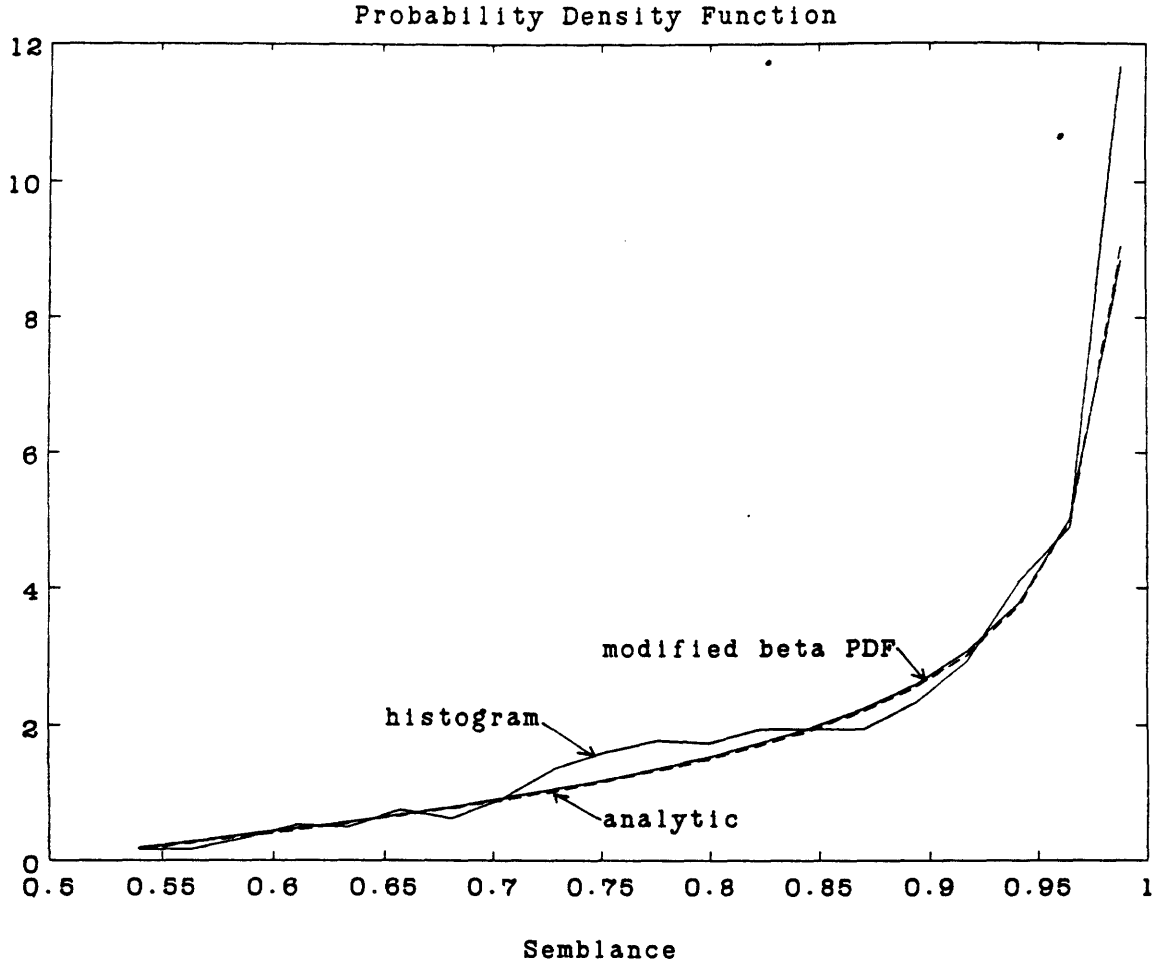

Figure 3.1: The plot of the analytically derived PDF of the semblance, a histogram made using 1000 samples and 20 bins, and the modified beta PDF fit to the samples. There are 2 channels(L) of 1024 samples(N) and both channels contain only WGN with variance(σ^2) 2.

3.2.4 Approximation of Probability Density Functions using the Beta Probability Density

The PDF of the semblance cannot be derived analytically for the general case. However, it can be accurately approximated by the beta probability density function as will be shown in this section. The approximation using the beta PDF is motivated by the fact that the beta PDF can match the finite range of the semblance exactly and its parameters can be adjusted to provide a close fit.

In statistics, the beta PDF has been frequently used to approximate PDFs with a finite range because many different finite distributions can be obtained by adjusting the parameters. The beta PDF is related to χ^2 density[29], gamma density[13,4,76], F-density[4,76], student's t-density[4,76], and *order statistics*[6,63,29]. In addition, when the samples have the unit Gaussian distribution, Murthy[46] states that the semblance has the beta distribution.

The standard beta PDF is defined for $0 \leq s \leq 1$ as[29,45,8,76]

$$p(s) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} s^{a-1}(1-s)^{b-1}$$

where the gamma function is $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ and a and b are free parameters. For the case where $x_l(n) \geq 0$ for $l = 1, \cdots, L$ and $n = -M, \cdots, M$, the lower bound of the semblance is $1/L$ by Property 3.2.2. Therefore, in this thesis, the beta random variable will be transformed to match this range exactly. Let x be the beta random variable and define a new random variable y such that $y = \frac{L-1}{L}x + \frac{1}{L}$. Because $0 \leq x \leq 1$, the new random variable has the desired range $\frac{1}{L} \leq y \leq 1$. The PDF of y, which will be referred to as the modified beta PDF, is defined as

$$p_\beta(y) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{L}{L-1}\right)^{a+b-1} (y-\frac{1}{L})^{a-1}(1-y)^{b-1} & \frac{1}{L} \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

Next, the parameters a and b will be determined by the ML estimation method using N independent samples. Given N independent samples of semblance, $S = (s_1, s_2, \cdots, s_N)^T$, the ML estimates of a and b are obtained by solving the likelihood equations:

$$\left.\frac{\partial \log p_\beta(S|a,b)}{\partial a}\right|_{a=\hat{a}} = 0 \quad \text{and} \quad \left.\frac{\partial \log p_\beta(S|a,b)}{\partial b}\right|_{b=\hat{b}} = 0$$

where

$$\log(p_\beta(S|a,b)) = N\log\frac{L}{L-1} + N(\log(\Gamma(a+b)) - \log(\Gamma(a)) - \log(\Gamma(b))) + (a-1)\sum_{i=1}^{N}\log\left(\frac{Ly_i - 1}{L-1}\right) + (b-1)\sum_{i=1}^{N}\log\left(\frac{L}{L-1}(1-y_i)\right).$$

Therefore, the likelihood equations become two coupled nonlinear equations which cannot be directly solved. The ML estimates are determined iteratively using the fact that the ML estimate maximizes the log likelihood function $\log(p_\beta(S|a,b))$. In this section, this multidimensional nonlinear optimization problem is solved by an iterative method called the downhill simplex method[55] because the derivatives of the log likelihood function are unavailable. The downhill simplex method performs well for this optimization problem because good initial estimates of a and b are obtained using the moments. The mean and the variance of the modified beta density (3.7) are

$$\text{mean}(y) = m_y = \frac{L-1}{L}\frac{a}{a+b} + \frac{1}{L}$$
$$\text{var}(y) = v_y = \left(\frac{L-1}{L}\right)^2 \frac{ab}{(a+b)^2(a+b+1)}.$$

These equations are used to express a and b in terms of the mean and the variance which are used as the initial values

$$\hat{b}_0 = \frac{(1-m')((1-m')m' - v')}{v'}$$
$$\hat{a}_0 = \hat{b}_0\frac{m'}{1-m'}$$

where $m' = \frac{L}{L-1}(m_y - \frac{1}{L})$ and $v' = \left(\frac{L}{L-1}\right)^2 v_y$. In the following, the mean, m_y, and the variance, v_y, will be replaced by the sample estimates and the iteration converges in a few steps.

Figure 3.2 shows the histogram of the semblance and the modified beta PDF approximation for a particular set of experiments. The histogram is made using 1000 semblance values and 20 bins. The semblances are generated using 8 channels of 1024 length data and a gate size of 5 samples. The amplitude of the narrowband component is 3.5. Figure 3.2(a) is computed using the semblance values generated with WGN variance of 2000 and the narrowband component in 4 out of 8 channels. Figure 3.2(b) is computed using the semblance values generated with WGN variance of 4000 and the narrowband component in all 8 channels. Figure 3.2(c) is

computed using the semblance values generated with WGN variance of 2000 and the narrowband component in all 8 channels. This figure strongly suggests that the modified beta PDF provides an excellent fit to the semblance PDF.

The previous discussion assumes that the modified beta density is an acceptable choice and Fig. 3.2 indicates that it is indeed a reasonable assumption. This assumption is tested by a goodness-of-fit test of the composite hypotheses:

$$\begin{cases} H_0: & \text{PDF of } s \text{ is } p_\beta(s|\hat{a},\hat{b}) \\ H_1: & \text{PDF of } s \text{ is not } p_\beta(s|\hat{a},\hat{b}). \end{cases}$$

An often used test statistic is[8,5]

$$T = \sum_{k=1}^{K} \frac{(N_k - n_k)^2}{n_k} \tag{3.8}$$

where N is the total number of samples, K is the number of bins, N_k is the number of samples in the kth bin, and n_k is the expected number of samples in the kth bin which is defined as

$$n_k = N \int_{x \in k\text{th bin}} p_\beta(x|\hat{a},\hat{b})dx.$$

The number of bins and the sizes of the bins are determined arbitrarily. However, K should be large and the bin sizes should be determined to have all $n_k's$ approximately the same. For this problem, K is specified to be 2^c for some constant c and the bins are determined such that

$$n_i \approx \frac{1}{K} \int_{s_{min}}^{s_{max}} p_\beta(x)dx$$

for all $i = 1, 2, \cdots, K$ and where s_{max} and s_{min} denote the maximum and the minimum sample values, respectively. The cumulative distribution function of the modified beta PDF is

$$B_\beta(x) = \int_{\frac{1}{L}}^{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{L}{L-1}\right)^{a+b-1} (y - \frac{1}{L})^{a-1}(1-y)^{b-1}dy.$$

By using the change of variables $t = \frac{L}{L-1}(y - \frac{1}{L})$, the above expression becomes

$$B_\beta(x) = \int_0^{\frac{L}{L-1}(x-\frac{1}{L})} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}dt = I_{\frac{L}{L-1}(x-\frac{1}{L})}(a,b)$$

where the incomplete beta function, $I_x(a,b)$, is defined as

$$I_x(a,b) = \int_0^x \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}dt.$$

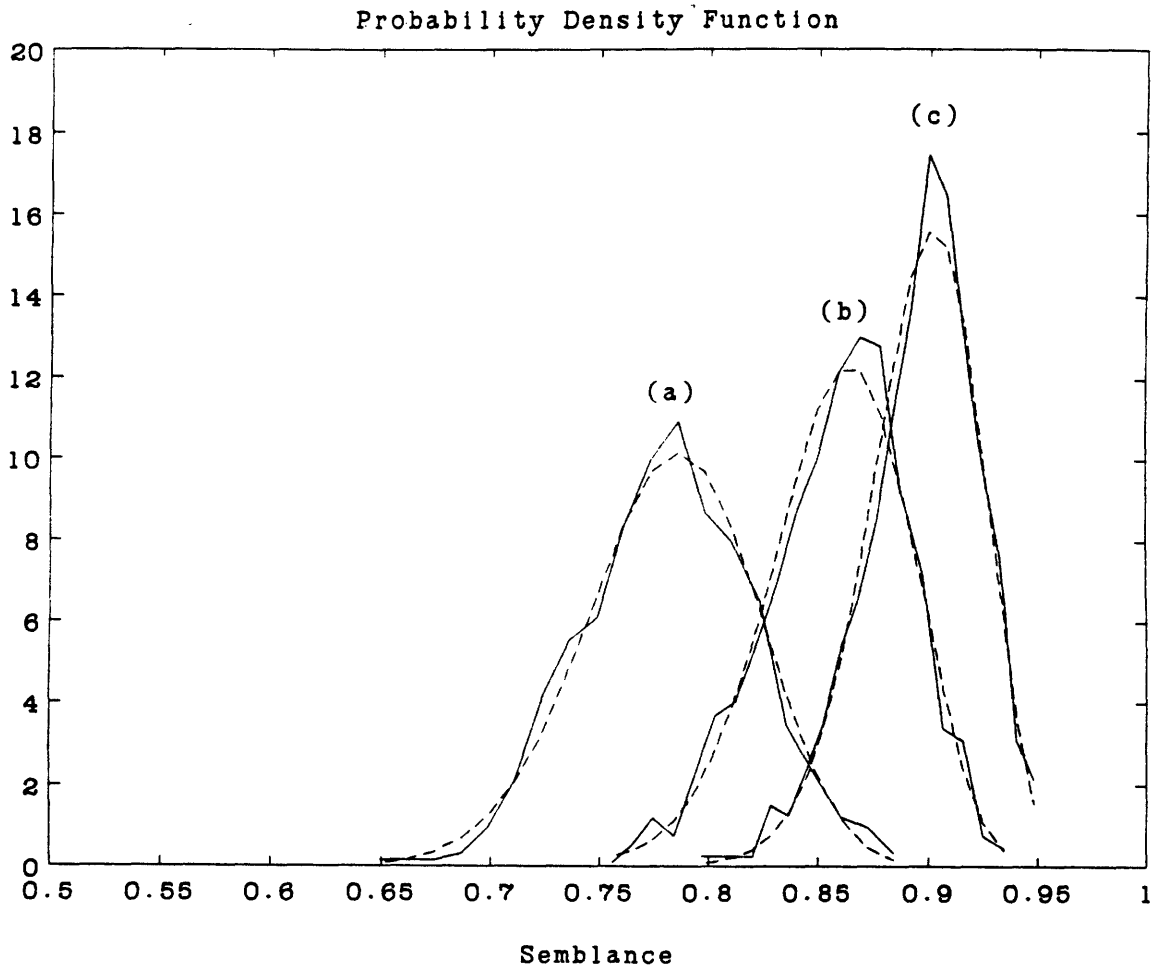

Figure 3.2: A histogram of the samples plotted in the dashed line and the modified beta PDF with the ML estimate parameters plotted in the solid line. The histogram is made with 1000 samples and 20 bins. The semblances are generated using 8 channels(L) of 1024 length data(N) and a gate size of 5 samples. The amplitude(A) of the narrowband component is 3.5; (a) curves are computed using the semblance values which are generated with WGN variance(σ^2) of 2000 and the narrowband component in 4 out of 8 channels; (b) curves are computed using the semblance values which are generated with WGN variance(σ^2) of 4000 and the narrowband component in all 8 channels; (c) curves are computed using the semblance values which are generated with WGN variance(σ^2) of 2000 and the narrowband component in all 8 channels.

Therefore, $n_i \approx \frac{1}{K}(B_\beta(s_{max}) - B_\beta(s_{min}))$. Using $\{s_1, s_2, \cdots, s_N\}$ samples, K bins are determined recursively as follows. First, x_1 is determined to satisfy

$$B_\beta(x_1) = \frac{(B_\beta(s_{max}) + B_\beta(s_{min}))}{2}.$$

In this section, the above nonlinear equation will be solved by finding the root of the equation

$$g(x) = B_\beta(x) - \frac{(B_\beta(s_{max}) + B_\beta(s_{min}))}{2}$$

by the bisection method using $s_{min} < x_1 < s_{max}$. After x_1 is determined, x_2 and x_3 are determined, separately, to satisfy

$$B_\beta(x_2) = \frac{(B_\beta(x_1) + B_\beta(s_{max}))}{2} \text{ and } B_\beta(x_3) = \frac{(B_\beta(s_{min}) + B_\beta(x_1))}{2}.$$

This process continues until all x's are determined, then the boundary locations of the bins are specified by the x's.

A reason for the popularity of the T statistic (3.8) is that under the H_0 hypothesis, T has χ^2 distribution with $K-3$ degrees of freedom, denoted $p_{\chi^2_{K-3}}(x)$, when N is reasonably large. If the probability of incorrectly rejecting the "modified beta PDF is correct" hypothesis, i.e. $prob(declare H_1 | H_0 \, is \, true)$, is specified to some value α, the threshold, γ, is determined by

$$\alpha = \int_\gamma^\infty p_{\chi^2_{K-3}}(x)dx.$$

The hypothesis that the samples are beta distributed is accepted if $T < \gamma$. For $K = 16$, $\alpha = 0.1$ requires $\gamma = 19.81$ and $\alpha = 0.01$ requires $\gamma = 27.69$. Table 3.1 shows the mean and the standard deviation of the test statistic, T, obtained using 100 trials with 1000 semblance samples(N), and 16 bins(K), for the χ^2 fit. The semblance values are generated by setting the gate size to 5 samples and varying the number of channels and the signal to WGN ratio. The values in the table confirm that the modified beta density provides an acceptable fit to the semblance values.

3.3 Numerical Evaluation of Detection Statistics

The probability of detection, denoted P_D, and the probability of false alarm, denoted P_F, are computed by

$$P_D = \int_\eta^\infty p_{z|H_1}(z|H_1)dz \text{ and } P_F = \int_\eta^\infty p_{z|H_0}(z|H_0)dz$$

		T	
channels	SNR	mean	std. dev.
4	0.1	13.87	4.84
4	12.5	16.36	5.87
8	0.1	14.12	5.54
8	12.5	13.37	5.63

Table 3.1: The mean and standard deviation of T of the modified beta PDF fit using 100 trials with $N = 1000$ and $K = 16$. If the probability of declaring that "the PDF of the semblance is not modified beta PDF" given that the PDF is truly beta PDF is set to 0.1 then the hypothesis that the samples are from the beta PDF is accepted when $T < 19.81$. If the probability is set to 0.01, then the hypothesis is accepted when $T < 27.69$.

where z denotes any detection statistic and η is the threshold. The ROC is specified by the (P_F, P_D) pairs. However, the analytical evaluation of the probability of detection and the probability of false alarm is often intractable because the conditional PDFs of the detection statistic, $p_{R|H_1}(R|H_1)$ and $p_{R|H_0}(R|H_0)$, are usually too complicated or are unknown. There are three ways of handling this difficulty. In the first method, the unknown PDFs can be approximated using the histogram, for example, then the approximate PDFs are integrated numerically. In the second method, the integrals involved in computing P_D and P_F are estimated using the Monte Carlo integration. In the third method, the probability of detection and the probability of false alarm are computed using Monte Carlo detection computations. When a large number of random samples are used, the second and the third methods become approximately the same. In computing the ROC of the average and the semblance, the Monte Carlo detection computation method is used. Because they are closely related, the Monte Carlo integration and the Monte Carlo detection computation methods are discussed next.

3.3.1 Monte Carlo Integration for Receiver Operating Characteristic

To compute P_F or P_D, an integral of the form

$$\alpha = \int_{\eta}^{\infty} p_z(z)dz \tag{3.9}$$

must be computed. For the probability of detection, $p_z(z) = p_{z|H_1}(z|H_1)$, and for the probability of false alarm, $p_z(z) = p_{z|H_0}(z|H_0)$. The above integral will be approximated using the Monte

Carlo simulations as[21,20,62]

$$\hat{\alpha} = \frac{M_\eta}{M} \tag{3.10}$$

where M_η is the number of samples from the PDF $p_z(z)$ whose value is greater than or equal to η and M is the total number of samples used.

To compute the bias and the variance of the above estimate, (3.10) is interpreted in the following way. The integral (3.9) can be written as

$$\alpha = \int_{-\infty}^{\infty} g(z) p_z(z) dz \tag{3.11}$$

with $g(z) = \begin{cases} 1 & \text{if } z \geq \eta \\ 0 & \text{if } z < \eta \end{cases}$. Using M independent samples $\{z_1, z_2, \cdots, z_M\}$ from the PDF, $p_z(z)$, the Monte Carlo estimate of (3.11) is given by

$$\hat{\alpha} = \frac{1}{M} \sum_{i=1}^{M} g(z_i) = \frac{M_\eta}{M} \tag{3.12}$$

which is (3.10). The bias is defined to be $\alpha - \mathbf{E}\{\hat{\alpha}\}$ and from (3.12)

$$\mathbf{E}\{\hat{\alpha}\} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{E}\{g(z_i)\} = \mathbf{E}\{g(z)\} = \int_{-\infty}^{\infty} g(z) p_z(z) dz = \alpha \tag{3.13}$$

where $\mathbf{E}\{.\}$ denotes expectation operation. Therefore, the Monte Carlo estimate is unbiased. Using the independence of the samples, the variance of the estimate is computed as

$$Var\{\hat{\alpha}\} = \frac{1}{M} Var\{g(z)\} = \frac{\alpha(1-\alpha)}{M}. \tag{3.14}$$

Therefore, the Monte Carlo estimate is consistent and the standard deviation of the estimate decreases as $\frac{1}{\sqrt{M}}$.[1] The convergence rate can be approximated using the Chebychev inequality or the central limit theorem[32].

The Monte Carlo estimation, given by (3.10), is the ML estimate given M observations. Specifically, it will be shown next that given M observations, the α estimated by (3.10) is the ML estimate of the true α defined in (3.9). Given that $P_F = \alpha$, the probability that M_η of M samples are greater or equal to η is determined to be

$$prob(M_\eta \text{ of } M \ z_i \geq \eta | P_F = \alpha) = \binom{M}{M_\eta} \alpha^{M_\eta} (1-\alpha)^{M-M_\eta}.$$

[1] There are several techniques of reducing the variance of the Monte Carlo estimate[60,3]. However, they require closed form expressions of the PDFs, hence they are not applicable.

