

Syracuse University

SURFACE

Electrical Engineering and Computer Science -
Dissertations

College of Engineering and Computer Science

2011

Decision-Making with Heterogeneous Sensors - A Copula Based Approach

Satish Giridhar Iyengar
Syracuse University

Follow this and additional works at: https://surface.syr.edu/eecs_etd



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Iyengar, Satish Giridhar, "Decision-Making with Heterogeneous Sensors - A Copula Based Approach" (2011). *Electrical Engineering and Computer Science - Dissertations*. 310.
https://surface.syr.edu/eecs_etd/310

This Dissertation is brought to you for free and open access by the College of Engineering and Computer Science at SURFACE. It has been accepted for inclusion in Electrical Engineering and Computer Science - Dissertations by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Abstract

Statistical decision making has wide ranging applications, from communications and signal processing to econometrics and finance. In contrast to the classical one source - one receiver paradigm, several applications have been identified in the recent past that require acquiring data from multiple sources or sensors. Information from the multiple sensors are transmitted to a remotely located receiver known as the fusion center which makes a global decision. Past work has largely focused on fusion of information from homogeneous sensors. This dissertation extends the formulation to the case when the local sensors may possess disparate sensing modalities. Both the theoretical and practical aspects of multimodal signal processing are considered.

The first and foremost challenge is to ‘adequately’ model the joint statistics of such heterogeneous sensors. We propose the use of copula theory for this purpose. Copula models are general descriptors of dependence. They provide a way to characterize the nonlinear functional relationships between the multiple modalities, which are otherwise difficult to formalize. The important problem of selecting the ‘best’ copula function from a given set of valid copula densities is addressed, especially in the context of binary hypothesis testing problems. Both, the training-testing paradigm, where a training set is assumed to be available for learning the copula models prior to system deployment, as well as generalized likelihood ratio test (GLRT) based fusion rule for the online selection and estimation of copula parameters are considered. The developed theory is corroborated with extensive computer simulations as well as results on real-world data.

Sensor observations (or features extracted thereof) are most often quantized before their transmission to the fusion center for bandwidth and power conservation. A detection scheme is proposed for this problem assuming uniform scalar quantizers at each sensor. The designed rule is applicable for both binary and multibit local sensor decisions. An alternative suboptimal but computationally efficient fusion rule is also

designed which involves injecting a deliberate disturbance to the local sensor decisions before fusion. The rule is based on Widrow's statistical theory of quantization. Addition of controlled noise helps to *linearize* the highly nonlinear quantization process thus resulting in computational savings. It is shown that although the introduction of external noise does cause a reduction in the received signal to noise ratio, the proposed approach can be highly accurate when the input signals have bandlimited characteristic functions, and the number of quantization levels is large.

The problem of quantifying neural synchrony using copula functions is also investigated. It has been widely accepted that multiple simultaneously recorded electroencephalographic signals exhibit nonlinear and non-Gaussian statistics. While the existing and popular measures such as correlation coefficient, corr-entropy coefficient, coh-entropy and mutual information are limited to being bivariate and hence applicable only to pairs of channels, measures such as Granger causality, even though multivariate, fail to account for any nonlinear inter-channel dependence. The application of copula theory helps alleviate both these limitations. The problem of distinguishing patients with mild cognitive impairment from the age-matched control subjects is also considered. Results show that the copula derived synchrony measures when used in conjunction with other synchrony measures improve the detection of Alzheimer's disease onset.

DECISION-MAKING WITH HETEROGENEOUS SENSORS
- A COPULA BASED APPROACH

By

Satish Giridhar Iyengar
B.E. University of Mumbai, 2002
M.S. Syracuse University, 2004

DISSERTATION

Submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Electrical Engineering
in the Graduate School of Syracuse University

August 2011

© Copyright by Satish Giridhar Iyengar, 2011.

All rights reserved.

Contents

| | |
|--|-----------|
| Abstract | i |
| List of Tables | viii |
| List of Figures | ix |
| Acknowledgements | xiii |
| 1 Introduction | 1 |
| 1.1 Literature Survey | 3 |
| 1.2 Thesis Organization and Main Contributions | 6 |
| 2 Statistical Modeling of Heterogeneous Signals | 9 |
| 2.1 Modeling Heterogeneous Data - Challenges | 10 |
| 2.1.1 Commonly Adopted Approaches for Modeling Heterogeneous Data | 12 |
| 2.2 Copula Theory | 14 |
| 2.2.1 Some Well-known Copula Families | 15 |
| 2.2.2 Estimation of Model Parameters | 17 |
| 2.2.3 Selection of the Copula Function | 24 |
| 2.3 Summary | 25 |
| 3 Hypothesis Testing With Heterogeneous Data | 27 |
| 3.1 Copula-Based Fusion Rule | 29 |
| 3.2 Copula Selection for Hypothesis Testing Problems | 31 |

| | | |
|----------|---|-----------|
| 3.2.1 | The Kullback-Leibler Divergence Criterion | 31 |
| 3.2.2 | The Area Under the Receiver Operating Characteristic Curve Criterion | 38 |
| 3.3 | Biometric Authentication: An Application | 43 |
| 3.3.1 | Fusion of Multiple Biometric Modalities | 44 |
| 3.3.2 | Experimental Results | 45 |
| 3.4 | Summary | 50 |
| 4 | Hypothesis Testing With Heterogeneous Data: Uncertain Parameters | 57 |
| 4.1 | (Misspecified) Generalized Likelihood Ratio Test Based Fusion Rule . | 59 |
| 4.1.1 | Fixed vs. Variable Threshold Design | 62 |
| 4.2 | Performance Analysis | 62 |
| 4.2.1 | Upper Bounds on Detection Power | 62 |
| 4.3 | Example | 66 |
| 4.4 | Application: Footstep Detection | 69 |
| 4.4.1 | Related Work and Our Approach | 70 |
| 4.4.2 | Data Acquisition and Preprocessing | 74 |
| 4.4.3 | Statistical Modeling and Detector Design | 76 |
| 4.4.4 | Results | 81 |
| 4.5 | Summary | 82 |
| 5 | Hypothesis Testing With Quantized Sensor Measurements | 86 |
| 5.1 | Problem Formulation | 89 |
| 5.2 | (Misspecified) GLRT Based Fusion of Soft Decisions | 91 |
| 5.3 | A Computationally Efficient Fusion Rule | 94 |
| 5.3.1 | Widrow's Statistical Theory of Quantization: A Review | 94 |
| 5.3.2 | Fusion Rule Based on the Pseudo-Quantization Noise Model . | 98 |

| | | |
|----------|---|------------|
| 5.3.3 | Loss in the Detection Power due to Quantization: PQN Model- Based Analysis | 102 |
| 5.4 | An Illustrative Example | 103 |
| 5.5 | Summary | 108 |
| 6 | Quantification of Neural Synchrony Using Copulas | 110 |
| 6.1 | Synchrony Measures: A Review | 112 |
| 6.1.1 | Granger Causality | 113 |
| 6.1.2 | Stochastic Event Synchrony | 114 |
| 6.2 | Copula based Synchrony Measure | 116 |
| 6.3 | Early Diagnosis of Alzheimer’s Disease | 119 |
| 6.3.1 | EEG Data | 119 |
| 6.3.2 | Computation of EEG Synchrony | 120 |
| 6.3.3 | Copula Selection | 121 |
| 6.3.4 | Classification Results | 124 |
| 6.4 | Summary | 125 |
| 7 | Conclusion | 127 |
| 7.1 | Future Research Directions | 128 |
| A | Optimal Likelihood Ratio Test For Example 3.1 | 131 |
| | Bibliography | 135 |

List of Tables

| | | |
|-----|---|-----|
| 2.1 | Archimedean Copula functions | 17 |
| 3.1 | AUC estimated using the Wilcoxon-Mann-Whitney statistic | 41 |
| 3.2 | Peak increase in P_D due to copula processing ($P_D^k - P_D^p$). Corresponding P_F for ‘Request for retrial’, ‘Censoring’ and ‘Accept H_0 ’ designs is 0.0026%, and that for ‘Accept H_1 ’ and ‘Random decision’ designs is 9.9%. | 49 |
| 4.1 | Mean and symmetric one standard deviation confidence intervals for $\hat{\beta}$ | 81 |
| 6.1 | Classification rates using linear and quadratic discriminant analysis with leave-one-out cross validation | 125 |

List of Figures

| | | |
|-----|---|----|
| 2.1 | Correlation coefficient is a weak measure of dependence | 13 |
| 3.1 | A multisensor system with common region of interest (ROI). Different shapes for the sensors denote different sensing capabilities. | 28 |
| 3.2 | Scatter plot of 500 realizations. Non-zero dependence between observations under H_1 (Column 1) is evident from the figure. | 35 |
| 3.3 | P_D vs. η obtained for different copula functions: Parameters for the marginal PDFs are $(\sigma_0^2 = 1, \sigma_1^2 = 1.1)$, $(\lambda_0 = 2.1, \lambda_1 = 2.2)$, $(b_0 = b_1 = 1, a_0 = 1.8, a_1 = 2)$. The shaded region represents the increase in the area (over the product model) when FGM copula is used to model dependence under H_1 | 36 |
| 3.4 | Multi-information estimates (averaged over 5000 Monte Carlo trials) with symmetric one standard deviation error bars, and Gain (Eq. (3.15)) computed using the trapezoidal rule, for several copula functions. The number of trials (in %) for each copula for which the improvability condition (3.12) was NOT satisfied is also indicated. | 37 |
| 3.5 | Monte Carlo based Receiver Operating Characteristics: Parameters for the marginal PDFs are $(\sigma_0^2 = 1, \sigma_1^2 = 1.1)$, $(\lambda_0 = 2.1, \lambda_1 = 2.2)$, $(b_0 = b_1 = 1, a_0 = 1.8, a_1 = 2)$ | 38 |
| 3.6 | ROC curves obtained when Clayton and Gumbel copula densities model H_1 and H_0 respectively. | 41 |

| | | |
|------|---|----|
| 3.7 | ROC curves obtained when Gaussian and Clayton copula densities model H_1 and H_0 respectively. | 42 |
| 3.8 | A multibiometric authentication system. Biometric signatures of disparate modalities such as face, iris and fingerprint are fused. | 45 |
| 3.9 | Scatter plot of the genuine and impostor scores from the two face matchers | 46 |
| 3.10 | Gaussian mixture models for match scores from the two face matchers (FM) | 52 |
| 3.11 | Receiver operating characteristic curves | 53 |
| 3.11 | Receiver operating characteristic curves | 54 |
| 3.12 | Mutual Information estimates averaged over thirty resamples for ‘request for retrieval’ design. The plot also shows symmetric one standard deviation error bars. | 54 |
| 3.13 | Probability of error vs. $Pr(H_1)$ | 55 |
| 3.13 | Probability of error vs. $Pr(H_1)$ | 56 |
| 4.1 | Scatter plot of 500 realizations. Non-zero dependence between $Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Z_2 \sim \text{Beta}(\alpha_1, \beta_1)$ (under H_1) is evident from the figure. 68 | 68 |
| 4.2 | Application of Theorem 4.1 for determining the detector thresholds. Simulation parameters are: $(\mu_0, \sigma_0^2) = (0, 1)$, $(\mu_1, \sigma_1^2) = (0.1, 1.1)$, $(\alpha_0, \beta_0) = (2, 2)$, $(\alpha_0, \beta_0) = (2.2, 2.2)$ and Kendall’s $\tau H_1 = 0.2$. The detector makes a decision after processing $L = 100$ samples. | 69 |
| 4.3 | Monte Carlo based receiver operating characteristic curves: Parameters for the marginal PDFs are $(\mu_0 = 0, \sigma_0^2 = 1)$ and $(\mu_1 = 0.1, \sigma_1^2 = 1.1)$ and $(\alpha_0 = 2, \beta_0 = 2)$ and $(\alpha_1 = 2.2, \beta_1 = 2.2)$ | 70 |
| 4.4 | Footstep data collection at the US Army Research Lab: Experimental setup | 74 |
| 4.5 | Footstep signals and their spectrograms due to normal walk | 75 |

| | | |
|------|--|----|
| 4.6 | Canonical correlation coefficients for noise alone and footsteps plus noise. The plot also shows symmetric one standard deviation error bars. | 76 |
| 4.7 | The probability density of the generalized Gaussian model for different values of the shape parameter, $\beta = 0.6, 1, 2$ and 10. All distributions are normalized to unit variance and have zero mean. | 78 |
| 4.8 | Histogram of the acoustic CCA variate, u_1 , together with the fitted generalized Gaussian density | 80 |
| 4.9 | Histogram of the seismic CCA variate, v_1 , together with the fitted generalized Gaussian density | 80 |
| 4.10 | Histogram of β estimated using the ‘noise-alone’ data | 81 |
| 4.11 | Receiver operating characteristic curves for the CCA-copula based detector | 84 |
| 4.11 | Receiver operating characteristic curves for the CCA-copula based detector | 85 |
| 5.1 | Distributed heterogeneous sensor network: A parallel architecture. . . | 87 |
| 5.2 | Input-output transfer function of a uniform scalar quantizer | 90 |
| 5.3 | Pseudo-quantization noise (PQN) model for quantization | 95 |
| 5.4 | Illustration of the quantization process in the CF domain: (a) CF of z_n ; (b) CF of q_n , the sinc function; (c) CF of $z_n + w_n$; (d) repetition of CF of $z_n + w_n$; the CF of the quantized variable is given by the summation of these repetitions. | 97 |
| 5.5 | A <i>controlled</i> noise d_n is added at the output of each sensor n . The approach, although suboptimal, greatly simplifies the fusion rule by avoiding the need to compute multidimensional integrals. The method can be considerably accurate so long as Widrow’s quantization Theorem I is satisfied and noise is appropriately designed. | 99 |

| | | |
|-----|---|-----|
| 5.6 | 'Filtering' the quantized signal with LPF-noise. The quantization step size, q_1 , is set to $0.3\sigma_1$: (a) CF of z_1 ; (b) CF of $z_1 + w_1$; (c) CF of u_1 ; (d) CF of the external LPF-noise, d_n ; (e) CF of $y_n = z_n + w_n + d_n$. | 105 |
| 5.7 | System probability of false alarm vs. Detector threshold. The figure shows that P_F does not vary across different quantization levels, 0.3σ and 0.6σ . | 107 |
| 5.8 | Monte Carlo based Receiver Operating Characteristics: Performance of the PQN model based fusion rule is very close to the analog transmission case. The PQN model based fusion rule is, thus, a good and computationally efficient approximation to the exact fusion rule derived in Section 5.2. | 108 |
| 6.1 | Bump modeling | 115 |
| 6.2 | Bump matching of two event sequences, E and E' . The pairs in the dashed square box show coincident bumps. | 115 |
| 6.3 | Clustering of the 21 EEG channels into 5 zones indicated by the colors and dashed lines, (1) frontal, (2) left temporal, (3) central, (4) right temporal, and (5) occipital | 120 |
| 6.4 | Normalized histogram of estimates of degrees of freedom (ν) computed for different model orders and frame lengths (L) | 122 |
| 6.5 | Difference of multi-information estimates based on the t and Gaussian copula functions. | 123 |

Acknowledgements

It is with great pleasure and a profound sense of reverence that I express my deepest gratitude and sincere thanks to my advisor, Dr. Pramod K. Varshney, for his guidance, support and encouragement during the course of this research. Apart from the technical guidance, I highly admire his systematic way of working and perpetual encouragement. I would also like to thank other members of my defense committee, Dr. Mark Glauser, Dr. Biao Chen, Dr. Thyagaraju Damarla, Dr. Kishen Mehrotra and Dr. Fred Schlereth, who carefully read through this dissertation and contributed to its improvement with helpful comments and suggestions.

I am highly indebted to Dr. Ruixin Niu for taking the time out of his busy schedule to listen to my research ideas. Conversations with him led to the results presented in Chapter 5. I also wish to thank Dr. Justin Dauwels for all the long distance discussions, and his contribution to the results of Chapter 6.

I owe much to the current and former members of the Sensor Fusion Lab who have been a constant source of encouragement and companionship. In particular, I have thoroughly enjoyed the weekly brainstorming sessions with Ashok Sundaresan, Arun Subramanian, Hao Chen and Swarnendu Kar.

Words cannot fully express my appreciation for my wife, Smita, who many-a-times had to be both, a mother and father, to our loving son Aarush, thus giving me enough time for research. This dissertation would not have been possible without her support and understanding.

Above all, I thank my parents, Mr. S. Giridhar and Dr. Rajani Giridhar, for their unconditional love and support in all my pursuits.

Chapter 1

Introduction

Multimodal/Heterogeneous signal processing refers to the joint analyses and fusion of data from a variety of sensors (e.g., acoustic, seismic, magnetic, video and infrared) to solve a common inference task. One of the prime examples of a multimodal system is the human brain, which integrates diverse sensory information such as sight, smell, touch, hearing and taste to infer about the surrounding environment. For example, speech perception is known to be a bimodal process that involves both auditory and visual inputs [69]. Visual cues such as lip movements of the speaker have shown to significantly improve speech intelligibility especially in environments when the auditory signal is compromised. Similarly, building an automated system that can consolidate information from diverse sources of information, could offer several advantages and new possibilities for performance improvement. For example, video surveillance cameras can be augmented with audio sensors to mitigate the effects of *screen flicker* noise. The acoustic modality is unaffected by the lighting conditions and can provide the much needed robustness. Thus, in general, noise in one modality (e.g. poor lighting conditions) may not affect the signal-to-noise ratio of the other modality (such as audio) and may result in enhanced overall system performance.

Additionally, much useful information can be extracted from the joint analysis of the different modalities that is otherwise unavailable if the signal modalities are considered independently. The use of multiple modalities may provide complementary information and thus increase the accuracy of the overall decision making process. For example, fusion of ‘functional’ images from positron emission tomography (PET) and ‘structural’ data from magnetic resonance imaging (MRI) has become commonplace in clinical practice and in preclinical biomedical research [22]. Even within MRI, structural images of high resolution are acquired along with low-resolution images of brain chemistry or blood flow. Thus, an efficient system design is the one which exploits both the *shared* and the *complementary* information contained in the multimodal signals.

While humans have a natural ability to handle and process such diverse streams of information, teaching machines to do the same remains a challenge. Some of these challenges are listed below:

1. Signals of different modalities are typically acquired at different rates. There may be differences in the processing times of individual modalities. The resulting temporal non-synchronicity between the signals complicates the fusion task. For example, the mean amplitude feature from the audio sensor data may span a large time interval, and may not be directly linked to the appearance and movement of an object in the visual image data. Thus, features from one sensor data stream may temporally lag those from another and the fusion center has to be made aware of this delay between the inputs.
2. There may be differences in the dimensionality as well as the support space across the different signals. In the audio-visual fusion example, the audio stream is a 1-D signal, while video signals are 3-D measurements.

3. Data provided by different sensors are generally incommensurate and are characterized by considerably different features. Thus, it is unclear how one would weight each modality in the fusion process.
4. Inter-modal correlations can be quite complex, and it is not straightforward to extract useful information shared by the different data modalities.
5. The ‘nice’ properties of linearity and Gaussianity may be too inaccurate to model the structural properties of the underlying phenomenon.
6. Fusion of heterogeneous signals is a complex problem, and the solutions proposed so far have largely been problem specific [6, 117]. Hence, they are not easily scalable in terms of the number and diversity of modalities used in the fusion process.

Thus, while signal processing for single modality signals is a well researched area, the problem of jointly analyzing heterogeneous signals is more complex and has not been investigated in detail. However, some advances have been reported in the past. We review these studies in the next section.

1.1 Literature Survey

Fusion of audio and visual information has been dealt with in studies involving audio-visual segmentation, retrieval and summarization of multimedia streams [109], joint audio-video tracking [9], audio-visual face and speech recognition [17], and audio-visual event detection in sports events [67]. Below, we discuss some of the noteworthy contributions.

Hershey and Movellan [49] were among the first to consider the problem of measuring audio-visual synchrony. They proposed to use mutual information to measure audio-visual synchrony. The authors compute mutual information between the energy

of an audio track and individual pixels in video under the assumption that both the audio and video features are jointly Gaussian. Slaney and Covell [91] generalized this approach and measure synchrony between audio features and video images (instead of single pixels). The authors use canonical correlation analysis (CCA) which allows the comparison of sequences of different dimensions, and thus deduce relationships between audio features and video frames. The application of CCA was also considered by Kidron et al. [63] for an audio-video localization problem. The goal is to localize pixels in images that correspond to the acquired audio signal. By exploiting the sparsity of cross-modal events, their algorithm is shown to provide a unique solution, and gives promising results even in the presence of visual distractions. In [84], the authors use CCA to extract lip features that are maximally correlated to the Mel-Frequency Cepstral Coefficients (MFCC) [81] extracted from the audio signal. Lip features such as 2D-discrete cosine transform (DCT) coefficients of the intensity based image, optical flow vectors within the lip region, and distances between pre-defined points on the lip contour were considered.

The problem of speaker detection is addressed in [24] where a time-delay neural network is used to learn the correlation between a single microphone signal and a camera image. Normalized cross correlation between consecutive images is employed as the video feature, while cepstral representation is used for the audio signal. Probabilistic models such as Hidden Markov Models (HMM) and Bayesian networks (e.g. [9]) have also been employed to represent a sequence of detected events in a video stream. These ideas have been extended to develop coupled HMMs to combine probabilistic models developed for each sensing modality [41]. However, the audio and video observations are typically assumed to be conditionally independent.

The methods discussed above are based on rather restrictive assumptions such as linearity, joint Gaussianity and statistical independence; they fail to extract any nonlinear dependence information between the different modalities. In order to ad-

dress these issues, information-theoretic approaches have been proposed in the recent past. Fisher and co-workers [39, 40] have developed a multimodal signal-level fusion framework based on a probabilistic generation model. The audio and video features, namely, audio periodograms and pixel intensities, are first projected onto a lower-dimensional subspace. The projection rules are restricted to be linear and are chosen so that they maximize the mutual information (instead of the correlation coefficient) between the projected audio and video representations. The audio-video joint density required to compute the maximally informative projections are estimated using the Parzen’s density estimator [76] instead of making the Gaussian assumption. A similar approach is proposed in [15] by Butz and Thiran. Correspondence between the audio and video signals is quantified by maximizing the ratio between the audio-visual mutual information and the audio-visual joint entropy. The distributions are again estimated using the nonparametric Parzen’s estimator. The use of Parzen windows to estimate mutual information, however, requires a considerable amount of data for reliable estimation, and thus may not scale very well with the number of sensors.

In contrast to the previous research efforts that mainly address the data association problem, we consider a more general formulation in this dissertation. While the main interest of the previous studies was in detecting changes in the statistical dependence between heterogeneous sources, our goal here is to detect both, the changes in statistical dependence (*common information*) between the heterogeneous sources, as well as variations in the marginal distributions (*complementary information*). Our solution is based on the *statistical theory of copulas* [70] which allows us to approximate joint distributions with *arbitrary* marginal distribution functions. Copulas allow modeling of the potential nonlinear relationships between multiple modalities, and thus do not require linearity and Gaussianity assumptions. Further, the parametric nature of our solution makes the algorithms easily scalable with respect to the number of sensors. No prior knowledge regarding signal models or functional relationship between

the modalities is assumed. Instead, joint distributions are approximated directly from the acquired data using copula theory.

1.2 Thesis Organization and Main Contributions

As discussed above, the problem of fusing or joint processing of heterogeneous data is difficult and the solutions proposed have largely been problem specific. The central goal of this dissertation is to develop a principled methodology for fusion of sensors with different sensing capabilities. Our objective is to model correlated multi-modal data, so that effective techniques can be developed for inference problems such as detection, estimation and tracking. This dissertation investigates the use of copula functions for modeling heterogeneous random vectors. Particularly, our focus is on developing a framework for general binary hypothesis testing problems.

More precisely, in Chapter 2, we discuss the issues involved in modeling heterogeneous random vectors, and propose a solution based on the *statistical theory of copulas*. Copula functions are more general descriptors of dependence (than the well-known Pearson's correlation coefficient ρ) and are good candidates for modeling complex relationships between multiple modalities such as audio and video. We show, in this chapter, how they possess all the *ingredients* necessary to model heterogeneous measurements. We give clear points of comparison between the proposed copula-based models and other approaches that are currently adopted for characterizing heterogeneous random vectors. Several methods for estimating copula parameters from the acquired data, and the estimation performance, are also reviewed in this chapter.

The problem of multisensor detection is formulated as a general two hypotheses testing problem in Chapter 3, and a copula-based test statistic is derived. Particularly important here is the choice of the copula function. Different copula functions would associate the same set of individual modalities to different joint distributions, thus

resulting in different detection performances. This important problem of identifying the best copula is addressed in this chapter. In addition to the simulation results, we also test our methodology on real-data and present results for a multibiometric fusion application.

Hypothesis testing in the presence of unknown model parameters is considered in Chapter 4. The use of the generalized likelihood ratio test (GLRT) is well-known [61] in the classical setting, where, the PDFs of the observations under both hypotheses are completely known but for some model parameters. We consider the case when one has only partial knowledge regarding the PDFs. Specifically, a copula model that would converge to the true dependence structure is unknown. We call the resultant test, the misspecified GLRT (mGLRT), and also propose methods for large sample performance analysis. As an application, we solve the problem of detecting the presence of a human using footstep signals from seismic and acoustic sensors. An approach based on CCA and copula theory is employed to derive mGLRT. Experimental results based on real data collected at the U.S. Army Research Lab (US-ARL) are presented which show significant improvements in detection performance for different walking styles.

A distributed detection problem is formulated in Chapter 5. Local measurements are first compressed using uniform scalar quantizers before their transmission to a fusion center. The fusion center employs a copula-based rule to fuse the received correlated multi-bit decisions. It is acknowledged that the exact fusion rule is computationally expensive especially when the number of sensors and/or the number of quantization levels increase. A suboptimal procedure based on Widrow's quantization model [112, 113] is proposed that requires injecting an artificial noise at the receiver before fusion. The proposed system greatly reduces the computational complexity of the system, and can be considerably accurate, especially when the number

of quantization levels is large. An upper bound on the loss of detection rate due to quantization, at a fixed probability of false alarm P_F , is also derived.

In Chapter 6, we address the problem of quantifying dependence between multiple time series. As a specific application, we study the phenomenon of synchronization between multiple simultaneously recorded electroencephalographic (EEG) signals. Quantification of synchrony between EEG channels is an important problem as it is known to be indicative of several neurophysiological disorders. Several measures have been proposed in the past to test if the EEG channels are inter-related, and also quantify the ‘strength’ of this relation. While some measures are limited to being bivariate, some others, even though multivariate, may fail to account for nonlinear dependencies. We exploit copula theory to alleviate both these limitations. The copula derived synchrony measures are then used for diagnosing Alzheimer’s disease at a very early stage. Results show that the use of the proposed synchrony measures in conjunction with other synchrony measures such as Granger causality and stochastic event synchrony improves the detection of Alzheimer’s disease onset.

Finally, in Chapter 7, we summarize the findings and results of this dissertation and draw some conclusions. Several directions and ideas for future research are also presented.

Chapter 2

Statistical Modeling of Heterogeneous Signals

Given a set of observations $\mathbf{z} = \{z_{1l}, z_{2l}, \dots, z_{Nl}\}_{l=0}^{L-1}$ generated from an unknown joint density $f(\mathbf{z})$, it is often required to represent the data by fitting a parametric model $\{\hat{f}(\mathbf{z}; \psi \in \Psi \subset \mathbb{R}^a)\}$ where ‘ a ’ denotes the dimensionality of the parameter vector ψ . This is the first step when developing algorithms for signal processing applications such as detection, estimation and tracking. An important aspect is the choice of the model $\{\hat{f}(\cdot; \Psi)\}$. It is necessary to consider the underlying physical principles, together with difficulties of measurement and observations (noise sources) when formulating a parametric probability model. Appropriate statistical techniques, such as the method of maximum likelihood (ML), can then be used for purposes of estimation and statistical inference.

Determining $\{\hat{f}(\mathbf{z}; \Psi)\}$, however, is not always straightforward. This is especially the case with data streams of disparate modalities (such as audio and video) that typically share complex relationships. As Fisher and Darrell [39] note, one of the main difficulties limiting the development of algorithms that could process heterogeneous data is the scarcity of ‘good’ models to describe their joint statistics. In this chapter,

an approach based on the statistical theory of copulas is proposed as a solution. It will be helpful for later discussions to partition the parameter vector ψ as below,

$$\psi = \begin{pmatrix} \psi_p = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_N \end{pmatrix} \\ \psi_d \end{pmatrix} \quad (2.1)$$

where $\psi_p \in \Psi_p \subset \mathbb{R}^{a_p}$ and $\psi_d \in \Psi_d \subset \mathbb{R}^{a_d}$ denote the marginal and dependence parameter vectors respectively. For example, for a bivariate Gaussian density, $f(z_1, z_2) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, $\psi_p = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)^T$ and $\psi_d = \rho$.

The remainder of the chapter is organized as follows. In Section 2.1, we identify the key requirements that a model needs to satisfy to ‘adequately’ model heterogeneous random vectors. Modeling approaches that are commonly adopted, and their limitations are also reviewed here. Section 2.2 discusses the theory of copulas. We begin with Sklar’s theorem which is central to copula theory, give examples and parametric forms of copula densities used in this dissertation in Section 2.2.1, and discuss approaches for estimating copula parameters from the acquired data in Section 2.2.2. Relationship between copula parameters and nonlinear rank dependence measures such as Kendall’s τ and Spearman’s ρ , are also discussed in this section. Motivation for developing methods for copula selection is provided in Section 2.2.3, before summarizing the chapter in Section 2.3.

2.1 Modeling Heterogeneous Data - Challenges

Heterogeneous data streams are not always commensurate. The physics governing each modality may be different and so may be their dimensionality, support and sampling rates during data acquisition. These differences require the formulation of

multivariate statistical models that allow for marginals that follow disparate distribution functions. A formal definition for heterogeneous observations follows.

Definition 2.1. *A random vector $\mathbf{Z} = \{Z_n\}_{n=1}^N$ governing the joint statistics of an N -variate data set is termed as heterogeneous if the marginals Z_n are non-identically distributed.¹*

Further, in most information fusion applications, the signals share a common source and thus may exhibit statistical dependence. Consider, for example, an acoustic sensor and a video camera monitoring a region for trespassers. Presence of a target may result in an increase in both the acoustic energy and the pixel intensities of the images acquired by the video camera. Both sensors provide information about the same event (and hence are statistically dependent) but in different *domains*. Thus, the following two requirements are expected of a *good* model for heterogeneous random variables:

1. The model should allow for disparate marginal distribution functions.
2. The model must also account for any statistical dependence that may be present between the disparate marginal distribution functions.

Section 2.2 reviews the theory of copulas and shows how copula densities possess properties that allow us to satisfy the above requirements. We first provide a brief discussion on some commonly adopted models. While these approaches are attractive due to their analytical tractability, each of them leads to suboptimal solutions. Assessment of these shortcomings will motivate the proposed copula based solution.

¹Definition 2.1 is, of course, inclusive of the special case when the marginals are identically distributed and/or are statistically independent. It also encompasses the case when the signals, although sharing a common modality (*e.g.*, two acoustic sources or two video sources), may exhibit different statistics (due to different locations, signal strengths, *etc.*). Hence, we prefer the term *heterogeneous* in this dissertation in place of *multimodal* as used by some others in the literature.

2.1.1 Commonly Adopted Approaches for Modeling Heterogeneous Data

The Product Model

Often statistical independence is assumed when dealing with heterogeneous data sets for mathematical tractability. The joint density $f(\mathbf{z}; \psi)$ is approximated as a product of the marginal densities,

$$\hat{f}(\mathbf{z}; \psi) = f_p(\mathbf{z}; \psi_p) = \prod_{n=1}^N f_n(z_n; \psi_n \in \Psi_n) \quad (2.2)$$

As is evident, the product model, $f_p(\mathbf{z}; \psi_p)$ **allows** for disparate marginals but completely neglects the dependence across the disparate marginals.

The Multivariate Gaussian Model

The use of the multivariate Gaussian density

$$\hat{f}(\mathbf{z}; \psi) = f_G(\mathbf{z}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\Sigma)} \exp \left[-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu}) \right] \quad (2.3)$$

to model multivariate random data is quite prevalent in the literature. Two limitations plague this approach:

1. Marginals are constrained to follow Gaussian distributions.
2. The correlation matrix Σ_c characterizes only the **linear** relationship and is thus a weak measure of dependence [68]. For example, in Fig. 2.1, although dependence between variables is evident in the scatter plot, the correlation coefficient ρ computed for this bivariate data is zero. Inter-modal interactions are usually much more complex (than just being linear).

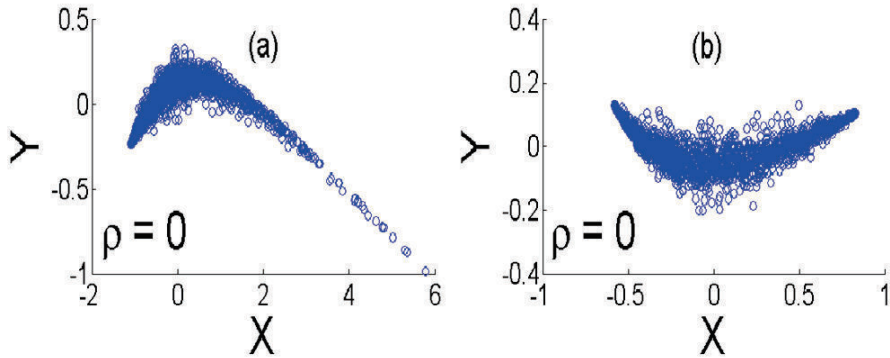


Figure 2.1: Correlation coefficient is a weak measure of dependence

Copulas are more general descriptors of dependence that can describe the functional relationship between multiple random variables more accurately as will be discussed in Section 2.2.

The Nataf Transform

Nataf transform, first introduced in the reliability literature [53], involves transforming the marginals to the standard normal space, $Y_n = \Phi^{-1}(F_n(Z_n; \psi_n))$ where Φ is the standard normal cumulative distribution function (CDF). After this transformation, the joint density $f(\mathbf{z})$ is given as

$$\hat{f}(\mathbf{z}; \psi) = f_{Nataf}(\mathbf{z}; \psi_p, \Sigma_c) = f_p(\mathbf{z}; \psi_p) \frac{\phi(y_1, \dots, y_N; \Sigma_c)}{\phi(y_1) \cdots \phi(y_N)} \quad (2.4)$$

where $\phi(\mathbf{y}; \Sigma_c)$ and $\phi(y_n)$ are the N-dimensional multivariate and univariate standard normal densities respectively. Σ_c is the covariance matrix in the transformed (normal) space. As will be evident in Section 2.2, Nataf transform is equivalent to using the Gaussian copula density. Thus, the copula approach is a general framework that includes the Nataf transform approach as a special case.

2.2 Copula Theory

We begin with the definition of a copula.

Definition 2.2. *An N -dimensional copula is an N -variate CDF on $[0, 1]^N$ whose univariate marginals are uniformly distributed on $[0, 1]$.*

Thus, copulas functions ‘couple’ multivariate joint distribution functions to their component marginal distribution functions [56, 65, 70]. The following theorem by Sklar is central to the statistical theory of copulas.

Theorem 2.1. *(Sklar’s Theorem, 1959)*

Let F be an N -dimensional CDF with continuous marginal CDFs F_1, F_2, \dots, F_N . Then there exists a unique copula C such that for all z_1, z_2, \dots, z_n in $[-\infty, \infty]$,

$$F(z_1, z_2, \dots, z_N) = C(F_1(z_1), F_2(z_2), \dots, F_N(z_N)) \quad (2.5)$$

Note that the marginal CDFs of the copula $C(u_1, u_2, \dots, u_N)$ are uniform due to the probability integral transform ($U_n = F_n(Z_n) \sim \mathcal{U}(0, 1)$). The joint probability density function (PDF) is obtained by taking the N^{th} order derivative of (2.5),

$$\begin{aligned} f(\mathbf{z}) &= \frac{\partial^N}{\partial z_1 \dots \partial z_N} (C(F_1(z_1), F_2(z_2), \dots, F_N(z_N))) \\ &= f_p(\mathbf{z}) c(F_1(z_1), \dots, F_N(z_N)) \end{aligned} \quad (2.6)$$

It is interesting to note the form of Eq. (2.6). The copula density *re-weights* the zero dependence or the product model $f_p(\mathbf{z})$ to incorporate statistical dependence between the variables.

Theorem 2.1 also admits the following converse when one wants to construct a statistical model by considering, separately, the univariate behavior of the components of a random vector and the dependence structure defined by some copula. The

converse is especially useful in practice when the true CDF F (and hence the true copula C) is unknown. Note that this property of separating the dependence structure from the univariate marginals is also well-suited for modeling heterogeneous random variables where a different distribution might be needed for each marginal (see Section 2.1).

Theorem 2.2. *If F_1, F_2, \dots, F_N are univariate marginal CDFs and if K is an N dimensional copula, then the function $F_k : \mathbb{R}^N \rightarrow [0, 1]$,*

$$F_k(z_1, \dots, z_N) = K(F_1(z_1), \dots, F_N(z_N)), \quad (2.7)$$

is a valid N -variate CDF with marginals F_1, F_2, \dots, F_N .

A copula based parametric model can be derived by taking the N^{th} order derivative of Eq. (2.7) to obtain

$$\hat{f}(\mathbf{z}; \psi) = f_p(\mathbf{z}; \psi_p)k(F_1(z_1; \psi_1), \dots, F_N(z_N; \psi_N); \psi_d). \quad (2.8)$$

A variety of copula functions, with different dependence properties, exist in the literature [65, 70]. The most popular ones out of these are the Student's t and Gaussian copula functions, and copula functions that belong to the Archimedean family. Details of these copula functions, their functional form and the parameters incorporating dependence, are summarized next.

2.2.1 Some Well-known Copula Families

Elliptical Copulas

Copulas of elliptically contoured distribution functions are known as elliptical copulas. The two well-known elliptical copulas are the Gaussian and the t copulas that are derived from the multivariate Gaussian and the Student's t distributed functions

respectively. Both specify dependence using the correlation matrix and are given as follows.

$$K_G(\cdot|\Sigma) = \Phi_\Sigma^G(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_N)) \quad (2.9)$$

where Φ_Σ^G denotes the multivariate Gaussian CDF with correlation matrix Σ and Φ denotes the univariate Gaussian CDF, and,

$$K_t(\cdot|\Sigma, \nu) = t_{\nu, \Sigma}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_N)) \quad (2.10)$$

where, $t_{\Sigma, \nu}$ is the multivariate Student-t distribution with correlation matrix Σ and ν degrees of freedom ($\nu \geq 3$) and t_ν denotes the univariate Student's t distribution with ν degrees of freedom. As $\nu \rightarrow \infty$, the t copula approaches the Gaussian copula function.

Archimedean Copulas

Archimedean copulas form an important family of copulas and are given as follows.

$$K(\cdot|\psi_d) = \zeta^{-1} \sum_{n=1}^N \zeta(u_i) \quad (2.11)$$

where, ζ is the so called generator function. By appropriately choosing the generator function, different copula functions with different dependence properties can be generated. The generator functions for Clayton, Frank, Gumbel and the product copulas and their parametric forms are shown in Table 2.1.

The product copula is a special case which results in the joint density equal to the product of the marginal PDFs, i.e., it corresponds to zero inter-modal dependence.

Table 2.1: Archimedean Copula functions

| Copula | Generator Function | Parametric Form | Range of ψ_d |
|---------|---|---|--------------------------------|
| Clayton | $\frac{1}{\psi_d} (u^{-\psi_d} - 1)$ | $\left(\sum_{n=1}^N u_i^{-\psi_d} - 1 \right)^{-\frac{1}{\psi_d}}$ | $[-1, \infty) \setminus \{0\}$ |
| Frank | $\frac{\exp^{-\psi_d} - 1}{\exp^{-\psi_d u} - 1}$ | $\frac{-1}{\psi_d} \log \left(1 + \frac{\prod_{n=1}^N (\exp^{-\psi_d u_n} - 1)}{\exp^{-\psi_d} - 1} \right)$ | $\mathbb{R} \setminus \{0\}$ |
| Gumbel | $-\log u^{\psi_d}$ | $\exp \left\{ - \left(\sum_{n=1}^N (-\log u)^{\psi_d} \right)^{\frac{1}{\psi_d}} \right\}$ | $[1, \infty)$ |
| Product | $-\log u$ | $\prod_{n=1}^N u_n$ | NA |

The Farlie-Gumbel-Morgenstern (FGM) Family

The FGM copula is given as

$$K(\cdot; \psi_d) = \left(\prod_{n=1}^N u_n \right) \left[1 + \sum_{n=2}^N \sum_{1 \leq j_1 < \dots < j_n \leq N} \psi_{d_{j_1 j_2 \dots j_n}} (1 - u_{j_1})(1 - u_{j_2}) \dots (1 - u_{j_n}) \right]. \quad (2.12)$$

It is the only copula that is quadratic in all its arguments u_n .

While many other copula families have been defined, the above mentioned copula functions are the ones considered in this work. However, the theory developed is applicable to all valid copula functions.

2.2.2 Estimation of Model Parameters

In this section, we discuss methods for estimating the model parameters $\psi = (\psi_p, \psi_d)^T$, while assuming full knowledge of the specific copula function that would best characterize the given data. The issue of selecting the copula function is taken up in Section 2.2.3.

Maximum likelihood estimation

The parameter vector ψ in Eq. (2.8) is typically unknown and is estimated from the acquired observations. Given L independent and identically distributed (i.i.d.) samples of the N -variate data, the ML estimate $\hat{\psi}_{mle}$ maximizes the log-likelihood function,

$$\begin{aligned} \mathcal{L}(\mathbf{z}; \psi) &= \sum_{l=1}^L \log \hat{f}_k(\mathbf{z}_l; \psi) & (2.13) \\ &= \sum_{l=1}^L \log \left\{ \left(\prod_{n=1}^N f_n(z_{nl}; \psi_n) \right) k(F_1(z_{1l}; \psi_1), \dots, F_1(z_{Nl}; \psi_N); \psi_d) \right\} & (2.14) \end{aligned}$$

Thus,

$$\begin{aligned} \hat{\psi}_{mle} &= \operatorname{argmax}_{\psi \in \Psi} \left\{ \sum_{l=1}^L \log k(F_1(z_{1l}; \psi_1), \dots, F_1(z_{Nl}; \psi_N); \psi_d) \right. \\ &\quad \left. + \sum_{l=1}^L \sum_{n=1}^N f_n(z_{nl}; \psi_n) \right\} & (2.15) \end{aligned}$$

Assuming that the usual *regularity conditions* [104] hold so that the asymptotic ML theory remains valid, the ML estimate $\hat{\psi}_{mle}$ is known to be consistent and asymptotically efficient. In the limit ($L \rightarrow \infty$),

$$\sqrt{L}(\hat{\psi}_L - \psi_0) \rightsquigarrow \mathcal{N}(0, \mathcal{F}^{-1}(\psi_0)), \quad (2.16)$$

where, ‘ \rightsquigarrow ’ denotes convergence in distribution, $\mathcal{F}(\psi_0)$ is the Fisher information matrix, and ψ_0 is the true value of ψ . However, Eq. (2.16) holds only when the underlying parametric model is *well-specified* (see Definition 2.3 below).

Definition 2.3. (*Well-specified model, White 1994 [110]*)

A parametric model $\{f(x; \Theta)\}$ is well-specified for a random variable X if there

exists a unique $\theta' \in \Theta$ such that $f(x; \theta') \in \{f(x; \Theta)\}$ corresponds to the true density of X . Otherwise, $\{f(x; \Theta)\}$ is said to be misspecified for X [57].

Quasi-Maximum Likelihood Estimation

When the model is misspecified, $\hat{\psi}_L$ is no longer consistent with the true value ψ_0 . Statistical properties of $\hat{\psi}_L$ in such cases can be studied using the quasi-maximum likelihood (QML) framework developed by White [110].

Let ψ_* , where

$$\psi_* = \operatorname{argmax}_{\psi \in \Psi} \mathbb{E}_f \left(\frac{1}{L} \sum_{l=1}^L \frac{\partial}{\partial \psi} \log \hat{f}_k(\mathbf{z}_l; \psi) \right), \quad (2.17)$$

be a point in the interior of a compact set $\Psi \subset \mathcal{R}^p$. The expectation in Eq. (2.17) is with respect to the true PDF $f(\mathbf{z})$. ψ_* is also known as the pseudo-true value associated with the model $\hat{f}_k(\mathbf{z}_l; \psi)$. Under the usual regularity conditions, the estimate $\hat{\psi}_L$ is determined by solving the following equation,

$$\xi_L(\psi) = \frac{1}{L} \sum_{l=1}^L \frac{\partial}{\partial \psi} \log \hat{f}_k(\mathbf{z}_l; \psi). \quad (2.18)$$

That is, $\xi_L(\hat{\psi}_L) = 0$. For simplicity of presentation, let ψ be a scalar parameter. By mean value theorem,

$$\begin{aligned} \sqrt{L} (\hat{\psi}_L - \psi_*) &= \frac{\sqrt{L} (\xi_L(\hat{\psi}_L) - \xi_L(\psi_*))}{\xi'_L(\tilde{\psi})} \\ &= -\frac{\sqrt{L} \xi_L(\psi_*)}{\dot{\xi}_L(\tilde{\psi})} \end{aligned} \quad (2.19)$$

where, $\tilde{\psi}$ is a point between $\hat{\psi}_L$ and ψ_* , and, $\dot{\xi}_L(\psi)$ denotes the derivative of $\xi_L(\psi)$ with respect to ψ . By Central limit theorem (CLT), the numerator $-\sqrt{L} \xi_L(\psi_*) = -\sqrt{L} \sum_l \frac{\partial}{\partial \psi} \log \hat{f}(\mathbf{z}_l; \psi)|_{\psi=\psi_*}$ converges to a normal distribution with mean zero and

variance $a_{\psi_*} = E_f \left[\frac{\partial}{\partial \psi} \log \hat{f}_k(\mathbf{z}; \psi) |_{\psi=\psi_*} \right]^2$. We now analyze the term in the denominator. Since $\psi_* < \tilde{\psi} < \hat{\psi}_L$ and $\tilde{\psi} \xrightarrow{P} \psi_*$ (Theorem 3.4 in [110]), we have $\dot{\xi}_L(\tilde{\psi}) \xrightarrow{P} b_{\psi_*} = E_f \left[\frac{\partial^2}{\partial \psi^2} \log \hat{f}_k(\mathbf{z}; \psi) |_{\psi=\psi_*} \right]$. Now, using Slutsky's theorem, $-\sqrt{L} \frac{\xi_L(\psi_*)}{\xi'_L(\tilde{\psi})} \xrightarrow{P} \frac{x}{b_{\psi_*}}$ where the random variable $x \sim \mathcal{N}(0, a_{\psi_*})$. Thus,

$$\sqrt{L} \left(\hat{\psi} - \psi_* \right) \rightsquigarrow \mathcal{N} \left(0, \frac{a_{\psi_*}}{b_{\psi_*}^2} \right) \quad (2.20)$$

The above result can be easily extended to the case when ψ is a p -dimensional vector ($p > 2$). In this case,

$$\sqrt{L} \left(\hat{\psi}_L - \psi_* \right) \rightsquigarrow \mathcal{N}(0, \Sigma). \quad (2.21)$$

The $p \times p$ covariance matrix, is given by

$$\Sigma = B_{\psi_*}^{-1} A_{\psi_*} B_{\psi_*}^{-1} \quad (2.22)$$

where

$$A_{\psi} \equiv E_f \left[\frac{\partial}{\partial \psi} \log \hat{f}_k(\mathbf{z}; \psi) \frac{\partial}{\partial \psi'} \log \hat{f}_k(\mathbf{z}; \psi) \right], \quad B_{\psi} \equiv E_f \left[\frac{\partial^2}{\partial \psi \partial \psi'} \log \hat{f}_k(\mathbf{z}; \psi) \right]. \quad (2.23)$$

When the model is well-specified and the regularity conditions hold, $B_{\psi_*} = -A_{\psi_*}$ and therefore, $\Sigma = -B_{\psi_*}^{-1} = \mathcal{F}^{-1}(\psi_0)$, as expected.

Method of Inference for Margins

Maximizing Eq. (2.15) can be computationally intensive especially for higher dimensions (i.e., when N is large) as it requires the joint estimation of the marginal and

copula parameters. A simpler approach known as the method of inference for the margins (IFM) [57] estimates ψ in two steps:

- Step 1: Estimation of the marginal parameters

$$\hat{\psi}_p = \operatorname{argmax}_{\psi_p \in \Psi_p} \sum_{n=1}^N \sum_{l=1}^L \log f_n(z_{nl}; \psi_n) \quad (2.24)$$

- Step 2: Estimation of the copula dependence parameter

$$\hat{\psi}_d = \operatorname{argmax}_{\psi_d \in \Psi_d} \sum_{l=1}^L \log k \left(F_1(z_{1l}; \hat{\psi}_1), \dots, F_N(z_{Nl}; \hat{\psi}_N); \psi_d \right) \quad (2.25)$$

where the marginal parameters in Eq. (2.15) have been replaced by their estimates computed in Step 1. From Eqs. (2.24) and (2.25), we have

$$\hat{\psi}_{ifm} = \left(\hat{\psi}_p, \hat{\psi}_d \right) \quad (2.26)$$

Note that, equivalence between the IFM and ML estimates does not hold in general. To see this, let \mathcal{L} , \mathcal{L}_n and \mathcal{L}_k denote the *full*, marginal and copula likelihood functions respectively. That is,

$$\mathcal{L}(\mathbf{z}) = \sum_{l=1}^L \log \hat{f}_k(z_{1l}, z_{2l}, \dots, z_{Nl}; \psi) \quad (2.27)$$

$$\mathcal{L}_n(\mathbf{z}) = \sum_{l=1}^L \log f_n(z_{nl}; \psi_n) \quad (2.28)$$

$$\mathcal{L}_k(\mathbf{z}) = \sum_{l=1}^L \log k(z_{1l}, z_{2l}, \dots, z_{Nl}; \psi_d) \quad (2.29)$$

Now, while $\hat{\psi}_{mle}$ is obtained by solving

$$\left(\frac{\partial \lambda}{\partial \psi_1}, \dots, \frac{\partial \lambda}{\partial \psi_N}, \frac{\partial \lambda}{\partial \psi_d} \right) = 0, \quad (2.30)$$

the IFM estimate is computed by solving

$$\left(\frac{\partial \lambda_1}{\partial \psi_1}, \dots, \frac{\partial \lambda_N}{\partial \psi_N}, \frac{\partial \lambda_k}{\partial \psi_d} \right) = 0. \quad (2.31)$$

It is also possible to consider a semi-parametric approach where a parametric form is assumed only for the dependence structure. The marginal CDFs in Eq. (2.25) are replaced with their empirical estimates. This approach is known as the canonical maximum likelihood (CML) method and the estimate is computed as below.

$$\hat{\psi}_{cml} = \operatorname{argmax}_{\psi_d \in \Psi_d} \sum_{l=1}^L \log k \left(\tilde{F}_1(z_{1l}), \dots, \tilde{F}_N(z_{Nl}); \psi_d \right). \quad (2.32)$$

where,

$$\tilde{F}_n(a) = \frac{1}{L} \sum_{l=1}^L \mathbb{I}(z_{nl} \leq a). \quad (2.33)$$

$\mathbb{I}(\mathcal{E})$ in Eq. (2.33) indicates the occurrence of event \mathcal{E} .

Chapters 3 and 4 of this dissertation consider the use of IFM for estimating model parameters. The CML approach is adopted in Chapter 6.

Estimation via nonparametric dependence measures

The copula dependence parameter ψ_d can also be estimated by exploiting its relation with several concordance measures such as Kendall's τ and Spearman's ρ [70]. Let (z_{1i}, z_{2i}) and (z_{1j}, z_{2j}) be two observations from a bivariate vector (Z_1, Z_2) . The observations are said to be concordant if $(z_{1i} - z_{1j})(z_{2i} - z_{2j}) > 0$ and discordant if $(z_{1i} - z_{1j})(z_{2i} - z_{2j}) < 0$.

The population version of Kendall's $\tau_{1,2}$ between the variables Z_1 and Z_2 can be expressed in terms of a copula function $K(\cdot; \psi_d)$ as

$$\tau_{1,2} = 4 \int_{u_1, u_2} K(u_1, u_2; \psi_d) dK - 1 \quad (2.34)$$

where $u_n = F_n(z_n) \sim \mathcal{U}(0, 1)$ (see Section 2.2). The above integral equation Eq. (2.34) can be used to express ψ_d in terms of $\tau_{1,2}$. For example, the following relation holds for elliptical copulas [70],

$$\psi_d = \sin\left(\frac{\pi\tau_{1,2}}{2}\right) \quad (2.35)$$

Given observations $\{z_{1l}, z_{2l}\}_{l=0}^{L-1}$, we first rank order them and obtain the sample estimate of Kendall's $\tau_{1,2}$,

$$\hat{\tau}_{1,2} = \frac{n_c - n_d}{n_c + n_d} \quad (2.36)$$

where, n_c and n_d are the number of concordant and discordant pairs respectively. An estimate of $\hat{\psi}_d$ can be obtained by using $\hat{\tau}_{1,2}$ in Eq. (2.34) (equivalently in Eq. (2.35) for elliptical copulas).

The extension to the N -variate case is straightforward in the case of elliptical copulas where each element of the $N \times N$ correlation matrix $\Sigma_{i,j}$ ($i, j = 1, 2, \dots, N$) can be obtained by computing the pair-wise Kendall's $\tau_{i,j}$.

Similar relations are available for Spearman's correlation coefficient, ρ^s , as well. The population version of ρ^s can be given in terms of the copula function $K(\cdot)$ as below [70]:

$$\rho_{1,2}^s = 12 \int_{u_1, u_2} u_1 \cdot u_2 dK(\cdot; \psi_d) - 3. \quad (2.37)$$

The above integral equation Eq. (2.37) can be used to express ψ_d in terms of $\rho_{1,2}^s$.

When $\rho_{1,2}^s$ is unknown, its sample estimate can be used in Eq. (2.37). Bivariate measurements (z_{1l}, z_{2l}) are first converted to rankings x_i and y_i . The sample estimate $\hat{\rho}^s$ is then given as [70],

$$\hat{\rho}_{1,2} = 1 - \frac{6 \sum d_i^2}{L(L^2 - 1)}, \quad (2.38)$$

where,

$$\begin{aligned} d_i &= a_i - b_i \\ &= \text{difference between the ranks of } z_1 \text{ and } z_2 \\ L &= \text{number of observations} \end{aligned}$$

In Chapter 4, we use Kendall's τ to generate a synthetic heterogeneous dataset.

2.2.3 Selection of the Copula Function

As discussed above, several copula functions, with different dependence characteristics, have been defined in the literature. Different copula functions associate the same set of marginals to different joint distributions. With such one-to-many possible mappings of the marginals, an important question that arises is: *How does one choose a particular copula from (say) a finite set \mathcal{A}_k of valid copula functions?* This is essentially a model selection problem. Approaches in the past have mostly focused on data modeling, i.e., select the copula that best fits the data. Several copula goodness-of-fit tests have been developed in the past (see [10, 43] for recent surveys). In this dissertation, we propose a more application-specific paradigm for copula selection (than data modeling as the primary goal).

Classification performance of a binary hypothesis test depends on the statistics of the detector functional and the resultant hyperplane separating the two hypotheses, rather than the density estimates of the observations under each hypothesis. Thus, although density estimates may be *coarse*, the resultant classification rates may still be within the required limits [89, pg. no. 124]. Thus, attempts to model densities with high accuracy may prove to be an *overkill* for hypotheses testing problems.

In Chapter 3, two approaches for copula selection are proposed. The first copula selection method is based on $D(f, \hat{f}_k)$, the Kullback-Leibler (KL) divergence between the true joint PDF and its copula-based estimate. We show that $D(f, \hat{f}_k)$ is related to the area under the probability of detection curve, and under certain conditions quantifies the ‘distance’ between the distributions of the test statistic under the two competing hypotheses. The second approach considers maximizing the area under the receiver operating characteristic curve (AUC), a global measure of classifier performance. Note that these methods (particularly the AUC based method) can be viewed as discriminative training methods [11, 71, 79, 101] that do not concentrate on modeling the data distribution under each hypothesis. Instead, the aim is to directly separate the competing hypotheses through design of effective decision boundaries.

2.3 Summary

Statistical signal processing tasks such as detection, estimation and tracking require complete specification of the joint distribution functions of the observed samples. However, in many cases, the derivation of the joint distribution functions becomes mathematically intractable. In problems such as processing of heterogeneous signals, random variables associated with each source may follow disparate distribution functions due to differences in physics governing each modality. Moreover, inter-modal interactions may be complex and well-known measures such as Pearson’s correlation

coefficient ρ may be insufficient to characterize statistical dependence across multiple modalities.

In this chapter, statistical theory of copulas especially in the context of modeling heterogeneous random variables has been discussed. Copula theory provides a re-parametrization of joint densities where the dependence structure is separated out from the constituent marginals; a key property useful for modeling heterogeneous random vectors. They are more general descriptors of statistical dependence and possess all the ingredients necessary to model heterogeneous data. Several copula functions have been defined in the literature. The performance of a copula based system depends strongly on the chosen copula density, and, it is important to derive formal methods for copula selection which will be done in the following chapters for binary hypothesis testing problems.

Chapter 3

Hypothesis Testing With Heterogeneous Data

In Chapter 2, the problem of modeling heterogeneous random vectors was considered. It was discussed how disparate marginals give rise to heterogeneity, and also, how copula functions allow construction of multivariate distributions with disparate marginals, while also incorporating the inter-modal dependence structure. In this chapter, the developed copula based statistical model is applied to solve a classification problem, one of the fundamental problems in statistical signal processing.

As shown in Fig. 3.1, a parallel network of N heterogeneous sensors monitor a common region of interest (ROI), and communicate their observations (or features extracted thereof) to a remotely located fusion center. We denote by $\bar{z}_n = (z_{n1}, z_{n2}, \dots, z_{nL})^T$, the vector of L observations (over time) received at the fusion center from sensor n , and consider the following binary hypothesis testing problem at the fusion center,

$$\begin{aligned} H_0 : (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_N) &\sim p_{\mathbf{z}}(\mathbf{z}; H_0) = g(\bar{z}_1, \dots, \bar{z}_N), \\ H_1 : (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_N) &\sim p_{\mathbf{z}}(\mathbf{z}; H_1) = f(\bar{z}_1, \dots, \bar{z}_N), \end{aligned} \tag{3.1}$$

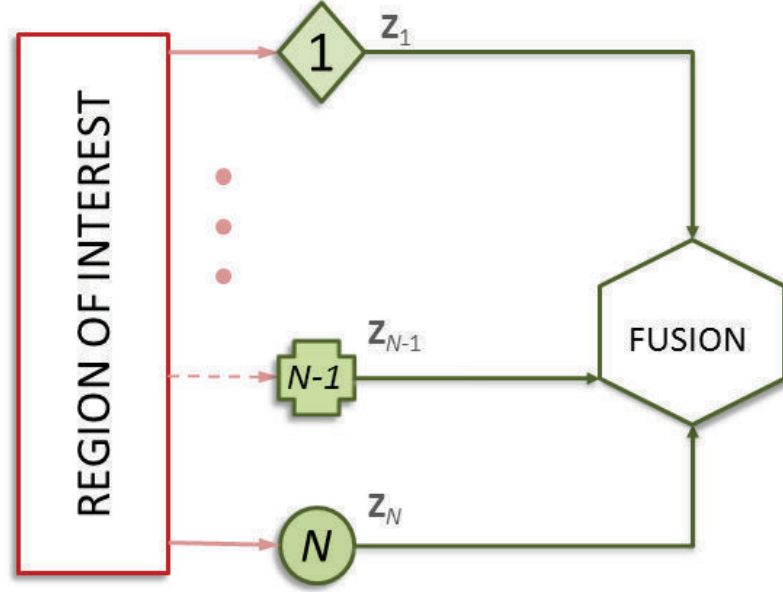


Figure 3.1: A multisensor system with common region of interest (ROI). Different shapes for the sensors denote different sensing capabilities.

where $g(\cdot)$ and $f(\cdot)$ denote the appropriate PDFs. The fusion center consolidates the data received from local sensors, and computes the test statistic, $\mathcal{T}(\mathbf{z})$, based on the received data for the presence (H_1) or absence (H_0) of a target. Two types of error are possible at the fusion center. *Type 1 error* or the probability of false alarm, P_F , is the probability of deciding in favor of H_1 when, in fact, H_0 is true, $P_F = Pr(\text{decide } H_1 | H_0)$. On the other hand, the probability of favoring H_0 when H_1 is true is called the *Type 2 error* or the probability of miss, $P_M = Pr(\text{decide } H_0 | H_1)$. Adopting the Neyman-Pearson approach, the goal of the fusion center is to maximize the detection probability, $P_D = 1 - P_M$, for a fixed $P_F = \alpha$.