To compute the ML estimate of α, the derivative of the log of the above expression with respect to α is taken to obtain

$$\frac{M_\eta}{\alpha} - (M - M_\eta)\frac{1}{1-\alpha} = 0.$$

Therefore, the ML estimate of α is $\frac{M_\eta}{M}$. The mean (3.13) and the variance (3.14) are easily verified using the moments of the binomial distribution as

$$\begin{aligned} \mathbf{E}\{\hat{\alpha}\} &= \sum_{k=0}^{M} \frac{k}{M}\binom{M}{k}\alpha^k(1-\alpha)^{M-k} = \alpha \\ Var\{\hat{\alpha}\} &= \sum_{k=0}^{M}(\frac{k}{M} - \alpha)^2\binom{M}{k}\alpha^k(1-\alpha)^{M-k} = \frac{\alpha(1-\alpha)}{M}. \end{aligned}$$

3.3.2 Monte Carlo Detection Computation

The previously described Monte Carlo integration method estimates the ROCs by approximating the integrals of the conditional densities (3.9) for each threshold value, η. This method computes the threshold, η', for a probability of false alarm, then a sample, z', is hypothesis tested by $z' \underset{H_0}{\overset{H_1}{\gtrless}} \eta'$. In this section, an alternative method which is referred to as the Monte Carlo detection computation method is discussed. A reference set containing M samples, $\{z_1, z_2, \cdots, z_M\}$, which are consistent with the H_0 hypothesis is computed. When a sample, z', is to be hypothesis tested, a new set $\{z_1, z_2, \cdots, z_M; z'\}$ is formed by appending the test sample to the reference set. Using a specified false alarm probability, which is denoted by $P_{F,0}$, a cutoff sample number is computed as

$$K = M \times P_{F,0}. \tag{3.15}$$

The hypothesis testing becomes

$$\begin{cases} \text{declare } H_1: & \text{if } z' \text{ is among } K \text{ largest of the new set} \\ \text{declare } H_0: & \text{if } z' \text{ is not among } K \text{ largest of the new set.} \end{cases}$$

This Monte Carlo detection computation is Barnard's Monte Carlo test[58,2,42,25,28] specialized for the ROC computation. This method and the method described in the previous section become identical as the number of samples increases. Because the size of the required number of random samples has not been analyzed non-asymptotically, an analysis of the required number of random samples is presented next.

The false alarm probability obtained by the above Monte Carlo detector, denoted $P_{F,MC}$, is determined using results from order statistics[6,34,76].

$$\begin{aligned} P_{F,MC} &= prob\{0 \text{ sample from } \{z_1, z_2, \cdots, z_M\} \text{ has value } \geq z'\} \\ &\quad + prob\{1 \text{ sample from } \{z_1, z_2, \cdots, z_M\} \text{ has value } \geq z'\} \\ &\quad + \cdots + prob\{K-1 \text{ samples from } \{z_1, z_2, \cdots, z_M\} \text{ has value } \geq z'\} \\ &= \sum_{r=0}^{K-1} \binom{M}{r} f^r (1-f)^{M-r} \end{aligned}$$

where $f = \int_{z'}^{\infty} p_{z|H_0}(z|H_0)dz$. The above equation is concisely expressed as

$$P_{F,MC} = 1 - I_f(K, M-K+1)$$

using the incomplete beta function[1][2]

$$I_f(K, M-K+1) = \sum_{r=K}^{M} \binom{M}{r} f^r (1-f)^{M-r}.$$

Unfortunately, the true false alarm probability, f, cannot be computed for a given z' because the conditional density, $p_{z|H_0}(z|H_0)$, is unknown. However, the above false alarm probability, $P_{F,MC}$, can be set to a desired probability of false alarm, $P_{F,0}$, and the corresponding true probability of false alarm can, then, be computed by solving

$$P_{F,0} = 1 - I_f(M \cdot P_{F,0},\ M(1-P_{F,0})+1)$$

for f using different values of M. Table 3.2 shows the true probability of false alarm, f, as a function of the number of samples, M, when the probability of false alarm for this Monte Carlo detector, $P_{F,MC}$, is set to the desired value. As expected, a smaller $P_{F,MC}$ requires more samples to make the value of the true probability of false alarm closer to the desired probability of false alarm. For $P_{F,MC} = 0.01$, 1000 samples of the Monte Carlo simulation provide an adequate estimate.

3.4 Examples of Receiver Operating Characteristic

The PDFs of the average and the semblance have been approximated. Because numerical evaluation of the required integrals is difficult, the ROCs of the statistics are evaluated

[2]The incomplete beta function is defined as $I_x(a,b) = \int_0^x \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}dt$.

$P_{F,MC} = 0.1$		$P_{F,MC} = 0.01$		$P_{F,MC} = 0.001$	
M	f	M	f	M	f
100	0.1383	100	0.0450	100	0.0329
200	0.1272	200	0.0327	200	0.0202
500	0.1172	500	0.0230	500	0.0107
1000	0.1122	1000	0.0186	1000	0.0068
2000	0.1086	2000	0.0158	2000	0.0046
5000	0.1054	5000	0.0135	5000	0.0029
10000	0.1038	10000	0.0124	10000	0.0018

Table 3.2: The true probability of false alarm(f) as a function of the number of samples(M) when the Monte Carlo detector's probability of false alarm($P_{F,MC}$) is set to the desired value($P_{F,0}$).

numerically using the Monte Carlo detection computation in this section.

The ROCs of the average and the semblance used as the detection statistics for various hypothesis testing problems are computed using the method described in Section 3.3.2. The ROCs are computed using 1000 random samples of H_1 data and 1000 samples of H_0 data. To estimate the probability of detection at some probability of false alarm value P_F, a cutoff sample number K of (3.15) is chosen to be $P_F \times 1000$. Each sample of the H_1 data set is checked to determine if it is larger than the Kth largest sample from the H_0 data set. Let M denote the number of samples of the H_1 data set which are larger than the Kth largest sample of the H_0 data set, then the estimated value of the probability of detection is $M/1000$. This process is repeated for different values of P_F to estimate the ROC curve.

Figure 3.3 shows the ROCs for the detection problem where the narrowband component exists in all channels under the H_1 hypothesis and the narrowband component exists in only some of the channels under the H_0 hypothesis. For this example, the received data contain a total of 8 channels and the length of the received data, N, is 1024 samples. The amplitude of the narrowband component, A, is 3.5 and the variance of WGN, σ^2, is 2000. Data for the H_1 hypothesis are generated by computing the magnitude of the complex Gaussian random variable. The real part of the random variable has $\frac{AN}{2}$ mean and $\frac{N\sigma^2}{2}$ variance and the imaginary part of the random variable has zero mean and $\frac{N\sigma^2}{2}$ variance. Data for the H_0 hypothesis are generated by computing the magnitude of a complex Gaussian random variable with zero mean and $\frac{N\sigma^2}{2}$ variance for both the real and the imaginary part.

The ROCs of the average are marked by 'a1', 'a2', $\cdots$, 'a7' where the integer indicates

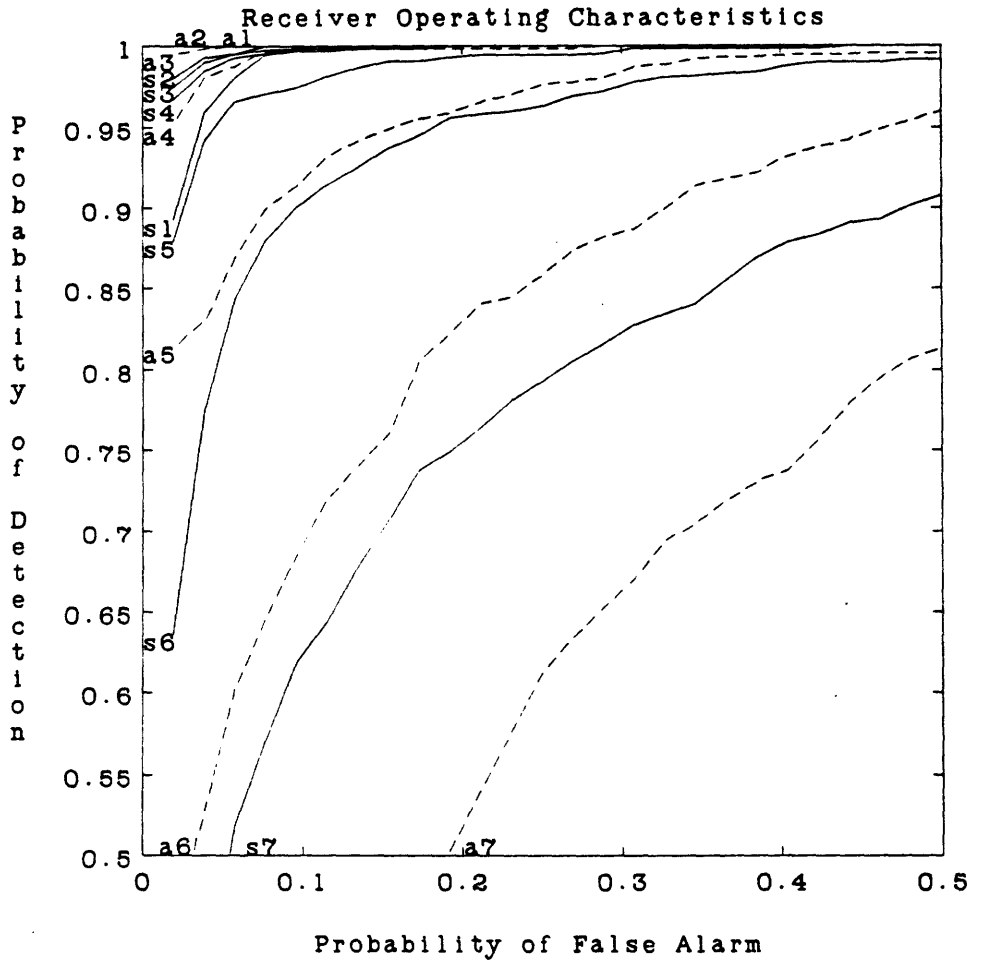

Figure 3.3: The ROCs of the average are plotted in the dashed line and the ROCs of the semblance are plotted in the solid line. The ROCs are computed by the Monte Carlo detection computation using 1000 samples. The length of data(N) is 1024 samples. The number of channels(L) is 8. The amplitude(A) of the narrowband component is 3.5. The variance of WGN(σ^2) is 2000. Under the H_1 hypothesis, the narrowband component exists in all channels. The curves marked by 'a1', 'a2', $\cdots$, 'a7' indicate the average ROC curves. For example, 'a7' indicates the ROC of the average for the case where 7 of 8 channels have the narrowband component under the H_0 hypothesis. Similarly, the curves marked by 's1', 's2', $\cdots$, 's7' indicate the average ROC curves. The gate length of the semblance is 1.

the number of channels containing the narrowband component under the H_0 hypothesis. For example, the ROC curve marked by 'a7' corresponds to the case where the received data contain the narrowband component in 7 of 8 channels under the H_0 hypothesis. The detection probability is high when the H_0 hypothesis data contain the narrowband component in less than half of the total number of channels. The curves marked by 'a1', 'a2', and 'a3' display this characteristic of the average. On the other hand, the curves marked by 'a5', 'a6', and 'a7' indicate that the detection performance of the average degrades noticeably when the H_0 hypothesis data contain the narrowband component in most of the channels.

The ROCs of the semblance are marked by 's1', 's2', $\cdots$, 's7' where the integer indicates the number of channels containing the narrowband component under the H_0 hypothesis. As was the case for the average, the probability of detection degrades as the number of channels containing the narrowband component increases. However, compared to the average, the semblance has a significantly improved probability of detection when many of the channels contain the narrowband component under the H_0 hypothesis as indicated by the 's5', 's6', and 's7' curves. In general, when more than half of the channels contain the narrowband component under the H_0 hypothesis, the semblance is the preferred detection statistic. Alternatively, when less than half of the channels contain the narrowband component, the average is the preferred detection statistic.

3.5 Alternate Multiple Hypothesis Testing Formulation

In the previous section, it was shown that the average and the semblance have different detection performance characteristics and the preference between the two statistics depends on the structure of the H_0 hypothesis data. In this section, a new multiple hypothesis testing formulation, which ensures that the narrowband components have the same amplitude, is developed for the detection problem.

The received data model (3.1) is modified to explicitly indicate the existence of the narrow-

band component with equal amplitude as

$$\begin{pmatrix} \underline{r}_1 \\ \vdots \\ \underline{r}_L \end{pmatrix} = \begin{pmatrix} b_1\underline{s}_1 + \underline{w}_1 \\ \vdots \\ b_L\underline{s}_L + \underline{w}_L \end{pmatrix} \tag{3.16}$$

where, as before, $\underline{s}_l$ is the narrowband component vector and $\underline{w}_l$ is WGN vector with zero mean and σ^2 variance. The scalar, b_l's, are either 0 or 1 to indicate the absence or presence of the narrowband component. To simplify the discussion, the frequency domain model of (3.16) is used in the following. Let $\underline{\rho} = (\rho_1, \rho_2, \cdots, \rho_L)^T$ with $\rho_l = \sum_{n=0}^{N-1} r_l(n)e^{-j\omega_l n}$, therefore, ρ_l is the Fourier transform evaluated at the frequency, ω_l. The above model is transformed to

$$\begin{pmatrix} \rho_1 \\ \vdots \\ \rho_L \end{pmatrix} = \begin{pmatrix} b_1\frac{AN}{2}e^{j\phi_1} + W_1 \\ \vdots \\ b_L\frac{AN}{2}e^{j\phi_L} + W_L \end{pmatrix}$$

where W_l denotes the Fourier transform of WGN at $\omega = \omega_l$. Because $\mathbf{E}(\rho_l) = \frac{AN}{2}$ and $\mathbf{Re}\{\rho_l\}$ and $\mathbf{Im}\{\rho_l\}$ are uncorrelated Gaussian densities with variance $\frac{N\sigma^2}{2}$, the joint PDF becomes

$$p_{\underline{\rho}|A,\underline{\phi}}(\underline{\rho}|A,\underline{\phi}) = \frac{1}{(\pi N\sigma^2)^L}\exp(-\frac{1}{N\sigma^2}(\underline{\rho} - \frac{AN}{2}\underline{u})^H\ (\underline{\rho} - \frac{AN}{2}\underline{u}))$$

where $(.)^H$ denotes the conjugate transpose and $\underline{u} = (b_1e^{j\phi_1}, b_2e^{j\phi_2}, \cdots, b_Le^{j\phi_L})^T$. The unknown parameters A, ϕ_1, ϕ_2, $\cdots$, and ϕ_L are replaced by the ML estimates which are computed by

$$\max_{A,\phi_1,\cdots,\phi_l} \log p_{\underline{\rho}|A,\underline{\phi}}(\underline{\rho}|A,\underline{\phi}) = \max_{A,\phi_1,\cdots,\phi_l} [-\frac{1}{N\sigma^2}(\underline{\rho} - \frac{AN}{2}\underline{u})^H\ (\underline{\rho} - \frac{AN}{2}\underline{u})] - L\log(\pi N\sigma^2).$$

The above maximization with respect to A and ϕ_l's results[33] in the estimates

$$\hat{\phi}_l = \arg\{\rho_l\} \quad \text{and} \quad \hat{A} = \frac{2}{N}\frac{\underline{\rho}^H\hat{\underline{u}}}{\hat{\underline{u}}^H\hat{\underline{u}}} = \frac{2}{N}\frac{\sum_{l=1}^{L} b_l|\rho_l|}{\sum_{l=1}^{L} b_l}.$$

This multiple hypothesis detector computes the log-likelihood of all possible hypotheses and then selects the hypothesis having the maximum value. Therefore, the selection is made by

$$\max_k \mathcal{L}_k = -\frac{1}{N\sigma^2}(\underline{\rho} - \frac{\hat{A}N}{2}\hat{\underline{u}})^H\ (\underline{\rho} - \frac{\hat{A}N}{2}\hat{\underline{u}}) - L\log(\pi N\sigma^2) + \gamma_k \tag{3.17}$$

where k indicates all hypotheses and γ_k's are the constants set to obtain the desired probability of false alarm. As with all multiple hypothesis testing methods, when the *a priori* probabilities

are unknown, γ_k's must be set experimentally which often makes multiple hypothesis testing methods difficult to implement.

Replacing the unknowns by the ML estimates in (3.17), the log-likelihood becomes

$$\begin{aligned}\mathcal{L}_k &= -\frac{1}{N\sigma^2}(\underline{\rho}^H\underline{\rho} - \frac{(\underline{\rho}^H\hat{\underline{u}})^2}{\hat{\underline{u}}^H\hat{\underline{u}}}) - L\log(\pi N\sigma^2) + \gamma_k \\ &= -\frac{1}{N\sigma^2}\underline{\rho}^H\underline{\rho}(1 - \frac{(\underline{\rho}^H\hat{\underline{u}})^2}{(\underline{\rho}^H\underline{\rho})(\hat{\underline{u}}^H\hat{\underline{u}})}) - L\log(\pi N\sigma^2) + \gamma_k.\end{aligned}$$

By removing the common terms which appear in all $\mathcal{L}_k$'s, the above log-likelihood further simplifies to

$$\mathcal{L}_k = -\underline{\rho}^H\underline{\rho}\,(1 - \frac{(\underline{\rho}^H\hat{\underline{u}})^2}{(\underline{\rho}^H\underline{\rho})(\hat{\underline{u}}^H\hat{\underline{u}})}) + \gamma_k'. \tag{3.18}$$

The first component of the above expression is $\underline{\rho}^H\underline{\rho} = \sum_{l=1}^{L}|\rho_l|^2$ which is the previously discussed average scaled by L. The second component is written as $1 - S$ with

$$S = \frac{(\underline{\rho}^H\hat{\underline{u}})^2}{(\underline{\rho}^H\underline{\rho})(\hat{\underline{u}}^H\hat{\underline{u}})} = \frac{(\sum_{l=1}^{L} b_l|\rho_l|)^2}{(\sum_{l=1}^{L} b_l)(\sum_{l=1}^{L}|\rho_l|^2)}$$

which can be interpreted as a generalization of the semblance defined by (3.3) for multiple hypotheses. Because b_l's indicate the existence of the narrowband component, if the narrowband component exists in all L channels, $S = \frac{(\sum_{l=1}^{L}|\rho_l|)^2}{L\sum_{l=1}^{L}|\rho_l|^2}$ which is (3.3) with $M = 0$.

The multiple hypothesis log-likelihood (3.18) is a function of the average and the generalized semblance S. In general, measurements of a hypothesis testing problem are completely summarized by the sufficient statistic. If the average and S are viewed as the "generalized sufficient statistics" of the detection problem, then all relevant quantities of the received data are summarized by the average and S. The detection statistic should, therefore, be a function of the average and S only. Therefore, the log-likelihood (3.18) suggests that a combination of the average and the semblance might be an effective detection statistic. The semblance rather than the generalized semblance S is used because a binary hypothesis test is desired. In Chapter 4, the combination of the average and the semblance are examined.

To conclude this discussion of the multiple hypothesis testing, an implementation of (3.18) is briefly discussed. As mentioned earlier, selecting the constants, γ_k's, of (3.18) is difficult but inherent in any multiple hypothesis testing method. A simple way of selecting the constants is discussed. For the L channel case, 2^L log-likelihoods must be to calculated. However, because

only the number, not the identity, of the channels containing the narrowband component is relevant, only L log-likelihoods need to be computed. Given L magnitudes of the Fourier transforms, which are denoted $|\rho_1|, |\rho_2|, \cdots, |\rho_L|$, they are sorted in ascending order and labeled $\tilde{\rho}_l$ such that $|\tilde{\rho}_1| \geq |\tilde{\rho}_2| \geq \cdots \geq |\tilde{\rho}_L|$. To check all $\mathcal{L}_k$'s which correspond to the hypotheses that any J of L channels contain the narrowband component, only the J largest $|\tilde{\rho}_l|$'s need to be used. Therefore, selecting the maximum of the log-likelihood (3.18) simplifies to

$$\max_{0 \leq J \leq L} \frac{1}{J}(\sum_{l=1}^{J} |\tilde{\rho}_l|)^2 + \gamma'_J = \mathcal{L}'_J.$$

The constants, γ'_Js, are set to remove the contribution due to WGN. When the received data contain only WGN,

$$\mathbf{E}(\mathcal{L}'_J) = \frac{1}{J}(JN\sigma^2 + J(J-1)\frac{\pi N\sigma^2}{4}) = N\sigma^2(\frac{\pi}{4}J + 1 - \frac{\pi}{4})$$

because $\mathbf{E}(|\tilde{\rho}_l|^2) = N\sigma^2$ and $\mathbf{E}(|\tilde{\rho}_l||\tilde{\rho}_k|) = \frac{\pi N\sigma^2}{4}$ from the fact that $|\tilde{\rho}_l|$ has a Rayleigh density. Therefore, to remove this bias, $\gamma'_J = -\alpha JN\sigma^2$, where α is an adjustable parameter. Finally, this multiple hypothesis testing detector computes

$$\max_{0 \leq J \leq L} \frac{1}{J}(\sum_{l=1}^{J} |\tilde{\rho}_l|)^2 - \alpha JN\sigma^2 \tag{3.19}$$

and decides that an emitter exists if $J = L$. This detector will be compared to the combined statistic in Chapter 4.

3.6 Summary

In this chapter, detection using the average or the semblance was discussed because both utilize data from all channels collectively. The PDFs of the statistics, which are necessary to compute the ROC, were approximated. However, they are numerically difficult to evaluate, hence it was necessary to compute the ROCs using the Monte Carlo simulations. The Monte Carlo detection computation was used to compute the ROCs. A conclusion made from the ROCs of the average and the semblance, as shown in Figures 3.3, is that the average performs better when less than four channels contain the narrowband component while the semblance performs better when more than four channels contain the narrowband component under the H_0 hypothesis. Neither statistic performs well for a wide range of the received data model.

In general, the average performs better than the semblance when more than half of the channels contain only wideband noise under the H_0 hypothesis. The semblance performs better than the average when more than half of the channels contain the narrowband component under the H_0 hypothesis. A new multiple hypothesis testing formulation of the detection problem was discussed to motivate the combined use of the average and the semblance which are interpreted as the "generalized sufficient statistics" of the detection problem. To maintain a good detection probability over a wide range of received data, a new class of detection statistics will be developed by combining the average and the semblance in Chapter 4.

Appendix 3.A

Semblance Properties

The semblance properties summarized in Section 3.2.2 are proved.

Property 3.2.1 $0 \leq s \leq 1$.

This follows from Cauchy's inequality: $(\sum_{l=1}^{L} a_l b_l)^2 \leq \sum_{l=1}^{L} a_l^2 \sum_{l=1}^{L} b_l^2$. Let $a_l = x_l(n)$ and $b_l = 1$ for $l = 1, \cdots, L$. Then the inequality simplifies to $(\sum_{l=1}^{L} x_l(n))^2 \leq L \sum_{l=1}^{L} x_l(n)^2$. Using the above inequality in the definition of the semblance

$$s = \frac{\sum_{n=-M}^{M} \left(\sum_{l=1}^{L} x_l(n)\right)^2}{L \sum_{n=-M}^{M} \sum_{l=1}^{L} x_l(n)^2} \leq \frac{\sum_{n=-M}^{M} \left(L \sum_{l=1}^{L} x_l(n)^2\right)}{L \sum_{n=-M}^{M} \sum_{l=1}^{L} x_l(n)^2} = 1.$$

The inequality $s \geq 0$ follows from the fact that the semblance is a ratio of two positive quantities. □

Property 3.2.2 *If* $x_l(n) \geq 0$ *for* $l = 1, \cdots, L$ *and* $n = -M, \cdots, M$ *then* $s \geq \frac{1}{L}$.

The semblance is written as

$$\begin{aligned} s &= \frac{\sum_{n=-M}^{M} \left(\sum_{l=1}^{L} x_l(n)^2 + 2 \sum_{l=1}^{L} \sum_{k=l+1}^{L} x_l(n) x_k(n)\right)}{L \sum_{n=-M}^{M} \sum_{l=1}^{L} x_l(n)^2} \\ &= \frac{1}{L} + \frac{2}{L} \frac{\sum_{n=-M}^{M} \sum_{l=1}^{L} \sum_{k=l+1}^{L} x_l(n) x_k(n)}{\sum_{n=-M}^{M} \sum_{l=1}^{L} x_l(n)^2} \geq \frac{1}{L} \end{aligned}$$

because the second term is always greater than or equal to zero. □

Property 3.2.3 *If the semblance of* $\{x_l(n) : l = 1, \cdots, L; n = -M, \cdots, M\}$ *is* s_x *then the semblance of* $\{c x_l(n) : l = 1, \cdots, L; n = -M, \cdots, M\}$ *for some constant* c *is also* s_x.

This follows directly from the definition. □

Property 3.2.4 *If* $x_l(n) = x(n)$ *for* $l = 1, \cdots, L$ *then* $s = 1$.

This follows directly from the definition. □

Property 3.2.5 *Let* $x_l(n) = c_l y(n)$ *for* $l = 1, \cdots, L$ *and* $n = -M, \cdots, M$ *where* c_l*'s are constants. The semblance is maximized and has the value of* 1 *if and only if* $c_l = c$ *for all* l.

If $c_l = c$ then the semblance is 1 from the above two properties. The maximising choice of c_l's is shown next.

$$s = \frac{\sum_{n=-M}^{M} \left(\sum_{l=1}^{L} c_l y(n)\right)^2}{L \sum_{n=-M}^{M} \sum_{l=1}^{L} c_l^2 y(n)^2} = \frac{1}{L}\frac{\alpha}{\beta}$$

with $\alpha = (\sum_{l=1}^{L} c_l)^2$ and $\beta = \sum_{l=1}^{L} c_l^2$. The maximizing c_l's are determined by solving a set of equations given by $\frac{\partial s}{\partial c_i} = 0$ for $l = 1, \cdots, L$.

$$\frac{\partial s}{\partial c_i} = \frac{1}{L}\frac{2\beta \sum_{l=1}^{L} c_l - 2\alpha c_i}{\beta^2} = 0.$$

Therefore, $c_i = \frac{\sum_{l=1}^{L} c_l^2}{\sum_{l=1}^{L} c_l}$ so they are independent of i and are equal. □

Property 3.2.6 *If $x_l(n) = s(n) + w_l(n)$ where $s(n)$ is the signal and $w_l(n)$ is zero mean noise, then as the number of channels increases $s \approx$ signal power/(signal power + noise power).*

Using the definition of semblance and replacing $x_l(n)$ by $s(n) + w_l(n)$,

$$s = \frac{\sum_{n=-M}^{M} \left(L^2 s(n)^2 + 2Ls(n) \sum_{l=1}^{L} w_l(n) + (\sum_{l=1}^{L} w_l(n))^2\right)}{L \sum_{n=-M}^{M} \left(Ls(n)^2 + 2s(n) \sum_{l=1}^{L} w_l(n) + \sum_{l=1}^{L} w_l(n)^2\right)}$$

but $\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} w_l(n) \to 0$. Therefore,

$$s \approx \frac{\sum_{n=-M}^{M} s(n)^2}{\sum_{n=-M}^{M} s(n)^2 + \frac{1}{L} \sum_{n=-M}^{M} \sum_{l=1}^{L} w_l(n)^2}.$$

□

Chapter 4

Combining Average and Semblance

In Chapter 3 it was shown that if the received data under the H_0 hypothesis contain only wideband noise, the generalized likelihood ratio test(GLRT) detection statistic for the hypothesis testing problem is the average. When the received data contain the narrowband component under the H_0 hypothesis, the average performs poorly because it is sensitive only to the total power. Thus, the average is an unsatisfactory statistic for the detection problem described by (2.1) in which all channels must contain the narrowband component in order to decide that an emitter is present. To improve the probability of detection for the cases in which the narrowband component exists under the H_0 hypothesis, the semblance was considered as a detection statistic. Because the semblance is sensitive to the existence of the narrowband component in all channels, it is suitable for the detection problem described by (2.1). It was shown in Section 3.4 that the semblance is particularly effective when the received data under the H_0 hypothesis have the narrowband component in more than half of the channels. Unfortunately, however, wideband noise alone can produce a high semblance value because the semblance is insensitive to the power of the signal. Therefore, in this chapter, a new detection statistic is developed to take advantage of the sensitivity of the semblance to the correlation between channels and the sensitivity of the average to the signal power. A new detection statistic will be obtained by combining the semblance and the average as motivated by the log-likelihood of Section 3.5. The log-likelihood function suggested a product of the average and the generalized semblance, however, the semblance will be used to obtain a binary hypothesis test.

A statistic formed by combining the semblance and the average should have good detection

performance over a wide range of received data under the H_0 hypothesis. It should perform as well as the average alone and the semblance alone for all possible received data. To determine a combined statistic, the approach taken in this chapter is to treat the semblance and the average as two random variables which are combined by a commutative operation. To weigh the average and the semblance equally, their PDFs are matched before the combination. The resultant combined statistic has superior performance as will be shown by its ROC.

An optimal detection statistic can be derived using the LRT by treating the average and the semblance as the measurement to a hypothesis testing problem. Unfortunately, the derivation yields a closed form solution only for a simple case. For this simple case, it will be shown that the combined statistic using the PDF matching performs as well as the LRT statistic. In addition, statistics are designed by treating the average and the semblance as two features of a pattern recognition problem and are compared to the combined statistic. This chapter concludes with a comparison of the ROCs of the different statistics.

4.1 Combining Average and Semblance Using Probability Density Function Matching

In this section, the average and the semblance are combined by requiring that both be used with equal weighting. This requirement suggests that the combining function, $f(a, s)$, should satisfy the following properties:

1. $f(a, s)$ should be symmetric, $f(a, s) = f(s, a)$.
2. $f(a, s)$ must change equally to equal changes in a or s, $f(a + \Delta, s) = f(a, s + \Delta)$
3. $f(a, s)$ must increase monotonically as a or s increases.

where a denotes the average and s denotes the semblance. The first and the second properties suggest that the combining function should not distinguish between the average and the semblance. The third property is required in order to use the threshold test for detection.

Before the above properties are used to determine $f(a, s)$, a general comment about any detection statistic which is determined by combining the average and the semblance is made. The ROC of the detection statistic, $z = f(a, s)$, is completely specified by the functional form

of $f(a,s)$. If the PDF of the average and the semblance, conditioned by the H_1 hypothesis or the H_0 hypothesis, are known, the conditional PDF of z is determined by

$$p_z(z|H_i)\,dz = \int\int_{R_z} p_{a,s|H_i}(a,s)da\,ds$$

for $i = 0$ or 1 where R_z is the region in $a-s$ plane such that $z < f(a,s) < z+dz$. When the combining function $f(a,s) = a+s$ is used, the PDF of z is the convolution of the PDF of a and the PDF of s assuming that the average and the semblance are independent. Using the fact that the PDF of s has a much narrower width compared to that of a as discussed in Chapter 3, the result of the convolution is approximately equal to the PDF of a. Therefore, the PDF of the combined statistic will be mostly determined by the PDF of the average. To ensure that the combined statistic will be equally determined by the average and the semblance, the PDF of the average will be mapped to match the PDF of the semblance before the combining operation is performed. The discussion on different mapping methods is deferred to Section 4.1.1 and 4.1.2. In the remainder of this section, a mapping is assumed to have been performed such that the PDF of the mapped average, $\tilde{a}$, is approximately the same as the PDF of the semblance. Next, a combining function using $\tilde{a}$ and s is developed by satisfying the above functional constraints.

The first property which must be satisfied by the combining function is the symmetry, $f(\tilde{a},s) = f(s,\tilde{a})$. This is satisfied by a wide variety of functions; two particularly simple functions are $f(\tilde{a},s) = g_1(\tilde{a}+s)$ and $f(\tilde{a},s) = g_2(\tilde{a}\cdot s)$ where $g_1(.)$ and $g_2(.)$ are arbitrary functions. The third property requires that $f(\tilde{a},s)$ must increase monotonically as a or s increases. Because both $\tilde{a}+s$ and $\tilde{a}\cdot s$ monotonically increase as a or s increases, $g_1(.)$ and $g_2(.)$ must be monotonically increasing functions. Because the detection statistic can be transformed by any monotonic function without affecting the ROC, the functions $g_1(.)$ and $g_2(.)$ can be removed. The combining functions which satisfy both the first and the third properties are $\tilde{a}+s$ and $\tilde{a}\cdot s$.

The second property requires that a change in z due to a Δ increase in the mapped average should be same as the change due to a Δ increase in the semblance. The probability of detection and the probability of false alarm are affected equally by the Δ increase in either $\tilde{a}$ or s because the effects of $\tilde{a}$ and s are summarized completely by $z_0 = f(\tilde{a_0}+\Delta, s_0) = f(\tilde{a_0}, s_0+\Delta)$, where

$\tilde{a}_0$ and s_0 are some particular values, and

$$P_{D,0} = \int_{z_0}^{\infty} p_{z|H_1}(z|H_1)dz \quad \text{and} \quad P_{F,0} = \int_{z_0}^{\infty} p_{z|H_0}(z|H_0)dz.$$

Because $\tilde{a} \cdot (s + \Delta) \neq (\tilde{a} + \Delta) \cdot s$, the combining function which satisfies all three properties is

$$f(\tilde{a}, s) = \tilde{a} + s.$$

Different methods of matching the PDFs will be discussed next.

4.1.1 Probability Density Function Matching by Transform Functions

In this section, transform functions which map the average to match the PDF of the semblance will be developed. As discussed in Section 3.1.2, when the channels contain only wideband noise, the PDF of the average can be approximated by the truncated Gram-Charlier series as

$$p_a(a) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(a-\bar{a})^2}{2\sigma^2})(1 + \frac{m_3}{6}((a-\bar{a})^3 - 3(a-\bar{a}))) \tag{4.1}$$

for $a \geq 0$ where $\bar{a} = \frac{L}{2}\sqrt{\pi N\sigma^2}$, $\sigma^2 = (2 - \frac{\pi}{2})\frac{NL\sigma^2}{2}$, and $m_3 = 3\sqrt{\frac{2\pi}{L}}\frac{1-L-L^2}{(2-\frac{\pi}{2})^{\frac{3}{2}}}$. L is the number of channels, N is the data length, and σ^2 is the variance of WGN. As developed in Section 3.2.4, the PDF of the semblance is approximated by the modified beta density

$$p_s(s) = \frac{\Gamma(c_a + c_b)}{\Gamma(c_a)\Gamma(c_b)} \left(\frac{L}{L-1}\right)^{c_a+c_b-1} (y - \frac{1}{L})^{c_a-1}(1-y)^{c_b-1} \tag{4.2}$$

for $\frac{1}{L} \leq s \leq 1$ where c_a and c_b are constants. A transfer function $h(a)$ with the property that the transformed variable $\tilde{a} = h(a)$ has a PDF which is equal to the PDF of s is desired. Unfortunately, a perfect transformation resulting in identical PDFs is difficult to determine. Hence, transformations which map the range of the average to the range of the semblance are examined to select the best transformation. Three possible transformations which map the range $[0, \infty)$ to $[\frac{1}{L}, 1]$ are:

1. arc-tangent transformation: $\tilde{a} = h_t(a) = \frac{2}{\pi}\frac{L-1}{L}\arctan(\alpha_t a) + \frac{1}{L}$
2. rational transformation: $\tilde{a} = h_r(a) = \frac{L-1}{L}\frac{\alpha_r a}{1+\alpha_r a} + \frac{1}{L}$
3. hyperbolic tangent transformation: $\tilde{a} = h_h(a) = \frac{L-1}{L}\tanh(\alpha_h a) + \frac{1}{L}$

where α_t, α_r, and α_h are free parameters adjusted to obtain the best match. The transformed PDF is determined by

$$p_{\tilde{a}}(\tilde{a}) = \frac{1}{\left|\frac{dh(a)}{da}\right|_{a^*}} p_a(a^*)$$

where a^* denotes a expressed in terms of $\tilde{a}$. Using (4.1) in the above equation, the arc-tangent transformation, $h_t(a)$, results in

$$p_{\tilde{a}}(\tilde{a}) = \frac{\pi}{2}\frac{L}{L-1}\frac{1}{\alpha_t}(1+\tan^2(\frac{\pi}{2}\frac{L\tilde{a}-1}{L-1}))\ p_a(\frac{1}{\alpha_t}\tan(\frac{\pi}{2}\frac{L\tilde{a}-1}{L-1}))$$

for $\frac{1}{L} \leq \tilde{a} \leq 1$. Similarly, the rational transformation, $h_r(a)$, results in

$$p_{\tilde{a}}(\tilde{a}) = \frac{L-1}{L}\frac{1}{\alpha_r(1-\tilde{a})^2}\ p_a(\frac{1}{\alpha_r}\frac{L\tilde{a}-1}{L(1-\tilde{a})})$$

and the hyperbolic tangent transformation, $h_h(a)$, results in

$$p_{\tilde{a}}(\tilde{a}) = \frac{L(L-1)\alpha_h}{(L-1)^2+(L\tilde{a}-1)^2}\ p_a(\frac{1}{\alpha_h}\tanh^{-1}(\frac{L\tilde{a}-1}{L-1}))$$

Unfortunately, they do not have the same functional form as the beta density (4.2). The free parameters, α_t, α_r and α_h, will be selected to transform the mean of a, denoted m_a, to the mean of s, denoted m_s. The parameter is determined to be

$$\alpha_t = \frac{1}{m_a}\tan(\frac{\pi}{2}\frac{L}{L-1}(m_s - \frac{1}{L}))$$

for the arc-tangent transformation,

$$\alpha_r = \frac{1}{m_a}\frac{Lm_s-1}{L(1-m_s)}$$

for the rational transformation, and

$$\alpha_h = \frac{1}{m_a}\tanh^{-1}(\frac{Lm_s-1}{L-1}) = \frac{1}{2m_a}\log\left(\frac{L(1+m_s)-2}{L(1-m_s)}\right)$$

for the hyperbolic tangent transformation. This simple method of determining the free parameter does not imply that the mean of the transformed variable, $\tilde{a}$, matches the mean of the semblance, s. Selecting α to satisfy $\mathbf{E}(\tilde{a}) = \mathbf{E}(s)$ is an intractable non-linear problem. Fortunately, however, this simple method of determining the free parameter is adequate for the purpose of combining two detection statistics as the detection results will demonstrate.

Transformation	Error		
	$\sum_n \|p_s(n) - p_{\bar{a}}(n)\|^2\Delta$	$\sum_n \|p_s(n) - p_{\bar{a}}(n)\|\Delta$	$\max_n \|p_s(n) - p_{\bar{a}}(n)\|$
rational	4.580	0.935	10.232
arctan	3.021	0.802	8.252
tanh	0.234	0.228	1.935

Table 4.1: Errors of the probability density function mapping by different transformations.

Figure 4.1 shows the histograms of the transformed averages. The figure shows that the hyperbolic tangent transformation produces the closest match to the PDF of the semblance. The transformations are compared numerically using three different error measures. The results are shown in Table 4.1 which verify that the hyperbolic tangent transformation has the smallest error for all three error measures. Therefore, in this thesis, the average will be mapped by the hyperbolic tangent transform before it is combined with the semblance, hence the combined statistic $h_h(a) + s$ will be used as the detection statistic.

4.1.2 Probability Density Function Matching based on Cumulative Distribution Function

In this section, an alternative mapping method is discussed. In this method, the PDFs are mapped by matching the cumulative distribution functions(CDFs). The random variable a can be mapped to obtain the desired PDF $p_s(s)$ by constructing a one-to-one mapping between the CDF of a and the CDF of s. A unique mapping exists because the CDFs are monotonically increasing functions. The CDF of the semblance, denoted $Q_s(s)$, is the incomplete beta function as developed in Chapter 3. Because the closed form expression of the CDF of the average, denoted $Q_a(a)$, is unavailable, it is approximated by integrating the histogram. The mapping is determined by selecting a value for a, denoted a_0, computing $Q_a(a_0)$, then determining s_0 such that $Q_s(s_0) = Q_a(a_0)$. The (a_0, s_0) pairs completely specify the mapping.

Figure 4.2 shows the histogram of the transformed average using CDF matching and the histogram of the semblance. This method theoretically provides an exact match and even when a finite number of samples is used, a good match is obtained. In general, this method is more sensitive to the data than the transform method of the previous section because the transform method uses only the mean of the data whereas the CDF matching method uses

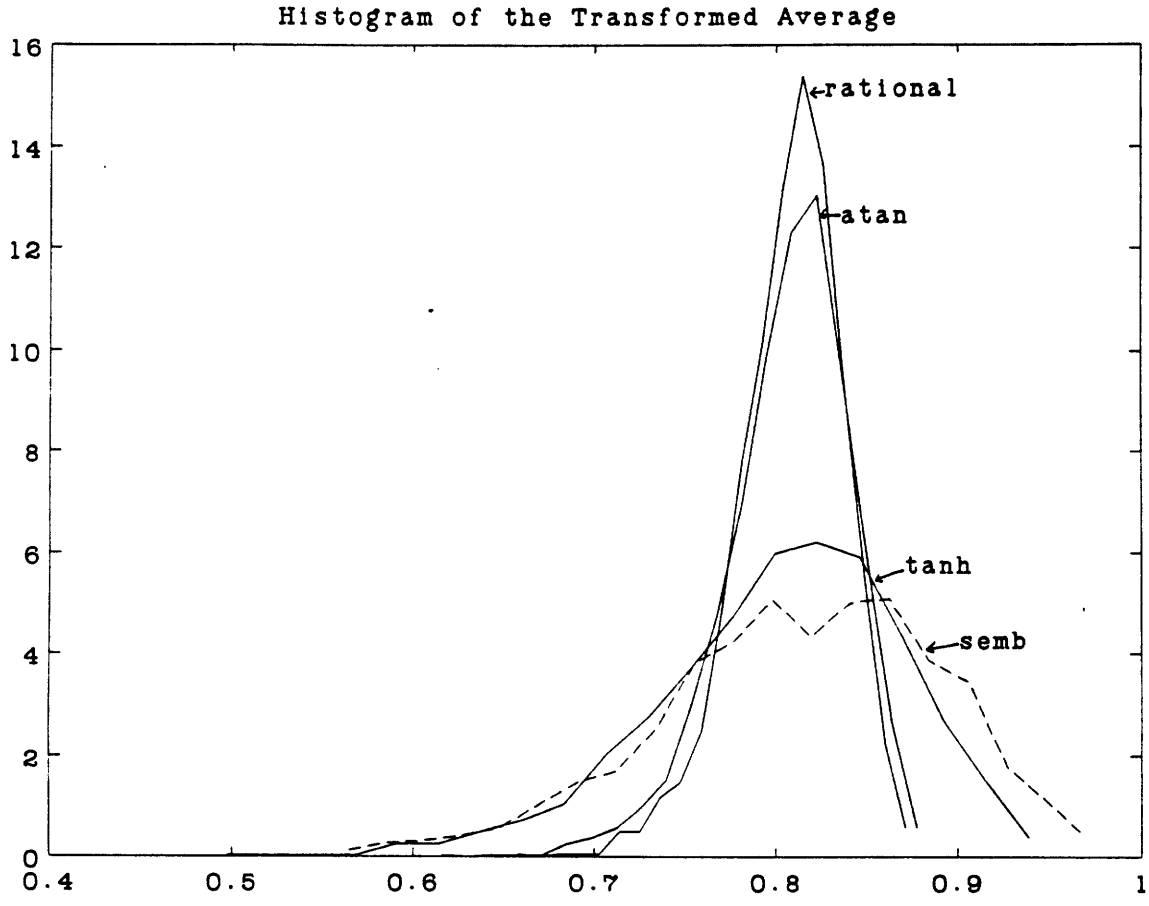

Figure 4.1: Histograms of the transformed averages using the arc-tangent transformation, denoted by 'atan', the rational transformation, denoted by 'rational', and the hyperbolic tangent transformation, denoted by 'htan', are plotted against a histogram of the semblance in the dashed line. The histograms are made using 2000 samples and 20 bins. The semblance and the average are computed using the jointly Gaussian random variables with zero mean and variance 10.

the histogram of the data. Because the transform method is more robust and the hyperbolic tangent transformation performs as well as the CDF matching method, the hyperbolic tangent transform method will used in the following.

4.2 Combining Average and Semblance Using Likelihood Ratio

In the previous section, the average and the semblance were combined using a function which satisfied a set of properties to achieve equal weighting of the statistics. In this section, the average and the semblance are combined by applying the LRT to a hypothesis testing problem which uses the average and the semblance as the measurements. Because the resulting statistic is optimal in the Neyman-Pearson sense, it will be used to evaluate the previous method for a simple case only because this optimal statistic is not known for the general case.