The remaining part of the chapter is organized as follows. The copula-based fusion rule is derived in Section 3.1. The important problem of selecting the best copula is considered in Section 3.2. Two selection rules, the KL divergence and the AUC-based copula selection rules are presented, along with simulation results. The method is also tested on real-data provided by the National Institute of Standards and Technology

(NIST), for the multibiometrics fusion problem. The chapter is summarized in Section 3.4.

3.1 Copula-Based Fusion Rule

Given the densities, $p_{\mathbf{z}}(\mathbf{z}; H_0)$ and $p_{\mathbf{z}}(\mathbf{z}; H_1)$, and P_F , decision rule at the fusion center can be derived using the Neyman-Pearson (NP) lemma, which states that, the likelihood ratio (LR) test, or equivalently, the log-likelihood ratio (LLR) test,

$$\mathcal{T}_{LLR}(\mathbf{z}) = \log \underbrace{\frac{p_{\mathbf{z}}(\mathbf{z}; H_1)}{p_{\mathbf{z}}(\mathbf{z}; H_0)}}_{LR} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (3.2)$$

results in the maximum P_D for a given P_F , where η is the detector threshold. P_D and P_F for a given test ‘ $\mathcal{T}(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$ ’ are given as,

$$P_D = \int_{\{\mathbf{z}: \mathcal{T}(\mathbf{z}) > \eta\}} p_{\mathbf{z}}(\mathbf{z}; H_1) d\mathbf{z}, \quad \text{and} \quad P_F = \int_{\{\mathbf{z}: \mathcal{T}(\mathbf{z}) > \eta\}} p_{\mathbf{z}}(\mathbf{z}; H_0) d\mathbf{z}. \quad (3.3)$$

We make the following assumptions in this chapter.

Assumption 3.1. *Each sensor n employs an analog compression rule $\gamma_n : \mathbb{R}^{D_n} \mapsto \mathbb{R}^{d_n}$, $d_n < D_n$, so that the compressed observations $z_n = \gamma_n(\cdot)$, $n = 1, 2, \dots, N$, follow continuous distribution functions.*

Note that $\gamma_n(\cdot)$ can be interpreted as a feature extraction step where data at each sensor is transformed so that only relevant information is extracted and retained. Any further processing is performed on the reduced representation rather than the full size input. This alleviates the well-known *curse of dimensionality*.

Assumption 3.2. *The N -variate sequences $\{z_{1l}, \dots, z_{Nl}\}_{l=1}^L$ are independent and identically distributed (i.i.d).*

Now, given the above assumptions, the PDFs $f(\cdot)$ and $g(\cdot)$ can be expressed in terms of their copula functions and Eq. (3.2) can be written as

$$\mathcal{T}_{LLR}(\mathbf{z}) = \log \frac{f(\bar{z}_1, \dots, \bar{z}_N)}{g(\bar{z}_1, \dots, \bar{z}_N)} \quad (3.4)$$

$$= \sum_{l=1}^L \log \left(\frac{f_p(\mathbf{z}_l) c_1(F_1(z_{1l}), \dots, F_N(z_{Nl}))}{g_p(\mathbf{z}_l) c_0(G_1(z_{1l}), \dots, G_N(z_{Nl}))} \right) \quad (3.5)$$

$$= \sum_{l=1}^L \left\{ \left(\sum_{n=1}^N \log \frac{f_n(z_{nl})}{g_n(z_{nl})} \right) + \log \frac{c_1(F_1(z_{1l}), \dots, F_N(z_{Nl}))}{c_0(G_1(z_{1l}), \dots, G_N(z_{Nl}))} \right\}, \quad (3.6)$$

where the copula density $c_i(\cdot)$ denotes the actual statistical dependence structure between the variables under the hypothesis H_i , $i \in \{0, 1\}$. The subscript ‘ p ’ in Eq. (3.5) is used to denote the product densities under the two hypotheses. $f_n(\cdot)$ (respectively $F_n(\cdot)$) denotes the marginal PDF (respectively CDF) of Z_n under the hypothesis H_1 . Similarly, their counterparts under H_0 are denoted as $g_n(\cdot)$ and $G_n(\cdot)$ respectively.

It is interesting to note the form of the test statistic in Eq. (3.5). The first term,

$$\mathcal{T}_p(\mathbf{z}) = \sum_{l=1}^L \sum_{n=1}^N \log \frac{f_n(z_{nl})}{g_n(z_{nl})} \quad (3.7)$$

corresponds to the differences in the marginal statistics of each modality (or the product distribution) across the two hypotheses while the cross-modal dependence and interactions are included in the second term. $\mathcal{T}_p(\cdot)$ in (3.7), is also the test statistic obtained when, (a) the variables Z_1, Z_2, \dots, Z_N are statistically independent, or, (b) when dependence between them is deliberately neglected for analytical simplicity or due to lack of knowledge of dependence structures $c_1(\cdot)$ and/or $c_0(\cdot)$. The latter case naturally results in a suboptimal performance. In contrast, we replace the unknown dependence structures $c_1(\cdot)$ and $c_0(\cdot)$ in Eq. (3.5) by copula densities $k_1(\cdot; \psi_d)$ and $k_0(\cdot; \lambda_d)$ chosen from a finite set \mathcal{A}_k of valid copula densities. The resulting test

statistic is given as,

$$\mathcal{T}_k(\mathbf{z}) = \sum_{l=1}^L \left\{ \left(\sum_{n=1}^N \log \frac{f_n(z_{nl})}{g_n(z_{nl})} \right) + \log \frac{k_1(F_1(z_{1l}), \dots, F_N(z_{Nl}); \psi_d)}{k_0(G_1(z_{1l}), \dots, G_N(z_{Nl}); \lambda_d)} \right\}, \quad (3.8)$$

where, the copula parameters ψ_d and λ_d , and the marginal PDFs (if unknown), are estimated from training data which is assumed to be available. Note that the copula approach allows us to simplify the complex task of multivariate density estimation. The problem can now be solved in two steps: (1) estimation of univariate marginal PDFs, and (2) selection of copula densities to approximate the unknown statistical dependence.

In the following, we assume that the marginal PDFs are either known or have been estimated from the given training data. We focus primarily on designing rules for copula selection. This allows us to test the efficacy of copula modeling relative to the simpler and more commonly adopted product model (that assumes inter-modal independence) in the context of binary classification. The assumption of known marginal PDFs allows us to compare the copula approach with the best performing product model (since there is no uncertainty in the marginal PDFs).

3.2 Copula Selection for Hypothesis Testing Problems

In this section, we derive methods for choosing copula densities to fuse local sensors' observations using Eq. (3.8).

3.2.1 The Kullback-Leibler Divergence Criterion

Assume that the goal is to select the copula density $k_1(\cdot)$, and thus estimate PDF hypothesized to be true under H_1 . Denote this PDF estimate based on the copula

density $k_1(\cdot)$ by $f_{k_1}(\cdot)$. Ideally, we wish to select the model that is *closest* to the true PDF. Measures belonging to the ϕ – divergence family [5] have been widely used to quantify the distance between two distributions (see [75] and references therein). Of all ϕ – divergence measures, the most popular is the KL divergence which, for two PDFs $p_X(x)$ and $q_X(x)$, is given as,¹

$$D(p, q) = \int p_X(x) \log \frac{p_X(x)}{q_X(x)} dx. \quad (3.9)$$

$D(p, q)$ measures how different $p_X(x)$ is relative to $q_X(x)$. It is important to note that $D(p, q)$ is not a true measure of distance since it is not symmetric ($D(p, q) \neq D(q, p)$), and, does not satisfy the triangular inequality. Nevertheless, it has useful properties: a) $D(p, q) \geq 0$, and b) $D(p, q) = 0$ iff $p = q$, and therefore, can be an effective quantification of model mismatch error.

The KL divergence between the *true* PDF, $f(\cdot)$, and its copula based estimate, $f_{k_1}(\cdot)$ is

$$\begin{aligned} D(f, f_{k_1}) &= \int_{\mathbf{z}} f(\mathbf{z}) \log \frac{f(\mathbf{z})}{f_{k_1}(\mathbf{z})} d\mathbf{z} & (3.10) \\ &= \int_{\mathbf{z}} f(\mathbf{z}) \log \frac{f_p(\mathbf{z})c_1(\cdot)}{f_p(\mathbf{z})k_1(\cdot)} d\mathbf{z} \\ &= \underbrace{\int_{\mathbf{z}} f(\mathbf{z}) \log c_1(\cdot) d\mathbf{z}}_{=\mathcal{I}_1(Z_1; \dots; Z_N)} - \underbrace{\int_{\mathbf{z}} f(\mathbf{z}) \log k_1(\cdot) d\mathbf{z}}_{=\mathbb{E}_f \log k_1(\cdot) = \mathcal{I}_{k_1}(Z_1; \dots; Z_N)}. & (3.11) \end{aligned}$$

The term $\mathcal{I}_{k_1}(\cdot)$ in Eq. (3.11) can be interpreted as a copula-based estimate of multi-information [94], $\mathcal{I}_1(\cdot)$. Note that $\mathcal{I}_1(\cdot)$ (which reduces to the well-known mutual information when $N = 2$) describes the complete nature of dependence between the variables and bounds the copula expectation term from above, i.e., $\mathcal{I}_{k_1}(\cdot) \leq \mathcal{I}_1(\cdot)$ since $D(f, f_{k_1}) \leq 0$. It is also clear that $\mathcal{I}_{k_1}(\cdot)$ should be strictly greater than zero for

¹The base of the logarithm in Eq. (3.9) is arbitrary. We use natural logarithm in our definition.

the copula-based model to be ‘closer’ to the true PDF than the product model, i.e.,

$$D(f, f_{k_1}) < D(f, f_p) \text{ iff } 0 < \mathcal{I}_{k_1} \quad (3.12)$$

Thus, of all the copula densities in set \mathcal{A}_k that satisfy Eq. (3.12), the copula density $k_1(\cdot)$, where,

$$k_1(\cdot) = \operatorname{argmax}_{k(\cdot) \in \mathcal{A}_k} \mathbb{E}_f \log k(\cdot) \quad (3.13)$$

$$\approx \operatorname{argmax}_{k(\cdot) \in \mathcal{A}_k} \frac{1}{L} \sum_{l=0}^{L-1} \log k(F_1(z_{1l}), \dots, F_N(z_{Nl})), \quad (3.14)$$

results in minimal mismatch (in terms of the KL divergence), and is thus the best choice to model the inter-modal dependence structure under H_1 . Note that it is not possible to evaluate Eq. (3.13) since the expectation is with respect to the true joint PDF $f(\mathbf{z})$ which is unknown. We, therefore, approximate it by taking the sample expectation as shown in Eq. (3.14).

An interesting link between the mismatch error $D(f, f_{k_1})$, and detection performance exists. In problems where the distribution under one of the hypotheses, (say) H_0 is known, it has been shown in [35] that the average loss in detection power, $\Delta_{loss} = \int \left(P_D(\eta) - \hat{P}_D(\eta) \right) d\eta$, when $f(\mathbf{z})$ is misspecified as $f_{k_1}(\mathbf{z})$ in Eq. (3.4) is equal to $D(f, f_{k_1})$. Note that Δ_{loss} is, in fact, a *global* measure obtained by integrating over all possible detector thresholds (η). In other words, Δ_{loss} is the amount by which the area under P_D , denoted as A_{P_D} , decreases due to mismatch in $f(\mathbf{z})$. The selection rule proposed in Eq. (3.13) minimizes Δ_{loss} . Also, it is easy to see that the performance gain (increase in area under P_D) achieved due to the use of the copula

density $k_1(\cdot)$ over the product model is,

$$\text{Gain} = \Delta_{loss}(f_p) - \Delta_{loss}(f_{k_1}) \quad (3.15)$$

$$= D(f, f_p) - D(f, f_{k_1})$$

$$= \mathcal{I}_1(\cdot) - \{\mathcal{I}_1(\cdot) - \mathcal{I}_{k_1}(\cdot)\}$$

$$= \mathcal{I}_{k_1}(\cdot). \quad (3.16)$$

For cases when the distribution of the test static under H_0 does not vary too much across different models in contention, Δ_{loss} measures the “separation” between $p_{\mathcal{T}_k}(t_k; H_0)$ and $p_{\mathcal{T}_k}(t_k; H_1)$, the PDFs of the test statistic under H_0 and H_1 respectively.

Next, we present an illustrative example.

Example 3.1

Consider a network of three ($N = 3$) heterogeneous sensors monitoring a region of interest (ROI) for the presence of a target. We assume that the heterogeneous observations, z_1, z_2 and z_3 , received at the fusion center under hypothesis H_i follow the Gaussian, exponential and beta distribution functions respectively,

$$Z_1 \sim \mathcal{N}(0, \sigma_i^2), \quad Z_2 \sim \text{exponential}(\lambda_i), \quad Z_3 \sim \text{beta}(a_i, b_i = 1), \quad (3.17)$$

where $\sigma_1^2 > \sigma_0^2$, $\lambda_1 > \lambda_0$ and $a_1 > a_0$. Further, without loss of generality, we assume $c_0(\cdot) = 1$ (the product copula under H_0), i.e., $g(\mathbf{z}) = \prod_{n=1}^3 g_n(z_n)$. A common random phenomenon causes a change in the statistics of the three sensors’ data and the sensor measurements are thus statistically dependent under hypothesis H_1 . In order to generate dependent data with the marginals given in Eq. (3.17), we define two auxiliary random variables $W_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $W_2 \sim \text{gamma}(\alpha_1, \beta_1 = \lambda_1)$. It follows

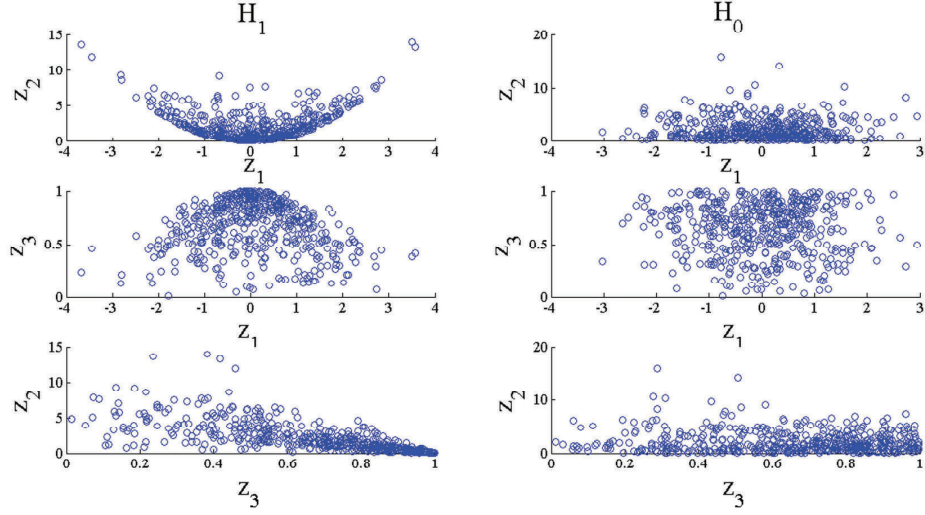


Figure 3.2: Scatter plot of 500 realizations. Non-zero dependence between observations under H_1 (Column 1) is evident from the figure.

then that

$$Z_2 = Z_1^2 + W_1^2 \sim \text{exponential}(\lambda_1 = 2\sigma_1^2) \quad (3.18)$$

$$Z_3 = \frac{W_2}{W_2 + Z_2} \sim \text{beta}(a_1 = \alpha_1, b_1 = 1). \quad (3.19)$$

Statistical dependence between sensor observations is evident from the scatter plot shown in Fig. 3.2. Their joint PDF can be derived as (see Appendix A),

$$f(z_1, z_2, z_3) = \frac{\mathbb{I}(z_2 > z_1^2)}{2\pi\Gamma(\alpha_1)\lambda_1^{1+\alpha_1}} \frac{z_2^{\alpha_1} z_3^{\alpha_1-1}}{\sqrt{z_2 - z_1^2} (1 - z_3)^{1+\alpha_1}} \exp\left(-\frac{z_2}{\lambda_1(1 - z_3)}\right),$$

where, $z_1 \in (-\infty, \infty)$, $z_2 \in (0, \infty)$ and $z_3 \in (0, 1)$, $\mathbb{I}(\mathcal{E})$ is an indicator of event \mathcal{E} and $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ is the gamma function. The optimal LR test at the fusion

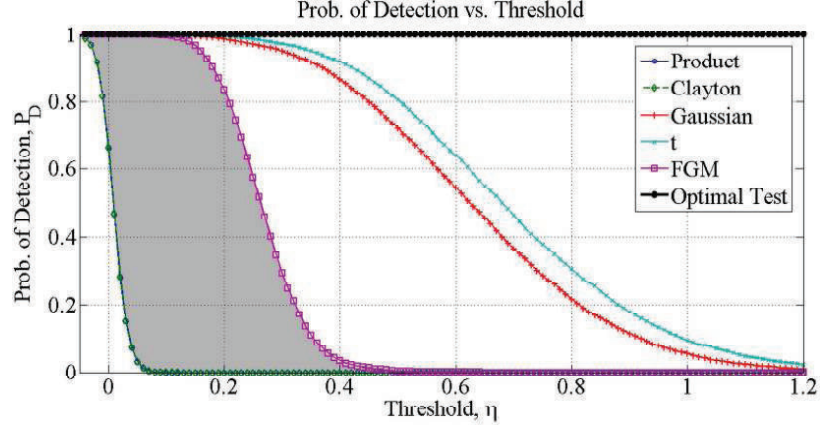


Figure 3.3: P_D vs. η obtained for different copula functions: Parameters for the marginal PDFs are $(\sigma_0^2 = 1, \sigma_1^2 = 1.1)$, $(\lambda_0 = 2.1, \lambda_1 = 2.2)$, $(b_0 = b_1 = 1, a_0 = 1.8, a_1 = 2)$. The shaded region represents the increase in the area (over the product model) when FGM copula is used to model dependence under H_1 .

center can now be obtained as (see Appendix A),

$$\mathcal{T}_{LR} = \frac{\prod_{l=1}^L \frac{z_{2l}^{a_1} z_{3l}^{a_1-1} (1-z_{3l})^{-(1+a_1)}}{2\pi\Gamma(a_1)\lambda_1^{1+a_1}\sqrt{z_{2l}-z_{1l}^2}} \exp\left(-\frac{z_{2l}}{\lambda_1(1-z_{3l})}\right)}{\prod_{l=1}^L \frac{1}{\sqrt{2\pi\sigma_0^2}} \frac{a_0}{\lambda_0} z_{3l}^{a_0-1} \exp\left(-\left(\frac{z_{1l}^2}{2\sigma_0^2} + \frac{z_{2l}}{\lambda_0}\right)\right)} \mathbb{I}(z_2 > z_1^2) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \quad (3.20)$$

In cases where PDF under H_1 is unknown, we use the copula based test given in (3.8) in place of Eq. (3.20) at the fusion center. We set the decision window, L , to 20 samples and plot P_D (averaged over 5000 Monte Carlo trials) as a function of the decision threshold η for several copula functions in Fig. 3.3. The plot for the optimal test is also included. The shaded region in the figure represents the increase in A_{P_D} (over the product model) when fusion rule based on the FGM copula is employed. In this problem, the t-copula yields the best performance as its corresponding P_D curve is the right most curve in Fig. 3.3. Further, as derived in Eq. (3.16), the increase in the area under the P_D vs. threshold curve, A_{P_D} , achieved by using the copula density $k_1(\cdot)$, is exactly equal to the multi-information, \mathcal{I}_{k_1} . This is confirmed in Fig. 3.4

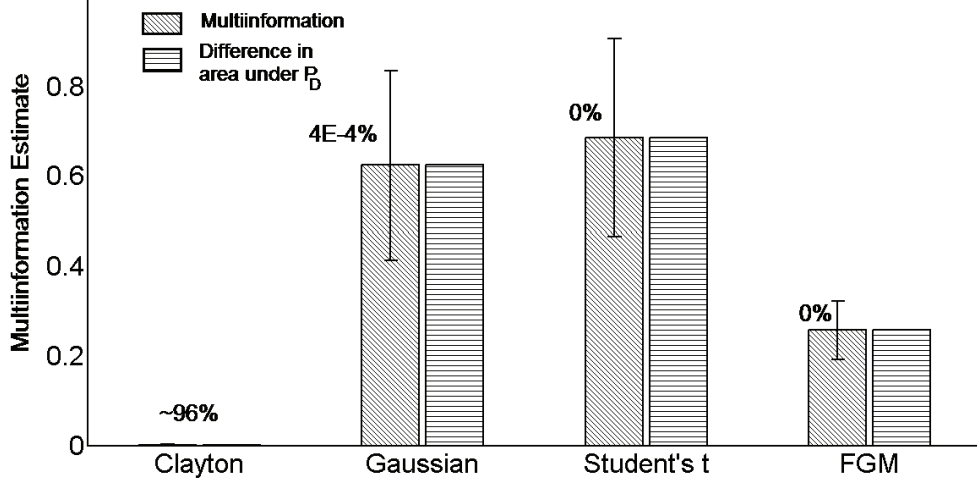


Figure 3.4: Multi-information estimates (averaged over 5000 Monte Carlo trials) with symmetric one standard deviation error bars, and Gain (Eq. (3.15)) computed using the trapezoidal rule, for several copula functions. The number of trials (in %) for each copula for which the improvability condition (3.12) was **NOT** satisfied is also indicated.

which shows that the sample based estimates of \mathcal{I}_{k_1} and ‘Gain’ (Eq. (3.15)) computed using the trapezoidal rule are more or less equal for all copula functions. The figure also shows that value of the multi-information estimate is the largest in the case of the Student’s t-copula, and thus the selection rule given by Eq. (3.14) would choose the Student’s t-copula as the model to characterize dependence between observations under H_1 . Receiver operating characteristic (ROC) curves obtained for the optimal test and tests using different copula densities are shown in Fig. 3.5. Note that the optimal test achieves zero false alarm rate for all detector thresholds. This is because, it is highly likely that $\mathbb{I}(z_{2l} > z_{1l}^2) = 0$ for at least one of the L observations belonging to H_0 . Thus, $P_F = P(\mathcal{T}_{LR}(\mathbf{z}_L; H_0) > \eta) \rightarrow 0$ as L increases.

The number of trials for which the improvability condition ($\mathcal{I}_{k_1} > 0$) was not satisfied is indicated in Fig. 3.4. For the Clayton copula model, \mathcal{I}_{k_1} was less than zero for 4833 of the 5000 trials ($\approx 96\%$ of the trials). For the other three copulas, \mathcal{I}_{k_1} was almost never less than zero during the 5000 trials. This is the cause for the Clayton copula based ROC to be lower than that obtained using the product

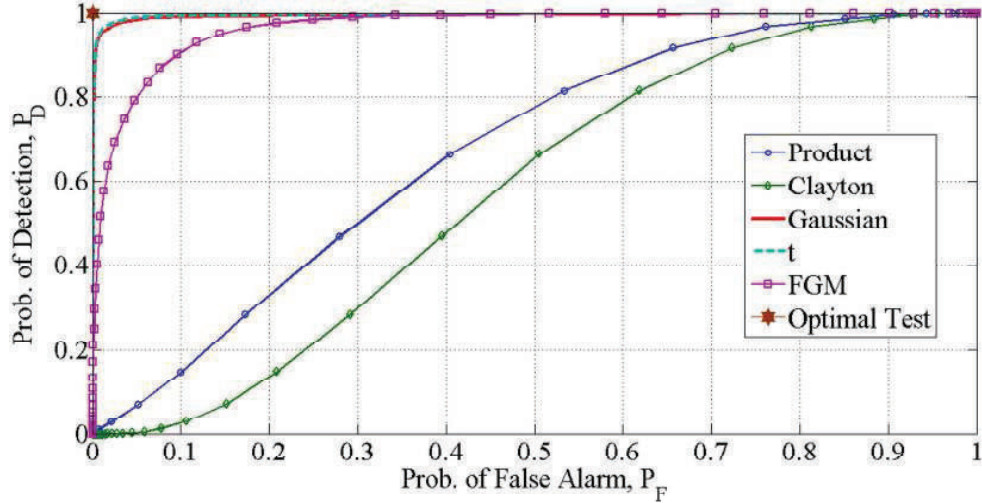


Figure 3.5: Monte Carlo based Receiver Operating Characteristics: Parameters for the marginal PDFs are $(\sigma_0^2 = 1, \sigma_1^2 = 1.1)$, $(\lambda_0 = 2.1, \lambda_1 = 2.2)$, $(b_0 = b_1 = 1, a_0 = 1.8, a_1 = 2)$.

model (Fig. 3.5). The mismatch between the actual dependence structure and the Clayton copula density is quite severe so that it results in a nonconcave ROC. It is thus important to develop good copula selection rules. As evident from Fig. 3.4, the selection rule (3.14), derived in this chapter correctly discards the Clayton copula and identifies the Students' t-copula as the best choice among the four copulas tested in this example.

3.2.2 The Area Under the Receiver Operating Characteristic Curve Criterion

Area under the ROC curve (AUC) is a *classification-specific* criterion, and has been used as a performance measure in many fields such as mathematical psychology [8], psychophysics [46], medical diagnostics, signal processing [47] and machine learning [13]. By definition,

$$\text{AUC} = \int_0^1 P_D(P_F) dP_F, \quad (3.21)$$

where, $0 \leq \text{AUC} \leq 1$ since it is a portion of a unit square. The measure is completely independent of the threshold η and summarizes the entire ROC in a single number. Further, AUC can also be related to the Bayesian probability of error, P_E . When the two hypotheses have equal probabilities of occurrence ($\pi_0 = \pi_1 = 0.5$), it has been shown in [87] that,

$$1 - \text{AUC} \leq P_E \leq \sqrt{\frac{1}{2} - \frac{\text{AUC}}{2}}. \quad (3.22)$$

Thus, AUC-based copula selection is applicable for both the Neyman-Pearson and Bayesian detection frameworks.

When the test statistic follows a Gaussian distribution under both hypotheses, that is,

$$p(t_k; H_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(t_k - \mu_i)^2}{2\sigma_i^2}\right), \quad i = 0, 1, \quad (3.23)$$

it was shown in [90] that

$$\text{AUC} = \Phi^{-1}\left(\frac{d_a}{\sqrt{2}}\right) \quad (3.24)$$

where Φ denotes the CDF of a standard normal distribution, and

$$d_a = \frac{\mu_{k_1} - \mu_{k_0}}{\sqrt{\frac{\sigma_{k_1}^2 + \sigma_{k_0}^2}{2}}}. \quad (3.25)$$

From Eq. (3.25), it is clear that d_a is a monotonic function of AUC. We, thus, have a selection rule where we choose copula densities $k_1(\cdot)$ and $k_0(\cdot)$ such that d_a is maximized. Note that the d_a based selection rule is more accurate in the case of large L when the PDFs of the test statistic under both hypotheses approach normality due to the well-known central limit theorem (CLT).

For the case when the test statistic under the two hypotheses do not follow Gaussian distributions, one can use the Wilcoxon-Mann-Whitney (WMW) statistic, a nonparametric approach to estimate AUC [8, 23]. Given $\{l_{1a}\}_{a=1}^{L_a}$ and $\{l_{0b}\}_{b=1}^{L_b}$, the log-likelihood ratios (LLRs) under H_1 and H_0 respectively, the WMW statistic (also known as the U-statistic) is given as

$$U = \frac{\sum_{a=1}^{L_a} \sum_{b=1}^{L_b} \mathbb{I}(l_{1a} > l_{0b})}{L_a L_b}. \quad (3.26)$$

Note that (3.26) is the estimate of the probability $Pr(\mathcal{T}_k(\mathbf{z}; H_0) < \mathcal{T}_k(\mathbf{z}; H_1)) = \text{AUC}$ [8].

Next, we present an illustrative example.

Example 3.2

We consider a two sensor example where the observations Z_1 and Z_2 follow Gaussian and Exponential distributions respectively under the two competing hypotheses. In order to generate dependent data Z_1 and Z_2 with prescribed marginals, we use the auxiliary random variable method described above (see Eq. (3.18)). Thus,

$$Z_1 \sim \mathcal{N}(0, \sigma_i^2) \quad Z_2 \sim \text{exponential}(\lambda_i = 2\sigma_i^2) \quad (3.27)$$

However, unlike Example 3.1, we consider dependence under both hypotheses and consider AUC based copula selection.

AUC estimate, $U_{(k_0, k_1)}$, for copula pairing $k_0(\cdot)$ and $k_1(\cdot)$, is evaluated using Eq. (3.26). We also calculate the difference $U_\delta = U_{(k_0, k_1)} - U_p$, where U_p is the AUC estimated using product copula under both H_0 and H_1 , i.e., $k_0(\cdot) = k_1(\cdot) = 1$. Table 3.1 contains $U_{(k_0, k_1)}$, and corresponding U_δ in parentheses, for several copula pairings. For example, when Clayton and Gumbel copula densities are used to model dependence under H_1 and H_0 respectively, AUC is approximately 64%. Further, a negative U_δ

Table 3.1: AUC estimated using the Wilcoxon-Mann-Whitney statistic

| Copula H_1 \ Copula H_0 | Clayton | Frank | Gaussian | Gumbel |
|-----------------------------|---------------------------------------|-------------------------|------------------------|--------------------------------------|
| Clayton | 0.7026 (0.0010743) | 0.69182 (-0.0097022) | 0.6782 (-0.023321) | 0.6392 (- 0.06232) |
| Frank | 0.70898 (0.0074573) | 0.7 (-0.0015188) | 0.6877 (-0.013825) | 0.6474 (-0.05412) |
| Gaussian | 0.72191 (0.020384) | 0.71288 (0.011353) | 0.70296 (0.0014352) | 0.66424 (-0.037283) |
| Gumbel | 0.70981 (0.0082926) | 0.69974 (-0.0017851) | 0.68781 (-0.013709) | 0.65761 (-0.043912) |

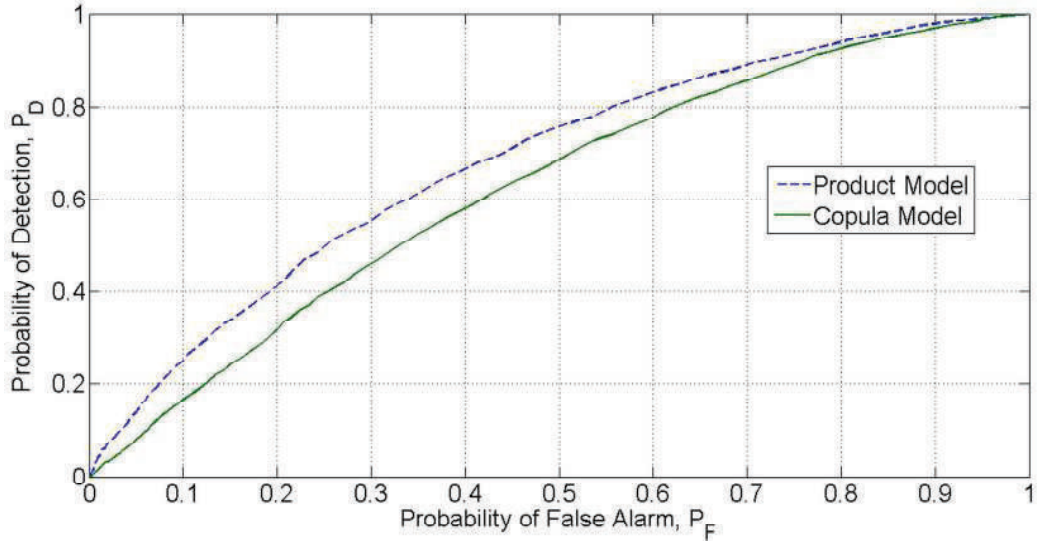


Figure 3.6: ROC curves obtained when Clayton and Gumbel copula densities model H_1 and H_0 respectively.