The detector which maximizes the probability of detection with a fixed probability of false alarm is given by the likelihood ratio test[72]

$$\Lambda(a,s) = \frac{p_{a,s|H_1}(a,s|H_1)}{p_{a,s|H_0}(a,s|H_0)}. \tag{4.3}$$

The joint PDF of the average and the semblance $p_{a,s}(a,s)$ is derived from $p_{\underline{x}_1,\cdots\underline{x}_L}(\underline{x}_1,\cdots\underline{x}_L)$ where $\underline{x}_l = (x_l(-M), x_l(-M+1),\cdots,x_l(M))^T$ is the data to be used in computing the average and the semblance. For a simple case of $L = 2$ and $M = 0$, the average and the semblance become

$$a = \frac{x_1(0)+x_2(0)}{2} \tag{4.4}$$

$$s = \frac{(x_1(0)+x_2(0))^2}{2(x_1(0)^2+x_2(0)^2)} \tag{4.5}$$

For the remainder of this section $x_1(0)$ and $x_2(0)$ are denoted by x_1 and x_2, respectively. The random variables, x_1 and x_2, have either Rician or Rayleigh density depending on the existence of the narrowband component and are independent. The joint PDF of the average and the semblance is expressed in terms of the joint PDF of x_1 and x_2 as

$$p_{a,s}(a,s) = \frac{p_{x_1,x_2}(\underline{x}_1^*)}{|J(\underline{x}_1^*)|} + \frac{p_{x_1,x_2}(\underline{x}_2^*)}{|J(\underline{x}_2^*)|} \tag{4.6}$$

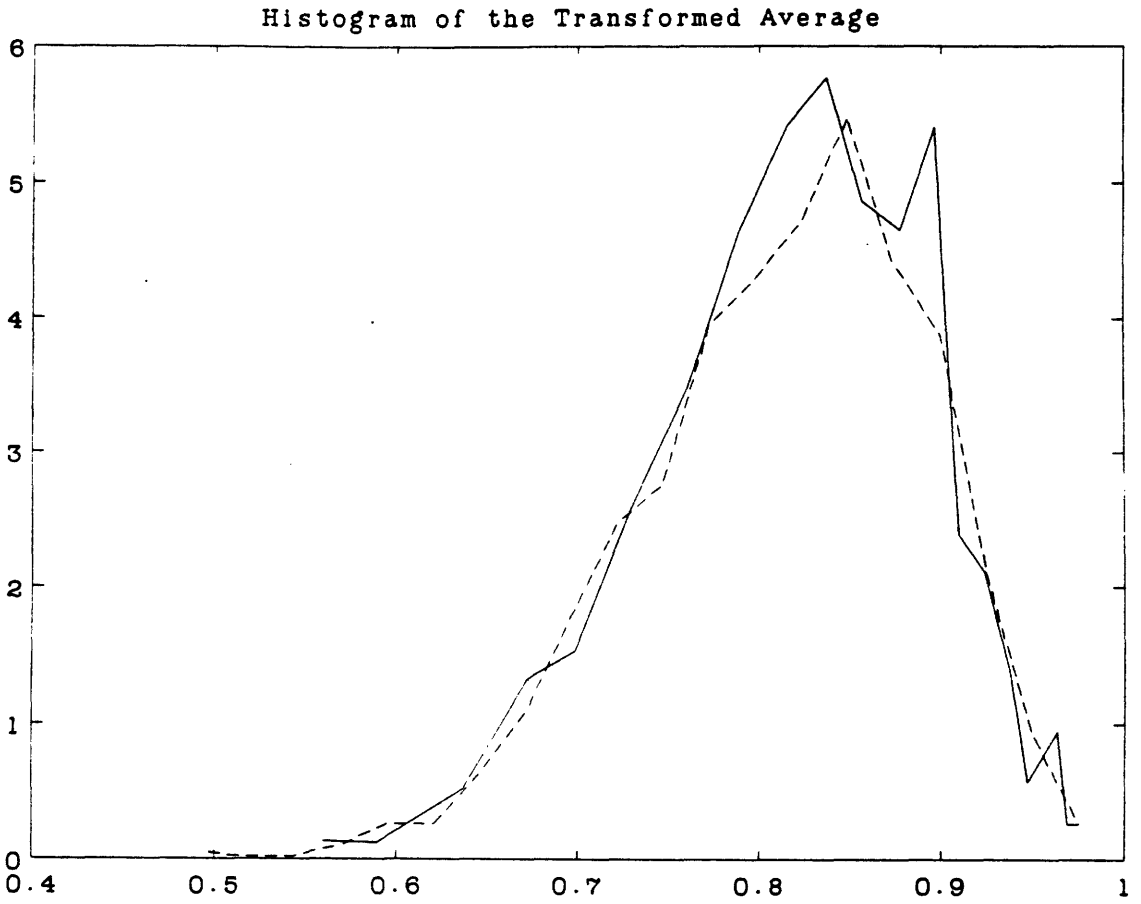

Figure 4.2: A histogram of the transformed average using the CDF matching method is plotted in the solid line and a histogram of the semblance is plotted in the dashed line. The histograms are made using 2000 samples and 20 bins. The semblance and the average are computed using the jointly Gaussian random variables with zero mean and variance 10.

where $\underline{x}_1^*$ and $\underline{x}_2^*$ describe x_1 and x_2 in terms of a and s. The Jacobian for this transformation is

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial a}{\partial x_1} & \frac{\partial a}{\partial x_2} \\ \frac{\partial s}{\partial x_1} & \frac{\partial s}{\partial x_2} \end{vmatrix} = \frac{(x_1 + x_2)(x_1^2 - x_2^2)}{2(x_1^2 + x_2^2)^2}. \tag{4.7}$$

From the average defined by (4.4), $x_2 = 2a - x_1$. It is used in the semblance defined by (4.5) to obtain

$$s = \frac{a^2}{x_1^2 - 2ax_1 + 2a^2}.$$

Using the quadratic formula, the above equation is solved for x_1 first, then x_2 is determined.

$$x_1 = a\left(1 \pm \sqrt{\frac{1-s}{s}}\right) \quad \text{and} \quad x_2 = a\left(1 \mp \sqrt{\frac{1-s}{s}}\right).$$

Therefore, $\underline{x}_1^* = \begin{pmatrix} a(1+\sqrt{\frac{1-s}{s}}) \\ a(1-\sqrt{\frac{1-s}{s}}) \end{pmatrix}$ and $\underline{x}_2^* = \begin{pmatrix} a(1-\sqrt{\frac{1-s}{s}}) \\ a(1+\sqrt{\frac{1-s}{s}}) \end{pmatrix}$ and from (4.7) the Jacobian becomes

$$|J(\underline{x}_1^*)| = |J(\underline{x}_2^*)| = \frac{s^2}{a}\sqrt{\frac{1-s}{s}}.$$

The joint PDF under the H_0 hypothesis is determined assuming that the narrowband component exists in only one channel or none. From Chapter 3, for the case of wideband noise only in both channels, the PDF of x_1 and x_2 are

$$p_{x_1}(x_1) = \frac{x_1}{\alpha^2}\exp(-\frac{x_1^2}{2\alpha^2}) \quad \text{and} \quad p_{x_2}(x_2) = \frac{x_2}{\alpha^2}\exp(-\frac{x_2^2}{2\alpha^2})$$

for $x_1 \geq 0$ and $x_2 \geq 0$, where $\alpha^2 = \frac{N\sigma^2}{2}$, N is the length of the received data, and σ^2 is the variance of the noise. Using (4.6), the joint PDF is determined to be

$$p_{a,s}(a, s|H_0) = \frac{2a^3}{\alpha^4}\sqrt{\frac{s}{1-s}}\frac{2s-1}{s^3}\exp(-\frac{a^2}{\alpha^2 s}) \tag{4.8}$$

for $a \geq 0$ and $\frac{1}{2} \leq s \leq 1$.

Because the narrowband component exists in both channels, the joint PDF under the H_1 hypothesis is determined using

$$p_{x_1}(x_1) = \frac{x_1}{\alpha^2}I_0(\beta x_1)\exp(-\frac{x_1^2+\gamma}{2\alpha^2}) \quad \text{and} \quad p_{x_2}(x_2) = \frac{x_2}{\alpha^2}I_0(\beta x_2)\exp(-\frac{x_2^2+\gamma}{2\alpha^2})$$

where $\alpha^2 = \frac{N\sigma^2}{2}$, $\beta = \frac{A}{\sigma^2}$, and $\gamma = (\frac{AN}{2})^2$. N is the length of the received data, A is the amplitude of the narrowband component, and σ^2 is the variance of the noise. Again, using

(4.6), the joint PDF is determined to be

$$p_{a,s}(a,s|H_1) = \frac{2a^3}{\alpha^4}\sqrt{\frac{s}{1-s}}\frac{2s-1}{s^3} I_0(\beta\tau_1) I_0(\beta\tau_2) \exp(-\frac{1}{\alpha^2}(\frac{a^2}{s}+\gamma)) \tag{4.9}$$

for $a \geq 0$ and $\frac{1}{2} \leq s \leq 1$ where $\tau_1 = a(1-\sqrt{\frac{1-s}{s}})$ and $\tau_2 = a(1+\sqrt{\frac{1-s}{s}})$.

If the narrowband component exists in only one channel under the H_0 hypothesis, then the joint PDF is determined to be

$$p_{a,s}(a,s|H_0) = \frac{a^3}{\alpha^4}\sqrt{\frac{s}{1-s}}\frac{2s-1}{s^3} (I_0(\beta\tau_1) + I_0(\beta\tau_2)) \exp(-\frac{1}{2\alpha^2}(\frac{2a^2}{s}+\gamma)) \tag{4.10}$$

for $a \geq 0$ and $\frac{1}{2} \leq s \leq 1$.

If the received data have only wideband noise in both channels under the H_0 hypothesis, the likelihood ratio (4.3) is determined using (4.9) for the H_1 hypothesis and (4.8) for the H_0 hypothesis to be

$$\Lambda(a,s) \propto I_0(\beta\tau_1) I_0(\beta\tau_2).$$

If the received data have the narrowband component in one channel under the H_0 hypothesis, the likelihood ratio (4.3) is determined, using (4.9) for the H_1 hypothesis and (4.10) for the H_0 hypothesis, to be

$$\Lambda(a,s) \propto \frac{I_0(\beta\tau_1)\, I_0(\beta\tau_2)}{I_0(\beta\tau_1) + I_0(\beta\tau_2)}.$$

The likelihood ratio value increases as the average or the semblance increases.

The above statistic is simplified using the approximation

$$\frac{I_0(\beta\tau_1)}{1+\frac{I_0(\beta\tau_1)}{I_0(\beta\tau_2)}} \approx I_0(\beta\tau_1)$$

because $\tau_2 > \tau_1 > 0$ and $\tau_2 \gg 1$. For the detection statistic, $I_0(\cdot)$ can be removed because it is monotonic. Finally, the detection statistic can be approximated by

$$a\left(1-\sqrt{\frac{1-s}{s}}\right). \tag{4.11}$$

Unfortunately, the likelihood ratio detection statistic for the general case is unknown as discussed in Appendix 4.A. In Section 4.4, this approximate LRT statistic will be compared to the combined statistic, $h_h(a) + s$.

4.3 Combining Average and Semblance Using Discriminant Functions

In this section, classification algorithms of pattern recognition theory will be used to determine the combining functions. The detection statistics will, then, be compared against the PDF matching statistic in Section 4.4. To employ the classifications algorithms, the average and the semblance are viewed as the features of a classification problem with two classes, H_0 and H_1. Two different discriminant functions will be designed using the Gaussian quadratic classifier method and the Fisher linear classifier method with quadratic features.

4.3.1 Gaussian Classifier

The Gaussian quadratic classifier is designed by assuming that the PDFs under the H_0 hypothesis and the H_1 hypothesis are Gaussian with unknown mean and covariance. The discriminant function is given by

$$\begin{aligned}\Lambda(\underline{f}) &= \ln\left(\frac{p_{\underline{f}|H_1}(\underline{f}|H_1)}{p_{\underline{f}|H_0}(\underline{f}|H_0)}\right) \\ &= (\underline{f}-\underline{m}_0)^T K_0^{-1}(\underline{f}-\underline{m}_0) - (\underline{f}-\underline{m}_1)^T K_1^{-1}(\underline{f}-\underline{m}_1) + \ln\frac{|K_0^{-1}|}{|K_1^{-1}|}\end{aligned}$$

where the feature vector is $\underline{f} = (a, s)^T$, $\underline{m}_0$ and $\underline{m}_1$ are the mean vector of the H_0 hypothesis and the H_1 hypothesis data, respectively, and K_0 and K_1 are the covariance matrix of the H_0 hypothesis and the H_1 hypothesis data, respectively. The above expression is rearranged to obtain the quadratic discriminant function

$$d_g(\underline{f}) = \underline{f}^T U \underline{f} + \underline{v}^T \underline{f} + w$$

with $U = K_0^{-1} - K_1^{-1}$, $v = 2(K_1^{-1}\underline{m}_1 - K_0^{-1}\underline{m}_0)$, and $w = \underline{m}_0 K_0^{-1}\underline{m}_0 - \underline{m}_1 K_1^{-1}\underline{m}_1 + \ln\frac{|K_0^{-1}|}{|K_1^{-1}|}$. The discriminant function describes a quadratic surface and its value increases as the average and the semblance increase.

4.3.2 Quadratic Classifier

A general quadratic discriminant function has the form

$$d_f(a, s) = w_1 s + w_2 a + w_3 as + w_4 s^2 + w_5 a^2 = \underline{w}^T \underline{f}',$$

where $\underline{w} = (w_1, w_2, w_3, w_4, w_5)^T$ is the desired weighting vector. The weights are computed using the Fisher linear discriminant analysis by defining a new feature vector, $\underline{f}' = (a, s, as, a^2, s^2)^T$. The Fisher linear discriminant function is obtained by maximizing the normalized inter-class distance[12], hence the cost function to be minimized is

$$C_f(\underline{w}) = \frac{\underline{w}^T S_B \underline{w}}{\underline{w}^T S_W \underline{w}}$$

where $S_B = (\underline{m}_0 - \underline{m}_1)(\underline{m}_0 - \underline{m}_1)^T$ and $S_W = \sum_{\underline{f}'_i \in H_1} (\underline{f}'_i - \underline{m}_1)(\underline{f}'_i - \underline{m}_1)^T + \sum_{\underline{f}'_i \in H_0} (\underline{f}'_i - \underline{m}_0)(\underline{f}'_i - \underline{m}_0)^T$. The desired weights are given by

$$\underline{w} = S_W^{-1}(\underline{m}_1 - \underline{m}_0).$$

A different quadratic classifier can be obtained by minimizing a different cost function. One possible alternate cost function is

$$C_p(\underline{w}) = - \sum_{\text{all } \underline{\tilde{f}}'} \underline{w}^T \underline{\tilde{f}}'$$

where $\underline{\tilde{f}}'$ denotes misclassified $\underline{f}'$. This cost function leads to the perceptron method[12]. The solution of the minimization problem can be determined by the gradient method. Because the resulting discriminant function is not significantly different from the Fisher discriminant function, it is not discussed further. In fact, all three classifiers share the required characteristic that the discriminant value increases as the average and the semblance increase. However, the exact form of the detection statistic is rather sensitive to the training data. The effectiveness of these as detection statistics will be examined in the next section by computing the ROCs.

4.4 Performance Comparison

The ROCs of the different statistics are determined in this section. To evaluate their effectiveness for the simple case of $L = 2$ and $M = 0$, the average, the semblance, and the detection statistic $h_h(a) + s$ are compared against the approximate optimal statistic (4.11). The ROCs are computed using 1000 trials of the simulated data with the following parameters: 2 channels ($L = 2$), semblance gate length of 1 ($M = 0$), narrowband component amplitude of 3.5 ($A = 3.5$), data length of 1024 samples ($N = 1024$), and WGN variance of 2000 ($\sigma^2 = 2000$).

Fig. 4.3 shows the ROCs for the case where the narrowband component exists in 1 of 2 channels under the H_0 hypothesis. The figure shows that the statistic $h_h(a) + s$ performs as well as the approximate optimal statistic. Next, more realistic cases are considered.

Figure 4.4 is computed using 5000 trials of the simulated data with the following parameters: 8 channels ($L = 8$), semblance gate length of 1 sample ($M = 0$), narrowband component amplitude of 2.9 ($A = 2.9$), data length of 1024 samples ($N = 1024$), and WGN variance of 2000 ($\sigma^2 = 2000$). It shows the ROCs for the case where the narrowband component exists in 2 of 8 channels under the H_0 hypothesis. Of course, the narrowband component exists in all 8 channels for the H_1 hypothesis as described by the received data model (2.1). Because the received data under the H_0 hypothesis is close to the wideband noise only case for which the average is the LRT statistic, the average performs well. As discussed in Chapter 3, the semblance does not perform as well as the average for this case. The figure also shows that the Fisher discriminant function statistic and the Gaussian quadratic discriminant function statistic perform significantly better than the semblance. The combined statistic, $h_h(a) + s$, performs as well as the average for this case. For comparison, the multiple hypothesis detection of Section 3.5 was also applied. By setting $\alpha = 0.1$ in (3.19), the probability of detection is estimated to be 0.63 when the probability of false alarm is 0.05 which indicates that this detector does not perform as well as the other statistics plotted in Fig. 4.4.

Figure 4.5 is computed using 5000 trials of the simulated data with the following parameters: 8 channels ($L = 8$), semblance gate length of 1 sample ($M = 0$), narrowband component amplitude of 3.2 ($A = 3.2$), data length of 1024 samples ($N = 1024$), and WGN variance of 2000 ($\sigma^2 = 2000$). It shows the ROCs for the case where the narrowband component exists in 4 of 8 channels under the H_0 hypothesis. The figure indicates that the Fisher discriminant function statistic and the Gaussian quadratic discriminant function statistic perform better than the average or the semblance alone. The combined statistic, $h_h(a) + s$, performs better than the other statistics. For comparison, the multiple hypothesis detection of Section 3.5 was also applied. By setting $\alpha = 0.01$ in (3.19), the probability of detection is estimated to be 0.72 when the probability of false alarm is 0.037. The combined statistic, $h_h(a) + s$, performs significantly better. For example, it has a probability of detection of 0.98 when the probability of false alarm is 0.02.

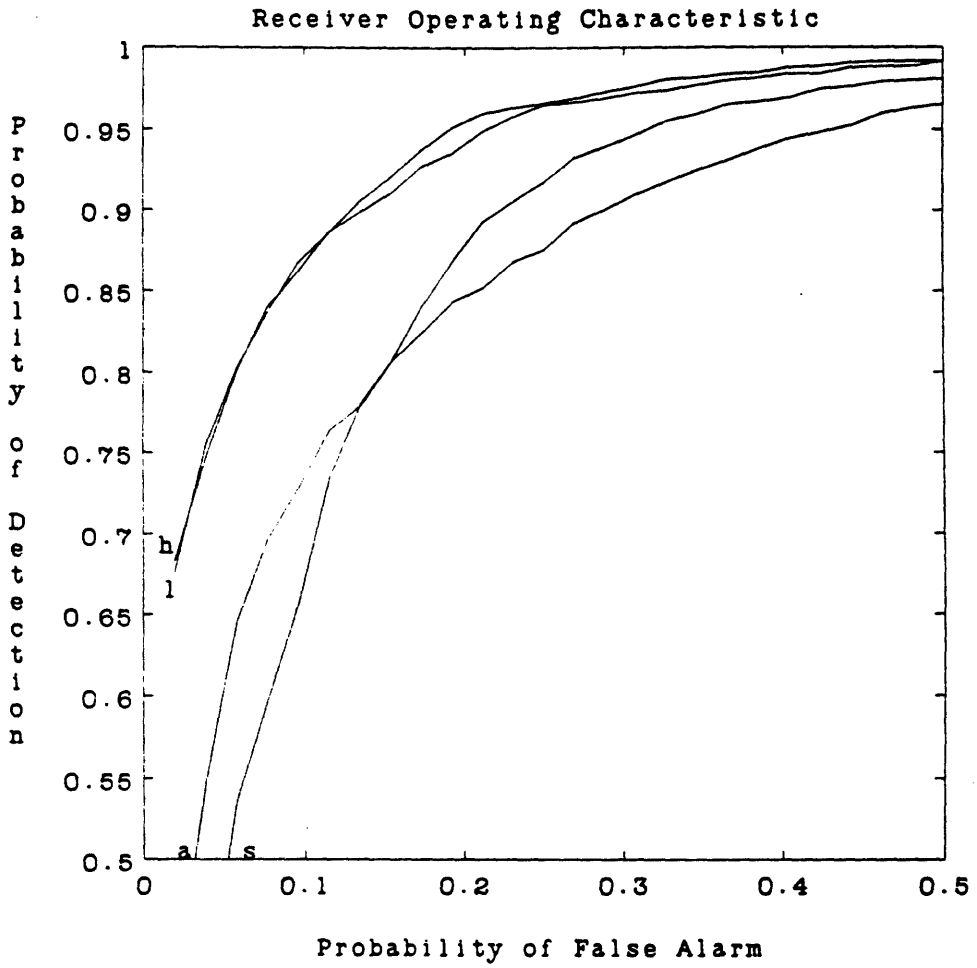

Figure 4.3: Comparison of ROCs of the statistics with the parameters: the number of channels(L) is 2; the semblance gate length is 1 ($M = 0$); the amplitude(A) of the narrowband component is 3.5; the length of data(N) is 1024 samples; the variance of WGN(σ^2) is 2000. 1000 trials are used. The narrowband component exists in 1 of 2 channels under the H_0 hypothesis. 'h' denotes $h_h(a) + s$, 'a' denotes average, 's' denotes semblance, and 'l' denotes the optimal statistic. The abscissa, indicating the probability of false alarm ranges, from 0 to 0.5 and the ordinate, indicating the probability of detection ranges, from 0.5 to 1.

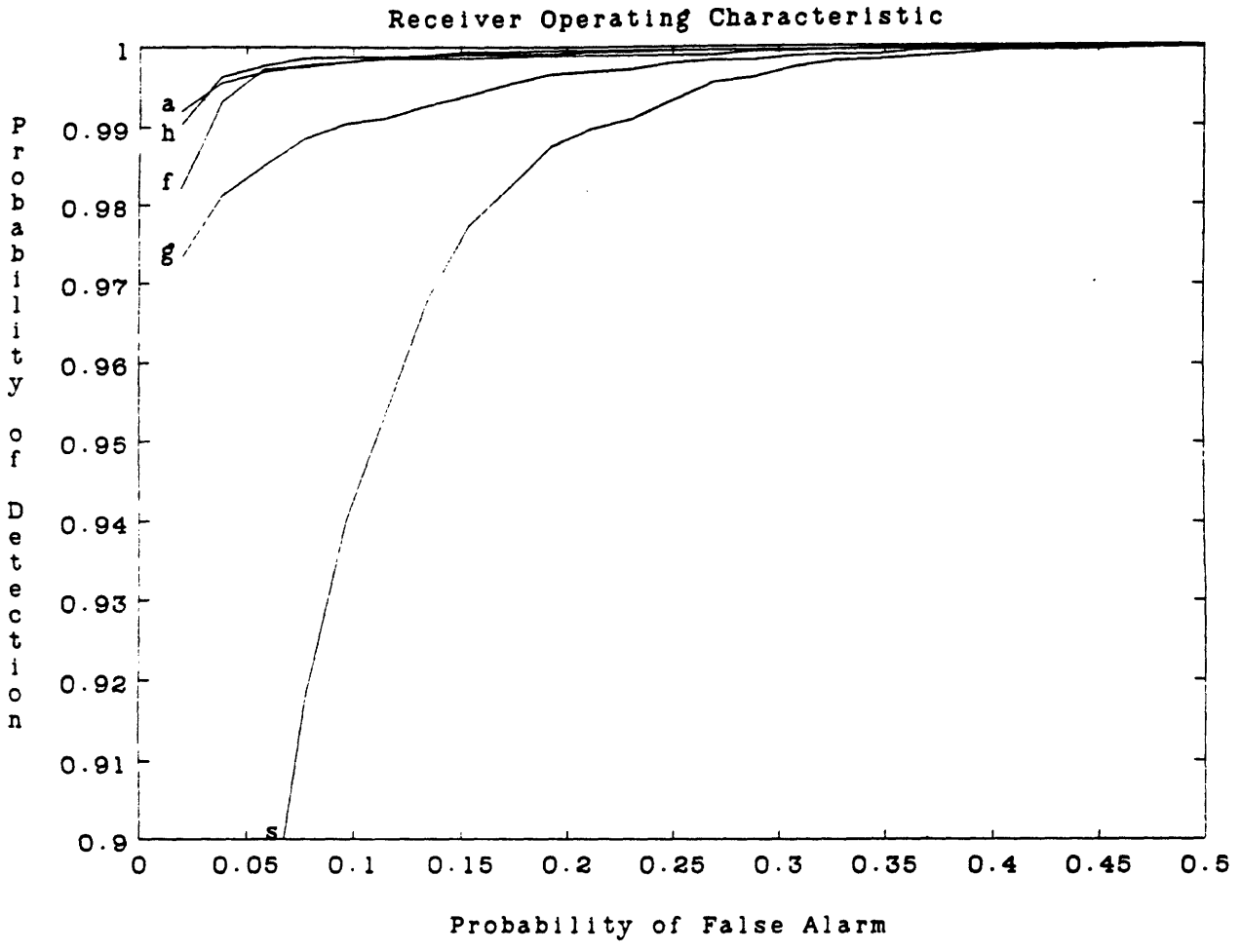

Figure 4.4: Comparison of ROCs of the statistics with the parameters: the number of channels(L) is 8; the semblance gate length is 1 ($M = 0$); the amplitude(A) of the narrowband component is 2.9; the length of data(N) is 1024 samples; the variance of WGN(σ^2) is 2000. 5000 trials are used. The narrowband component exists in 2 of 8 channels under the H_0 hypothesis. 'h' denotes $h_k(a) + s$, 'f' denotes Fisher discriminant function, 'g' denotes Gaussian discriminant function, 'a' denotes average, and 's' denotes semblance. The abscissa, indicating the probability of false alarm, ranges from 0 to 0.5 and the ordinate, indicating the probability of detection, ranges from 0.9 to 1.