(≈ -0.062) suggests that the product model outperforms the copula approach. This is confirmed in Fig. 3.6 which shows the ROCs obtained using both the copula and product models. On the contrary, it can be seen from Table 3.1, that the copula approach, when Gaussian and Clayton copula densities model H_1 and H_0 respectively, performs better than the product model (in terms of AUC). The corresponding ROC is shown in Fig. 3.7.

Remark. *It is important to note that, while the KL divergence and AUC are attractive because of their simplicity, they are still heuristic measures and their optimization may not always result in the best P_D vs. P_F tradeoff. As discussed in Section 3.2.1, the KL*

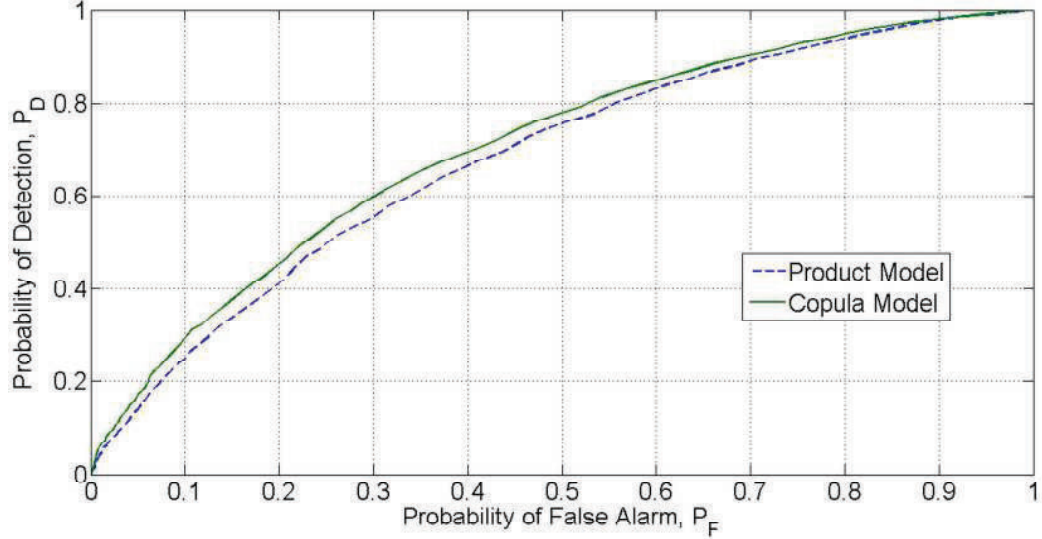


Figure 3.7: ROC curves obtained when Gaussian and Clayton copula densities model H_1 and H_0 respectively.

divergence is not a true distance metric. It is asymmetrical ($D(f, f_{k_1}) \neq D(f_{k_1}, f)$), and also does not satisfy the triangular inequality. This limits the utility of A_{P_D} , the area under the P_D curve, when used to compare different detectors. For example, the perceptible gap between the ROCs obtained using the product and Clayton copula models (as well as its non concavity) (Fig. 3.5) is not evident in Fig. 3.3. Nevertheless, the A_{P_D} based comparison is a useful approach due to its reduced complexity while still providing a systematic KL divergence based approach for copula selection.

Similarly, it may be useful in some applications to focus on the area computed for a portion of the ROC (as opposed to the entire ROC):

$$AUC_\alpha = \int_0^\alpha P_D(P_F) dP_F, \quad \alpha \ll 1$$

It is likely that in many cases a closed form for AUC_α is not available. One may need numerical integration in such cases.

In the next section, we apply the copula based selection framework to a real-world application.

3.3 Biometric Authentication: An Application

The process of authentication involves verifying the identity of a person claiming access to one or more resources of a system. Authentication systems can be based on passwords, security tokens or biometrics or combinations of them [72]. Passwords are words, phrases or alphanumeric personal identification numbers (PINs) that serve as short form indicators of a person's identity. They are usually created by the authorized users during the enrollment or registration phase (e.g., creating computer user accounts) and are kept secret from others. Security tokens, on the other hand, are physical devices that the users are required to carry in order to be allowed access to the system. More recent designs of automatic human recognition systems involve the use of features such as face, fingerprints, iris or behavioral traits such as gait or rate of keystrokes, etc. For example, in building access control applications, a person's face may be matched to templates stored in a database consisting of all enrolled users. Decision to allow or deny entry is then taken based on the similarity score generated by the face matching algorithm. Such security systems that rely on biometrics have several advantages over the more conventional ones (passwords or security tokens). For example, a PIN, if leaked, may be used by an unauthorized person causing serious security concerns. However, a person's physical signature belongs only to that individual and is extremely difficult if not impossible to emulate it. Further, biometric systems may be more convenient and user-friendly as there is no password to remember or any token to carry. See [72] for a more detailed comparison of the three approaches especially in terms of achievable security, convenience of use and the overall cost of the system.

3.3.1 Fusion of Multiple Biometric Modalities

A biometric authentication task is essentially a binary hypotheses testing problem where,

$$\begin{cases} H_0 : \text{claimant an impostor} \\ H_1 : \text{claimant a genuine user} \end{cases}$$

are the two competing hypotheses. While biometric authentication systems have several advantages over the more conventional systems, they are still not devoid of limitations. Biometric traits such as face and voice change with age. One may be required to update the systems' database to counter this time variability. Environmental noise and noise in the acquisition system further affect the accuracy and reliability of the system. Overlap between physical features or inter-class similarity (e.g., twins with identical facial features) limits the system's ability to distinguish between the classes. There also exist intra-class variations due to differences between the acquired biometric signature of an individual requesting the access and his/her template registered in the database. Apart from noise sources stated above, these differences may also stem from the psychological and behavioral variations of an individual at different instances of time. One method to overcome these limitations is to consider combining multiple sources of information. It may include fusing observations of disparate modalities (e.g., voice and face) or multiple features (extracted from the same biometric trait), multiple classifiers or multiple samples of the same source. This method of fusing several biometric sources is called multi-biometrics. Figure 3.8 shows a multimodal biometric system which considers fusion of disparate biometric signatures such as face, iris and fingerprints. We also note that, although the figure shows fusion of multimodal observations, the copula framework is general enough to handle other problems where heterogeneity in the acquired data is due to

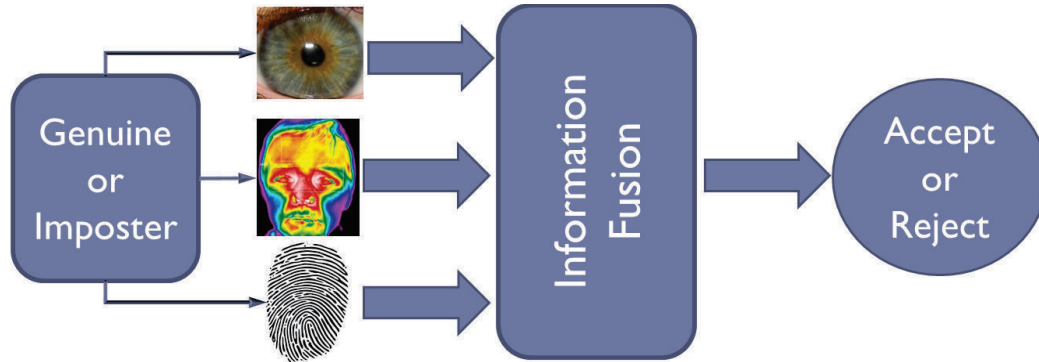


Figure 3.8: A multibiometric authentication system. Biometric signatures of disparate modalities such as face, iris and fingerprint are fused.

multiple samples, algorithms or multiple classifiers, i.e., the variables $\{Z_n\}_{n=1}^N$ may also denote multiple samples or multiple features (extracted from the same modality) or output of multiple algorithms which are combined (jointly processed) at a fusion center.

3.3.2 Experimental Results

The performance of the copula based test is evaluated on NIST-BSSR 1 [2], a publicly available database that includes similarity scores from two commercial face recognizers and one fingerprint system, and is partitioned into three sets. We use the NIST-face dataset which consists of match scores from three thousand subjects in our experiment. Two samples (match scores) are available per subject, thus resulting in a total of 2×3000 genuine and $2 \times 3000 \times 2999$ impostor scores. The scores are heterogeneous as the two face recognizers use different algorithms to match the frontal face images of the subjects under test. Face matcher 1 quantifies the similarity between two images using a scale from zero to one while the second matcher outputs score values ranging from zero to one hundred with higher values indicating a better match between the face image of the subject under test and the template recorded previously and stored in the database. A scatter plot of both the genuine and impostor scores is shown in Fig. 3.9.

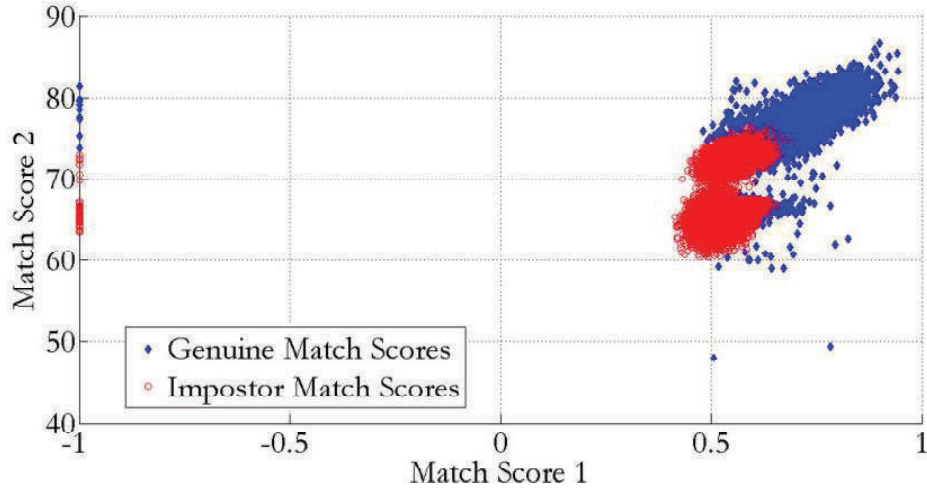


Figure 3.9: Scatter plot of the genuine and impostor scores from the two face matchers

The scatter plot shows that sometimes a score of -1 is reported by face matcher 1. Surprisingly, this is true even with some genuine match scores, i.e., when the acquired face image of a subject was matched to his/her own template in the database. This may have been due to errors during data acquisition, image registration or feature extraction due to the poor quality of the image. Negative one thus serves as an indicator to flag the incorrect working of the matcher. The fusion performance will thus depend on how this anomaly is handled in the decision making. For example, the fusion center can be designed to respond in one of the following ways upon the reception of the error flag,

1. *Request for a retrial*: The fusion center does not make a global decision upon receiving the error flag. Instead the subject claiming an identity is requested to provide his/her biometric measurement again. The design ensures that scores are always in the valid range ($z_1 \in [0, 1]$). We emulate this by deleting all the users whose match scores were reported as -1 and present results using (2×2992) genuine and $(2 \times 2992 \times 2991)$ impostor scores. However, there may be applications where the system does not have the liberty to request for a retrial and the fusion center has to make a decision after each match.

2. *Censoring* (face matchers that generate the error flag): In our example, face matcher 1 generates the error flag. Upon the reception of the error flag ($z_1 = -1$), the terms of the log likelihood ratio test that depend on z_1 are discarded. Thus, the first and the third terms in

$$\mathcal{T}_k(z_1, z_2) = \underbrace{\log \frac{f(z_1; \psi_1)}{g(z_1; \lambda_1)}}_{=0} + \log \frac{f(z_2; \psi_2)}{g(z_2; \lambda_2)} + \underbrace{\log \frac{k_1(F_1(z_1; \psi_1), F_2(z_2; \psi_2); \psi_d)}{k_0(G_1(z_1; \lambda_1), G_2(z_2; \lambda_2); \lambda_d)}}_{=0} \quad (3.28)$$

are set to zero.

3. *Accept H_0* : The system decides in favor of H_0 when one or more of the face matchers generate an error flag. This design is conservative and is thus suitable for applications where one desires minimal false alarm rates.
4. *Accept H_1* : The system decides in favor of H_1 when one or more error flags are generated.
5. *Random decision*: Toss a fair coin to decide between H_0 and H_1 .

Data (under each hypothesis) is partitioned into two subsets of equal size where the first subset is the training set used for model fitting and the second is the testing set used for system's performance evaluation. Thirty different training-testing sets (resamples or trials) are obtained from the same data by randomizing the partitioning process described above. We denote each resample/trial by $r \in 1, 2, \dots, 30$. System performance averaged across all r trials is then obtained. We now describe the training and the performance evaluation steps in detail.

Given the training set, the joint PDF (of scores from the two face matchers) is estimated by first modeling the marginal PDFs and then estimating the parameters of the selected copula densities $k_1(\cdot)$ and $k_0(\cdot)$. A Gaussian mixture model (GMM)

is used to fit the scores generated by both face matchers [38]. Fig. 3.10 shows the estimated marginal PDFs for both impostor and genuine scores. This is followed by the copula modeling step. We consider five families of copula functions, namely, the Gaussian copula, the Student’s t-copula, Clayton copula, Frank copula and the Gumbel copula. ML estimates of parameters of each copula function are obtained using the IFM approach discussed in Section 2.2.2, and copula functions are compared on the basis of the resultant detection performance. Specifically, the AUC criterion is used as the performance metric and we choose $k_1(\cdot)$ and $k_0(\cdot)$ so that AUC (Eq. (3.26)) is maximized. This results in the choice of the Gumbel and the Frank copula densities for modeling the dependence structure between the two matching scores under H_0 and H_1 respectively.

Model fitting is then followed by performance evaluation using the testing data. False alarm and detection rates are computed for each r ($P_F^{(r)}, P_D^{(r)}$). A mean ROC curve is obtained by averaging $P_F^{(r)}$ and $P_D^{(r)}$ across the thirty training-testing sets, i.e., $P_F = \text{MEAN}_r(P_F^{(r)})$ and $P_D = \text{MEAN}_r(P_D^{(r)})$. We show performance results for all five designs in Fig. 3.11.

The product fusion rule performs better than each of the individual face matchers, as expected. The copula based processing, in addition to the differences in the marginal PDFs, also exploits the differences in the dependence structures across the two hypotheses. This is evident from Fig. 3.12 which shows the average (over all thirty resamples) mutual information between the genuine and impostor scores for the ‘request for retrial’ design. Further, of all the five designs, the ‘request for a retrial’ approach achieves the best performance. This is because the method ensures that input data are of good quality. The performance of ‘censoring’ and ‘accept H_0 ’ are similar. With a few terms set to zero in the ‘censoring’ design, the rule decides in favor of H_0 in most cases (when the error flag is received). The ‘accept H_1 ’ and ‘random decision’ methods are more liberal (or favor H_1) and thus show increased

Table 3.2: Peak increase in P_D due to copula processing ($P_D^k - P_D^p$). Corresponding P_F for ‘Request for retrial’, ‘Censoring’ and ‘Accept H_0 ’ designs is 0.0026%, and that for ‘Accept H_1 ’ and ‘Random decision’ designs is 9.9%.

| Fusion rule | Increase in Average P_D | 95% Confidence interval |
|---------------------|---------------------------|-------------------------|
| Request for retrial | 3.31% | (2.82%, 3.81%) |
| Censoring | 4.40% | (3.73%, 5.07%) |
| Accept H_0 | 4.38% | (3.71%, 5.05%) |
| Accept H_1 | 0.64% | (0.59%, 0.67%) |
| Random decision | 0.71% | (0.66%, 0.77%) |

false-alarm rates. ROCs in Fig. 3.11 show the superiority of each of the above copula based fusion designs over the product rule. Peak improvements in the average P_D computed as the difference between the copula based average P_D denoted as (P_D^k) and that achieved using the product rule, (P_D^p), are given in Table 3.2 for all the five designs. We note here that [25] addressed the biometrics scores fusion problem using copulas but observed no improvement over the product fusion rule. The authors modeled the marginal PDFs as a mixture of discrete and continuous components to account for the error flags (negative ones). However, copula methods require the marginal distributions to be strictly continuous. Further, their analysis was limited to the use of Gaussian copula densities which may be insufficient to model the inter-modal dependence. In this chapter, we have employed different approaches to handle error flags and have considered the use of a more general family of copula functions with the potential of improving system performance. These reasons could explain the differences between our results and those in [25].

We now consider the Bayesian framework [61] where one is able to assign or know a priori, the probabilities of occurrence, $Pr(H_0)$ and $Pr(H_1)$ ($= 1 - Pr(H_0)$), for the two hypotheses H_0 and H_1 respectively. The objective of a Bayesian classifier is to minimize the probability of error (more generally, the Bayes risk function), i.e.,

$$P_E = \min (Pr(H_0|z), Pr(H_1|z)), \quad (3.29)$$

where $Pr(H_0|z)$ and $Pr(H_1|z)$ are the posterior probabilities of the two hypotheses.

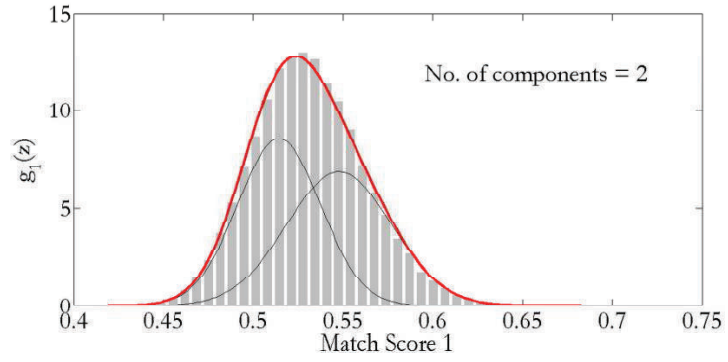
In Fig. 3.13, we plot P_E averaged over the thirty resamples as a function of the prior probability $Pr(H_1)$ for all five strategies. We see that the copula-based test achieves the best performance over the entire range of priors.

3.4 Summary

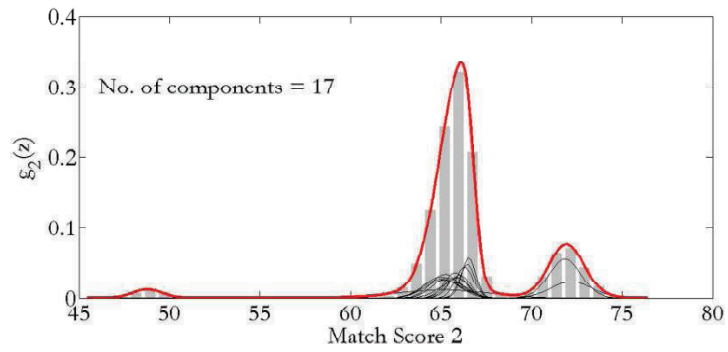
In this chapter, a binary hypothesis testing problem with heterogeneous observations was considered, and a test based on copula theory was obtained. The copula-based test is able to exploit both, the differences in the marginal PDFs and the differences in the dependence structures across the two hypotheses. The problem of identifying the best copula from a given library of valid copula densities was also considered. Two classification-specific methods, the KL divergence and the AUC-based methods were presented, that quantify classification performance enhancement due to copula processing. Under certain conditions, the KL divergence between the true PDF and the copula approximated PDF relates to the area under the P_D curve. The selection methods (particularly the AUC based approach) can be viewed as a discriminative training method [11,71,79,101] that aims to directly maximize classification accuracy rather than model the observations accurately. As shown in [11,79], discriminative training methods can partially compensate for the unavoidable mismatch, which is always present between the hypothesized model and the true distribution, and thus helps achieve better performance. Illustrative examples were provided for both copula selection methods.

The proposed approach was also successfully tested on real-data using the NIST database for a biometric authentication application. The dataset consists of similarity scores from two commercial face recognizers that employ different algorithms to match the frontal face images of the subjects under test. Thus, in this example, the

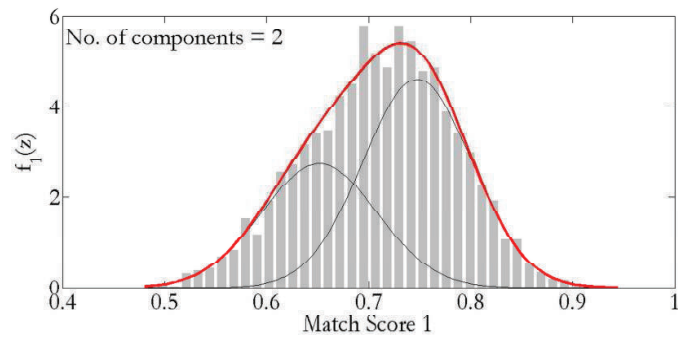
heterogeneity in the scores was due to the use of multiple algorithms. Results for both, the Neyman-Pearson and the Bayesian detection frameworks, were presented.



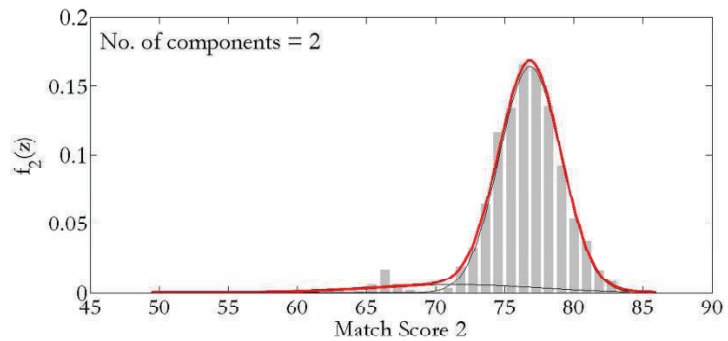
(a) FM C Impostor scores



(b) FM G Impostor scores

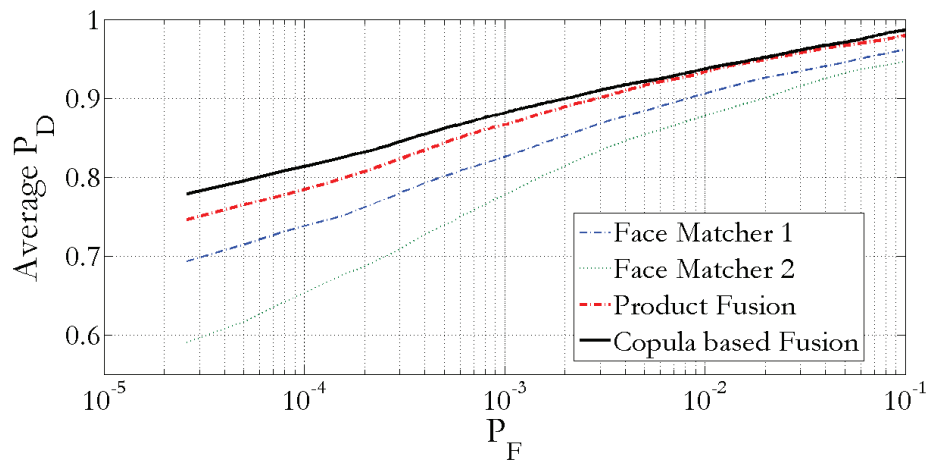


(c) FM C Genuine scores

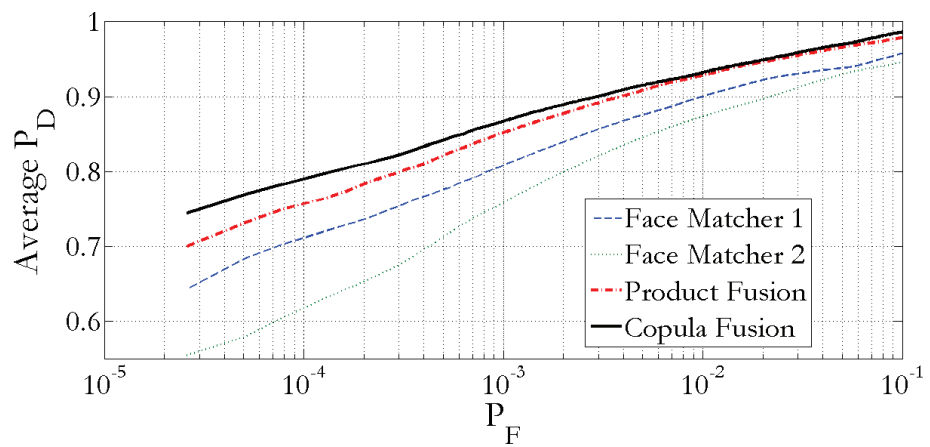


(d) FM G Genuine scores

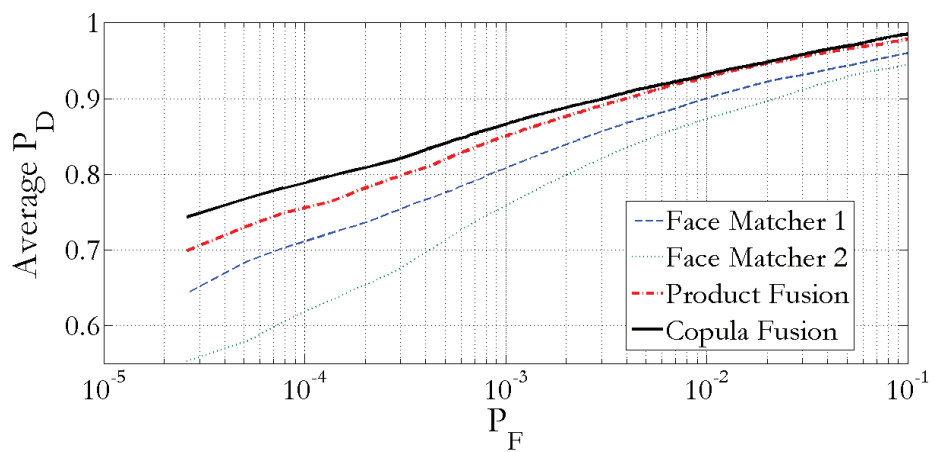
Figure 3.10: Gaussian mixture models for match scores from the two face matchers (FM)



(a) Request for retrieval

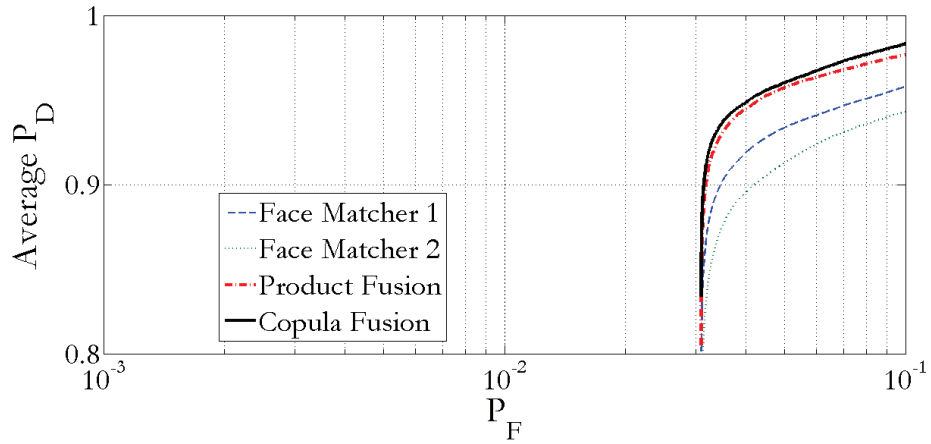


(b) Censoring

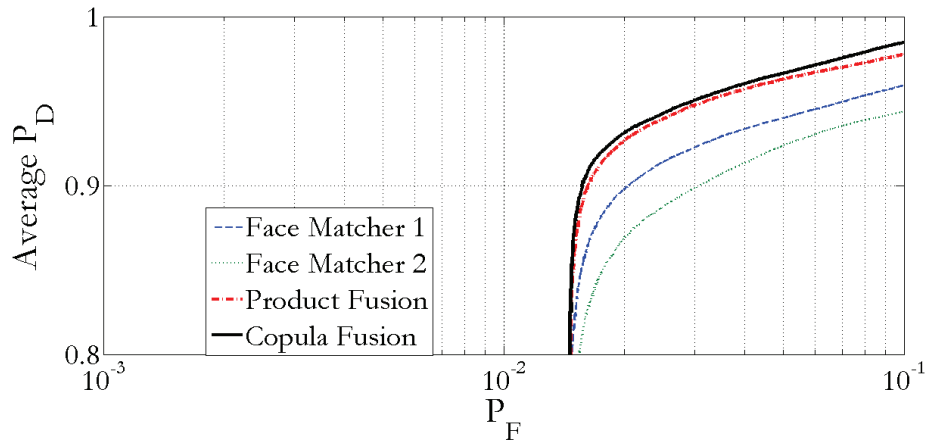


(c) Accept H_0

Figure 3.11: Receiver operating characteristic curves



(d) Accept H_1



(e) Random Decision

Figure 3.11: Receiver operating characteristic curves

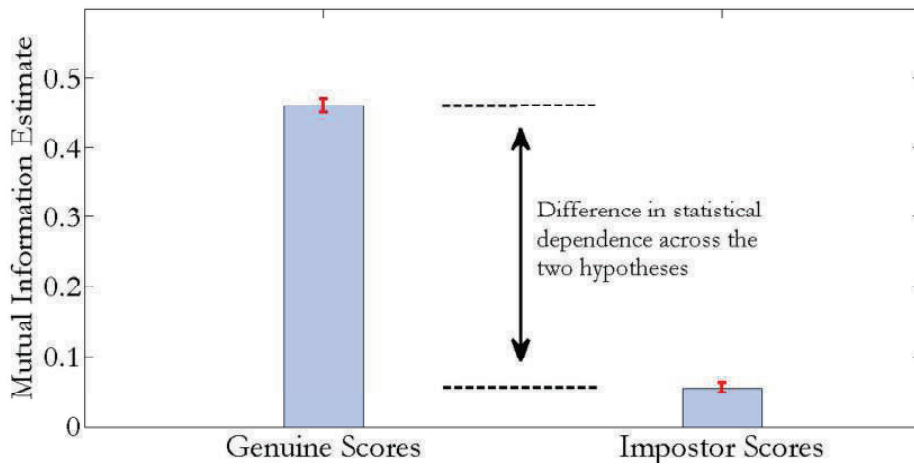
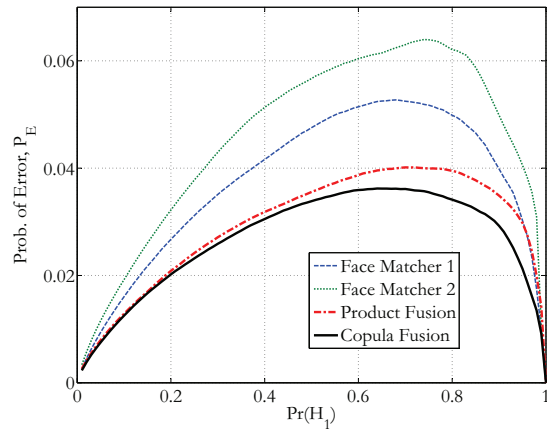
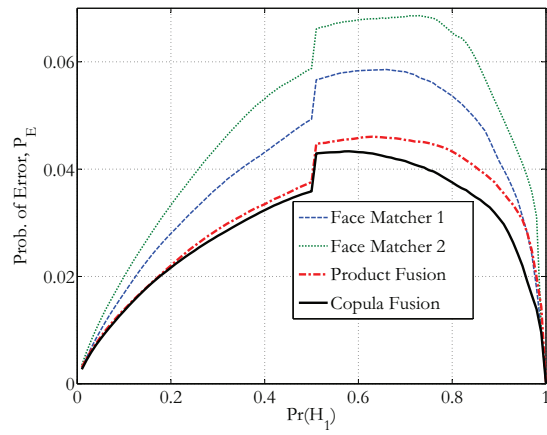


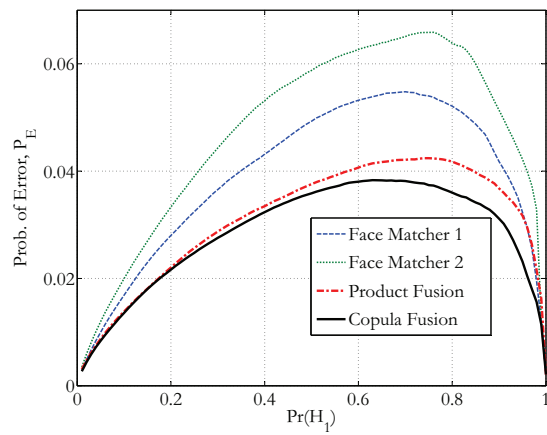
Figure 3.12: Mutual Information estimates averaged over thirty resamples for ‘request for retrieval’ design. The plot also shows symmetric one standard deviation error bars.