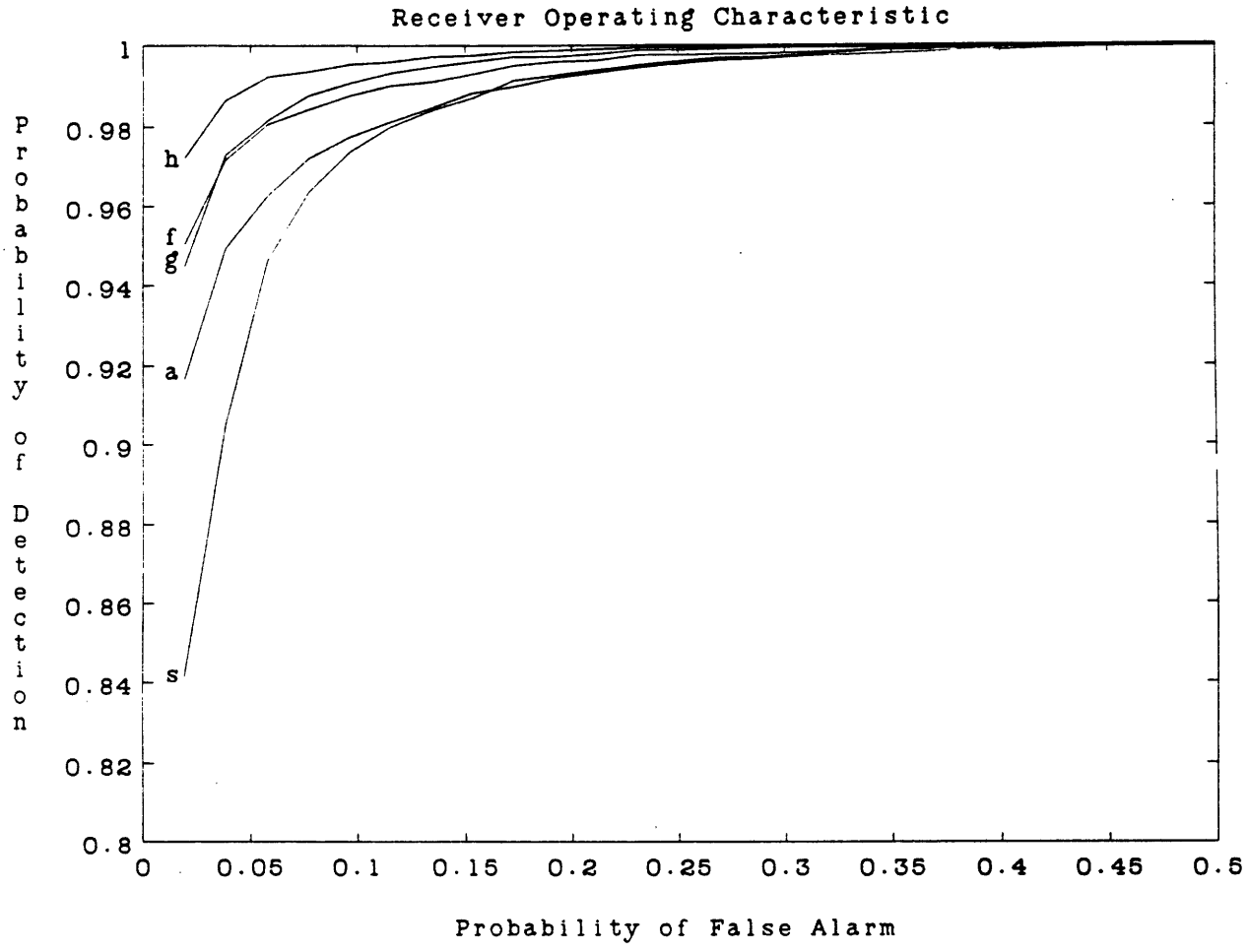

Figure 4.5: Comparison of ROCs of the statistics with the parameters: the number of channels(L) is 8; the semblance gate length is 1 ($M = 0$); the amplitude(A) of the narrowband component is 3.2; the length of data(N) is 1024 samples; the variance of WGN(σ^2) is 2000. 5000 trials are used. The narrowband component exists in 4 of 8 channels under the H_0 hypothesis. 'h' denotes $h_h(a) + s$, 'f' denotes Fisher discriminant function, 'g' denotes Gaussian discriminant function, 'a' denotes average, and 's' denotes semblance. The abscissa, indicating the probability of false alarm, ranges from 0 to 0.5 and the ordinate, indicating the probability of detection, ranges from 0.8 to 1.

Fig. 4.6 is computed using 5000 trials of the simulated data with the following parameters: 8 channels ($L = 8$), semblance gate length of 1 sample ($M = 0$), narrowband component amplitude of 3.5 ($A = 3.5$), data length of 1024 samples ($N = 1024$), and WGN variance of 2000 ($\sigma^2 = 2000$). It shows the ROCs for the case where the narrowband component exists in 6 of 8 channels under the H_0 hypothesis. The dashed line on the figure is the ROC curve of the multiple hypothesis detection of Section 3.5 obtained by varying α of (3.19). It clearly demonstrates the superiority of the combined statistic using the hyperbolic tangent transform, $h_r(a) + s$.

These figures indicate that the combined statistic, $h_h(a) + s$, performs well for a wide range of the H_0 hypothesis data and always as well as the average or the semblance alone. Therefore, $h_h(a) + s$ will be used as the detection statistic for the multichannel detection problem discussed in the next chapter.

4.5 Summary

In this chapter, a new detection statistic was developed by combining the average and the semblance. Transformations to match the PDF of the average to the PDF of the semblance were analyzed. It was concluded that the hyperbolic tangent function provides an accurate transformation. An alternative method based on the CDF matching was also discussed, however, it is more sensitive the samples used.

The combined statistic using the PDF matching by the hyperbolic tangent transformation was compared with the LRT statistic for the simple case. It was shown that the combined statistic performs as well as the optimal detection statistic. For more complicated cases, a closed form solution of the LRT statistic is unknown. Therefore, comparisons were made using the discriminant functions. The ROCs of the statistics were compared to demonstrate that the hyperbolic tangent transformation combination has a superior performance over a wide range of the H_0 hypothesis data.

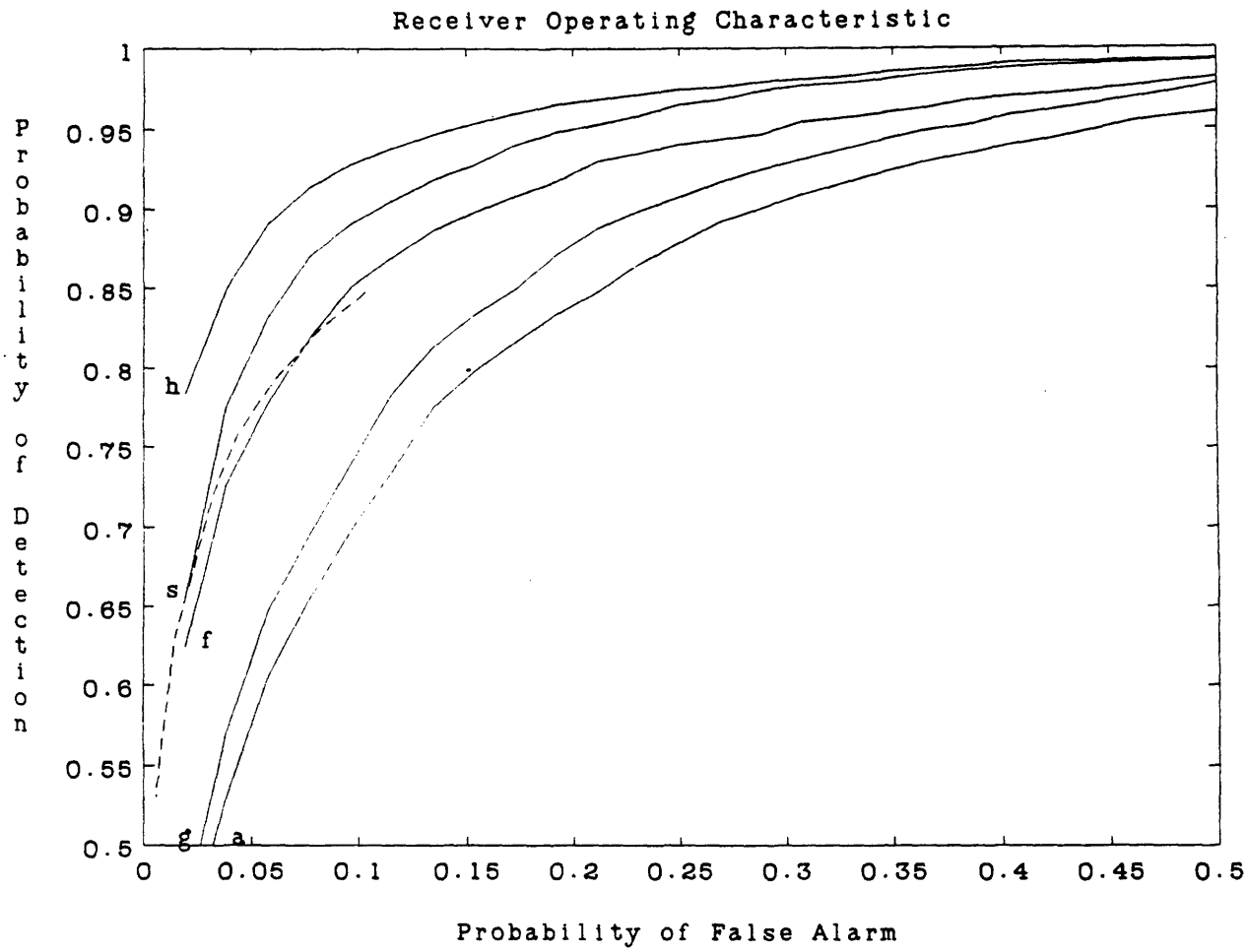

Figure 4.6: Comparison of ROCs of the statistics with the parameters: the number of channels(L) is 8; the semblance gate length is 1 ($M = 0$); the amplitude(A) of the narrowband component is 3.5; the length of data(N) is 1024 samples; the variance of WGN(σ^2) is 2000. 5000 trials are used. The narrowband component exists in 6 of 8 channels under the H_0 hypothesis. 'h' denotes $h_h(a) + s$, 'f' denotes Fisher discriminant function, 'g' denotes Gaussian discriminant function, 'a' denotes average, and 's' denotes semblance. The dashed line is the multiple hypothesis test detector of Section 3.5. The abscissa, indicating the probability of false alarm, ranges from 0 to 0.5 and the ordinate, indicating the probability of detection, ranges from 0.5 to 1.

Appendix 4.A

Combining Average and Semblance using the Likelihood Ratio Test: General Case

The likelihood ratio

$$\Lambda(a,s) = \frac{p_{a,s|H_1}(a,s|H_1)}{p_{a,s|H_0}(a,s|H_0)}$$

for an arbitrary number of channels and an arbitrary length of semblance gate is discussed. In Section 4.2, the simple case was examined. In this appendix, the general case is examined to show that a closed, analytical expression of the LRT statistic is unobtainable. The average, a, is defined as

$$a = \frac{1}{L}\sum_{l=1}^{L} x_l(0)$$

and the semblance, s, is defined as

$$s = \frac{\sum_{n=-M}^{M}\left(\sum_{l=1}^{L} x_l(n)\right)^2}{L\sum_{n=-M}^{M}\sum_{l=1}^{L} x_l(n)^2}$$

where L is the number of channels, $2M+1$ is the semblance gate size and $x_l(n)$ denotes the nth sample of the lth channel data. The joint PDF $p_{a,s}(a,s)$ is computed using $p_{\underline{X}}(\underline{X})$ where $\underline{X} = \begin{pmatrix} x_1(0) \\ x_2(0) \\ \underline{X}' \end{pmatrix}$ and $\underline{X}'$ is a vector containing all $x_l(n)$s for $l = 1,2,\cdots,L$ and $n = -M,-M+1,\cdots,M$ except $x_1(0)$ and $x_2(0)$. The random variables, $x_l(n)$s, are assumed to be independent. A new random vector is defined as

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \underline{Y}' \end{pmatrix} = \begin{pmatrix} a \\ s \\ \underline{Y}' \end{pmatrix} = \begin{pmatrix} g_1(\underline{X}) \\ g_2(\underline{X}) \\ \underline{X}' \end{pmatrix} \tag{4.A.1}$$

Separating $\underline{Y}$ into two groups is useful because $\underline{Y}'$ represents unwanted random variables. The PDF for the new random variables $\underline{Y}$ is determined by

$$p_{\underline{Y}}(\underline{Y}) = \frac{p_{\underline{X}}(\underline{X}_1^*)}{|J(\underline{X}_1^*)|} + \frac{p_{\underline{X}}(\underline{X}_2^*)}{|J(\underline{X}_2^*)|} \tag{4.A.2}$$

where $\underline{X}_1^*$ and $\underline{X}_2^*$ express $\underline{X}$ in terms of $\underline{Y}$ from (4.A.1). The Jacobian for the transformation (4.A.1) is

$$J(\underline{X}) = \begin{pmatrix} \frac{1}{L} & \frac{1}{L} & \frac{1}{L} & \frac{1}{L} & \frac{1}{L} & \cdots & \frac{1}{L} \\ b_1 - b_2 x_1 & b_1 - b_2 x_2 & b_1 - b_2 x_3 & b_1 - b_2 x_4 & b_1 - b_2 x_5 & \cdots & b_1 - b_2 x_{2LM} \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

where $b_1 = \frac{2}{L}\frac{\sum_{n=-M}^{M}\sum_{l=1}^{L} x_l(n)}{\sum_{n=-M}^{M}\sum_{l=1}^{L} x_l(n)^2}$, $b_2 = \frac{2}{L}\frac{\sum_{n=-M}^{M}(\sum_{l=1}^{L} x_l(n))^2}{(\sum_{n=-M}^{M}\sum_{l=1}^{L} x_l(n)^2)^2}$. Hence,

$$|J(\underline{X})| = \frac{b_2}{L}|x_1(0) - x_2(0)|. \tag{4.A.3}$$

Using the definition of the average,

$$x_2(0) = Ly_1 - x_1(0) - \sum_{l=3}^{L} x_l(0) \tag{4.A.4}$$

and using the definition of the semblance

$$y_2 = \frac{L^2 y_1^2 + \sum_{\substack{n=-M \\ n\neq 0}}^{M}(\sum_{l=1}^{L} x_l(n))^2}{L(x_1(0)^2 + (Ly_1 - x_1(0) - \sum_{l=3}^{L} x_l(0))^2 + \sum_{\substack{n=-M \\ n\neq 0}}^{M}\sum_{l=1}^{L} x_l(n)^2)}.$$

The above equation is solved for $x_1(0)$ using the quadratic formula, then $x_2(0)$ is determined from (4.A.4).

$$x_1(0) = (\frac{L}{2}y_1 + \frac{1}{2}w) \pm \sqrt{v} \tag{4.A.5}$$

$$x_2(0) = (\frac{L}{2}y_1 - \frac{3}{2}w) \mp \sqrt{v} \tag{4.A.6}$$

with

$$v = y_1(\frac{L}{2y_2} - \frac{L^2}{4}) - \frac{L}{2}y_1 w - \frac{w^2}{4} - \frac{1}{2}\sum_{\substack{n=-M \\ n\neq 0}}^{M} x_l(n)^2 + \frac{1}{2L^2 y_2^2}\sum_{\substack{n=-M \\ n\neq 0}}^{M}(\sum_{l=1}^{L} x_l(n))^2$$

and

$$w = \sum_{\substack{n=-M \\ n\neq 0}}^{M} x_l(n).$$

Unfortunately, replacing (4.A.5) and (4.A.6) in (4.A.3) reveals that $|J(\underline{X}_1^*)| \neq |J(\underline{X}_2^*)|$. The expression (4.A.2) cannot be simplified further and the desired joint PDF must be determined by

$$p_{a,s}(a,s) = \int_{-\infty}^{\infty} p_{\underline{Y}}(\underline{Y}) d\underline{Y}' = \int_{-\infty}^{\infty} \left(\frac{p_{\underline{X}}(\underline{X}_1^*)}{|J(\underline{X}_1^*)|} + \frac{p_{\underline{X}}(\underline{X}_2^*)}{|J(\underline{X}_2^*)|} \right) d\underline{Y}'$$

where the integration represents $L(2M+1)-2$ dimensional integral. This cannot be evaluated analytically and, therefore, the likelihood ratio detection statistic for the general case cannot be expressed in a closed form.

Chapter 5

Gravitational Wave Signal Detection

In the previous chapters, the multichannel detection problem, which has the requirement that the narrowband component must exist in all channels in order to decide that an emitter exists, was examined. A critical aspect of the problem is that the signal-like narrowband noise can exist when an emitter is absent. In Chapter 4 an improved detection statistic was developed by combining the transformed average and the semblance. In this chapter the combined statistic will be applied to the gravitational wave(GW) signal detection problem.

This chapter begins with the background of the gravitational wave signal processing problem in Section 5.1. The received periodic GW data can be viewed as a multichannel data and the detection of a GW emitter becomes the multichannel detection problem discussed in Chapter 2. The goal of GW signal processing is to accurately and efficiently detect frequency-modulated, periodic gravitational wave signals with unknown emitter location and frequency. Because the parameters of the emitter are unknown, GW signal detection algorithms first hypothesize an emitter location and frequency, then process the received data to detect the hypothesized emitter. The algorithm must test all directions and all frequencies which is a computationally demanding task. In this chapter, an algorithm will be developed which tests all possible frequencies simultaneously for a given emitter location.

As will be shown in Section 5.2, selecting the location and frequency of an emitter completely

determines the instantaneous frequency of the received signal, and hence uniquely specifies the narrowband component in each channel. Before the detection statistics can be applied, periodic GW signal detection algorithms must estimate the narrowband component of the hypothetical emitter in each channel. This entails computing the Fourier transform magnitude of the data in each channel at the frequency predicted by the emitter location and is due to the inherent structure of the data.

A formal description of the GW signal detection problem viewed as a multichannel detection problem is presented in Section 5.3. The algorithm developed by Livas[39], which is described in Section 5.4, resamples the received data and then uses the FFT to compute the magnitudes at all frequencies for a given emitter location. The detection statistic employed by the Livas algorithm is the average. In Section 5.5, the combined detection statistic developed in Chapter 4 is employed because the GW signal detection problem fits the multichannel detection problem description. In addition, a more accurate algorithm which avoids the resampling operation of the Livas algorithm is developed. This new algorithm uses the chirp z-transform to compute the required Fourier transform magnitudes in each channel. In Section 5.6, the new algorithm is used to process the GW data and compared against the Livas algorithm.

5.1 Background

When matter in the universe moves, the changes in its gravitational field are propagated throughout space and give rise to gravitational waves[61,67]. Gravitational radiation or gravitational waves are predicted to propagate at the speed of light by the general theory of relativity. Recently, several attempts have been made to detect gravitational waves from astrophysical emitters[68] because the detection of gravitational wave emissions will aid in understanding astrophysical phenomena.

There are two types of gravitational wave receivers. The first type is a bar antenna which uses a large homogeneous cylinder of material such as aluminum. The bar is acoustically isolated to oscillate only by passing gravitational waves. The second type of GW receiver is a laser interferometric antenna which uses a laser beam to measure changes in the separation distance of free masses caused by passing gravitational waves. In addition to the current international gravitational wave detection effort, an ambitious new system called the Laser Interferometer

Gravitational Wave Observatory is proposed[73]. It consists of two physically separate sites to provide redundant measurements and isolate local disturbances.[1] In addition, the receivers will be much more sensitive than existing systems.

There are three types of gravitational waves: periodic, impulsive, and stochastic. Periodic gravitational waves are generated by, for example, the orbiting motion of binary star systems and are steady-state signals. Impulsive gravitational waves are generated by, for example, collisions of a black hole with other objects or collapses of supernovae and are transient signals which last only a few milliseconds. Stochastic gravitational waves are generated by the superposition of many weak overlapping impulsive gravitational radiations and are background radiation signals with no clear time or frequency domain structure. A search for impulsive signals has been made by Dewey[9] using a search algorithm based on correlating the received data with a set of templates. Because impulsive signals have only a short duration, continuous on-line processing is required. A search for periodic signals has been made by Livas[39] using an algorithm which is based on threshold testing the average of the Fourier transform magnitudes of the received data.

In this chapter, the periodic GW signal detection problem is examined. Although the GW signal exists for all time, it is measured in short bursts because of the constraints imposed by hardware and local disturbances. For convenience, each burst of data is treated as the output of a channel of a multichannel receiver. Due to the relative motion between the emitter and the receiver, the received GW signal experiences frequency modulation. In addition, the received data is contaminated by local noises, some of which are narrowband. Fortunately, however, local narrowband noises exist intermittently or have a constant frequency over all channels. The frequency variation of the GW signal over the channels and the existence of the GW signal in all channels are the crucial features which are exploited to discard the false alarms caused by the local narrowband noises. Thus, the critical requirements for detection are the existence of the signal in all channels and the changes in the instantaneous frequency of the signal.

[1] In case of impulsive GW signals, multiple sites will provide emitter location information.

5.2 Description of Received Gravitational Wave Signal

In this section[2] a detailed derivation of the instantaneous frequency of the received GW signal is presented. If a periodic GW signal emitter[7,61,68,67,73], such as a binary star system, emits a periodic signal of frequency f_0, the frequency of the received signal is

$$f_r = f_0\gamma_0(1 + \frac{\vec{v}}{C} \cdot \vec{r}) \tag{5.1}$$

where $\gamma_0 = \frac{1}{\sqrt{1-(|\vec{v}|/C)^2}}$, $\vec{v}$ is the relative velocity vector between the emitter and the receiver, C is the speed of light, and $\vec{r}$ is the unit vector from the receiver to the emitter. The relative velocity vector, $\vec{v}$, is the sum of the rotational motion vector of the earth, the orbital motion vector of the earth around the solar system barycenter, and the motion vector of the solar system barycenter through the galaxy.[3] If the velocity vector remains constant during the measurement, its effect on the received signal is a simple shift in frequency. However, if the velocity vector changes appreciably during the measurement, its effect on the received signal is a more complicated form of frequency modulation. The solar system barycenter motion vector does not change appreciably over the measurement period of the data used in this chapter. Therefore, its contribution to the received signal is an overall frequency shift, and the solar system barycenter motion is ignored. The relative velocity vector, $\vec{v}$, is

$$\vec{v} = \vec{v}_o + \vec{v}_r \tag{5.2}$$

where $\vec{v}_o$ is the orbital motion vector of the earth around the solar system barycenter and $\vec{v}_r$ is the rotation motion vector of the earth. The total velocity is small compared to C because

$$|\vec{v}| = |\vec{v}_o + \vec{v}_r| \leq |\vec{v}_o| + |\vec{v}_r| \approx 3.01 \times 10^6 cm/sec.$$

Thus, $\gamma_0 = \frac{1}{\sqrt{1-(|\vec{v}|/C)^2}} \approx 1$ and (5.1) becomes

$$f_r = f_0(1 + \frac{\vec{v}}{C} \cdot \vec{r}).$$

Figure 5.1 shows the coordinate system used to determine $\vec{v}$ and $\vec{r}$. As shown in Fig. 5.1a,

[2] This section is based on discussions with Dr. Jeffery Livas and on his thesis[39].

[3] Inter-galactic motion could be added to this list but the current gravitational wave receivers are not sensitive enough for the signals emitted from other galaxies. Therefore, inter-galactic motion is ignored.

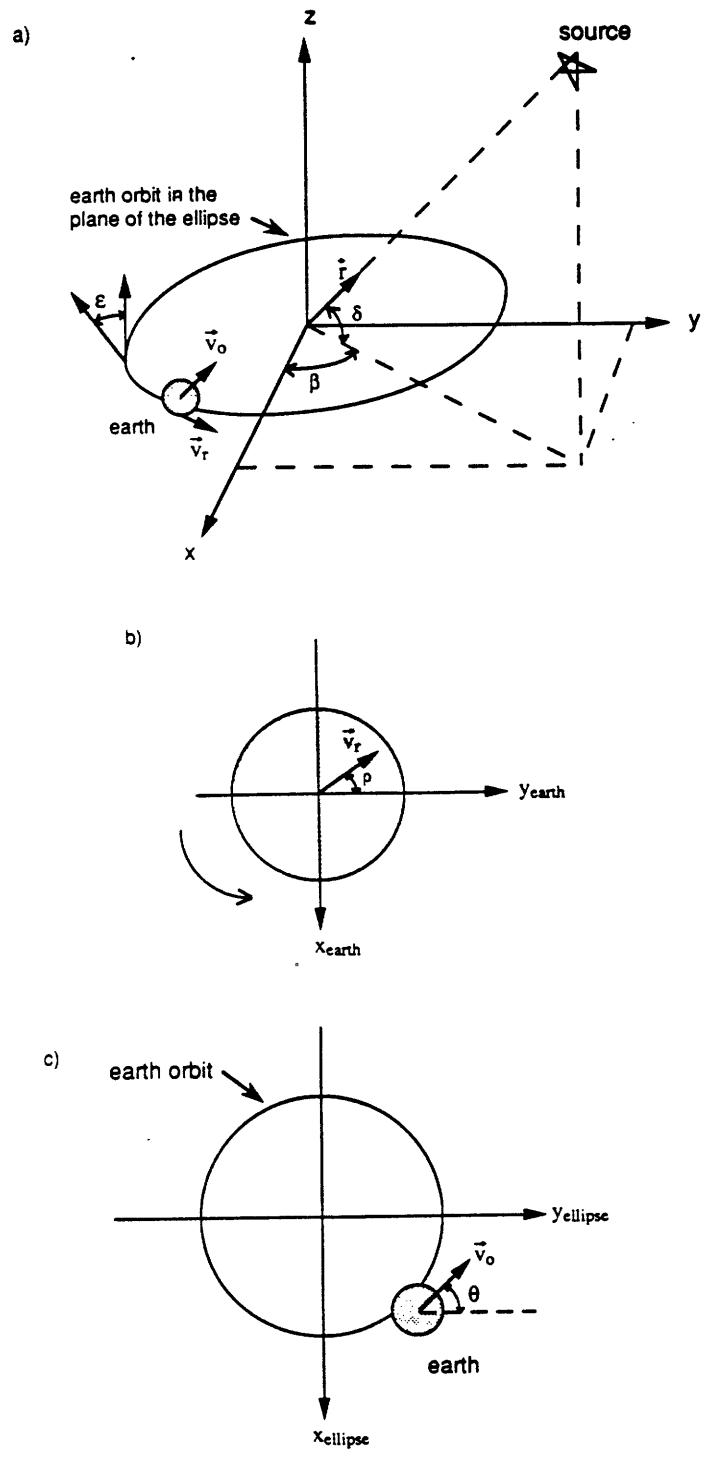

Figure 5.1: (a) Solar system barycenter coordinate system for the earth's orbit and the emitter. (b) Top view of the earth's axis for the earth's rotation. (c) Top view of the plane of the ellipse for the earth's orbit.

the earth's orbit is assumed to be circular in the plane of the ellipse. The earth's rotation is assumed to be circular about the earth's spin axis which is the z-axis in the figure. From Fig. 5.1b, $\vec{v}_r = |\vec{v}_r|(-\sin\rho, \cos\rho, 0)$ where ρ is the rotational angle. Additionally, the receiver's location on the earth must be included. Therefore, $\vec{v}_r = |\vec{v}_r|\ \cos\lambda\ (-\sin\rho, \cos\rho, 0)$ where λ is the latitude of the receiver location.

From Fig. 5.1c, the earth's orbital velocity vector in the plane of the ellipse, which is denoted by $\vec{v'}_o$, is $\vec{v'}_o = |\vec{v'}_o|(-\sin\theta, \cos\theta, 0)$. To determine the earth's orbital velocity in the solar system barycenter coordinate (x, y, z), the plane of the ellipse must be rotated about the x axis. The angle of rotation ϵ is shown in Fig. 5.1a. This angle is formed by the plane of the ellipse and the $x-y$ plane, or, equivalently, the tilt angle of the earth's spin axis. Because the rotation should move the plane of the ellipse onto the $x-y$ plane, the rotation matrix is

$$R_\epsilon = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\epsilon & \sin\epsilon \\ 0 & -\sin\epsilon & \cos\epsilon \end{pmatrix}.$$

Therefore, $\vec{v}_o = \vec{v'}_o\ R_\epsilon = |\vec{v}_o|\ (-\sin\theta,\ \cos\theta\cos\epsilon,\ \cos\theta\sin\epsilon)$ and from (5.2) the total relative velocity is

$$\vec{v} = |\vec{v}_o|\ (-\sin\theta,\ \cos\theta\cos\epsilon,\ \cos\theta\sin\epsilon) + |\vec{v}_r|\ \cos\lambda\ (-\sin\rho, \cos\rho, 0). \tag{5.3}$$

The unit vector from the receiver to the emitter is approximately equal to the unit vector from the solar system barycenter to the emitter because the distance from the earth to the solar system barycenter is negligible. From Fig. 5.1a, the distance vector is

$$\vec{r} = (\cos\delta\cos\beta, \cos\delta\sin\beta, \sin\delta)$$

which is used in (5.3) to obtain

$$\begin{aligned} \vec{v}\cdot\vec{r} &= |\vec{v}_o|(\cos\theta(\cos\epsilon\cos\delta\sin\beta + \sin\epsilon\sin\delta) - \sin\theta\cos\delta\cos\beta) \\ &\quad + |\vec{v}_r|\cos\lambda\cos\delta(\cos\rho\sin\beta - \sin\rho\cos\beta) \\ &= |\vec{v}_o| G\cos(\theta+\phi) + |\vec{v}_r|\cos\lambda\cos\delta\sin(\beta-\rho) \end{aligned} \tag{5.4}$$

where

$$G = \sqrt{(\cos\delta\cos\beta)^2 + (\cos\epsilon\cos\delta\sin\beta + \sin\epsilon\sin\delta)^2}$$

speed of light	$C = 2.99792458 \times 10^{10} cm/sec$
earth's radius	$R_e = 6.37103 \times 10^{8} cm$
earth's rotational angular speed	$\omega_r = 7.292115855 \times 10^{-5} radians/sec$
earth's rotational velocity ($\|\vec{v}_r\| = R_e \cdot \omega_r$)	$\|\vec{v}_r\| = 3.4326 \times 10^{4} cm/sec$
earth's rotational velocity at latitude λ	$\|\vec{v}_r(\lambda)\| = \cos\lambda \cdot\ 3.4326 \times 10^{4} cm/sec$
earth's orbital radius	$R_o = 1.474204307 \times 10^{13} cm$
earth's orbital angular speed	$\omega_o = 1.99098659 \times 10^{-7} radians/sec$
earth's orbital velocity ($\|\vec{v}_o\| = R_o \cdot \omega_o$)	$\|\vec{v}_o\| = 2.9816 \times 10^{6} cm/sec$
earth's spin axis tilt	$\epsilon = 0.4080\ radians$
latitude of Cambridge, MA	$\lambda = 0.739496\ radians$
right ascension of galactic center	$\beta = 4.63822\ radians$
declination of galactic center	$\delta = -0.5044\ radians$
initial phase of orbital velocity	$\theta_0 = 4.416635728\ radians$
initial phase of rotational velocity	$\rho_0 = 3.159827708\ radians$

Table 5.1: Constants used in the gravitational wave signal detection

and

$$\tan\phi = \frac{\cos\delta\cos\beta}{\cos\epsilon\cos\delta\sin\beta + \sin\epsilon\sin\delta}.$$

Finally, using (5.4) in (5.1) and showing the time dependence of f_r due to the earth's motion, the frequency of the received signal is

$$f_r(t) = f_0\left(1 + \frac{|\vec{v}_o|}{C}G\cos(\phi + \theta_0 + \omega_o t) + \frac{|\vec{v}_r|}{C}\cos\lambda\cos\delta\sin(\beta - \rho_0 - \omega_r t)\right) \qquad (5.5)$$

where θ_0 is the initial phase of the orbital velocity, ρ_0 is the initial phase of the rotational velocity, ω_o is the earth's orbital angular speed, and ω_r is the earth's rotational angular speed. The constants used in the calculations are collected in Table 5.1.

The above equation shows that the frequency of the received signal is completely specified by the emitter frequency, f_0, and the emitter location angles, δ and β. By assuming the location to be the galactic center and the frequency to be 1 KHz for a hypothesized emitter, an example of the received signal frequency $f_r(t) - f_0$ is plotted in Fig. 5.2.

5.3 Detection Using the Measured Data

The data used in this chapter were measured by J. Livas using the MIT 1.5 meter prototype interferometric gravitational wave receiver[9,39,75]. The data were collected between 12PM and 9AM on each night between June 3 - 10, 1985. The data span eight days but the actual

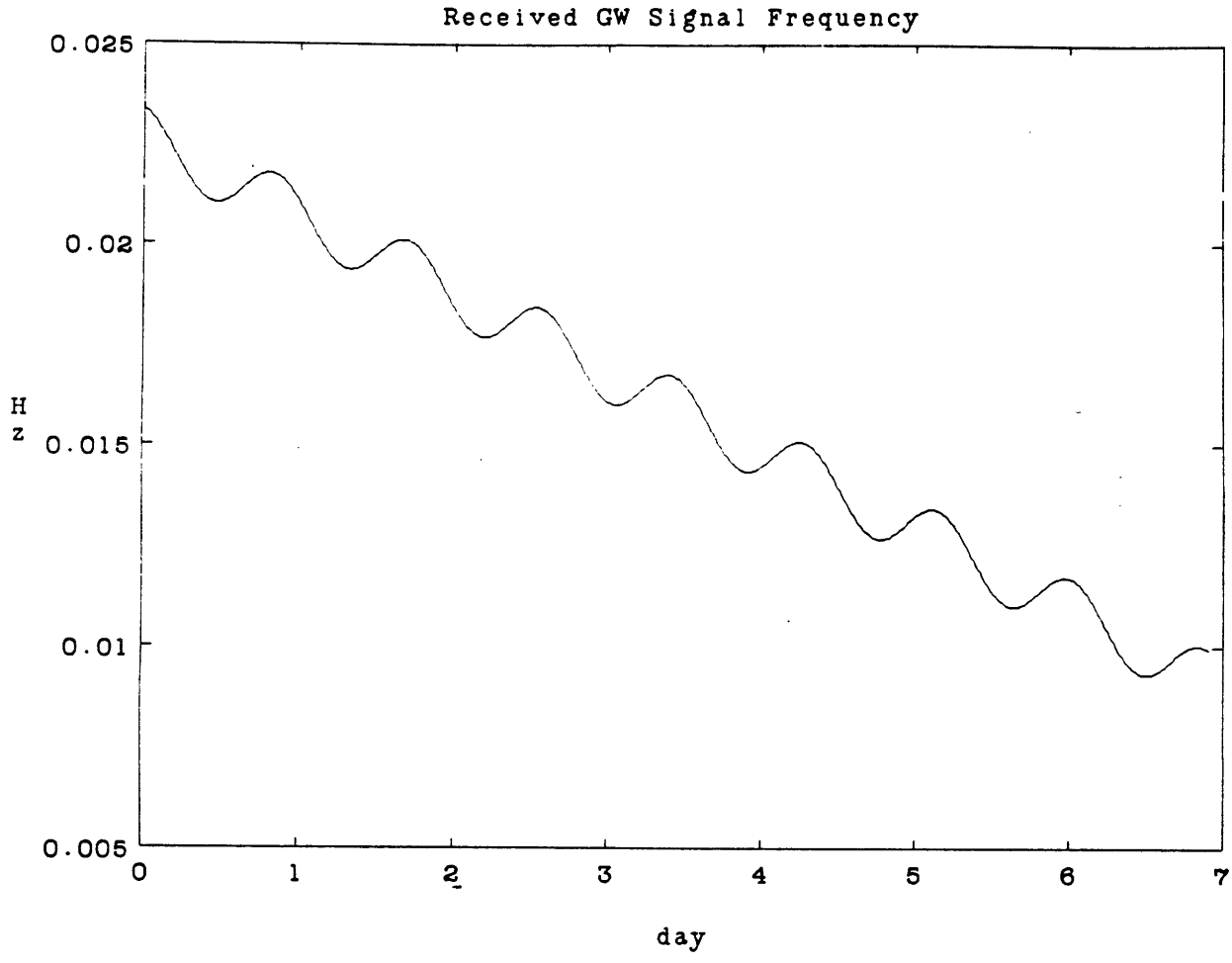

Figure 5.2: The difference between the instantaneous frequency of the received signal and the frequency of the transmitted signal, $f_r(t) - f_0$, for a hypothesized 1 KHz emitter located at the galactic center.

Data segment label	*Starting time*
J4D	1:59:47.1 June 5, 1985
J4F	2:45:38.7 June 5, 1985
J7D	2:59:22.1 June 8, 1985
J8A	22:35:49.1 June 8, 1985
J8I	2:45:46.6 June 9, 1985
J8K	3:36:04.0 June 9, 1985
J8M	4:27:56.5 June 9, 1985
J9C	4:38:01.4 June 10, 1985

Table 5.2: The starting times of the gravitational wave data set.

measurements cover only a small portion of the total time because the measurements were made continuously for approximately 15 minutes. The sampling rate was 20K samples/second and a 12 bit A/D converter was used. The data set used in this chapter consists of eight 15 minute segments as shown on Table 5.2.

The data in the lth segment span $T_l \leq t \leq T_l + \Delta_T$, where T_l is the starting time and Δ_T is the duration of the measurement. Because $\omega_o \times 15min \approx 0$ and $\omega_r \times 15min \approx 0$, for $\Delta_T = 15$ minutes,

$$|f_r(t) - f_r(t + \Delta_T)| \ll 1.$$

Therefore, the instantaneous frequency (5.5) of the narrowband component in the lth segment is approximated by a constant

$$f_r(T_l) = f_0 \left(1 + \frac{v_o}{C} G \cos(\phi + \theta_0 + \omega_o T_l) + \frac{v_r}{C} \cos\lambda \cos\delta \sin(\beta - \rho_0 - \omega_r T_l)\right). \tag{5.6}$$

The narrowband component contained in the lth segment, which is viewed as the lth channel of a multichannel receiver, is

$$s_l(n) = A\cos(2\pi f_r(T_l)\, n + \phi_l).$$

Using this interpretation of the measured data, the GW signal detection problem becomes a special case of the multichannel detection problem described in Chapter 2. A GW signal emitter is present only if the narrowband component, $s_l(n)$, exists for all channels $l = 1, 2, \cdots, L$. One inherent feature of the GW signal detection problem is that the unknown frequency of the signal in each channel is not estimated, rather it is predicted by the emitter location and frequency. However, because the emitter location and frequency are unknown, both are hypothesized.

For each hypothesized emitter, the detection algorithm must compute the Fourier transform magnitudes at the frequencies predicted by (5.6) in each channel, then compute a detection statistic using the magnitudes. In the following sections, an existing algorithm is reviewed and an improved algorithm is developed.

5.4 Previous Detection Algorithm

In this section, the periodic GW signal detection algorithm introduced by Livas[39] is described.[4] In Section 5.6, it will be compared to the new algorithm which will be developed in the next section. This algorithm selects an emitter location, then the Fourier transform magnitudes of a hypothesized emitter are computed at the frequencies specified by (5.6). The detection statistic is computed by averaging the Fourier transform magnitudes. To reduce computation, it tests all frequencies simultaneously for a given emitter location as described next.

The phase of the received signal is given by

$$\phi(t) = 2\pi \int_{T_1}^{t} f_r(\tau)d\tau + \phi_0 = 2\pi f_0 \int_{T_1}^{t} g(\tau, \delta, \beta)d\tau + \phi_0$$

where ϕ_0 is a constant and from (5.6)

$$g(t, \delta, \beta) = 1 + \frac{v_o}{C} G \cos(\phi + \theta_0 + \omega_o t) + \frac{v_r}{C} \cos\lambda \cos\delta \sin(\beta - \rho_0 - \omega_r t).$$

A new variable t' is defined as

$$t' = \int_{T_1}^{t} g(\tau, \delta, \beta)d\tau.$$

Using the approximation (5.6), t' for the lth channel becomes

$$t' = g(T_l, \delta, \beta)\,(t - T_1) \tag{5.7}$$

for $T_l \leq t \leq T_l + \Delta_T$. If the signal is sampled uniformly in t', the frequency modulation will be compensated for, hence, resulting in a sinusoid with a constant frequency. The algorithm

[4]Livas introduced two algorithms. The first algorithm, which is named *all sky search*, computes the FFT of each channel data segment separately. The computed Fourier transform magnitudes are examined to select a set of peaks which are sorted according to physically realizable doppler slope. These sorted peaks are used as the discriminant in detection. Because the probability of detection of this algorithm is inferior to his second algorithm, which is named *single direction search*, hence only the second algorithm is considered here.

suggested by Livas implements the resampling in the following manner. The lth channel data, $r_l(t)$, which is uniformly sampled in t, is viewed as $r_l(t')$, which is sampled non-uniformly in t'. The uniformly sampled values in t' are computed by interpolating the non-uniform samples. Because the exact bandlimited interpolation requires an excessive number of computations, Livas approximated it using a cubic spline interpolation[39]. The resampling normalizes the frequencies for all f_0's simultaneously. Hence, all possible emitter frequencies can be tested by computing one FFT of the resampled data of each channel. The average of the Fourier transform magnitudes is used as the detection statistic by Livas. To summarize, the Livas algorithm chooses a hypothetical emitter location by specifying δ and β, and then resamples the data in each channel using (5.7). The FFT of the resampled data is computed and the magnitudes are averaged. Each bin of the averaged FFT magnitudes is hypothesis tested for the existence of the GW emitter. However, as discussed in Chapter 4, the detection performance of the average is often unsatisfactory for the multichannel detection problem if the requirement for detection is that the narrowband component must be present in all channels. The probability of detection can be improved if the combined statistic developed in Chapter 4 is used. Before the detection statistics are compared, a new algorithm for computing the Fourier transform magnitudes is developed next.

5.5 New Algorithm for Estimating the Narrowband Component

In this section, a new GW signal detection algorithm is developed. This new algorithm employs the multichannel detection statistic developed in Chapter 4 because the GW signal detection problem fits the multichannel signal detection model as discussed in Section 5.3. Specifically, $h_h(a) + s$ is used as a detection statistic where a is the average, s is the semblance, and $h_h(a)$ is the hyperbolic tangent transformation function of Section 4.1.1. The detection statistics are computed using the estimated narrowband component of each channel, therefore, an improved estimation algorithm which uses the chirp z-transform is developed next. It computes the exact frequency samples with approximately the same computational complexity as the Livas algorithm.