(a) Request for retrial

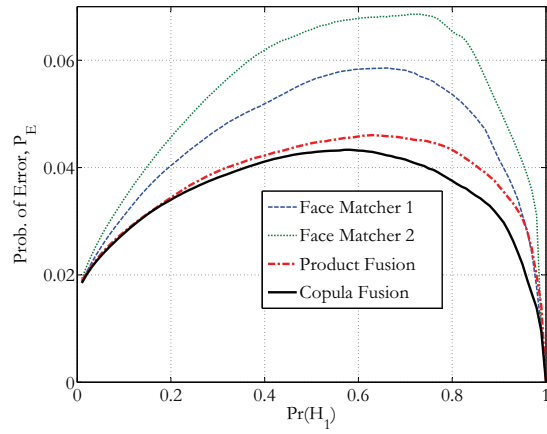


(b) Censoring

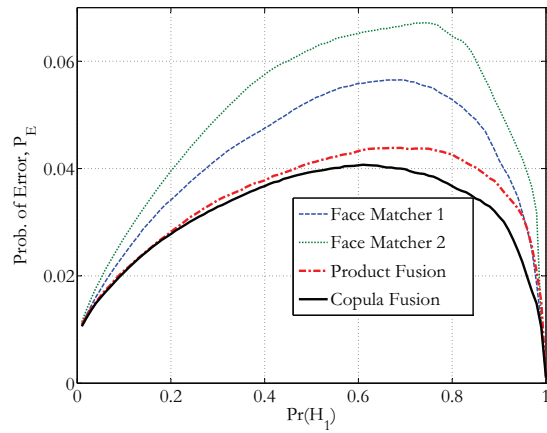


(c) Accept H_0

Figure 3.13: Probability of error vs. $Pr(H_1)$



(d) Accept H_1



(e) Random Decision

Figure 3.13: Probability of error vs. $Pr(H_1)$

Chapter 4

Hypothesis Testing With Heterogeneous Data: Uncertain Parameters

In Chapter 2, it was discussed how one could use copula theory to construct statistical models for heterogeneous random vectors. A binary hypothesis testing problem was then formulated and a copula based-test for the same was derived in Chapter 3. It was, however, assumed that data models under both the competing hypotheses could be constructed offline, i.e., prior to system deployment. Selection of the copula function and estimation of its parameters could be performed using a training dataset. Enough data samples were assumed to be available so that the copula parameters, as well as those of the marginal PDFs (if unknown) could be consistently estimated. However, in some applications, it may not be possible to learn the PDFs of sensor observations prior to system deployment. This could be due to several reasons. For example, acquisition of training samples is often very expensive. In some applications, especially biomedical, collection of data under *controlled* conditions may not even be possible.

In this chapter, we address the multisensor detection problem with a parallel network of heterogeneous sensors shown in Fig. 3.1, and derive a copula-based testing procedure to be employed at the fusion center. Unlike the methodology developed in Chapter 3, the method discussed in this chapter considers online (real-time) selection and estimation of copula models. In addition to the copula parameters, there could also be uncertainties associated with the marginal PDFs. We express these uncertainties by including parameters ψ_n and λ_n in the description of the marginal PDFs (e.g., $f_n(z; \psi_n)$ and $g_n(z; \lambda_n)$, $\forall n$). While the parameters $\psi_p = (\psi_1, \dots, \psi_N)^T$ and $\lambda_p = (\lambda_1, \dots, \lambda_N)^T$ are unknown and are to be estimated in real-time, it is assumed that the forms of the marginal PDFs are completely known.

Assumption 4.1. *The parametric models for marginal densities $\{f_n(z; \psi_n)\}$ and $\{g_n(z; \lambda_n)\}$ are well-specified (see Definition 2.3)*

However, no assumption such as above is made for copula densities. It is clear that the problem described above can be formulated as a composite hypothesis test [61] at the fusion center. However, unlike the classical formulation which assumes the knowledge of the parametric model up to certain parameters, the problem considered here requires online estimation of both, the PDF family and its parameters.

The two main approaches to composite hypothesis testing are the Bayesian methodology and the generalized likelihood ratio test (GLRT). Under the Bayesian approach, one has access to the prior PDFs, $p(\psi)$ and $p(\lambda)$, of the unknown parameters $\psi = (\psi_p, \psi_d)^T$ and $\lambda = (\lambda_p, \lambda_d)^T$ respectively. Using the priors, the unknown

parameters in the LLRT statistic can be integrated out as below,

$$\begin{aligned} \mathcal{T}_B(\mathbf{z}) &= \log \frac{\int_{\psi} f(\mathbf{z}; \psi) p(\psi) d\psi}{\int_{\lambda} g(\mathbf{z}; \lambda) p(\lambda) d\lambda} \quad (4.1) \\ &= \log \frac{\int_{\psi=(\psi_p, \psi_d)} f(\mathbf{z}; \psi_p) c_1 \{F_1(z_1; \psi_1), \dots, F_N(z_N; \psi_N); \psi_d\} p(\psi) d\psi}{\int_{\lambda=(\lambda_p, \lambda_d)} g(\mathbf{z}; \lambda_p) c_0 \{G_1(z_1; \lambda_1), \dots, G_N(z_N; \lambda_N); \lambda_d\} p(\lambda) d\lambda} \quad (4.2) \end{aligned}$$

Alternatively, the unknown parameters can be replaced by their ML estimates computed from the acquired data, to compute the likelihood ratio. The resultant test is known as the generalized likelihood ratio test (GLRT) [61, 66] given as,

$$\mathcal{T}_G(\mathbf{z}) = \log \frac{\max_{\psi \in \Psi} f(\mathbf{z}; \psi_p) c_1 \{F_1(z_1; \psi_1), \dots, F_N(z_N; \psi_N); \psi_d\}}{\max_{\lambda \in \Lambda} g(\mathbf{z}; \lambda_p) c_0 \{G_1(z_1; \lambda_1), \dots, G_N(z_N; \lambda_N); \lambda_d\}} \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (4.3)$$

We construct a similar test in the next section.

4.1 (Misspecified) Generalized Likelihood Ratio Test Based Fusion Rule

Although the GLRT does not always guarantee optimality (optimal for all null and alternative hypotheses), it is often employed in composite hypothesis testing problems due to its simplicity. Replacing the true but unknown copula densities $c_1(\cdot)$ and $c_0(\cdot)$ in Eq. (4.3) by copula densities $k_1(\cdot; \psi_d)$ and $k_0(\cdot; \lambda_d)$, we employ the following test

at the fusion center,

$$\begin{aligned}
\mathcal{T}_k(\mathbf{z}) &= \log \frac{\max_{\Psi_p, k_1(\cdot; \Psi_d) \in \mathcal{A}_k} f_p(\mathbf{z}; \psi_p) k_1(F_1(z_1; \psi_1), \dots, F_N(z_N; \psi_N); \psi_d)}{\max_{\Lambda_p, k_0(\cdot; \Lambda_d) \in \mathcal{A}_k} g_p(\mathbf{z}; \lambda_p) k_0(G_1(z_1; \lambda_1), \dots, G_N(z_N; \lambda_N); \lambda_d)} \\
&= \sum_{l=1}^L \sum_{n=1}^N \log \frac{f_n(z_{nl}; \hat{\psi}_n)}{g_n(z_{nl}; \hat{\lambda}_n)} + \\
&\quad \sum_{l=1}^L \log \frac{k_1^*(F_1(z_{1l}; \hat{\psi}_1), \dots, F_N(z_{Nl}; \hat{\psi}_N); \hat{\psi}_d)}{k_0^*(G_1(z_{1l}; \hat{\lambda}_1), \dots, G_N(z_{Nl}; \hat{\lambda}_N); \hat{\lambda}_d)} \underset{H_0}{\overset{H_1}{\geq}} \eta', \tag{4.4}
\end{aligned}$$

where L is the number of i.i.d N – variate observations received at the fusion center. Note that the maximization in Eq. (4.4) is over the unknown parameters as well as the copula densities belonging to the set \mathcal{A}_k . Unlike the classical formulation which would have required the knowledge of the true copula densities upto their parameters, we allow for the case when the set \mathcal{A}_k may not be inclusive of the true models $c_1(\cdot)$ and/or $c_0(\cdot)$. Thus, the copula densities $k_1^*(\cdot)$ and $k_0^*(\cdot)$ chosen after maximization may still be misspecified. We, therefore, call the test a *misspecified* GLRT (mGLRT). Thus, $(\hat{\psi}_n, \hat{\lambda}_n)$ and $(\hat{\psi}_d, \hat{\lambda}_d)$ in Eq. (4.4) denote the ML and QML estimates (see Section 2.2.2) of (ψ_n, λ_n) and (ψ_d, λ_d) respectively.

Following the NP formulation, the next step is to determine the threshold η' in Eq. (4.4) so that the false alarm rate is constrained ($P_F \leq \alpha$). This, however, requires the knowledge of $p(t_k|H_0)$, the PDF of the test statistic under the null hypothesis. Since the postulated statistical models $\{f(\mathbf{z}; \psi \in \Psi \subset \mathbb{R}^a)\}$ and $\{g(\mathbf{z}; \lambda \in \Lambda \subset \mathbb{R}^b)\}$ are only approximations of the true underlying distributions, it is not possible to derive the exact distribution of the test statistic under either hypothesis. However, some advancement is possible when L is large. The following theorem, due to Wilks [114], makes this feasible.

Theorem 4.1. *Suppose the following conditions hold:*

C 1. $\{f(\mathbf{z}; \Psi \subset \mathbb{R}^a)\}$ and $\{g(\mathbf{z}; \Lambda \subset \mathbb{R}^b)\}$ are well-specified under H_0 ,

C 2. $\{g(\mathbf{z}; \Lambda)\}$ is nested in $\{f(\mathbf{z}; \Psi)\}$, i.e., $g(\mathbf{z}; \lambda) \in \{f(\mathbf{z}; \Psi)\}$, $\forall \lambda \in \Lambda$.

Then the modified test statistic, $2\mathcal{T}_k(\cdot)$, converges to a chi-square distribution with $\nu (= a - b)$ degrees of freedom under the null hypothesis.

$$2\mathcal{T}_k(\mathbf{z}) \rightsquigarrow \chi_{a-b}^2, \quad \text{under } H_0 \quad (4.5)$$

From Eq. (4.5), we have

$$\begin{aligned} P_F &= Pr\left(2\mathcal{T}_k(\cdot) > 2\eta'; H_0\right) \\ &\stackrel{L \rightarrow \infty}{=} Q_{\chi_{a-b}^2}\left(2\eta'\right) \end{aligned} \quad (4.6)$$

where $Q_{\chi_\nu^2}(\cdot)$ denotes the right-tail probability of a chi-squared random variable with ν degrees of freedom. One can thus obtain the threshold η' so that P_F is constrained to a desired level $\alpha \in (0, 1)$ as below:

$$\eta' = \frac{Q_{\chi_{a-b}^2}^{-1}(\alpha)}{2} \quad (4.7)$$

The assumption of a well-specified $g(\cdot; \lambda)$ is reasonable for many applications. For example, it is always possible to collect enough training data under H_0 (when there is no signal present), so that $g(\cdot)$ can be consistently estimated. The formulation is related to the classical universal hypothesis testing problem, where one has to decide whether the samples are drawn from a fixed distribution [66]; we however limit the class of distributions under H_1 to the family $\{f(\mathbf{z}; \Psi)\}$, and assume that it converges to $g(\cdot)$ under H_0 . This problem is also known in the engineering literature as *anomaly detection* where the (nominal) null hypothesis is known while the (anomalous) alternative hypothesis is unknown.

4.1.1 Fixed vs. Variable Threshold Design

Note that the distribution of $2\mathcal{T}_k(\mathbf{z}; \hat{\lambda}, \hat{\psi})$ under the null hypothesis depends only on the model complexities of $\{f(\cdot; \Psi)\}$ and $\{g(\cdot; \Lambda)\}$, i.e., the number of uncertain parameters a and b . If the set \mathcal{A}_k consists of copula densities with parameters of different dimensions, the threshold η' must be adjusted accordingly (see Eq. (4.7)) to maintain a constant false alarm probability. Alternatively, one could restrict the set \mathcal{A}_k to include copula models with equal complexity to avoid the extra step of varying η' in real-time.

4.2 Performance Analysis

In the classical GLRT setting, it has been shown that under certain conditions, the modified test statistic, $2\mathcal{T}(\cdot)$ under H_1 , converges asymptotically to a noncentral chi-squared-distribution, where, the non-centrality parameter is a function of the true parameter ψ [61]. Thus, given ψ , one can exactly compute the detection probability, P_D , at a given false-alarm rate. The above convergence result does not extend to the misspecified GLRT formulation considered here. However, upper bounds on P_D at a given P_F can be derived as we show below. Computation of these bounds require the knowledge of only the first and second order moments of the mGLRT statistic.

4.2.1 Upper Bounds on Detection Power

We derive upper bounds on P_D , assuming the knowledge of only the first two moments, the mean μ_i and the variance σ_i^2 , of the mGLRT statistic under each hypothesis H_i . Approximate expressions for μ_i and σ_i^2 are also derived which allow computation of the moments given the marginal statistics and the copula densities $k_1(\cdot; \psi_d^*)$ and $k_0(\cdot; \lambda_d^*)$, where ψ_d^* and λ_d^* , are the pseudo-true values associated with the respective copula densities (see Eq. (2.17)). An important point to note here is that while the

bounds remain valid, the efficacy of these bounds to quantify detection performance depends on their tightness to the true P_D .

Proposition 4.1. *(Bound based on the Cantelli inequality)*

The detection power, P_D , of mGLRT is bounded from above as

$$P_D(\alpha) \leq \frac{1}{1 + \left(\frac{Q_{\chi_{a-b}^2}^{-1}(\alpha) - \mu_1}{\sigma_1} \right)^2}, \quad \mu_1 < Q_{\chi_{a-b}^2}^{-1}(\alpha), \quad (4.8)$$

where α is the probability of false alarm, and μ_1 and σ_1^2 are the mean and variance of the modified mGLRT statistic, $2\mathcal{T}_k(\cdot)$, under the hypothesis H_1 .

Proof. The upper bound on P_D is the result of the direct application of Cantelli's inequality (one-sided Tchebycheff inequality) [74].

Given a random variable X with mean μ_1 and variance σ_1^2 , we have the following bound due to Cantelli's inequality,

$$Pr(X - \mu_1 > a\sigma_1) \leq \frac{1}{1 + a^2}, \quad a > 0. \quad (4.9)$$

Setting $X = 2\mathcal{T}_k(\mathbf{z})$ and $a\sigma_1 + \mu_1 = 2\eta'$ in Eq. (4.9), we have

$$Pr\left(2\mathcal{T}_k(\mathbf{z}) > 2\eta'; H_1\right) \leq \frac{1}{1 + \left(\frac{2\eta' - \mu_1}{\sigma_1} \right)^2}, \quad \text{where } \mu_1 < 2\eta'. \quad (4.10)$$

Using Eq. (4.7) in the above equation gives the desired result. \square

The bound derived above can be improved further if the PDF, $p(t_k; H_1)$, is assumed to be unimodal.

Proposition 4.2. *(Bound based on the Vysochanskii-Petunin inequality)*

Given that the PDF, $p(t_k; H_1)$, is unimodal, the detection power, P_D , of the mGLRT

is bounded from above as

$$P_D(\alpha) \leq \max \left\{ \frac{4}{9} \frac{1}{1 + \left(\frac{Q_{\chi_{a-b}^2}^{-1}(\alpha) - \mu_1}{\sigma_1} \right)^2}, \frac{4}{3} \frac{1}{1 + \left(\frac{Q_{\chi_{a-b}^2}^{-1}(\alpha) - \mu_1}{\sigma_1} \right)^2} - \frac{1}{3} \right\}, \mu_1 < Q_{\chi_{a-b}^2}^{-1}(\alpha), \quad (4.11)$$

where α is the probability of false alarm, and μ_1 and σ_1^2 are the mean and variance of the modified mGLRT statistic, $2\mathcal{T}_k(\cdot)$, under the hypothesis H_1 .

Proof. Given a random variable X with mean μ_1 and variance σ_1^2 , we have the following bound due to the one-sided Vysochanskii-Petunin inequality [108].

$$Pr(X - \mu_1 \geq a) \leq \max \left\{ \frac{4}{9} \frac{\sigma_1^2}{\sigma_1^2 + a^2}, \frac{4}{3} \frac{\sigma_1^2}{\sigma_1^2 + a^2} - \frac{1}{3} \right\}, \quad a > 0. \quad (4.12)$$

The desired result can now be obtained by following the steps given for Proposition 4.1. □

Taylor Series Approximation for μ_i and σ_i^2

The first and second order statistics, μ_1 and σ_1^2 , of the mGLRT statistic, required to compute the upper bounds derived above can be approximated using Taylor's theorem.

Denote by $\theta_i = (\theta_{1i}, \theta_{2i}, \dots, \theta_{Ni})^T$, the mean vector of sensor observations under the hypothesis H_i . Suppose that the test statistic $\mathcal{T}_k(\mathbf{z}; H_i)$, is sufficiently 'smooth' so that it is differentiable to the required order. Then, the first-order Taylor series of $\mathcal{T}_k(\mathbf{z}; H_i)$ about θ_i is

$$\mathcal{T}_k(\mathbf{z}; H_i) = \mathcal{T}_k(\theta_i) + \sum_{n=1}^N \dot{\mathcal{T}}_{k_n}(\theta_i) (z_{ni} - \theta_{ni}) + \text{Remainder}, \quad (4.13)$$

where

$$\dot{\mathcal{T}}_{k_n}(\theta_i) = \left. \frac{\partial}{\partial z_n} \mathcal{T}_k(\mathbf{z}) \right|_{z_1=\theta_{1i}, \dots, z_N=\theta_{Ni}}. \quad (4.14)$$

Dropping the remainder and taking the expectation on both sides of Eq. (4.13), we have,

$$\begin{aligned} \mu'_i &\approx \mathcal{T}_k(\theta_i) + \sum_{n=1}^N \dot{\mathcal{T}}_{k_n}(\theta_i) \underbrace{E(z_{ni} - \theta_{ni})}_{=0} \\ &= \mathcal{T}_k(\theta_i). \end{aligned} \quad (4.15)$$

Thus,

$$\begin{aligned} \mu'_1 &= E\{\mathcal{T}_k(\mathbf{z}; H_1)\} \\ &\approx L \left[\sum_{n=1}^N \log \frac{f_n(\theta_{n1})}{g_n(\theta_{n1})} + \log \frac{k_1\{F_1(\theta_{11}), \dots, F_N(\theta_{N1}); \psi_{d^*}\}}{k_0\{G_1(\theta_{11}), \dots, G_N(\theta_{N1}); \lambda_{d^*}\}} \right], \end{aligned} \quad (4.16)$$

where ψ_{d^*} and λ_{d^*} denote the pseudo-true values,

$$\psi_{d^*} = \operatorname{argmax}_{\psi_d \in \Psi_d} \log f_p(\mathbf{z}) k_1(\cdot; \psi_d), \quad (4.17)$$

$$\lambda_{d^*} = \operatorname{argmax}_{\lambda_d \in \Lambda_d} \log g_p(\mathbf{z}) k_0(\cdot; \lambda_d). \quad (4.18)$$

Similarly, we can approximate the variance under each hypothesis H_i by

$$\sigma_i'^2 \approx E [\mathcal{T}_k(\mathbf{z}; H_i) - \mathcal{T}_k(\theta_i)]^2 \quad (4.19)$$

$$\approx E \left(\left(\sum_{n=1}^N \dot{\mathcal{T}}_{k_n}(\theta_i)(z_{ni} - \theta_{ni}) \right)^2 \right) \quad (4.20)$$

$$= \sum_{n=1}^N \dot{\mathcal{T}}_{k_n}(\theta_i)^2 \text{Var}(z_n; H_i) + 2 \sum_{n>m} \dot{\mathcal{T}}_{k_n}(\theta_i) \dot{\mathcal{T}}_{k_m}(\theta_i) \text{Cov}(z_n, z_m; H_i). \quad (4.21)$$

The mean, μ_1 , and the variance, σ_1^2 , of the modified test statistic, $2\mathcal{T}_k(\cdot)$ can be approximated as,

$$\mu_1 \approx 2\mu_1' \quad \text{and} \quad \sigma_1^2 \approx 4\sigma_1'^2 \quad (4.22)$$

Thus, given the first and second order statistics of the marginals, and the first order pair-wise correlation between the variables, we can compute μ_1 and σ_1^2 . These values can be subsequently used with Propositions 4.1 or 4.2 to derive the respective upper bounds on P_D .

Next, we present an illustrative example.

4.3 Example

Consider a parallel network of two heterogeneous sensors ($N = 2$). Measurements z_1 and z_2 at the fusion center are set to normal and beta distributions respectively, i.e., $Z_1 \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and $Z_2 \sim \text{Beta}(\alpha_i, \beta_i)$, $i \in \{0, 1\}$. Without loss of generality (*w.l.o.g.*), we assume that the variables Z_1 and Z_2 are statistically independent and completely known under H_0 , i.e.,

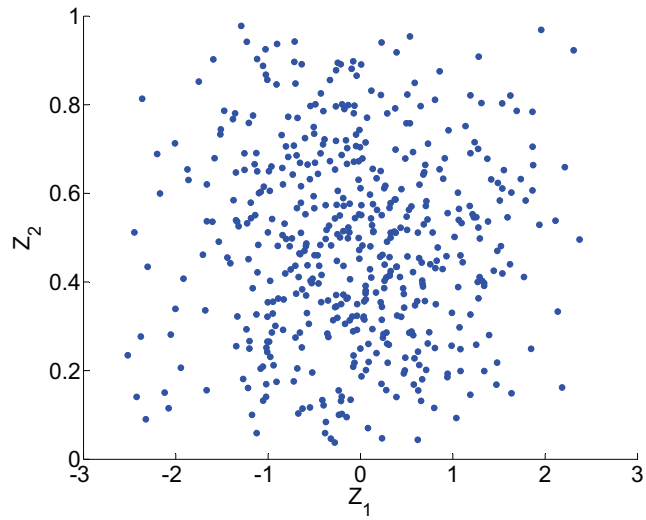
$$g(z_1, z_2; \lambda) = g_1(z_1; \mu_0, \sigma_0^2) g_2(z_2; \alpha_0, \beta_0), \quad (4.23)$$

and ‘ b ’ is equal to zero. A Student’s t-copula with four degrees of freedom and Kendall’s $\tau = 0.2$ is used to generate correlated sensor observations under H_1 . To do this, we first obtain the copula dependence parameter ψ_d given Kendall’s τ (see Eq. (2.35)). The fully specified Student’s t-copula is then used to generate L i.i.d realizations of a uniformly distributed bivariate random vector, $\mathbf{U}_l = [U_{1l} \ U_{2l}]_{l=0, \dots, L-1}$. Each sample is finally transformed using the inverse CDF function to obtain the desired marginal PDFs,

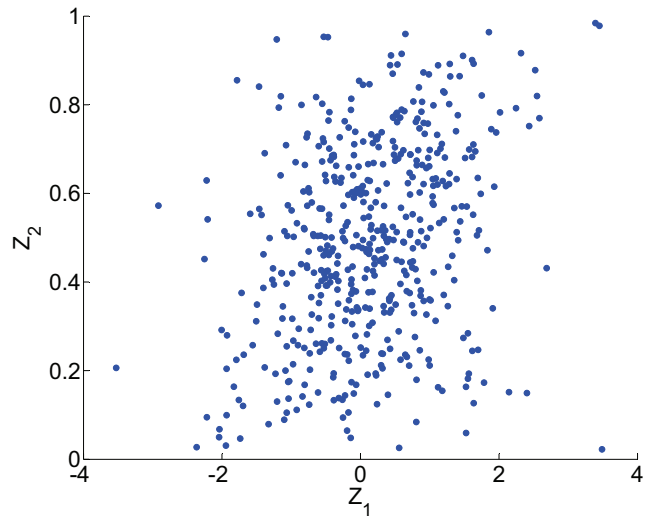
$$z_{nl} = F_n^{-1}(u_{nl}), \quad \forall n, l \quad (4.24)$$

Fig. 4.1 shows the scatter plot of the bivariate random vector (500 realizations). Dependence between the two marginals under H_1 is evident from this figure. It must also be noted that Kendall’s τ is preserved under the inverse CDF transform due to its invariance property under strict monotonic mappings, i.e., $\tau_Z = \tau_U$ [70].

We consider three models, the FGM, Frank and Gaussian copula functions, in the set \mathcal{A}_k and evaluate the performance of the mGLRT derived in Eq. (4.4). Each of these copula functions converge to the independence model under H_0 , i.e., $k(\cdot; \hat{\psi}_d) \xrightarrow{H_0} 1$ for all the three copula densities. Thus, Theorem 4.1 is applicable and can be used to determine the detector threshold. Denoting by $dim(x)$, the dimensionality of parameter x , we have $a = dim(\psi_1) + dim(\psi_2) + dim(\psi_d) = 2 + 2 + 1 = 5$. Thus, $P_F = Q_{\chi^2_5}(2\eta')$ by Theorem 4.1. Fig. 4.2 shows a plot of this theoretical P_F . Also, shown are simulated false-alarm values obtained using 10000 Monte-carlo trials with L set to 100 samples. A good match between the theoretical and simulation P_F values is evident from the figure. The ROC curves for the product and copula-based models are shown in Fig. 4.3. It can be seen from the figure that the mGLRT clearly outperforms the test based on the product model



(a) Hypothesis H_0



(b) Hypothesis H_1

Figure 4.1: Scatter plot of 500 realizations. Non-zero dependence between $Z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Z_2 \sim \text{Beta}(\alpha_1, \beta_1)$ (under H_1) is evident from the figure.

In the next section, we apply the copula-based framework to a real-world application.

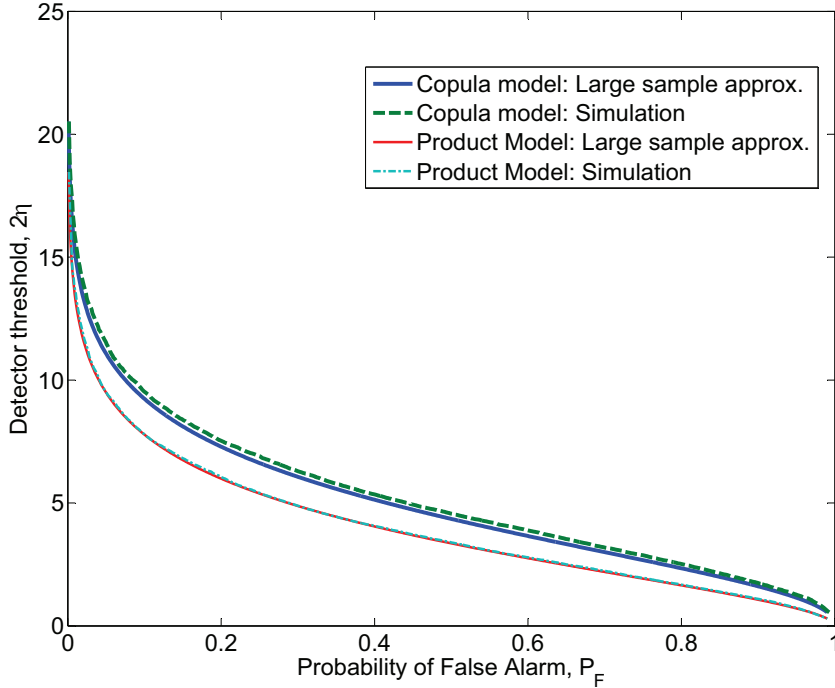


Figure 4.2: Application of Theorem 4.1 for determining the detector thresholds. Simulation parameters are: $(\mu_0, \sigma_0^2) = (0, 1)$, $(\mu_1, \sigma_1^2) = (0.1, 1.1)$, $(\alpha_0, \beta_0) = (2, 2)$, $(\alpha_1, \beta_1) = (2.2, 2.2)$ and Kendall's $\tau|H_1 = 0.2$. The detector makes a decision after processing $L = 100$ samples.