When the emitter is located at (δ_1, β_1), the frequency of the lth channel data is given by (5.6) which was denoted as $f_r(t) = f_0\, g(t, \delta_1, \beta_1)$. If a particular frequency, f_1, is selected as the frequency in the first channel, $f_1 = f_r(T_1)$, then the emitter frequency, f_0, is

$$f_0 = \frac{f_1}{g(T_1, \delta_1, \beta_1)}.$$

Thus the frequency in the lth channel must be

$$f(T_l) = f_0\, g(T_l, \delta_1, \beta_1) = f_1 \frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)}.$$

Similarly, if $f_1 + \Delta_f$ is selected as the frequency in the first channel, then the corresponding emitter frequency is

$$f_0 = \frac{f_1 + \Delta_f}{g(T_1, \delta_1, \beta_1)} = \frac{f_1}{g(T_1, \delta_1, \beta_1)} + \frac{\Delta_f}{g(T_1, \delta_1, \beta_1)}$$

and the frequency in the lth channel must be

$$f(T_l) = \frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)} f_1 + \frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)} \Delta_f.$$

Therefore, if the frequency is incremented by Δ_f in the first channel, it is incremented by

$$\frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)} \Delta_f$$

in the lth channel. This observation about the increase in the frequency sampling step is used to test all possible emitter frequencies simultaneously.

The N-length FFT of the first channel data is computed which assumes that $\Delta_f = \frac{2\pi}{N}$. The required frequency sampling interval in the lth channel must be

$$\frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)} \Delta_f = \frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)} \frac{2\pi}{N}.$$

Unfortunately, the frequency values of the above sampling interval cannot be computed by the FFT. However, an efficient computation of the frequency samples is still possible by the chirp z-transform[50] which computes the samples in the z-plane as

$$X(z_k) = \sum_{n=0}^{N-1} x(n) A^{-n} W^{nk}$$

for $k = 0, 1, \cdots, M-1$. To compute the lth channel frequency samples, let

$$A = 1 \quad \text{and} \quad W = \exp(j \frac{g(T_l, \delta_1, \beta_1)}{g(T_1, \delta_1, \beta_1)} \Delta_f).$$

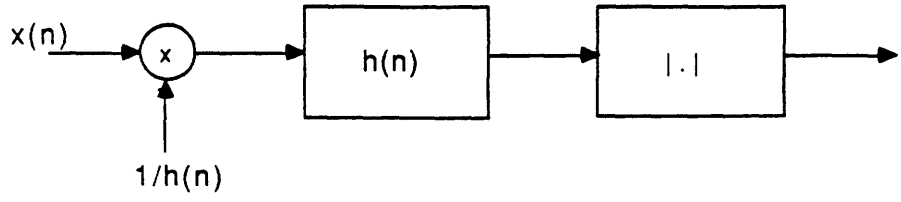

Figure 5.3: Computation of the Fourier transform magnitude using the chirp z-transform where $h(n) = W^{-n^2/2} = \exp(-j\frac{\Delta_f}{2}\frac{g(T_l,\delta_1,\beta_1)}{g(T_1,\delta_1,\beta_1)}n^2)$.

Because only the Fourier transform magnitude values are required, the post-multiplication in the chirp z-transform computation is unnecessary. Figure 5.3 shows the method based on the chirp z-transform.

The number of multiplications required in the chirp z-transform based computation method is approximated next. The N-length FFT of the first channel requires $\frac{N}{2}\log_2 N$ multiplications. Additionally, there are $L-1$ remaining channels each of which must be chirp z-transformed. This requires $(L-1)\times(N+N\log_2(2N))$ multiplications, hence the total number of multiplications is

$$(\frac{1}{2}+(L-1))N\log_2 N+(L-1)N \;\propto\; O(N\log_2 N).$$

In comparison, the Livas method requires

$$\frac{LN}{2}\log_2 N+4(L-1)N \;\propto\; O(N\log_2 N)$$

because L FFTs are computed and each cubic spline requires $4N$ multiplications[55,64]. Thus the chirp z-transform based algorithm requires approximately the same number of multiplications but computes the Fourier transform magnitudes at the exact frequency locations.

Because the frequency sampling interval of the lth channel is scaled by $g(T_l,\delta_1,\beta_1)/g(T_1,\delta_1,\beta_1)$, not all frequency samples can be used if the sampling theorem is to be satisfied. By assuming

$$\frac{g(T_l,\delta_1,\beta_1)}{g(T_1,\delta_1,\beta_1)} \leq 1,$$

the largest FFT bin number of the first channel which can be used without violating the sampling theorem is determined. Let the highest frequency used in the lth channel be π. Its

corresponding frequency in the first channel is $k\frac{2\pi}{N}$ where k satisfies

$$k\frac{2\pi}{N}\frac{g(T_l,\delta_1,\beta_1)}{g(T_1,\delta_1,\beta_1)} = \pi.$$

Therefore, the highest allowed FFT bin in the first channel is

$$k = \frac{N}{2}\frac{g(T_l,\delta_1,\beta_1)}{g(T_1,\delta_1,\beta_1)}$$

and only $\frac{g(T_l,\delta_1,\beta_1)}{g(T_1,\delta_1,\beta_1)}$ of the total spectrum should be used to avoid aliasing.

When a new emitter location, which is specified by (δ_2,β_2), is to be tested, the frequency samples must be recomputed. However, if the frequency samples can be approximated with the samples located within one half of the frequency sampling step size, a significant number of computations can be eliminated. In the lth channel, the emitter located at (δ_1,β_1) requires a frequency sampling step size of $\mu_1\Delta_f$ and the emitter located at (δ_2,β_2) requires a step size of $\mu_2\Delta_f$ where

$$\mu_1 = \frac{g(T_l,\delta_1,\beta_1)}{g(T_1,\delta_1,\beta_1)}\Delta_f \quad \text{and} \quad \mu_2 = \frac{g(T_l,\delta_2,\beta_2)}{g(T_1,\delta_2,\beta_2)}\Delta_f.$$

Let k be the highest frequency sample number allowed by the sampling theorem. The same frequency samples can be used for both emitter locations if

$$|k\mu_1\Delta_f - k\mu_2\Delta_f| < \frac{\mu_1\Delta_f}{2}$$

is satisfied. Because $k = \frac{N}{2\mu_1}$, the required condition on μ_1 and μ_2 becomes

$$|\mu_1 - \mu_2| < \frac{\mu_1^2}{N}.$$

After the Fourier transform magnitudes at the frequency locations are estimated, the detection statistic is computed by combining the transformed average and the semblance.

5.6 Comparison

The GW signal detection algorithms use the Fourier transform magnitudes to compute a detection statistic which is, then, threshold tested to decide the existence of a GW emitter. The Fourier transform magnitudes in each channel are estimated at the predicted frequencies using the location of the hypothesized emitter. The Livas algorithm employs the average of

the Fourier transform magnitudes computed using cubic spline approximation as the detection statistic. The new algorithm employs $h_h(a)+s$ as the detection statistic where the average and the semblance are computed using the Fourier transform magnitudes at the exact frequency locations using the chirp z-transform algorithm. In this section, the new algorithm is compared to the Livas algorithm. The accuracy of the magnitude estimates and the effectiveness of the detection statistics are separately compared. The computation of the Fourier transform magnitudes is analyzed first, then the use of different detection statistics is discussed.

The error in the computed Fourier transform magnitudes due to the cubic spline approximation is difficult to determine generally. However, because the GW signal is assumed to be a sinusoid in each channel, the cubic spline approximation error can be experimentally determined by processing a test signal $\cos(g_0 f_0 \pi t)$ where $t = n\Delta$ for $n = 0, 1, \cdots, N-1$, with the sampling interval, Δ. The frequency, f_0, denotes the emitter frequency and g_0 denotes a particular value of $g(t, \delta, \beta)$ obtained by specifying t to be some T_l and choosing emitter location angles, δ_0 and β_0. The Livas demodulation algorithm resamples the signal uniformly in t', which is defined to be $t' = g_0\ (t - T_l)$, using the cubic spline algorithm with zero second derivative condition on both boundaries[55]. After the data at $t' = n\Delta$ for $n = 0, 1, \cdots, N-1$ are approximated, the Fourier transform of this resampled data is computed. The magnitude value at the frequency, f_0, is compared to the correct value of $N/2$.

Table 5.3 shows the computed Fourier transform magnitudes normalized by the correct value. These values are dependent on the frequency, f_0, the data length, N, and $g_0 = g(t, \delta, \beta)$. However, this table represents a typical case in which the following parameters are used: $f_0 = 0.5$ where 1.0 is the highest frequency; $N = 1024$ or 4096; and $g_0 = 0.995, 0.996, \cdots, 1.005$. Table 5.3 also shows the Fourier transform magnitudes computed with the new algorithm based on the chirpz-transform with $W = \exp(j\, g_0 \frac{2\pi}{N})$. As expected, the computed magnitude of both algorithms is exact when $g_0 = 1$. When $g(t, \delta, \beta) < 1$, the cubic spline algorithm has significant error while the chirp z-transform maintains its accuracy. In fact, the new chirp z-transform based algorithm is always more accurate than the Livas algorithm in computing the required Fourier transform magnitudes.

Next, the detection statistics used by the GW signal detection algorithms are compared. The detection statistics of the Livas algorithm and the new algorithm are compared by assuming

	computed magnitude/correct magnitude			
	$N = 1024$		$N = 4096$	
$g(t, \delta, \beta)$	cubic spline	chirp z-tranf.	cubic spline	chirp z-tranf.
0.995	0.9925	1.0009	2.2348	1.0001
0.996	1.0069	1.0000	1.6248	1.0000
0.997	0.9768	1.0009	1.2401	1.0000
0.998	0.9857	1.0000	1.0523	1.0000
0.999	0.9860	1.0009	0.9908	1.0000
1.000	1.0000	1.0000	1.0000	1.0000
1.001	0.9857	1.0009	0.9857	1.0000
1.002	0.9857	1.0000	0.9856	1.0000
1.003	0.9856	1.0009	0.9855	1.0000
1.004	0.9855	1.0000	0.9853	1.0000
1.005	0.9854	1.0009	0.9852	1.0001

Table 5.3: Comparison of the Fourier transform magnitudes of a sinusoid computed using the cubic spline algorithm and the chirp z-transform algorithm. The normalized frequency of the sinusoid is 0.5 and the computed magnitude normalized by the correct magnitude is shown.

that the computed Fourier transform magnitudes are accurate. However, because the magnitude estimate of the Livas algorithm is not always accurate, the actual performance of the Livas algorithm could be worse than the results indicated by the following discussion. The following comparisons are made using the measured MIT data. Because the data apparently do not contain a real GW signal, a synthetic GW signal is added to test the algorithms. The first comparison is made by assuming that the location of the emitter is known. The detection statistics are computed for all possible emitter frequencies. The second comparison is made using the receiver operating characteristic curves. They will show that the new algorithm has superior performance.

For the first comparison, the emitter location of the added synthetic signal is assumed to be known for both the Livas algorithm and the new algorithm. Figures 5.4 and 5.5 show the detection statistics of the Livas algorithm and the new algorithm, respectively. The detection statistic value at each candidate emitter frequency is plotted. The emitter frequency of the added signal is $\pi/8$ and the amplitude is 2.5. To generate the figures, 1024 samples of the data from the channels J8A, J8I, J8K, and J8M of Table 5.2 are used. The semblance is computed with the gate length of 3 samples. The figures indicate that the added signal is more conspicuous in the output of the new algorithm and, hence, is easier to detect. The Livas algorithm results

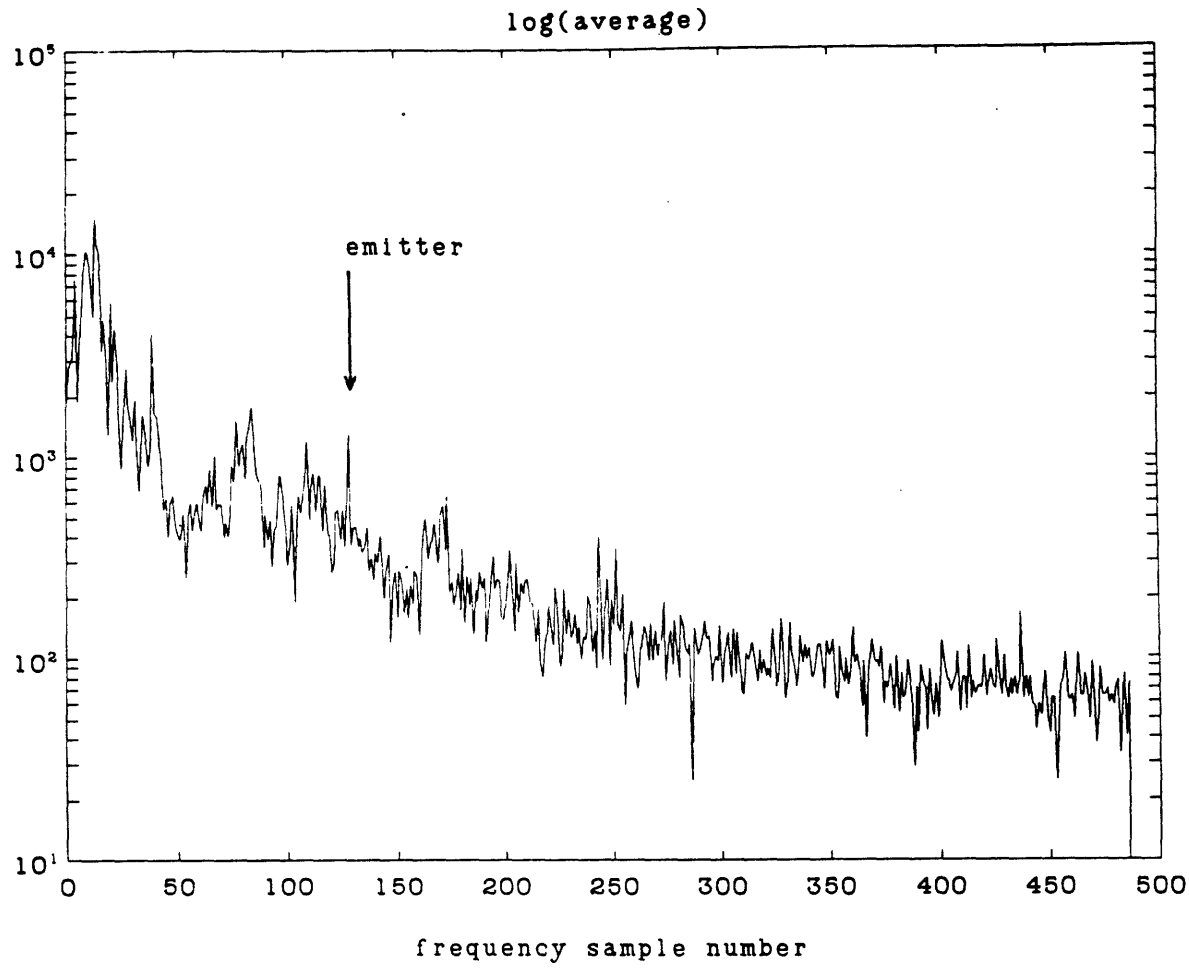

Figure 5.4: The detection statistic of the Livas algorithm as a function of frequency for a known emitter location. The frequency of the added signal is $\pi/8$ and the amplitude is 2.5. Input data length is 1024 samples and the data from the channels J8A, J8I, J8K, and J8M are used.

in numerous strong peaks due to the low frequency noises.

The previous comparison was made by inspecting the detection statistics of the two algorithms. The performance can be quantified by computing the ROCs. To compute the ROCs, a synthetic signal with an arbitrarily selected emitter location and frequency is added to the measured data. The detection statistic computed using the correct emitter frequency and location becomes a sample for the H_1 hypothesis. The detection statistic computed using an incorrect emitter location becomes a sample for the H_0 hypothesis. To generate many samples of the H_1 hypothesis, the frequency of the added signal is varied. For Fig. 5.6, 1024 different frequencies

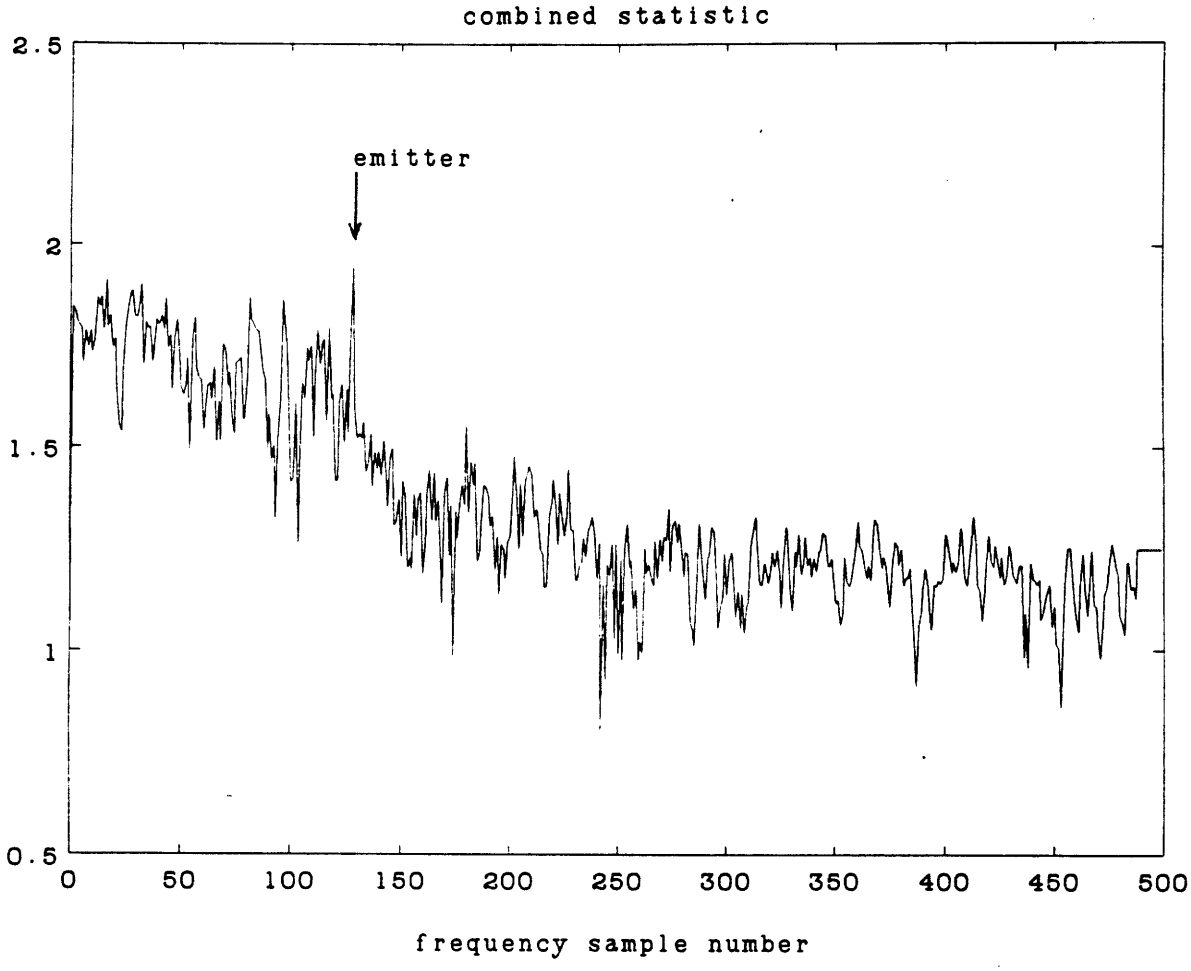

Figure 5.5: The detection statistic of the new algorithm as a function of frequency for a known emitter location. The frequency of the added signal is $\pi/8$ and the amplitude is 2.5. Input data length is 1024 samples; the semblance gate length is 3; and the data from the channels J8A, J8I, J8K, and J8M are used.

are used. The amplitude of the added sinusoidal signal is 0.5 and 2048 samples of the data from the channels J8A, J8I, J8K, and J8M of Table 5.2 are used. The semblance is computed with a gate length of 3 samples. To generate the H_0 hypothesis data, 4 incorrect locations are used. Figure 5.6 shows the ROCs which indicate that the new algorithm performs significantly better than the Livas algorithm.

5.7 Summary

In this chapter, the periodic GW signal detection problem was discussed. A detailed description of the received GW signal was presented and the GW signal detection problem was formulated as the multichannel detection problem of Chapter 2. The detection algorithms estimate the Fourier transform magnitudes at the frequency locations predicted by the emitter location and frequency. The estimated Fourier transform magnitudes are used to compute a detection statistic which is threshold tested to determine the presence of the GW signal emitter. The algorithm developed by Livas was described. It resamples the data using cubic spline approximation and then uses the FFT. The average of the magnitudes is used as the detection statistic.

A new algorithm was developed which has improvements on both the Fourier transform magnitude computation and the detection statistic. It computes the Fourier transform magnitudes at the exact frequencies using the chirp-z transform. The detection statistic used by the new algorithm is the combined average and semblance using the PDF matching method of Chapter 4. Finally, these two algorithms were compared to show that the new algorithm has improved performance.

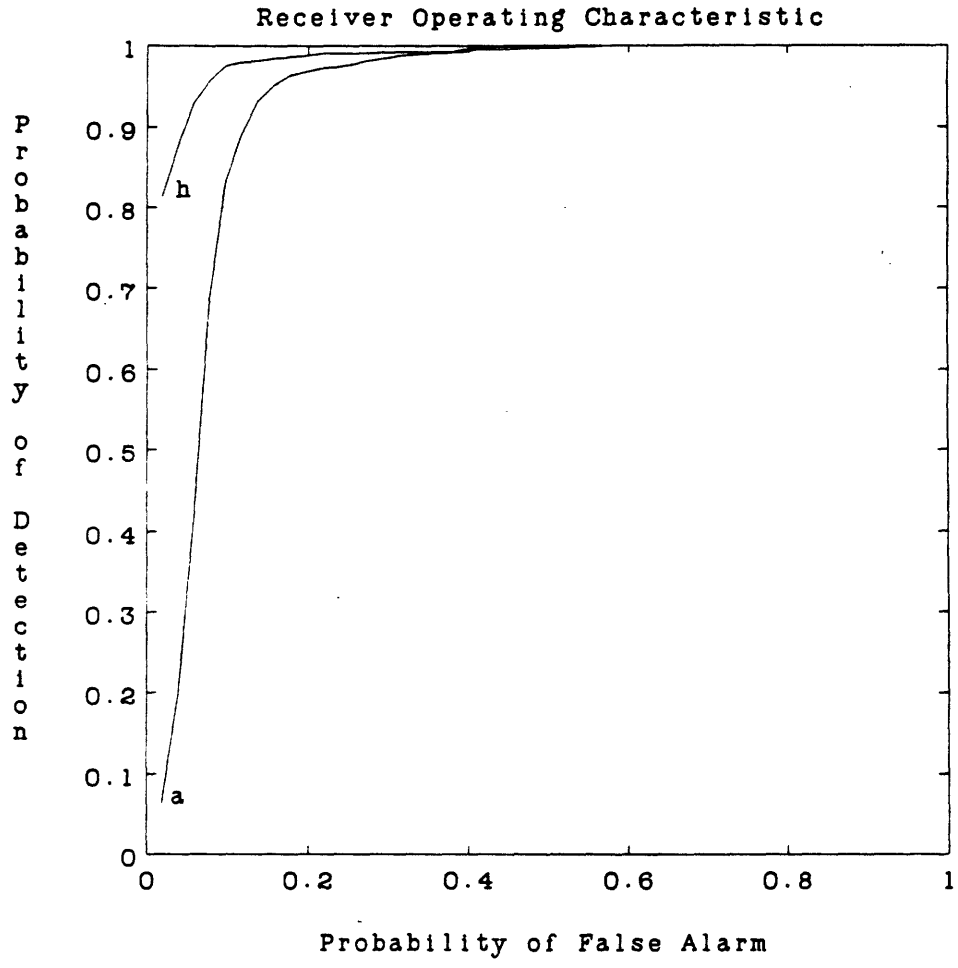

Figure 5.6: Receiver operating characteristic curves of the Livas algorithm and the new algorithm. Input data length is 2048 samples; the amplitude of the added sinusoid is 0.5; the semblance gate length is 3 samples; and the channels J8A, J8I, J8K, and J8M are used. 1024 different emitter frequencies are used for the H_1 hypothesis data and 4 incorrect directions are used for the H_0 hypothesis data. The average statistic, used by Livas, is marked by 'a' and the combined statistic using the hyperbolic tangent transformation $h_h(a) + s$, used by the new algorithm, is marked by 'h'. The abscissa, indicating the probability of false alarm, ranges from 0 to 1 and the ordinate, indicating the probability of detection, ranges from 0 to 1.

Chapter 6

Summary and Future Research

6.1 Summary

A multichannel signal detection problem was analyzed in this thesis. The critical aspect of the detection problem considered is that the narrowband component must exist in all channels to decide that an emitter is present. Even when an emitter is not present, any number of channels, but not all channels, can contain the narrowband component. In Chapter 2, this problem was formulated as a series of single channel binary hypothesis tests which separately decide the existence of the narrowband component in each channel. The single channel binary hypothesis testing problem was examined in detail to derive a detection statistic and its ROC. Because the solution of the single channel binary hypothesis testing problem is efficiently implemented by the FFT, the effect of quantized FFT coefficients on detection was analyzed to show that the probability of detection is not significantly affected if a reasonable number of quantization bits are used.

Even though the detection based on a series of binary hypothesis tests is straightforward, it is unsatisfactory because data from different channels are not used collectively. Therefore, this detector is not sensitive to the requirement that all channels must contain a narrowband component with the same amplitude. In order to collectively use the data from different channels, two statistics were considered in Chapter 3. The first statistic analyzed was the average. It was shown that the average is the likelihood ratio statistic for a special case of the multichannel detection problem. The PDF of the average was approximated using the Gram-Charlier series.

Because the detection performance of the average applied to the multichannel detection problem can be improved, the semblance was analyzed as a detection statistic. The semblance measures the coherence between the channels and thus it is well matched to the requirement of the multichannel detection problem. The PDF of the semblance was analytically derived for a simple case and was accurately approximated using the modified beta PDF for the general case.

The effectiveness of the average and the semblance as the statistic for the multichannel signal detection problem was measured by computing their respective ROCs using the Monte Carlo detection method. The ROC curves indicated that the average performs better than the semblance if most of the channels contain only wideband noise when the emitter is absent. On the other hand, the semblance performed better than the average if most of the channels contain the narrowband component when the emitter is absent. Thus, neither statistic performed well for a wide range of received data.

To find a statistic which has an improved detection performance for a wider range of received data, a new detection statistic based on combining the average and the semblance was developed in Chapter 4. Before the average and the semblance are combined, the average is transformed in order to weight the average and the semblance equally. The transformed average was combined with the semblance using addition because it satisfied the functional constraints imposed on the combining function. This new statistic was compared with the likelihood ratio statistic for the simple case and with the discriminant function statistics for the general case. For the simple case, the combined detection statistic performed as well as the optimal statistic. In all cases, the combined statistic was shown to perform as well or better than the other statistics.

In Chapter 5, the combined detection statistic was applied to the gravitational wave(GW) detection problem. A review of the received GW signal structure was given and the GW signal detection problem was formulated as the multichannel detection problem. The GW signal detectors estimate the Fourier transform magnitude at the frequencies which are specified by the hypothesized emitter location. The estimated magnitudes are used to compute the detection statistic and then decide the presence of the GW emitter. The algorithm proposed by Livas employed the average of the magnitudes estimated using the cubic-spline based approximate resampling method as the detection statistic. A new algorithm was developed which employed

the combined statistic developed in Chapter 4 as the detection statistic. The Fourier transform magnitudes were computed at the exact frequencies using the chirp z-transform based algorithm. Examples of the GW detection using the MIT data were presented to demonstrate that the new algorithm performs better.

6.2 Suggestions for Future Research

The semblance was used for the multichannel signal detection problem because it measures the coherence between channels. This property of the semblance might also be useful in a different type of detection problem. In this new detection problem, the received data contain only wideband noise when an emitter is absent. If an emitter is present, at least one channel must contain the narrowband component in additive wideband noise. Therefore, the detection problem is formulated as the following hypothesis testing problem:

$$\begin{cases} H_1 \text{ (emitter is present)} : & \underline{R} \neq \begin{pmatrix} \underline{w}_1 \\ \vdots \\ \underline{w}_L \end{pmatrix} \\ H_0 \text{ (emitter is absent)} : & \underline{R} = \begin{pmatrix} \underline{w}_1 \\ \vdots \\ \underline{w}_L \end{pmatrix} \end{cases}$$

where $\underline{R}$ denotes the received data and $\underline{w}_l$ is the lth channel wideband noise vector. This type of detection problem is encountered if an emitter fades in and out.

A straightforward solution to this detection problem is to test each channel for the existence of the narrowband component. If any one channel contains the narrowband component, then the hypothesis H_1 is decided. This solution does not use the channels collectively and does not explicitly require equal signal amplitude, hence its detection performance can be improved upon if the multichannel structure of the data is exploited. For example, if two channels are

used collectively and the data fit the model

$$\begin{cases} H_1 \text{ (emitter is present)} : & \underline{R} = \begin{pmatrix} \underline{s}_1 + \underline{w}_1 \\ \underline{s}_2 + \underline{w}_2 \end{pmatrix} \\ H_0 \text{ (emitter is absent)} : & \underline{R} = \begin{pmatrix} \underline{w}_1 \\ \underline{w}_2 \end{pmatrix} \end{cases}$$

where $\underline{s}_l$ is the lth channel narrowband component for $l = 1, 2$, then the probability of detection using the above model is better than the probability of detection of the detector which tests each channel separately. The probability of detection improves as the number of channels used increases. Of course, the difficulty is in deciding which channels to use for the hypothesis test.

The semblance might be useful in deciding which channels to use. The semblance will be used to screen the data to remove the channels containing wideband noise only and then the detector will use the remaining channels. The screening of the channels might be accomplished by computing the semblance of subsets of the channels and removing the channels which lower the semblance value when included in the semblance computation. After the channels containing wideband noise are removed, the detection problem becomes a conventional signal-in-noise versus noise-only detection problem for which the average is the likelihood ratio statistic. This proposed use of semblance is speculative and requires further investigation to assess its effectiveness.

Another future research area is the analysis of the combining operation. Two different statistics were combined using a function which satisfied a list of intuitive functional constraints in Section 4.1. The function $\tilde{a} + s$, where $\tilde{a}$ denotes the transformed average and s denotes the semblance, was used because it was simple yet effective. The most general forms of the function which satisfy the first functional constraint, $f(\tilde{a}, s) = f(s, \tilde{a})$, of Section 4.1 are

$$g_1(h_1(\tilde{a}), h_1(s)) + g_1(h_1(s), h_1(\tilde{a})) \quad \text{and} \quad g_2(h_2(\tilde{a}), h_2(s)) \times g_2(h_2(s), h_2(\tilde{a}))$$

for arbitrary functions $g_1(\cdot)$, $g_2(\cdot)$, $h_1(\cdot)$, and $h_2(\cdot)$. However these functions are difficult to manipulate to satisfy the other functional constraints of Section 4.1. Therefore, new combining functions which result from optimizing some other criteria might be useful in evaluating the detection statistic developed and used in this thesis.

The likelihood ratio detection statistic was derived only for the simple case. Approximations of the likelihood ratio detection statistic for the general case will be useful in further evaluation of the combined detection statistic used in this thesis. In Section 4.2, the likelihood ratio statistic was given by

$$\Lambda(a,s) = \frac{p_{a,s|H_1}(a,s|H_1)}{p_{a,s|H_0}(a,s|H_0)}.$$

If the conditional PDFs, $p_{a,s}(a,s|H_1)$ and $p_{a,s}(a,s|H_0)$, are approximated, the likelihood ratio detection statistic can be estimated using the approximated PDFs. The following are some of the applicable approximation methods.

- Gram-Charlier series expansion[16,34,30]: This method uses the estimated moments to approximate the unknown density[1]. The first term of the expansion is the two-dimensional Gaussian density and the subsequent terms involve higher moments.

- Orthogonal basis function expansion[44,17]: A set of basis function is formed using two sets of orthogonal bases. The basis functions are defined by $\{\phi_{kl}(a,s) = \psi_k^1(a) \times \psi_l^2(s) \text{ for } k,l = 1,2,\cdots\}$ where, for example, $\psi_k^1(a)$ is Laguerre function[15] and $\psi_l^2(s)$ is Lagendre function[15] mapped to match the range of the semblance. The PDF is approximated as $p_{a,s}(a,s) \approx \sum_k \sum_l c_{kl}\ \phi_{kl}(a,s)$ with $c_{kl} = \frac{1}{N}\sum_{n=1}^{N} \phi_{kl}(a_n,s_n)$ where N is the total number of samples and a_n and s_n are nth average sample and semblance sample, respectively.

- Generalized histogram[66,53,44,17]: The unknown density is approximated as $p_{a,s}(a,s) \approx \frac{1}{N}\sum_{n=1}^{N} \frac{1}{W} h(\frac{\|(a,s)-(a_n,s_n)\|}{W})$ where $\|.\|$ denotes a distance measure, N is the total number of samples, and a_n and s_n are nth average sample and semblance sample, respectively. The kernel $h(\cdot)$ and the weighting $\frac{1}{W}$ are general functions satisfying the following conditions to obtain an asymptotically unbiased and consistent estimate. The requirements on $h(\cdot)$ are: it has unit area; it is absolutely integrable; its supremum of the absolute value is finite; and $\lim_{t\to\infty} |t \cdot h(t)| = 0$. The requirements on W, which is a function of the number of samples, are $\lim_{N\to\infty} W(N) = 0$ and $\lim_{N\to\infty} N \times W(N) = \infty$, where N is the total number of samples.

[1] One dimensional case was discussed in Section 3.1.2.

- **Synthesis using marginal distribution functions[54,14,41,30]: The following observations about the distribution functions of the marginal PDFs are used to approximate the distribution function of the unknown joint density.**

$$\begin{aligned} P_{a,s}(a,s) &\leq \min(P_a(a), P_s(s)) \\ P_{a,s}(a,s) &\geq P_a(a) + P_s(s) - 1 \end{aligned}$$

where $P_{a,s}(a,s)$ is the distribution function of the unknown joint PDF of the average and the semblance and $P_a(a)$ and $P_s(s)$ are the distribution function of the average and the semblance, respectively. The distribution function, $P_{a,s}(a,s)$, is approximated as a combination of $P_a(a)$ and $P_s(s)$. Two examples among many possibilities are

$$P_{a,s}(a,s) \approx \lambda \min(P_a(a), P_s(s)) + (1-\lambda)(P_a(a) + P_s(s) - 1) \quad \text{for } 0 \leq \lambda \leq 1$$

and

$$P_{a,s}(a,s) \approx P_a(a)P_s(s)(1 + \alpha(1 - P_a(a))(1 - P_s(s))) \quad \text{for some } \alpha.$$

This method might be useful because the marginal densities have been approximated in Chapter 3.

These methods and perhaps others should be analyzed to determine the most accurate approximation of the conditional PDFs. The detection statistic based on the approximated PDFs should be compared to the combined statistic.

Finally, more extensive processing of the measured gravitational wave data is necessary. The new algorithm was applied to the GW signal detection problem only to demonstrate its usefulness. It was shown to be as efficient as the existing algorithm, however refinements are required to achieve real-time or near real-time processing speed which is important for future gravitational wave detection systems.

Bibliography

[1] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions. Applied Mathematics Series 55*, National Bureau of Standards, June 1964.

[2] Z. W. Birnbaum. Computers and unconventional test-statistics. In F. Proschan and R. J. Serfling, editors, *Reliability and Biometry: Statistical Analysis of Lifelength*, pages 441–458, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974.

[3] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation.* Springer-Verlag, New York, 1983.

[4] K. V. Bury. *Statistical Models in Aplied Science.* John Wiley & Sons, Inc., New York, 1975.

[5] W. J. Conover. *Practical Nonparametric Statistics.* John Wiley & Sons, Inc., New York, second edition, 1971.

[6] H. A. David. *Order Statistics.* John Wiley & Sons, Inc., New York, 1970.

[7] P. C. W. Davies. *The Search for Gravity Waves.* Cambridge University Press, Cambridge, England, 1980.

[8] M. H. DeGroot. *Probability and Statistics.* Addison-Wesley Publishing Co., Reading, MA, 1975.

[9] D. Dewey. *A Search for Astronomical Gravitational Radiation with an Interferometric Broad Band Antenna.* PhD thesis, Massachusetts Institute of Technology, Feb. 1986.

[10] J. D. DiFranco and W. L. Rubin. *Radar Detection.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.

[11] E. J. Douze and S. J. Laster. Statistics of semblance. *Geophysics*, 44(12):1999–2003, Dec. 1979.

[12] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, Inc., New York, 1973.

[13] M. Dwass. *Probability and Statistics.* W. A. Benjamin, Inc., New York, 1970.

[14] D. J. G. Farlie. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47(3):307–323, 1960.

[15] L. E. Frank. *Signal Theory.* Dowden & Culver, Inc., Stroudsburg, PA, 1981.

[16] T. C. Fry. *Probability and Its Engineering Uses.* D. Van Nostrand Co., Princeton, N.J., 1965.

[17] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, New York, 1972.

[18] J. D. Gibson and J. L. Melsa. *Introduction to Nonparametric Detection with Applications.* Academic Press, New York, 1975.

[19] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore,MD, 1983.

[20] J. H. Halton. A retrospective and prospective survey of the Monte Carlo method. *SIAM Review*, 12(1):1–63, Jan. 1970.

[21] J. M. Hammersely and D. C. Handscomb. *Monte Carlo Methods.* Chapman and Hall, Ltd., London, 1964.

[22] C. W. Helstrom. *Statistical Theory of Signal Detection.* Pergamon Press, New York, 1968.

[23] U. Heute. Results of a deterministic analysis of FFT coefficient errors. *Signal Processing*, 3:321–331, 1981.

[24] U. Heute and H. Schuessler. FFT-accuracy – new insights and a new point-of-view. In *Procedings of ICASSP*, pages 631–6341, Boston, MA, 1983.

[25] A. C. A. Hope. A simplified Monte Carlo significance test procedure. *Journal of Royal Statistical Society*, B 30(3):582–598, 1968.

[26] D. James. Quantization errors in the fast Fourier transform. *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-23(3):227–283, June 1975.

[27] G. M. Jenkins and D. G. Watts. *Spectral Analysis and Its Applications.* Holden-Day Inc., San Francisco, CA, 1968.

[28] K. Jöckel. Finite sample properties and asymptotic efficiency of Monte Carlo tests. *The Annals of Statistics*, 14(1):336–347, 1986.

[29] N. L. Johnson and S. Kotz. *Distributions in Statistics, Continuous Univariate Distributions - 2.* Volume 3, John Wiley & Sons, Inc., New York, 1972.

[30] N. L. Johnson and S. Kotz. *Distributions in Statistics, Multivariate Distributions.* Volume 4, John Wiley & Sons, Inc., New York, 1972.

[31] T. H. Joo and A. V. Oppenheim. Effects of FFT coefficient quantization on sinusoidal signal detection. In *Procedings of ICASSP*, pages 1818–1821, New York, 1988.

[32] M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods: Volume I Basics.* John Wiley & Sons, Inc., New York, 1986.

[33] S. M. Kay. *Modern Spectral Estimation: Theory and Application.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.

[34] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics- Distribution Theory.* Volume 1, Charles Griffin & Company Limited, London, second edition, 1963.

[35] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics- Inference and Relationship.* Volume 2, Charles Griffin & Company Limited, London, third edition, 1973.

[36] C. V. Kimball and T. L. Marzetta. Semblance processing of borehole acoustic array data. *Geophysics*, 49(3):274–281, March 1984.

[37] W. Knight and R. Kaiser. A simple fixed-point error bound for the fast Fourier transform. *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-27(6):615–620, Dec. 1979.

[38] E. L. Lehmann. *Testing Statistical Hypothesis.* John Wiley & Sons, Inc., New York, second edition, 1986.

[39] J. C. Livas. *Upper Limits for Gravitational Radiation from Some Astrophysical Sources.* PhD thesis, Massachusetts Institute of Technology, May 1987.

[40] J. K. Marcum. A statistical theory of target detection by pulsed radar. *IRE Transaction of Information Theory*, IT-6(2):59–267, Apr. 1960.

[41] K. V. Mardia. A translation family of bivariate distributions and Frechet's bounds. *Sankhya: The Indian Journal of Statistics: Series A*, 119–122, 1969.

[42] F. H. C. Marriott. Barnard's Monte Carlo tests: how many simulations? *Appl. Statist.*, 28(1):75–77, 1979.

[43] W. H. Mayne. Common reflection point horizontal data stacking techniques. *Geophysics*, XXVII(II):927–938, Dec. 1962.

[44] W. S. Meisel. *Computer-Oriented Approach to Pattern Recognition.* Academic Press, New York, 1972.

[45] P. A. P. Moran. *Introduction to Probability Theory.* Oxforn University Press, New York, 1968.

[46] V. K. Murthy. On the distribution of the sum of circular serial correlation coefficients and the effect of non-normality on its distribution(abstract). *Annals of Mathematical Statistics*, 31:239–240, 1960.

[47] O. E. Naess and L. Bruland. Stacking methods other than simple summation. In A. A. Fitch, editor, *Developments in Geophysical Exploration Methods*, chapter 6, pages 189–223, Elsevier Applied Science Publishers, New York, 1985. Vol. 6.

[48] N. S. Neidell and M. T. Taner. Semblance and other coherency measures for multichannel data. *Geophysics*, 36(3):482–497, June 1971.

[49] B. Noble. *Applied Linear Algebra.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1969.

[50] A. V. Oppenheim and R. W. Schafer. *Digital Signal Processing.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

[51] L. Palmer. Coarse frequency estimation using the discrete Fourier transform. *IEEE Trans. Information Theory*, IT-20:104–109, Jan. 1974.

[52] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill Book Company, New York, second edition, 1984.

[53] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[54] R. L. Plackett. A class of bivariate distributions. *American Statistical Association Journal*, 516–522, June 1965.

[55] W. H. Press, B. P. Flannery, S. T. Teukolsky, and W. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, England, 1988.

[56] S. O. Rice. Mathematical analysis of random noise. In N. Wax, editor, *Selected Papers on Noise and Stochastic Processes*, pages 133–294, Dover Publications, Inc., New York, 1954.

[57] D. Rife and R. Borystein. Single-tone parameter estimation from discrete-time observations. *IEEE Trans. Information Theory*, IT-20(5):591–598, Sep. 1974.

[58] B. D. Ripley. *Stochastic Simulation.* John Wiley & Sons, Inc., New York, 1987.

[59] E. A. Robinson. *Migration of Geophysical Data.* International Human Resources Development Corporation, Boston, MA, 1983.

[60] R. Y. Rubinstein. *Simulation and the Monte Carlo Method.* John Wiley & Sons, Inc., New York, 1981.

[61] S. L. Shapiro, R. F. Stark, and S. A. Teukolsky. The search for gravitational waves. *American Scientist*, 73:248–257, May 1985.

[62] Y. A. Shreider, editor. *The Monte Carlo Method: The Method of Statistical Trials.* Pergamon Press, New York, 1966.

[63] M. D. Springer. *The Algebra of Random Variables.* John Wiley & Sons, Inc., New York, 1979.

[64] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Wellesley, MA, 1986.

[65] M. T. Taner and F. Koehler. Velocity spectra – digital computer derivation and applications of velocity functions. *Geophysics*, 34(6):859–881, Dec. 1969.

[66] R. A. Tapia and J. R. Thompson. *Nonparametric Probability Density Estimation.* The Johns Hopkins University Press, Baltimore, MD, 1978.

[67] K. S. Thorne. Gravitational radiation. In S. W. Hawking and W. Israel, editors, *Three Hundred Years of Gravitation*, Cambridge University Press, Cambridge, England, 1987.

[68] K. S. Thorne. Gravitational-wave research: current status and future prospects. *Reviews of Modern Physics*, 52(2):285–297, Apr. 1980. Part I.

[69] D. Tufts, H. Hersey, and W. Mosier. Effects of FFT coefficient quantization on bin frequency response. *Proceedings of IEEE(Letter)*, 60:146–147, Jan. 1972.

[70] G. L. Turin. Error probabilities for binary symmetric ideal reception through nonselective slow fading and noise. *Proceedings of the IRE*, 1603–1619, Sep. 1958.

[71] G. L. Turin. An introduction to matched filters. *IRE Transaction of Information Theory*, 311–329, June 1960.

[72] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* Wiley & Sons, Inc., New York, 1968.

[73] R. Vogt, R. Drever, K. Thorne, and R. Weiss. *Caltech/MIT Project for a Large Interferometer Gravitational Wave Observatory.* NSF Proposal, 1987.

[74] C. Weinstein and A. Oppenheim. Effects of finite register length in digital filtering and the fast Fourier transform. *Proceedings of the IEEE*, 60:957–976, 1972.

[75] R. Weiss. Electromagnetically coupled broadband gravitational antenna. *MIT RLE Quarterly Progress Report*, 105:54, 1972.

[76] S. S. Wilks. *Mathematical Statistics.* John Wiley & Sons, Inc., New York, 1962.