4.4 Application: Footstep Detection

General security applications often require monitoring of the indoor/outdoor environments using multiple sensors, for the purpose of personnel monitoring and surveillance. The goals include detection, localization and tracking of personnel moving in a sensor field, thus invoking algorithms that perform data or feature-level fusion. While video is the most common mode of sensing for zone monitoring, in some scenarios (e.g. military applications), a direct line-of-sight (LOS) may not be available. The deployed sensors may also be required to possess long sensing range capabilities. An unattended ground sensor (UGS) network of acoustic and seismic sensors is usually employed in such scenarios [37, 92].

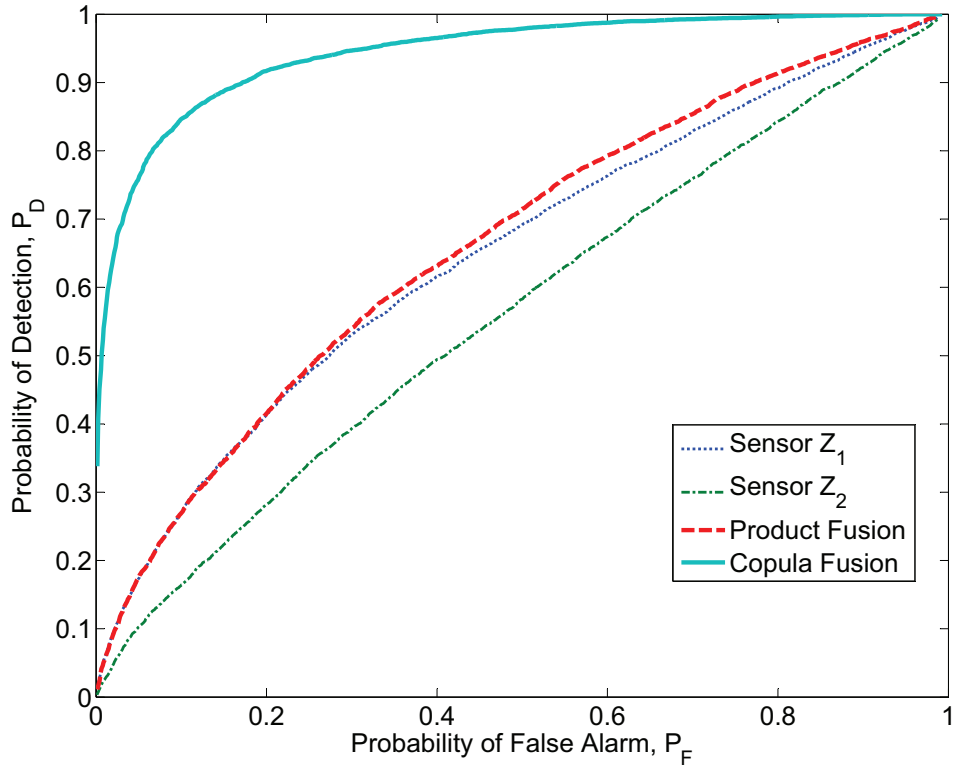


Figure 4.3: Monte Carlo based receiver operating characteristic curves: Parameters for the marginal PDFs are $(\mu_0 = 0, \sigma_0^2 = 1)$ and $(\mu_1 = 0.1, \sigma_1^2 = 1.1)$ and $(\alpha_0 = 2, \beta_0 = 2)$ and $(\alpha_1 = 2.2, \beta_1 = 2.2)$.

We consider, in this work, a personnel detection problem using an acoustic-seismic sensor network that is capable of sensing the vibrations and sound signatures resulting from human footsteps. Data (or features extracted) from all the sensors are fused to arrive at a final decision regarding the presence or absence of human activity. Both acoustic and seismic sensors have an added advantage of being passive and inconspicuous, inexpensive and easy to install.

4.4.1 Related Work and Our Approach

The topic of detecting footsteps using acoustic and/or seismic sensors has been considered in the past. Succi et al. [96, 97] have considered the problem of footstep

signal detection and personnel tracking using seismic sensors. They have suggested the use of kurtosis as a test statistic, which is motivated by the idea that the presence of footsteps would increase the “peakedness” of the underlying probability distribution resulting in higher kurtosis. A scheme based on autoregressive (AR) modeling of acoustic and seismic time series has been proposed by Bland [12]. Ekimov and Sabatier [36] have reported experimental results for vibration and sound characteristics with different walking styles in an indoor environment. Wavelet based heuristics have been proposed in [116]. Dilabar et al. [32] have considered the use of a neural network to detect and classify perimeter intrusion using geophones. Houston et al. [52] have developed a cadence based method for footstep detection.

In this work, we formulate the detection problem as a binary hypothesis testing problem where the hypothesis H_1 indicates the presence of human footsteps while H_0 corresponds to background noise alone. A novel approach based on canonical correlation analysis (CCA) and copula theory is proposed to fuse the acquired acoustic and seismic signals. From the given time series, localized frequency content is estimated using the short-time Fourier transform (STFT), and the acoustic and seismic STFT matrices, $\mathbf{X}_{r \times L}$ and $\mathbf{Y}_{s \times L}$ respectively, are obtained. Column l of \mathbf{X} and \mathbf{Y} contains the absolute values of complex discrete Fourier transform (DFT) coefficients for the time bin l . Assuming the DFT coefficients to be i.i.d. over time, CCA is used to project the coefficients onto a space of lower dimensionality to obtain new variates, \mathbf{u} (acoustic) and \mathbf{v} (seismic); $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, $d = \min(r, s)$.¹ The variates are processed through a likelihood ratio detector, and a final decision regarding the presence or absence of footsteps is obtained.

Application of CCA allows us to operate in the space where the inter-modal correlation is “emphasized”. In addition, for most practical scenarios, the first few pairs of canonical variates are able to incorporate a major proportion of correlation structure

¹The i.i.d. assumption implies $\mathbf{X} \in \mathbb{R}^r$ and $\mathbf{Y} \in \mathbb{R}^s$

present in the data. Thus, further reduction in dimensionality can be achieved by discarding the variates with insignificant correlation values. We note here that the use of CCA for multimodal signal processing problems has been considered in the past. For example, Kidron et al. [62] have considered the use of CCA for audio-video fusion, and, Sargin et al. [84] use CCA to integrate audio and lip-motion features. However, the formulation in these approaches has largely been that of data association. The algorithms focus only on the ‘common information’ between different modalities. In this work, we present a framework that could account for the information specific to individual modalities (complementary information) in addition to the information ‘common’ to all modalities. The basics of CCA are briefly discussed below; more details can be found in [44, 54].

Canonical Correlation Analysis

Denote by $\mathbf{X} = (X_1, \dots, X_r)^T$ and $\mathbf{Y} = (Y_1, \dots, Y_s)^T$, the two multidimensional variables. CCA considers the linear relationship between \mathbf{X} and \mathbf{Y} . The method seeks to replace the variables, \mathbf{X} and \mathbf{Y} , by d pairs of new variables,

$$(u_i, v_i), \quad i = 1, 2, \dots, d, \quad d \leq \min(r, s), \quad (4.25)$$

where

$$\begin{aligned} u_j &= \mathbf{a}_j^T \mathbf{X} = a_{1j}X_1 + a_{2j}X_2 + \dots + a_{rj}X_r \\ v_j &= \mathbf{b}_j^T \mathbf{Y} = b_{1j}Y_1 + b_{2j}Y_2 + \dots + b_{sj}Y_s \end{aligned} \quad (4.26)$$

$j = 1, 2, \dots, d$, are linear projections of \mathbf{X} and \mathbf{Y} respectively. The j^{th} pair of coefficient vectors, $\mathbf{a}_j = (a_{1j}, \dots, a_{rj})^T$ and $\mathbf{b}_j = (b_{1j}, \dots, b_{sj})^T$ are chosen so that, (a) the pairs (u_i, v_i) are ranked according the correlation $\rho_j = \text{corr}(u_j, v_j)$, so that

$1 > \rho_1 \geq \rho_2 \dots \geq \rho_d > 0$, and, (b) both u_j and v_j are uncorrelated with the previously derived variates u_k and v_k respectively, $k < j$.

One can use singular value decomposition (SVD) to obtain the new vector variates,

$$\mathbf{u} = \mathbf{A}\mathbf{X}, \quad \mathbf{v} = \mathbf{B}\mathbf{Y} \quad (4.27)$$

where each row of the matrices $\mathbf{A}_{d \times r}$ and $\mathbf{B}_{d \times s}$ jointly corresponds to one pair of coefficient vectors. Let

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

denote the collection of the $r + s$ variables partitioned into two disjoint subsets, where \mathbf{X} and \mathbf{Y} are jointly distributed with zero mean and covariance,

$$\mathbf{R}_{ZZ} = \mathbb{E}(\mathbf{Z}\mathbf{Z}^T) = \begin{pmatrix} \mathbf{R}_{XX} & \mathbf{R}_{XY} \\ \mathbf{R}_{YX} & \mathbf{R}_{YY} \end{pmatrix}. \quad (4.28)$$

The superscript ‘T’ in Eq. (4.28) denotes the matrix-transpose operation. Now, to solve for the canonical correlations and variates, SVD of the coherence matrix, $\mathbf{C}_h = \mathbf{R}_{XX}^{-\frac{1}{2}} \mathbf{R}_{XY} \mathbf{R}_{YY}^{-\frac{T}{2}}$, is obtained so that

$$\mathbf{C}_h = \mathbf{L}\mathbf{K}\mathbf{M}^T, \quad (4.29)$$

where \mathbf{L} and \mathbf{M} are unitary matrices [85]. The diagonal matrix $\mathbf{K}_{d \times d}$ is the canonical correlation matrix, and the canonical variates are given as,

$$\mathbf{u} = \underbrace{\mathbf{L}^T \mathbf{R}_{XX}^{-\frac{1}{2}}}_{\mathbf{A}} \mathbf{X}, \quad \mathbf{v} = \underbrace{\mathbf{M}^T \mathbf{R}_{YY}^{-\frac{1}{2}}}_{\mathbf{B}} \mathbf{Y}. \quad (4.30)$$

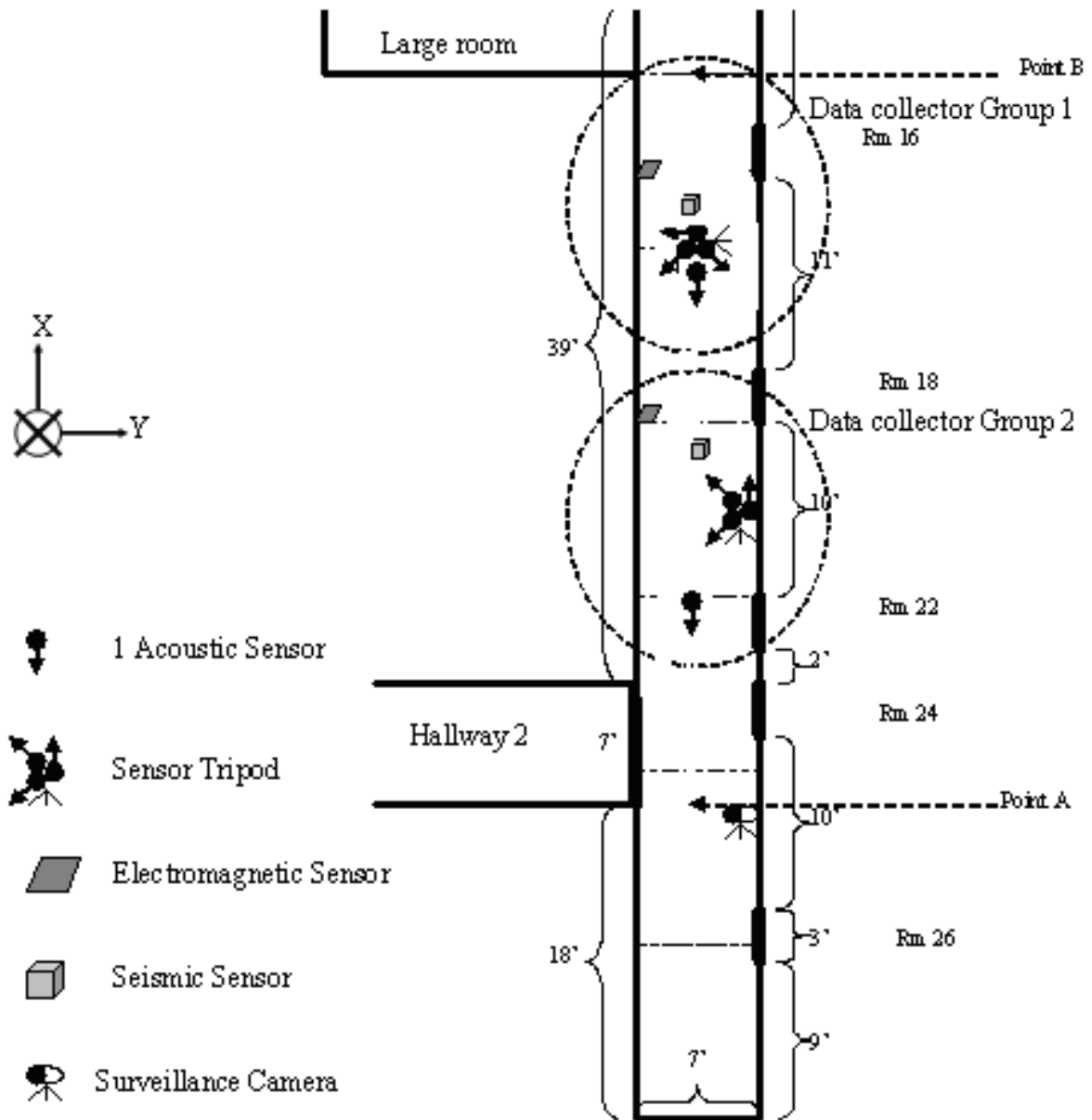


Figure 4.4: Footstep data collection at the US Army Research Lab: Experimental setup

4.4.2 Data Acquisition and Preprocessing

In this section, we describe experiments for footstep data collection and also discuss how the signals are processed before they are input to the likelihood ratio detector.

Experiments were conducted at the U.S. Army Research Lab in the basement of a building. Sensors of diverse modalities such as acoustic, seismic, magnetic, electro-

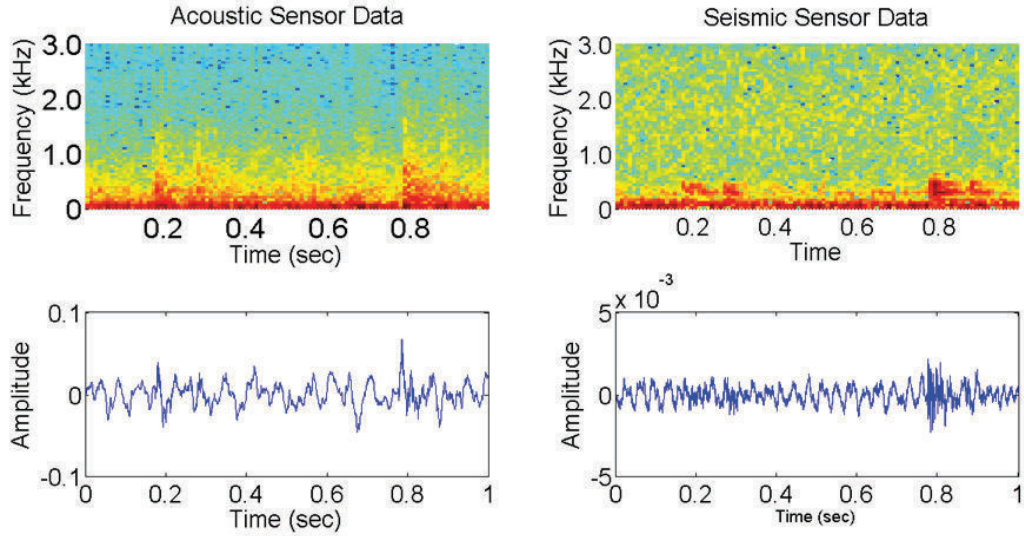


Figure 4.5: Footstep signals and their spectrograms due to normal walk

static and passive infrared (PIR) were placed along a long hallway (see Fig. 4.4) [95] to record movements of various classes of people (e.g., male, female, single, group etc.). In this work, we have used data from acoustic and seismic sensors to detect the presence of a single person. Different styles of walking such as normal walk, brisk walk and running were considered. In addition, an experiment for background noise characterization was also included wherein all sensors monitored the hallway in the absence of any deliberate activity for approximately two minutes.

Acoustic and seismic signals were collected at 16384 Hz and 8192 Hz respectively, and were resampled at an equal rate of 6000 Hz. The detector collects the data for one second before making a decision. A Hamming window of length 20 millisecond with 50% overlap is used alongwith 128-point fast Fourier transform (FFT) to compute the STFT matrices. Fig. 4.5 shows a typical plot of footstep signals due to normal walk along with their spectrograms.

Nineteen acoustic bands ($r = 19$, 187 - 1013 Hz) and ten seismic bands ($s = 10$, 0 - 468 Hz) are then linearly combined to obtain the CCA variates, \mathbf{u} and \mathbf{v} . In Fig. 4.6, we compare the canonical correlation coefficients estimated for different walking

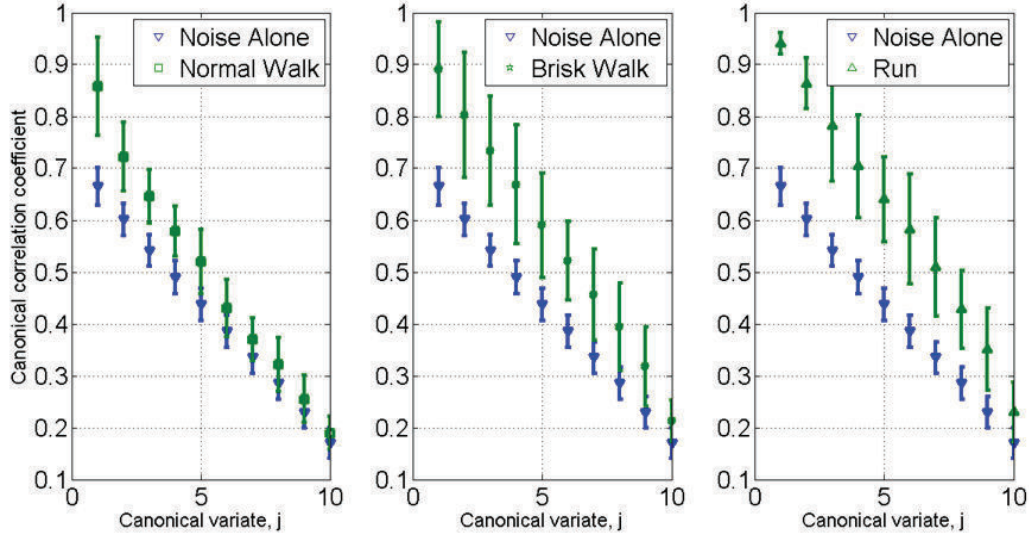


Figure 4.6: Canonical correlation coefficients for noise alone and footsteps plus noise. The plot also shows symmetric one standard deviation error bars.

styles. Data is divided into one second frames and average correlation coefficients are shown along with symmetric one standard deviation error bars. The plot shows a systematic increase in the statistical dependence among the variates, due to the presence of footstep signals. As expected, correlation is highest for signals corresponding to the running activity and lowest in the absence of footstep signals. This trend is true for all the ten ($d = 10$) possible variates.

4.4.3 Statistical Modeling and Detector Design

In order to reduce the detector complexity, we discard all but the first significant pair (u_1, v_1) , and employ the mGLRT (see Eq. (4.4)) at the fusion center,

$$\mathcal{T}(\mathbf{u}_1, \mathbf{v}_1) = \sum_{l=1}^L \log \frac{\max_{\Psi_1, \Psi_2, k_1(\cdot; \Psi_d) \in \mathcal{A}_k} f(u_{1l}; \psi_1) f(v_{1l}; \psi_2) k_1(\cdot; \psi_d)}{g(u_{1l}; \lambda_1) g(v_{1l}; \lambda_2) k_0(\cdot; \lambda_d)} \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (4.31)$$

Note that, in Eq. (4.31), we have assumed full knowledge of $g(u_1, v_1; \lambda)$, the joint PDF of the variates under H_0 . This is a reasonable assumption provided the background noise does not exhibit high frame-to-frame variability as one can always perform noise

characterization before system deployment. We now discuss statistical modeling of the variates under both H_1 and H_0 .

We consider the generalized Gaussian (GG) distribution as a candidate family for modeling the univariate marginals. A random variable X is said to be distributed as a generalized Gaussian if it has the following PDF,

$$p(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^\beta}, \quad (4.32)$$

where $\Gamma(\cdot)$ is the Gamma function, i.e., $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$, $z > 0$, $\alpha = \sigma \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}}$ is a scaling factor that allows the variance to be σ^2 , and β is the shaping factor. The shape of the GG density for different values of β is shown in Fig. 4.7. The family allows to parametrize a continuum of symmetric, platykurtic densities ($\beta > 2$), and a continuum of symmetric, leptokurtic densities ($0 < \beta < 2$). Also, the well-known Laplacian and the Gaussian distributions belong to the GG family with $\beta = 1$ and $\beta = 2$ respectively. In the limiting cases, when $\beta \rightarrow \infty$, $p(x)$ converges to a uniform distribution, whereas for $\beta \rightarrow 0+$, Eq. (4.32) approaches an impulse function.

The use of GG distribution is quite prevalent in the signal processing literature due to its flexibility to model data with different shapes. For example, it has been used in speech modeling [64], wavelet-based texture retrieval [33], image and video coding [58, 88] and many other applications. As we show below, the CCA variates can be accurately modeled using GGD; the noise-only variates are characterized by the standard Gaussian distribution ($\beta = 2$), while the presence of footsteps results in a more peaked distribution given by reduced values of β (< 2).

Several approaches such as method of moments (MM) [7, 103], entropy matching [4] or the ML framework [33] can be adopted to estimate the parameters of a GGD. Following [33], we use the ML approach and consider the following log-likelihood

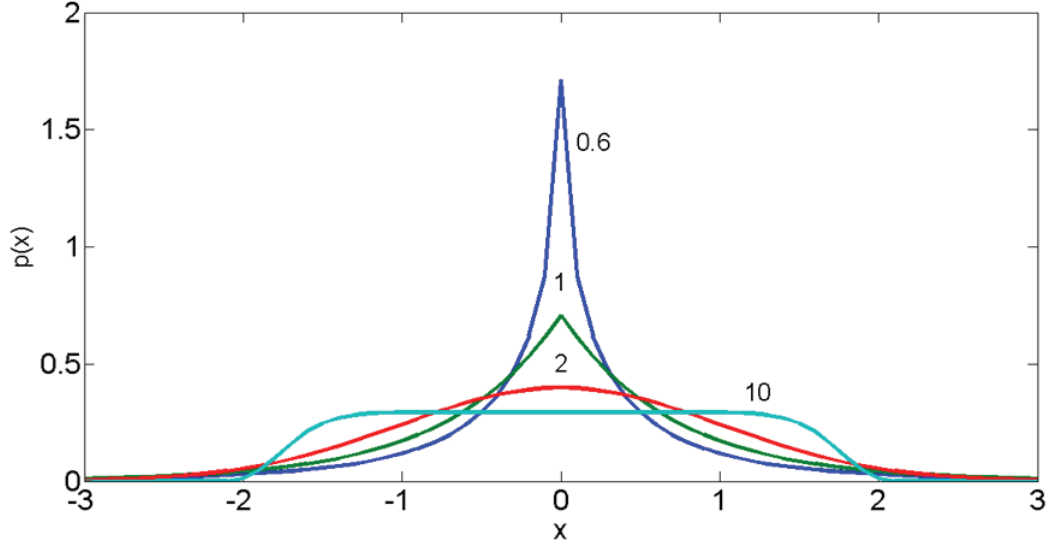


Figure 4.7: The probability density of the generalized Gaussian model for different values of the shape parameter, $\beta = 0.6, 1, 2$ and 10 . All distributions are normalized to unit variance and have zero mean.

function,

$$\mathcal{L}(X; \alpha, \beta) = \sum_{l=1}^L \log p(x_l; \alpha, \beta). \quad (4.33)$$

The associated Euler-Lagrange (EL) equations can be derived by setting the derivatives, $\frac{\partial}{\partial \alpha} \mathcal{L}(\mathbf{x}; \alpha, \beta)$ and $\frac{\partial}{\partial \beta} \mathcal{L}(\mathbf{x}; \alpha, \beta)$, to zero. That is,

$$\begin{aligned} \frac{\partial}{\partial \alpha} \mathcal{L}(\mathbf{x}; \alpha, \beta) &= \sum_{l=1}^L \frac{\partial}{\partial \alpha} \log p(x_l; \alpha, \beta) = 0 \\ \Rightarrow \sum_{l=1}^L \left\{ \frac{\partial}{\partial \alpha} \log \frac{\beta}{2\alpha \Gamma(\frac{1}{\beta})} - \frac{\partial}{\partial \alpha} \left(\frac{|x_l|}{\alpha} \right)^\beta \right\} &= 0 \\ \Rightarrow -\frac{L}{\alpha} + \sum_{l=1}^L \frac{\beta |x_l|^\beta \alpha^{-\beta}}{\alpha} &= 0, \end{aligned} \quad (4.34)$$

and,

$$\begin{aligned}
\frac{\partial}{\partial \beta} \mathcal{L}(\mathbf{x}; \alpha, \beta) &= \sum_{l=1}^L \frac{\partial}{\partial \beta} \log p(x_l; \alpha, \beta) = 0 \\
\Rightarrow \sum_{l=1}^L \left\{ \frac{\partial}{\partial \beta} \log \frac{\beta}{2\alpha\Gamma(1/\beta)} - \frac{\partial}{\partial \beta} \left(\frac{|x_l|}{\alpha} \right)^\beta \right\} &= 0 \\
\Rightarrow \sum_{l=1}^L \left\{ \frac{\Gamma(1/\beta)}{\beta} \frac{\partial}{\partial \beta} \frac{\beta}{\Gamma(1/\beta)} - \frac{\partial}{\partial \beta} \left(\frac{|x_l|}{\alpha} \right)^\beta \right\} &= 0 \\
\Rightarrow \frac{L}{\beta} + \frac{L\Psi(1/\beta)}{\beta^2} - \sum_{l=1}^L \left(\frac{|x_l|}{\alpha} \right)^\beta \log \left(\frac{|x_l|}{\alpha} \right) &= 0, \tag{4.35}
\end{aligned}$$

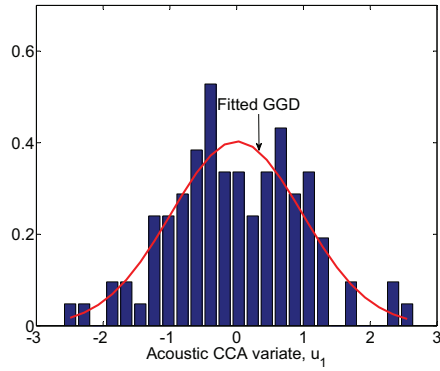
where $\Psi(\cdot)$ is the digamma function ($\Psi(z) = \frac{\dot{\Gamma}(z)}{\Gamma(z)}$), are solved to obtain the ML estimates of α and β . Given $\hat{\beta}$, Eq. (4.34) defines a unique estimator $\hat{\alpha} = \left(\frac{\hat{\beta}}{L} \sum_{l=1}^L |x_l|^{\hat{\beta}} \right)^{\frac{1}{\hat{\beta}}}$ which on substitution in Eq. (4.35) gives the following equation,

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{l=1}^L |x_l|^{\hat{\beta}} \log |x_l|}{\sum |x_l|^{\hat{\beta}}} + \frac{\log \left(\frac{\hat{\beta}}{L} \sum |x_l|^{\hat{\beta}} \right)}{\hat{\beta}} = 0. \tag{4.36}$$

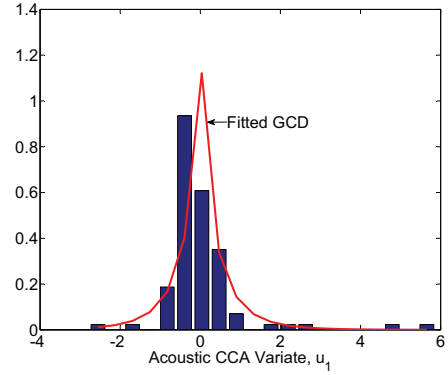
Equation (4.36) has been shown to have a unique root at least in probability [105] which corresponds to the ML estimate of β . The equation has to be solved numerically, and we adopt the Newton-Raphson iterative procedure proposed in [33].

Fig. 4.8(a) shows histogram of the CCA variate u_1 computed from a one second noise-only frame, together with the fitted GG density. A similar plot for footstep signals is shown in Fig. 4.8(b). As expected, the presence of footsteps results in a more ‘peakier’ distribution. The GGD fits were equally good for the seismic CCA variate, v_1 , as well (see Fig. 4.9).

It is interesting to note that the value of β estimated for the noise-only frame is very close to 2 for both, u_1 and v_1 , suggesting a Gaussian model for the variates under H_0 . Fig. 4.10 which analyzes the range of β computed from multiple one second noise-only frames, shows that this is indeed the case. The histogram of β estimates,

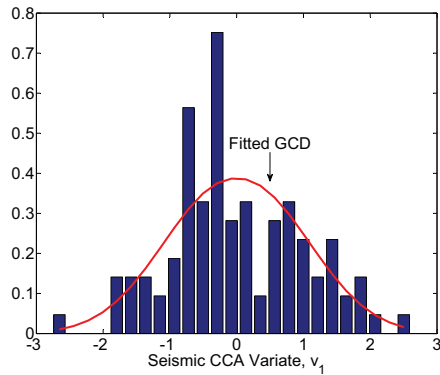


(a) Noise alone: $(\hat{\alpha} = 1.4, \hat{\beta} = 1.98)$

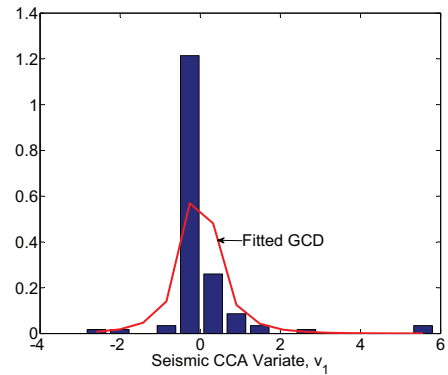


(b) Running: $(\hat{\alpha} = 0.29, \hat{\beta} = 0.72)$

Figure 4.8: Histogram of the acoustic CCA variate, u_1 , together with the fitted generalized Gaussian density



(a) Noise alone: $(\hat{\alpha} = 1.46, \hat{\beta} = 2.10)$



(b) Running: $(\hat{\alpha} = 0.5, \hat{\beta} = 0.9)$

Figure 4.9: Histogram of the seismic CCA variate, v_1 , together with the fitted generalized Gaussian density

for both the acoustic and the seismic variates, peaks at $\beta = 2$ with a low frame-to-frame variance of 0.25 and 0.36 respectively. In order to increase our confidence in the estimated values, we have used the moving block bootstrap technique [118] to generate Q overlapping frames each one second in duration to plot the histograms. The Gaussanity of the variates is further confirmed by the Kolmogorov-Smirnov (KS) goodness-of-fit test [66]. All Q frames passed the test for normality when tested at

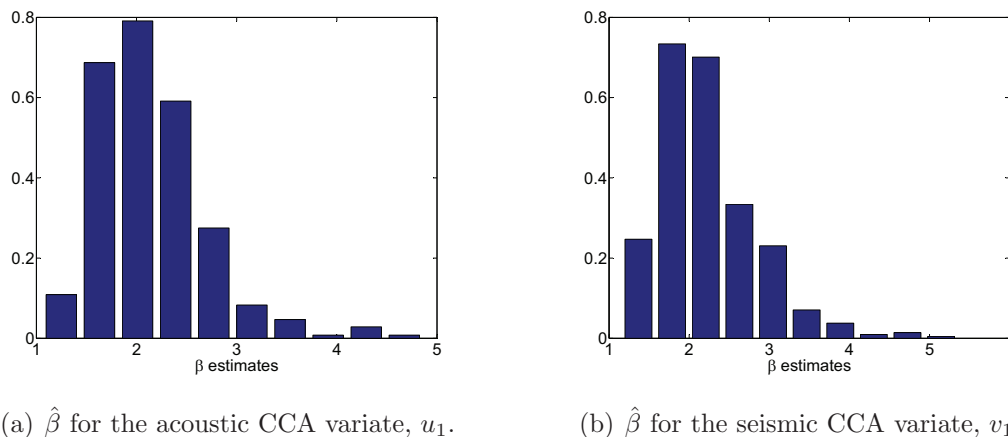


Figure 4.10: Histogram of β estimated using the ‘noise-alone’ data

Table 4.1: Mean and symmetric one standard deviation confidence intervals for $\hat{\beta}$

| Modality | Noise Alone | Normal Walk | Brisk Walk | Run |
|----------|---------------------|---------------------|---------------------|---------------------|
| Acoustic | 2.13 (± 0.5) | 1.1 (± 0.34) | 1.08 (± 0.29) | 1.0 (± 0.16) |
| Seismic | 2.19 (± 0.62) | 1.14 (± 0.35) | 1.13 (± 0.35) | 0.95 (± 0.15) |

0.05 significance level. The mean and symmetric one standard deviation confidence intervals for the $\hat{\beta}$ for all the activities considered here, are shown in Table 4.1.

We now focus on choosing the copula density, $k_0(\cdot; \lambda_d)$, to characterize the dependence structure between the variates u_1 and v_1 under H_0 . Using the KL divergence (see Section 3.2.1) as a criterion for model selection, we observe that the Gaussian copula with $\lambda_d = 0.66$ best fits the data set. Thus, $g(u_1, v_1) \sim \mathcal{N}[(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}); (\begin{smallmatrix} 1.0 & 0.66 \\ 0.66 & 1.0 \end{smallmatrix})]$. The copula model for the data under H_1 is estimated online under the mGLRT framework (see Eq. (4.31)).

4.4.4 Results

In this section, we investigate the performance of the proposed copula based approach for footstep detection. Probability of detection, P_D , was estimated using the footstep

recordings when a single person traversed from the origin, A , to the destination, B , and back by (a) walking at a normal pace (for approximately 15 seconds), (b) walking at a relatively brisk pace (for about 12 seconds) and (c) running for about 8 seconds. A decision window of length 6000 samples was used. Moving block bootstrap technique was used to generate Q_1 overlapping resamples under H_1 . Thus,

$$\hat{P}_D = \frac{\# \text{ decisions in favor of } H_1}{Q_1}. \quad (4.37)$$

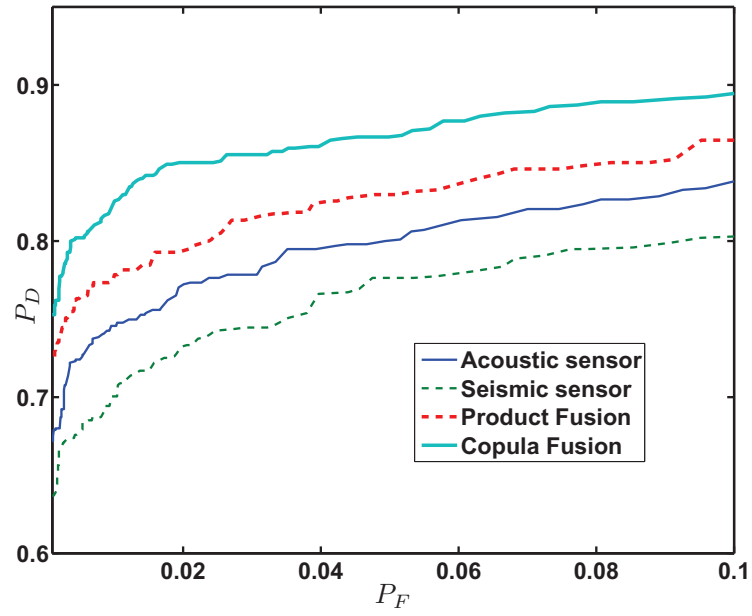
Probability of false alarm, P_F , was estimated from the two minute noise-only recordings following the same procedure. The ROC plots for the three walking styles are shown in Fig. 4.11.

As evident from the figure, the CCA-copula detector outperformed the product rule in all the three cases.

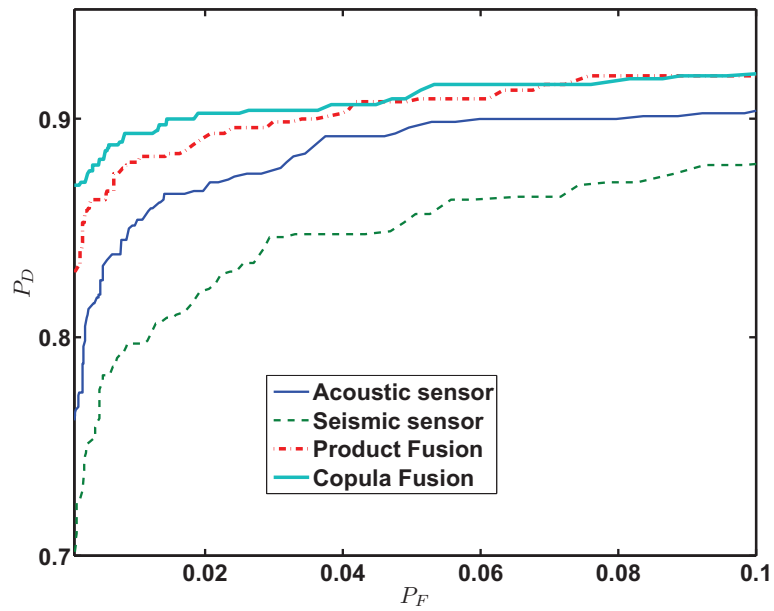
4.5 Summary

In this chapter, a composite hypothesis testing problem with heterogeneous observations was formulated, and a GLRT based test for the same was derived. It was assumed that the joint PDF under the null hypothesis H_0 , and the marginal PDFs under both hypotheses are well-specified. However, no such assumptions were made for the copula density used to approximate the joint PDF of observations under H_1 . We, therefore, call the test a misspecified GLRT. Copula selection was embedded in the estimation step of the GLRT. When conditions are satisfied so that Wilks theorem is applicable, it becomes possible to determine the detector threshold and also compute upper bounds on P_D , at a fixed value of P_F . The developed theory was then applied to solve the problem of detecting human footsteps using a network of acoustic and seismic sensors. A novel approach based on CCA and copula theory was presented. Application of CCA allows one to operate in the space where the inter-

modal correlation is “emphasized” while also achieving dimensionality reduction. It was shown that the CCA variates of both, the acoustic and the seismic signals, could be modeled using a generalized gaussian density. Different styles of walking were considered and it was shown that the proposed CCA-copula detector led to significant improvement over the product fusion rule which fails to characterize inter-modal correlations.

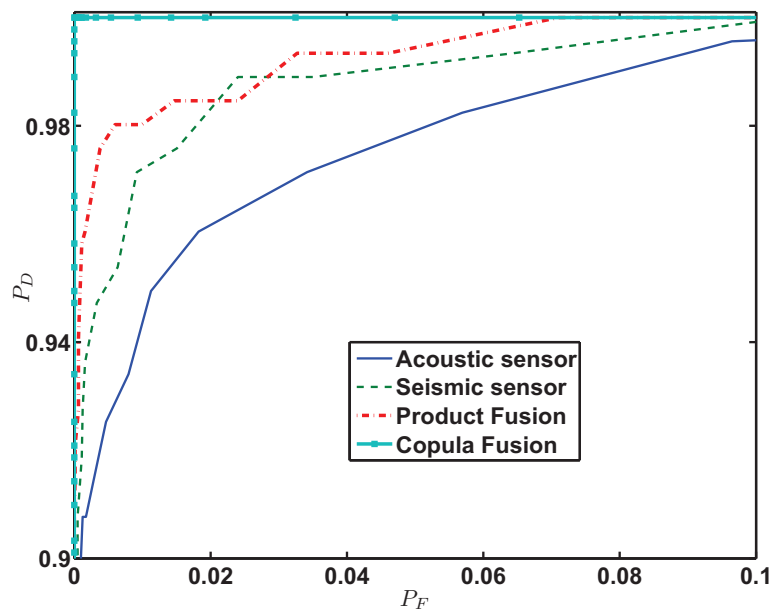


(a) Normal Walk



(b) Brisk walk

Figure 4.11: Receiver operating characteristic curves for the CCA-copula based detector



(c) Running

Figure 4.11: Receiver operating characteristic curves for the CCA-copula based detector

Chapter 5

Hypothesis Testing With Quantized Sensor Measurements

In the earlier chapters, a multisensor detection problem was considered where sensors of disparate modalities transmit their observations (or features extracted thereof) to a fusion center. While designing the copula based fusion rule, it was assumed that the fusion center had access to the exact real-valued (analog) version of the locally processed data. However, in many cases (e.g. wireless sensor networks), there could be limitations on both the transmission power and the bandwidth available for sensor-to-fusion center communication. It may, thus, be necessary to quantize the information at each sensor before its transmission to the fusion center. The goal of this chapter is to design a decision fusion rule for combining quantized heterogeneous information, thereby extending the formulation in the preceding chapter to a distributed setting (see Fig. 5.1).

Distributed detection has long been an active and important research area [106]. The design of a distributed detection system involves designing the local and fusion center decision rules under different criteria and constraints [19, 51, 106]. Under the assumptions that local observations are conditionally independent given the hy-

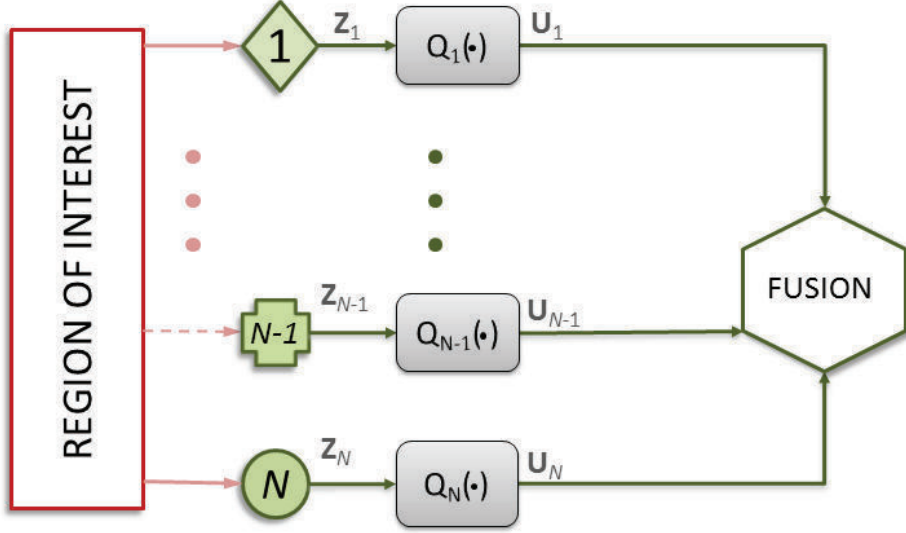


Figure 5.1: Distributed heterogeneous sensor network: A parallel architecture.

pothesis and the fusion center receives the local sensor outputs without any loss, the optimality of the LRT for local sensor decision rules under the Bayesian criterion and the Neyman-Pearson criterion have been proved in [50] and [99]. However, the problem becomes highly complex when conditional independence does not hold. The LR based decision rules at the local sensors may no longer result in an optimal system design [3, 115]. It has also been shown that distributed detection with dependent observations is an NP-complete problem; it cannot be solved using a polynomial time algorithm [73, 100]. The problem is usually simplified by constraining the local sensors to be binary quantizers. In [34], Drakopoulos and Lee have derived a rule for fusing correlated decisions under the assumption that the correlation coefficients between the sensor decisions are known and local sensor thresholds generating the correlated decisions are given. Kam et. al [59], employed another approach, namely, the Bahadur-Lazarsfeld series expansion of PDFs to derive the optimum fusion rule for correlated local decisions. It was assumed that the joint distribution of sensor observations was completely known. In this chapter, we present another approach based on copula theory for fusing statistically dependent decisions from the local heterogeneous sensors. The application of copula theory for fusing correlated decisions

has been recently considered in [98]. The local sensors were binary quantizers, and it was assumed that the *true* copula generating data under each hypothesis is known *a priori*. In this work, we relax this assumption, and also extend the formulation to multibit quantizers. A copula based GLRT (mGLRT), similar to the one designed in Chapter 4, is employed. An alternative suboptimal but computationally efficient fusion rule is also proposed which involves deliberately adding an external noise to the quantized observations before fusion. While noise is generally perceived as an unwanted signal, interestingly, several studies have shown that addition of *controlled* noise could in fact be *beneficial* in some cases. For example, dithering, the process of adding noise to the signal before quantization has been shown to improve signal quality and mitigate the artifacts introduced due to quantization [45, 82, 86]. Also, it has been observed by many researchers that some types of signals get amplified by a nonlinear system when noise is added to the input signal (see [20] and references therein). This phenomenon is popularly known as stochastic resonance (SR). Here, we use this approach of adding external noise to address the issue of computational complexity.

One of the main limitations of the copula-based rule for fusing discrete decisions is the considerable increase in computational complexity as the number of sensors and/or quantization levels increases. As the authors in [98] note, implementation of a copula based GLRT for N binary quantizers requires one to compute 2^N N -dimensional integrals, and optimize over an $N(N - 1)/2$ dimensional space for ML estimation (of elliptical copula parameters). The system proposed here greatly simplifies the fusion rule, and can be considerably accurate especially when the number of quantization levels is large. The key is to choose a suitable noise PDF for the external noise to be added. We call this noise, the low pass filter noise (LPF-noise), for reasons that will become clear later. Our approach is based on Widrow's additive quantization noise model [111, 113] and is similar to Gustafsson and Karlson [48] who

have considered the problem of estimating a deterministic parameter in noise using quantized observations. However, unlike [48] where the authors propose to inject the artificial dither noise before quantization, we add the deliberate disturbance post quantization and at the fusion center. As we show later, the addition of noise after quantization is equivalent to low pass filtering in the characteristic function (CF) domain, unlike dithering which amounts to anti-alias filtering [48, 112].

The rest of the chapter is organized as follows. We formulate the problem in Section 5.1 and define the input-output transfer function of the quantizers used in this chapter. A copula based rule for fusing correlated local sensor decisions is derived in Section 5.2. Section 5.3 addresses the issue of computational complexity associated with the fusion rule derived in the previous section. An alternative suboptimal but computationally efficient fusion rule based on Widrow’s pseudo-quantization noise model (PQN) is proposed. It is shown that under some conditions, the PQN model-based fusion rule can be highly accurate, while also reducing the computational complexity. An illustrative example is presented in Section 5.4 to elucidate the theory presented in the previous sections. Section 5.5 is a brief summary of the chapter.

5.1 Problem Formulation

The problem of signal detection is formulated as a binary hypothesis test where the hypothesis H_1 indicates the presence of a signal, while H_0 indicates the absence of the same. As in Chapter 4, observations at each sensor n are assumed to be i.i.d with PDFs $f_n(z_n; \psi_n)$ and $g_n(z_n; \lambda_n)$ under H_1 and H_0 respectively. We assume that the marginal PDFs are well-specified under both hypotheses; however no knowledge is available regarding the dependence structure between the heterogeneous sensed data streams. We approximate the unknown dependence structures using copula functions. Sensor observations are further passed through uniform scalar quantizers

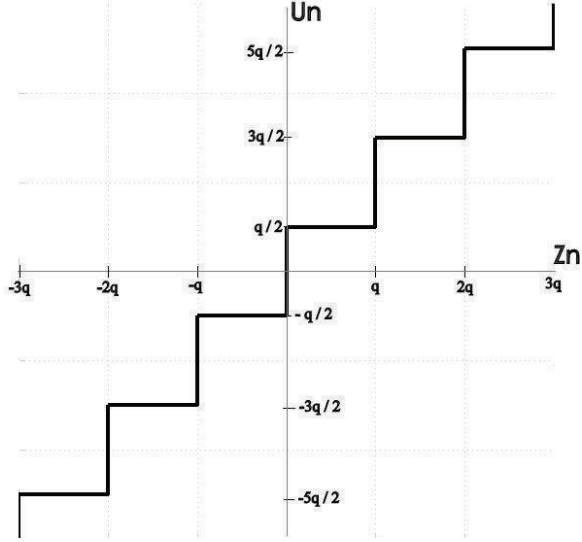


Figure 5.2: Input-output transfer function of a uniform scalar quantizer

(see Fig. 5.1) before their transmission to a remotely located fusion center. The input-output transfer function of the quantizer at each sensor is shown in Fig. 5.2. Thus, the quantizer output, during any time interval $1 \leq l \leq L$, can be given as

$$u_{nl} = Q_m(z_{nl}) = \begin{cases} -m_n q_n + \frac{q_n}{2}, & z_{nl} < -m_n q_n, \\ q_n \lfloor \frac{z_{nl}}{q_n} \rfloor + \frac{q_n}{2}, & -m_n q_n < z_{nl} \leq m_n q_n, \\ m_n q_n - \frac{q_n}{2}, & z_{nl} \geq m_n q_n, \end{cases} \quad (5.1)$$

where, q_n and $2m_n$ correspond to the quantizer step size and the number of quantization levels respectively, at sensor n . Further, $\lfloor x \rfloor$ stands for the floor operation that denotes an integer smaller than or equal to x . The quantized value at sensor n can be represented with an integer $i_n = -m_n + 1, -m_n + 2, \dots, m_n$.

Observations thus received at the fusion center are used to estimate the unknown model parameters, and a GLRT based fusion rule is employed for global decision making. In addition to estimating the model parameters, the selection of copula densities is also performed in real-time (see Section 4.1). Sensor observations and hence their quantized versions are assumed to be i.i.d in time, and, the focus is

on designing a fusion rule that could exploit the spatial correlation between sensor decisions for improved detection performance.

5.2 (Misspecified) GLRT Based Fusion of Soft Decisions

In the following, we consider a two-sensor network noting that the extension to more number of sensors is straightforward.

Under hypothesis H_1 , the probability that the data $R_l = (u_{1l}, u_{2l})$ received at the fusion center at the time instant l takes a specific value $(i_1 q_1 - \frac{q_1}{2}, i_2 q_2 - \frac{q_2}{2})$ is

$$P_{i_1, i_2} = \int_{(i_1-1)q_1}^{i_1 q_1} \int_{(i_2-1)q_2}^{i_2 q_2} f(z_1, z_2) dz_2 dz_1, \quad (5.2)$$

where $f(z_1, z_2)$ is the true but unknown joint PDF of unquantized sensor observations under H_1 . Now, approximating the dependence structure using a copula density (say) $k_1 \{F_1(z_1; \psi_1), F_2(z_2; \psi_2); \psi_d\}$ contained in some set \mathcal{A}_k of valid copula densities, we have

$$\begin{aligned} \hat{P}_{i_1, i_2}(\psi) &= \int_{(i_1-1)q_1}^{i_1 q_1} \int_{(i_2-1)q_2}^{i_2 q_2} f_1(z_1; \psi_1) f_2(z_2; \psi_2) k_1(F_1(z_1; \psi_1), F_2(z_2; \psi_2); \psi_d) dz_2 dz_1 \\ &= K_1 \{F_1(i_1 q_1; \psi_1), F_2(i_2 q_2; \psi_2); \psi_d\} - \\ &\quad K_1 \{F_1((i_1 - 1)q_1; \psi_1), F_2((i_2 - 1)q_2; \psi_2); \psi_d\}, \end{aligned} \quad (5.3)$$

where $\psi = (\psi_1, \psi_2, \psi_d)^T \in \Psi \subset \mathbb{R}^a$ are the unknown parameters that will be estimated from the received data, $K_1\{\cdot\}$ is the copula CDF and $F_n(\cdot)$ is the CDF of Z_n under hypothesis H_1 . The dependence of $\hat{P}_{i_1, i_2}(\psi)$ on $K_1\{\cdot\}$ is not made explicit for notational convenience.

The likelihood function of the data R_l under hypothesis H_1 can now be written as

$$\hat{P}(R_l; \psi, H_1) = \prod_{i_1} \prod_{i_2} \left[\hat{P}_{i_1, i_2}(\psi) \right]^{\delta(u_{1l} - i_1 q_1 + \frac{q_1}{2}, u_{2l} - i_2 q_2 + \frac{q_2}{2})}, \quad (5.4)$$

where $\delta(\cdot)$ is the two-dimensional Kronecker-delta function defined as

$$\delta(x, y) = \begin{cases} 1, & x, y = 0 \\ 0, & x, y \neq 0. \end{cases} \quad (5.5)$$

The log-likelihood function of R_l is, therefore,

$$\log \hat{P}(R_l; \psi, H_1) = \sum_{i_1} \sum_{i_2} \delta \left(u_{1l} - i_1 q_1 + \frac{q_1}{2}, u_{2l} - i_2 q_2 + \frac{q_2}{2} \right) \log \hat{P}_{i_1, i_2}(\psi). \quad (5.6)$$

Similarly, the likelihood function of R_l under H_0 , when a copula density $k_0(G_1(z_1; \lambda_1), F_2(z_2; \lambda_2); \lambda_d) \in \mathcal{A}_k$ is used to approximate the joint distribution under H_0 , can be derived as,

$$\log \hat{P}(R_l; \lambda, H_0) = \sum_{i_1} \sum_{i_2} \delta \left(u_{1l} - i_1 q_1 + \frac{q_1}{2}, u_{2l} - i_2 q_2 + \frac{q_2}{2} \right) \log \hat{Q}_{i_1, i_2}(\lambda), \quad (5.7)$$

where

$$\begin{aligned} \hat{Q}_{i_1, i_2}(\lambda) &= K_0 \{G_1(i_1 q_1; \lambda_1), G_2(i_2 q_2; \lambda_2); \lambda_d\} - \\ &K_0 \{G_1((i_1 - 1)q_1; \lambda_1), G_2((i_2 - 1)q_2; \lambda_2); \lambda_d\}. \end{aligned} \quad (5.8)$$

$\lambda = (\lambda_1, \lambda_2, \lambda_d)^T \in \Lambda \subset \mathbb{R}^b$ are the unknown parameters associated with the model under H_0 , $K_0\{\cdot\}$ is the copula CDF and $G_n(\cdot)$ is the CDF of Z_n under hypothesis H_0 .

With Eqs. (5.6) and (5.7), it is straightforward to derive the mGLRT to be employed at the fusion center,

$$\mathcal{T}_k^{\text{q}}(\mathbf{u}_1, \mathbf{u}_2) = \log \frac{\max_{k_1(\cdot) \in \mathcal{A}_k, \Psi} \prod_l \hat{P}(R_l; \psi, H_1)}{\max_{k_0(\cdot) \in \mathcal{A}_k, \Lambda} \prod_l \hat{P}(R_l; \lambda, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (5.9)$$

which results in

$$\sum_l \sum_{i_1} \sum_{i_2} \delta \left(u_{1l} - i_1 q_1 + \frac{q_1}{2}, u_{2l} - i_2 q_2 + \frac{q_2}{2} \right) \log \frac{\hat{P}_{i_1, i_2}^*(\hat{\psi})}{\hat{Q}_{i_1, i_2}^*(\hat{\lambda})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (5.10)$$

where $\hat{P}_{i_1, i_2}^*(\hat{\psi})$ and $\hat{Q}_{i_1, i_2}^*(\hat{\lambda})$ correspond to the copula functions $K_1^*(\cdot)$ and $K_0^*(\cdot)$ respectively, which maximize the terms on the left hand side of Eq. (5.9). The superscript ‘q’ in Eq. (5.9) denotes that the detector receives quantized data at its input.

When the conditions of Theorem 4.1 are satisfied, the modified test statistic $2\mathcal{T}_k^{\text{q}}(\mathbf{u}; H_0)$ converges asymptotically ($L \rightarrow \infty$) to a chi-squared distribution with $\nu = a - b$ degrees of freedom. Thus, the probability of false alarm, P_F^{q} , is

$$P_F^{\text{q}} = Pr \left(2\mathcal{T}_k^{\text{q}}(\cdot) > 2\eta; H_0 \right) \stackrel{L \rightarrow \infty}{=} Q_{\chi_{a-b}^2} (2\eta), \quad (5.11)$$

where $Q_{\chi_{\nu}^2}(\cdot)$ denotes the right-tail probability of a chi-squared random variable with ν degrees of freedom. One can thus obtain the threshold η so that P_F^{q} is constrained to a desired level $\alpha \in (0, 1)$ as below:

$$\eta = \frac{Q_{\chi_{a-b}^2}^{-1}(\alpha)}{2}. \quad (5.12)$$

Next, we derive a suboptimal but computationally much simpler test as an alternative to Eq. (5.9).

5.3 A Computationally Efficient Fusion Rule

The copula-based fusion approach involves solving multi-dimensional integrals to evaluate the joint probabilities (see Eq. (5.3)). Given N sensors each with a binary quantizer, there are 2^N possible pairs of joint decisions, and the probability of each pair requires the evaluation of an N -dimensional integral. Thus, the application of the mGLRT becomes highly prohibitive as the number of sensors N and/or quantization levels increases due to the increased computational complexity. In this section, we derive a computationally efficient test at the fusion center that is based on Widrow's pseudo-quantization noise (PQN) model. The test, although suboptimal, can be highly accurate especially for fine quantization systems ($q_n \rightarrow 0$). We first review Widrow's statistical theory of quantization.

5.3.1 Widrow's Statistical Theory of Quantization: A Review

The statistical theory of quantization was developed by Widrow and co-workers [111–113]. They interpreted quantization of a random variable as sampling of its PDF, and showed that the PDF of the quantized signal is the convolution of the input signal PDF with a rectangular pulse function followed by conventional sampling. Thus, the PDF of the quantizer output, u_{nl} , at sensor n and at any time instant, can be given as

$$p_{U_n}(z) = \left(p_{W_n}(z) \star p_{Z_n}(z) \right) \cdot c_{\delta'_n}(z), \quad (5.13)$$

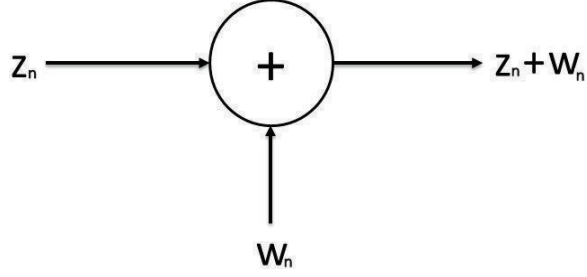


Figure 5.3: Pseudo-quantization noise (PQN) model for quantization

where $p_n(z)$ is the PDF of the random variable at the input Z_n , $p_{W_n}(z)$ denotes the rectangular pulse function,

$$p_{W_n}(z) = \begin{cases} \frac{1}{q_n}, & -q_n/2 < z < q_n/2 \\ 0, & \text{elsewhere,} \end{cases} \quad (5.14)$$

and $c_{\delta'_n}(z)$ denotes the impulse train,

$$c_{\delta'_n}(z) = \sum_{i=-\infty}^{\infty} q_n \delta' \left(z - i_n q_n - \frac{q_n}{2} \right). \quad (5.15)$$

The ‘ \star ’ in Eq. (5.13) denotes the convolution operation, and $\delta'(\cdot)$ in Eq. (5.15) is the dirac-delta function. This process of convolution followed by conventional sampling is popularly known as “area sampling” [113]. Also, note that $p_{W_n}(\cdot)$ is also the PDF of a uniform random variable, $W_n \sim \mathcal{U} \left(-\frac{q_n}{2}, \frac{q_n}{2} \right)$.

Area sampling motivates the PQN model shown in Fig. 5.3. The idea is to model the quantization process as the addition of a uniformly distributed external noise w_n with PDF $p_{W_n}(\cdot)$ given in Eq. (5.14). It is important to note that the external noise w_n is different from the actual quantization noise $\epsilon_n = u_n - z_n$. While w_n is statistically independent of the input z_n , there is a deterministic relationship between ϵ_n and z_n . However, under certain conditions, it becomes possible to substitute the

highly nonlinear quantization process by a simple linear additive uniform noise (AUN) model. These conditions can be explained in the characteristic function (CF) domain.

The CF of a random variable X is obtained by taking the Fourier transform of its PDF $p_X(x)$,

$$\phi_X(v) = \int_{-\infty}^{\infty} p_X(x) \exp^{jvx} dx = E [\exp^{jvx}]. \quad (5.16)$$

Taking the Fourier transform of Eq. (5.13), one obtains the CF of output variable U_n ,

$$\begin{aligned} \phi_{U_n}(v) &= \sum_{i_n=-\infty}^{\infty} e^{-j i_n \frac{2\pi}{q_n} \frac{q_n}{2}} \phi_{Z_n} \left(v + i_n \frac{2\pi}{q_n} \right) \text{sinc} \left(\frac{q_n(v + i_n \frac{2\pi}{q_n})}{2} \right) \\ &= \sum_{i_n=-\infty}^{\infty} (-1)^{i_n} \phi_{Z_n} \left(v + i_n \frac{2\pi}{q_n} \right) \text{sinc} \left(\frac{q_n(v + i_n \frac{2\pi}{q_n})}{2} \right) \end{aligned} \quad (5.17)$$

where $\phi_{Z_n}(v)$ is the CF of the input Z_n and $\text{sinc}(v) = \frac{\sin(v)}{v}$. Fig. 5.4 shows the operations in the 'frequency' domain. Note that the central lobe ($i_n = 0$ in Eq. (5.17)),

$$\phi_{Z_n+W_n}(v) = \phi_{Z_n}(v) \cdot \text{sinc} \left(\frac{q_n v}{2} \right), \quad (5.18)$$

corresponds to the CF one would obtain by adding an independent and uniformly distributed random variable W_n to the input Z_n . It is clear from Fig. 5.4 (d) that, in addition to the error introduced due to the addition of uniform noise, quantization also causes an aliasing error due to overlapping lobes of $\phi_{Z_n+W_n}(v)$.¹

However, if the input PDF is band-limited so that $\phi_{Z_n}(v) = 0$ for $|v| > \frac{\pi}{q_n}$, then the replicated versions of $\phi_{Z_n+W_n}(v)$ do not overlap and, in principle, the original PDF can be reconstructed from the knowledge of $p_{U_n}(\cdot)$. This is Widrow's first quantization theorem:

¹Note that unlike the central lobe, the repetitions have an additional phase shift (Eq. (5.17)), i.e., the odd numbered lobes ($i_n = \pm 1, \pm 3, \dots$) are, in fact, the negative of the central lobe. This is not shown in Fig. 5.4 (d) for simplicity.

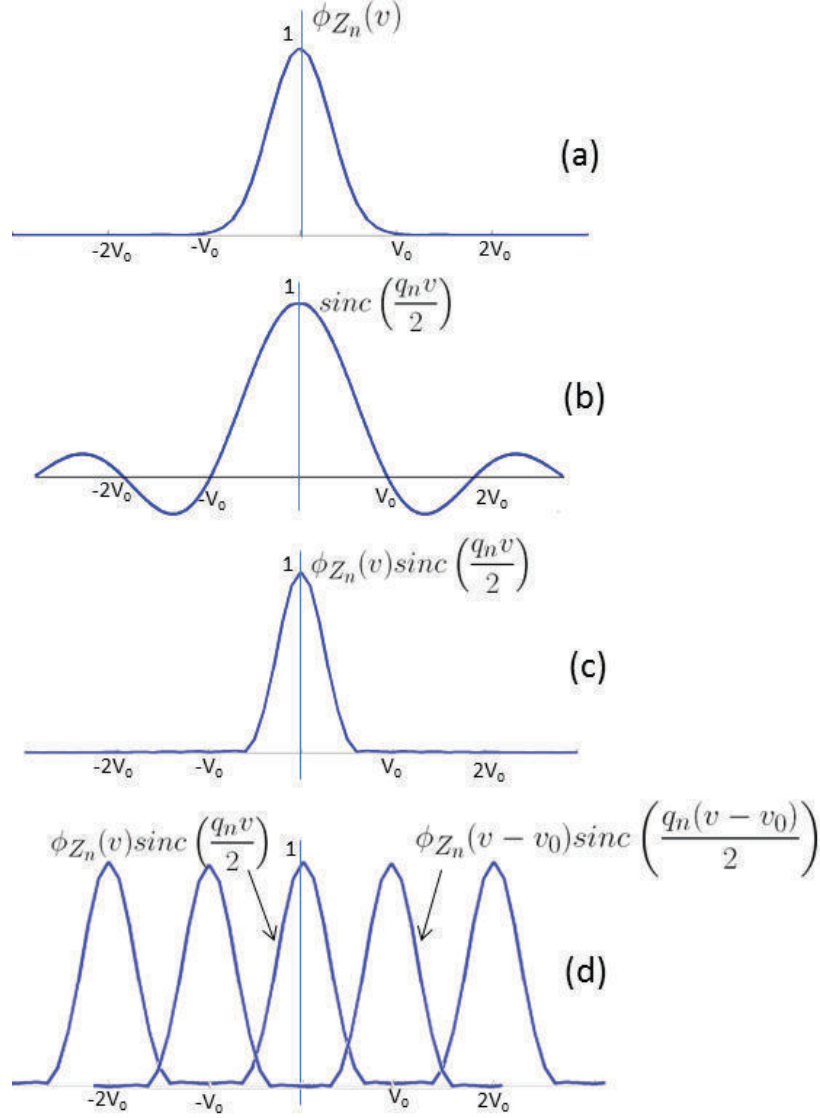


Figure 5.4: Illustration of the quantization process in the CF domain: (a) CF of z_n ; (b) CF of q_n , the sinc function; (c) CF of $z_n + w_n$; (d) repetition of CF of $z_n + w_n$; the CF of the quantized variable is given by the summation of these repetitions.

Theorem 5.1. (*Widrow's Quantization Theorem I*)

If the CF of the input variable Z_n is bandlimited so that

$$\phi_{Z_n}(v) = 0, \quad |v| > \frac{\pi}{q_n}, \quad (5.19)$$

then the different lobes in $\phi_{U_n}(v)$ do not overlap, and in principle, the original PDF $p_n(z_n)$ (before quantization) can be recovered from the PDF of U_n .

When $\phi_{Z_n}(v) = 0$ for $|v| > \frac{2\pi}{q_n}$ so that the derivatives of $\phi_{U_n}(v)$ computed at $v = 0$ are not affected due to the overlap of CF lobes, then the moments of Z_n can be recovered from those of U_n . This is Widrow's second theorem:

Theorem 5.2. (*Widrow's Quantization Theorem II*)

If the CF of Z_n is bandlimited so that

$$\phi_{Z_n}(v) = 0, \quad |v| > \frac{2\pi}{q_n}, \quad (5.20)$$

then the moments of Z_n can be derived from the moments of U_n .

In the following, it is assumed that Theorem 5.1 (and hence Theorem 5.2) holds, and derive a fusion rule based on the PQN model.

5.3.2 Fusion Rule Based on the Pseudo-Quantization Noise Model

As discussed previously, the high complexity in computing the mGLRT statistic for quantized observations stems from the need for computing multi-dimensional integrals. We propose to simplify the fusion process by adding *controlled noise* to the observations received at the fusion center. The system is shown in Fig. 5.5. An externally generated noise, d_n , with PDF $p_{D_n}(d_n)$ is added to the quantized observations from each sensor n before fusing them for making a global decision. Denote the new observations by $y_n = u_n + d_n$ whose CF is given by

$$\phi_{Y_n}(v) = \phi_{U_n}(v) \cdot \phi_{D_n}(v). \quad (5.21)$$

One can choose the noise source with a bandlimited CF to filter out the replicas in $\phi_{U_n}(v)$. This is analogous to low pass filtering in signal processing. We, therefore, call the noise D_n , the LPF-noise. As shown in Fig. 5.4 (d), an ideal noise source

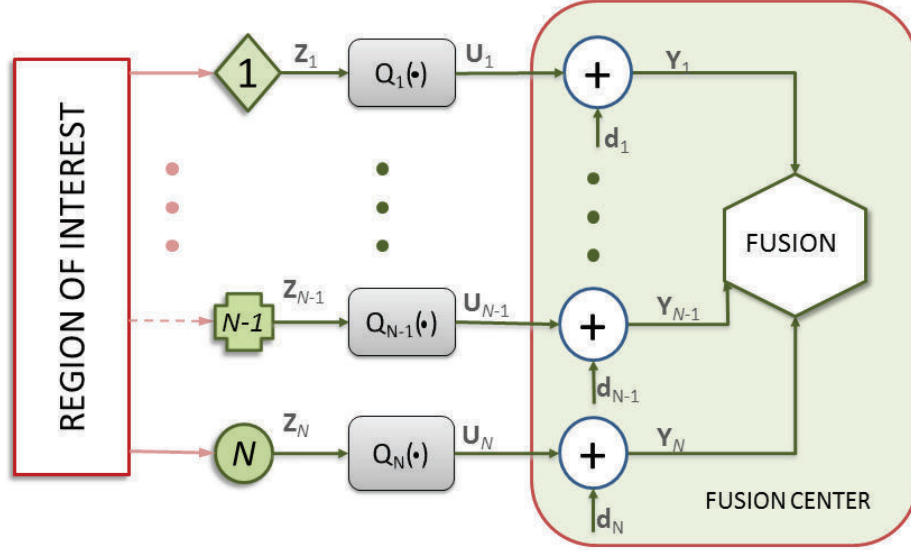


Figure 5.5: A *controlled* noise d_n is added at the output of each sensor n . The approach, although suboptimal, greatly simplifies the fusion rule by avoiding the need to compute multidimensional integrals. The method can be considerably accurate so long as Widrow's quantization Theorem I is satisfied and noise is appropriately designed.

would be one with a rectangular CF in the pass-band, $-\frac{\pi}{q_n} \leq v \leq \frac{\pi}{q_n}$, (also see Fig. 5.6). However, a rectangular function in the CF domain corresponds to a PDF whose shape corresponds to a sinc function, an invalid PDF. Note that this is similar to the non-realizability of an ideal low pass filter in signal processing. One needs to carefully design D_n so that it causes minimal distortion while transforming a discrete-valued random variable, U_n , into a continuous variable, Y_n . As long as the input variable Z_n satisfies Widrow's first quantization theorem (Theorem 5.1) under both H_1 and H_0 , we have,

$$Y_n = Z_n + W_n + D_n. \quad (5.22)$$

Thus, under hypothesis H_1 , the PDF of data, y_{nl} , at time instant l is

$$p_{Y_n}(y_{nl}; \psi_n, H_1) = p_{Z_n}(y_{nl}; H_1) \star p_{W_n}(y_{nl}) \star p_{D_n}(y_{nl}) \quad (5.23)$$

$$= f_n(y_{nl}; \psi_n) \star p_{W_n}(y_{nl}) \star p_{D_n}(y_{nl}). \quad (5.24)$$

Using a copula density (say) $k_1(\cdot; \psi_d) \in \mathcal{A}_k$ to estimate the dependence structure between sensor observations, the joint PDF of the data $\mathbf{y}_l = (y_{1l}, y_{2l})$ can now be approximated as

$$\hat{p}_{\mathbf{Y}}(y_{1l}, y_{2l}; \psi, H_1) = \left\{ \prod_{n=1}^2 p_{Y_n}(y_{nl}; \psi_n, H_1) \right\} k_1 \{F_{Y_1}(y_{1l}; \psi_1), F_{Y_2}(y_{2l}; \psi_2); \psi_d\} \quad (5.25)$$

where

$$F_{Y_n}(y) = \int_{-\infty}^y p_{Y_n}(t; \psi_n, H_1) dt \quad (5.26)$$

denotes the CDF of Y_n under H_1 .

Similarly, the joint PDF of the data under H_0 can be approximated as

$$\hat{p}_{\mathbf{Y}}(y_{1l}, y_{2l}; \lambda, H_0) = \left\{ \prod_{n=1}^2 p_{Y_n}(y_{nl}; \lambda_n, H_0) \right\} k_0 \{G_{Y_1}(y_{1l}; \lambda_1), G_{Y_2}(y_{2l}; \lambda_2); \lambda_d\}, \quad (5.27)$$

where

$$p_{Y_n}(y_{nl}; \lambda_n, H_0) = p_{Z_n}(y_{nl}; H_0) \star p_{W_n}(y_{nl}) \star p_{D_n}(y_{nl}) \quad (5.28)$$

$$= g_n(y_{nl}; \lambda_n) \star p_{W_n}(y_{nl}) \star p_{D_n}(y_{nl}), \quad (5.29)$$

$k_0(\cdot; \lambda_d) \in \mathcal{A}_k$ is the copula density used to estimate the dependence structure of sensor observations under H_0 , and

$$G_{Y_n}(y) = \int_{-\infty}^y p_{Y_n}(t; \lambda_n, H_0) dt \quad (5.30)$$

denotes the CDF of Y_n when the underlying hypothesis is H_0 .

With Eqs. (5.25) and (5.27), it is now straightforward to derive the mGLRT

$$\mathcal{T}_k^{pqn}(\mathbf{y}) = \log \frac{\max_{k_1(\cdot) \in \mathcal{A}_k, \Psi} \prod_{l=1}^L p_{\mathbf{Y}}(y_{1l}, y_{2l}; \psi, H_1)}{\max_{k_0(\cdot) \in \mathcal{A}_k, \Lambda} \prod_{l=1}^L p_{\mathbf{Y}}(y_{1l}, y_{2l}; \lambda, H_0)} \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (5.31)$$

The test derived above involves continuous-valued variables and thus does not involve computation of multidimensional integrals. This greatly simplifies the test. The reduced complexity is, however, at the expense of decreased signal-to-noise ratio due to the injection of noise d_n at the fusion center. The addition of external noise facilitates *filtering out the baseband* $\phi_{Z_n+W_n}(v)$ from the received quantized observations $\phi_{U_n}(v)$. This noise should be designed so that it destroys as little information as possible while filtering the required signal.

When the conditions of Theorem 4.1 hold, it is clear that $2\mathcal{T}_k^{pqn}(\mathbf{y}; H_0)$ converges asymptotically ($L \rightarrow \infty$) to a chi-squared distribution with $\nu = a - b$ degrees of freedom. Thus, the probability of false alarm, P_F^Y , is

$$\begin{aligned} P_F^Y &= Pr(2\mathcal{T}_k^{pqn}(\cdot) > 2\eta; H_0) \\ &\stackrel{L \rightarrow \infty}{\underset{=}{\approx}} Q_{\chi_{a-b}^2}(2\eta). \end{aligned} \quad (5.32)$$

One can thus obtain the threshold η so that P_F^Y is constrained to the desired level $\alpha \in (0, 1)$ as in Eq. (5.12). Note that the distribution of $2\mathcal{T}_k^{pqn}(\mathbf{y}; H_0)$ depends only on the dimensions of the parameter spaces under H_1 and H_0 . Thus, the false alarm probability (corresponding to a test threshold η) does not change due to quantization of the local measurements. However, quantization does cause a loss of detection power as we show in the following section.

5.3.3 Loss in the Detection Power due to Quantization: PQN Model-Based Analysis

Given that, (a) the signal is weak, i.e., $p_{\mathbf{Y}}(y_1, y_2; H_1) \approx p_{\mathbf{Y}}(y_1, y_2; H_0)$, (b) the MLE attains its asymptotic PDF, and (c) the composite hypothesis test can be recast as a test of parameters, the PDF of $2\mathcal{T}^{pqn}(\mathbf{y}; H_1)$ converges asymptotically to a noncentral chisquared distribution [61, p. 205]. For example, when the PDF under H_0 is completely known ($b = 0$) and is nested in $p_{\mathbf{Y}}(y_1, y_2; H_1)$, and one of the copula densities in the set \mathcal{A}_k corresponds to the true dependence structure under H_1 , we have

$$2\mathcal{T}^{pqn}(\mathbf{y}; H_1) \rightsquigarrow \chi_a^2(\xi_Y), \quad (5.33)$$

where ‘ \rightsquigarrow ’ in the above equation denotes the convergence in distribution. The non-centrality parameter ξ_Y is

$$\xi_Y = (\psi - \lambda)^T \mathcal{F}_Y(\lambda; \mathbf{q}) (\psi - \lambda), \quad (5.34)$$

where we explicitly note the dependence of the Fisher information matrix, $\mathcal{F}_Y(\cdot; \mathbf{q})$, on the quantization steps $\mathbf{q} = (q_1, q_2, \dots, q_N)$. Also, ψ in Eq. (5.34), denotes the true value of the parameters under H_1 .

From Eqs. (5.33) and (5.34),

$$P_D^Y = Q_{\frac{\alpha}{2}} \left(\sqrt{\xi_Y}, \sqrt{\eta} \right), \quad (5.35)$$

where P_D^Y denotes the detection power of the PQN model-based test, and $Q_M(\cdot)$ is the Marcum-Q function. The detector threshold η in Eq. (5.35) is set to constrain P_F^Y to a desired value $\alpha \in (0, 1)$ using Eq. (5.12) with b set to zero. Note that the detection probability is directly proportional to the non-centrality parameter, and hence the Fisher information evaluated at λ , the parameter under H_0 . Using the

Fisher information inequality [31, Theorem 13], one can show that

$$\mathcal{F}_Y(\lambda) \leq \mathcal{F}_Z(\lambda), \quad (5.36)$$

where $\mathcal{F}_Z(\lambda)$ is the Fisher information corresponding to the case when observations are transmitted to the fusion center without quantization. Thus, the detection probability, P_D , for the analog transmission case is higher than P_D^Y by the quantity P_D^{loss} , where

$$\begin{aligned} P_D^{loss}(\mathbf{q}) &= P_D - P_D^Y(\mathbf{q}) \\ &= Q_{\frac{\alpha}{2}}\left(\sqrt{\xi}, \sqrt{\eta}\right) - Q_{\frac{\alpha}{2}}\left(\sqrt{\xi_Y(\mathbf{q})}, \sqrt{\eta}\right). \end{aligned} \quad (5.37)$$

where

$$\xi = (\psi - \lambda)^T \mathcal{F}_Z(\lambda) (\psi - \lambda). \quad (5.38)$$

Strictly speaking, $P_D^{loss}(\mathbf{q})$ derived above, is an upper bound on the loss in the detection rate due to quantization. This is because the quantity also includes the effects of external noise.

Next, we present a numerical illustration.

5.4 An Illustrative Example

We consider the following hypothesis testing problem:

$$\begin{aligned} H_0 &: Z_1, Z_2 \sim \mathcal{N}(0, 0, 10, 10, 0) \\ H_1 &: Z_1, Z_2 \sim \mathcal{N}(1, 1, 10, 10, 0.2) \end{aligned} \quad (5.39)$$

where,

$$\begin{aligned} \mathcal{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) &:= \frac{1}{2\pi(1-\rho^2)\sigma_1\sigma_2} \\ &\cdot \exp\left(\frac{-1}{2(1-\rho^2)} \left[\left(\frac{z_1 - \mu_1}{\sigma_1}\right)^2 - \right. \right. \\ &\left. \left. 2\rho \left(\frac{z_1 - \mu_1}{\sigma_1}\right) \left(\frac{z_1 - \mu_1}{\sigma_1}\right) + \left(\frac{z_2 - \mu_2}{\sigma_2}\right)^2 \right] \right), \end{aligned} \quad (5.40)$$

is the usual bivariate Gaussian density function. We assume, without loss of generality, that the PDF under H_0 is completely known and so are the variance parameters σ_1^2 and σ_2^2 under both H_0 and H_1 . The observations, $\{z_{1l}, z_{2l}\}_{l=1}^L$, at the two local sensors are passed through uniform scalar quantizers before their transmission to the fusion center. Thus, the fusion center has access only to the quantized measurements, $\mathbf{u} = \{u_{1l}, u_{2l}\}_{l=1}^L$, to make a global decision in favor of one of the two hypotheses. The GLRT based fusion rule for this problem is the same as the one derived in Eq.(5.10), with the joint probabilities,

$$\begin{aligned} \hat{P}_{i_1, i_2}^* (\mu_1, \mu_2, \rho) &= C_\rho^\mathcal{N} \{ \Phi_{\mu_1, 10}^\mathcal{N} (i_1 q_1), \Phi_{\mu_2, 10}^\mathcal{N} (i_2 q_2) \} - \\ &C_\rho^\mathcal{N} \{ \Phi_{\mu_1, 10}^\mathcal{N} ((i_1 - 1)q_1), \Phi_{\mu_2, 10}^\mathcal{N} ((i_2 - 1)q_2) \} \end{aligned} \quad (5.41)$$

and

$$\hat{Q}_{i_1, i_2}^* = \Phi_{0, 10}^\mathcal{N} (i_1 q_1) \cdot \Phi_{0, 10}^\mathcal{N} (i_2 q_2) - \Phi_{0, 10}^\mathcal{N} ((i_1 - 1)q_1) \cdot \Phi_{0, 10}^\mathcal{N} ((i_2 - 1)q_2), \quad (5.42)$$

where $C_\rho^\mathcal{N}(\cdot)$ is a Gaussian copula with correlation coefficient ρ , and $\Phi_{\mu, \sigma^2}^\mathcal{N}$ is a univariate Gaussian CDF with mean μ and variance σ^2 .

An alternative suboptimal but computationally efficient test was derived in Section 5.3.2 based on Widrow's PQN model for quantization. We evaluate its performance using the Gauss-Gauss example presented here. It is important to point out that

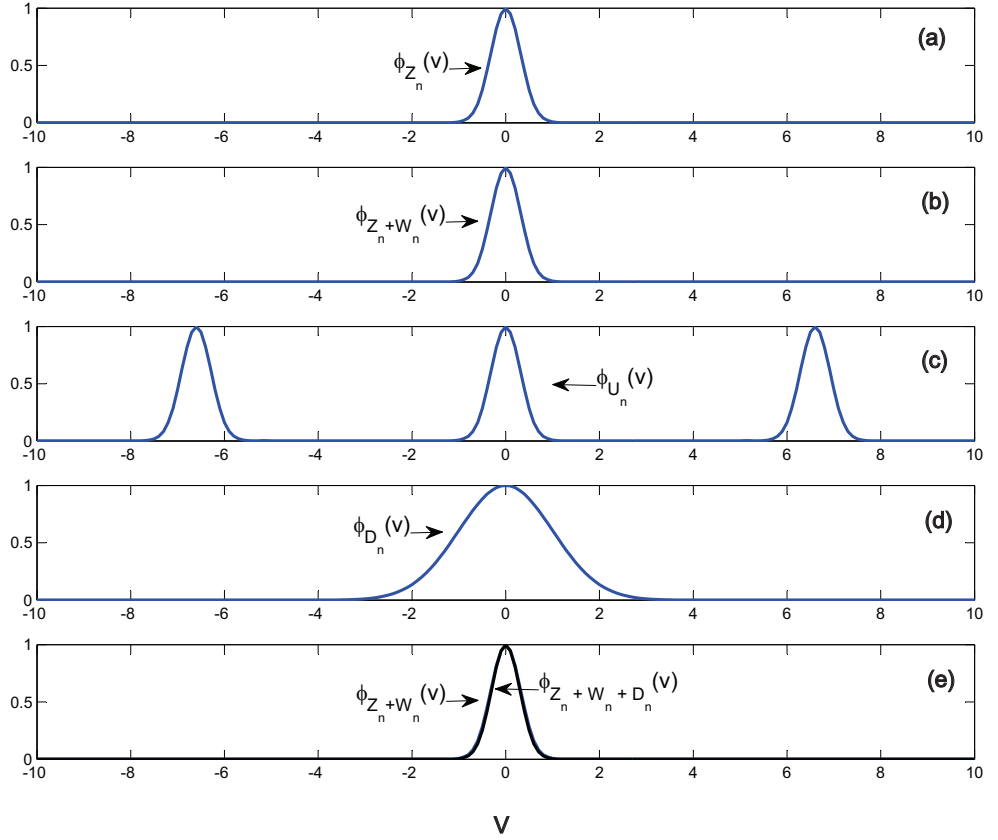


Figure 5.6: 'Filtering' the quantized signal with LPF-noise. The quantization step size, q_1 , is set to $0.3\sigma_1$: (a) CF of z_1 ; (b) CF of $z_1 + w_1$; (c) CF of u_1 ; (d) CF of the external LPF-noise, d_n ; (e) CF of $y_n = z_n + w_n + d_n$.

although the Gaussian CF is not perfectly bandlimited, a property necessary for applying of the PQN based fusion rule, they are very close to being bandlimited for all practical purposes.² Fig. 5.6 shows the PQN processing in the CF domain with the quantization step size set to one-tenth of the input standard deviation, i.e., $q_1 = 0.3\sigma_1$. The CF of the input z_1 is shown in Fig. 5.6(a). Addition of the quantization noise w_1 is equivalent to multiplication of $\phi_{Z_1}(v)$ (shown in Fig. 5.6(a)) with a sinc function, $\text{sinc}\left(\frac{q_1 v}{2}\right)$. The resultant CF, $\phi_{Z_1+W_1}(v)$, is shown in Fig. 5.6(b). This CF is repeated

²As discussed in Section 5.3.2 the PQN model based fusion rule assumes that the input signals, $\{z_n\}_{n=1}^N$, satisfy Widrow's first quantization theorem (Theorem 5.1).

and summed in Fig. 5.6(c) which represents the CF of the quantized signal, u_1 (see Eq. (5.17)). The CF of the LPF-noise, D_1 , a standard Gaussian distributed variable in this example, is shown in Fig. 5.6(d). It is clear that multiplication of $\phi_{U_1}(v)$ with $\phi_{D_1}(v)$ which is equivalent to addition of random variables, d_1 and $z_1 + w_1$, filters the signal so that only the main lobe ($v = 0$) of the input signal is retained (Fig. 5.6(e)). Since the LPF-noise is different from an ideal one with a rectangular CF, the signal, $z_n + w_n$, undergoes some distortion while being ‘filtered’. However, this distortion is almost imperceptible as evident from Fig. 5.6(e).

The PDF of the transformed variable, $Y_n = Z_n + W_n + D_n$, under the hypothesis H_1 is given by

$$p_{Y_n}(y_n; \mu_n, H_1) = p_{Z_n+D_n}(y_n) \star p_{W_n}(y_n) \quad (5.43)$$

$$= \mathcal{N}(\mu_n, \sigma_n^2 + \sigma_{d_n}^2) \star \mathcal{U}\left(-\frac{q_n}{2}, \frac{q_n}{2}\right) \quad (5.44)$$

$$= \frac{1}{q_n} \left[\Phi_{\mu,11}^{\mathcal{N}}\left(y_n + \frac{q_n}{2}\right) - \Phi_{\mu,11}^{\mathcal{N}}\left(y_n - \frac{q_n}{2}\right) \right] \quad (5.45)$$

Similarly, under H_0 , we have

$$p_{Y_n}(y_n, H_0) = p_{Z_n+D_n}(y_n) \star p_{W_n}(y_n) \quad (5.46)$$

$$= \mathcal{N}(0, \sigma_n^2 + \sigma_{d_n}^2) \star \mathcal{U}\left(-\frac{q_n}{2}, \frac{q_n}{2}\right) \quad (5.47)$$

$$= \frac{1}{q_n} \left[\Phi_{0,11}^{\mathcal{N}}\left(y_n + \frac{q_n}{2}\right) - \Phi_{0,11}^{\mathcal{N}}\left(y_n - \frac{q_n}{2}\right) \right] \quad (5.48)$$

Having derived the marginal PDFs above, we now approximate the joint PDFs under both hypotheses using copula functions, and subsequently derive the PQN model based fusion rule using Eq. (5.31). For the sake of simplicity, we use the Gaussian and product copula densities to approximate the joint distributions under H_1 and H_0 respectively. Note that the transformed observations under H_0 are truly independent, and thus the conditions of Theorem 4.1 hold. That is, the distribution

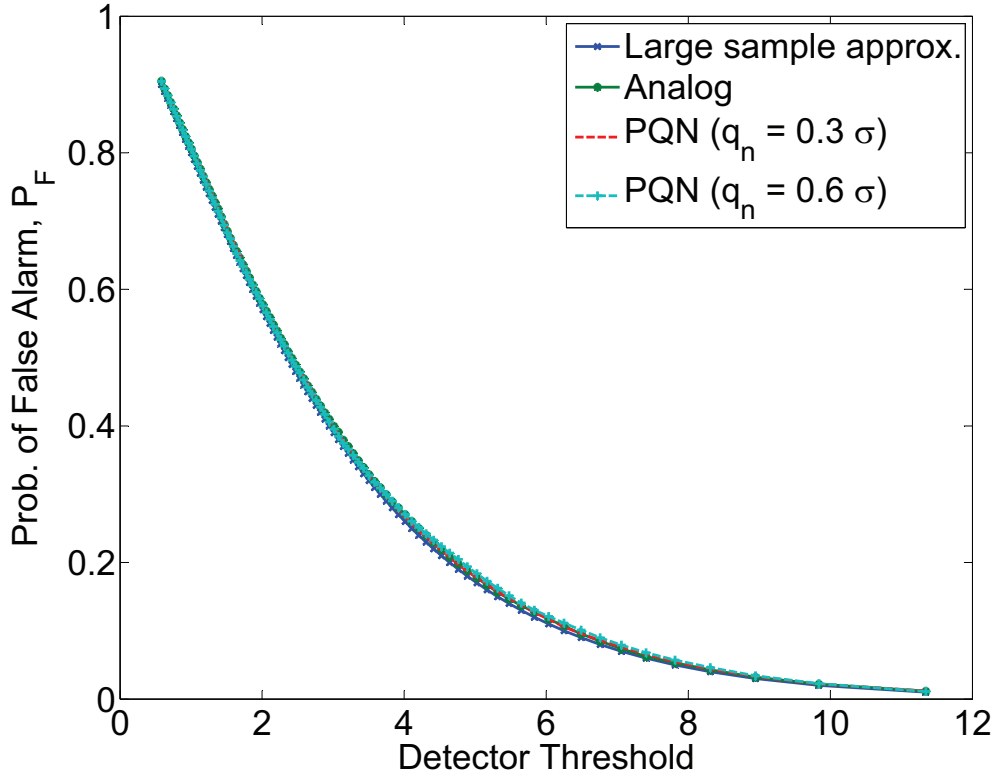


Figure 5.7: System probability of false alarm vs. Detector threshold. The figure shows that P_F does not vary across different quantization levels, 0.3σ and 0.6σ .

of the PQN model based test statistic converges asymptotically to a chi-squared distribution with 3 degrees of freedom for the example considered here. Thus, $P_F = Q_{\chi_3^2}(2\eta)$. Fig. 5.7 shows a plot of this theoretical P_F along with simulated false alarm values obtained using 10,000 Monte-carlo trials with L set to 30 samples. A good match between the theoretical and simulation P_F values across the two different quantization step sizes is evident from the figure. The ROC curves are shown in Fig. 5.8. It is evident from the figure that the performance of the PQN based fusion rule is very close to the analog transmission case which serves as an upper bound on detection performance. It can thus be concluded that the PQN model based fusion rule provides a good approximation to the copula based decision fusion rule derived in Section 5.2 while also being computationally efficient.

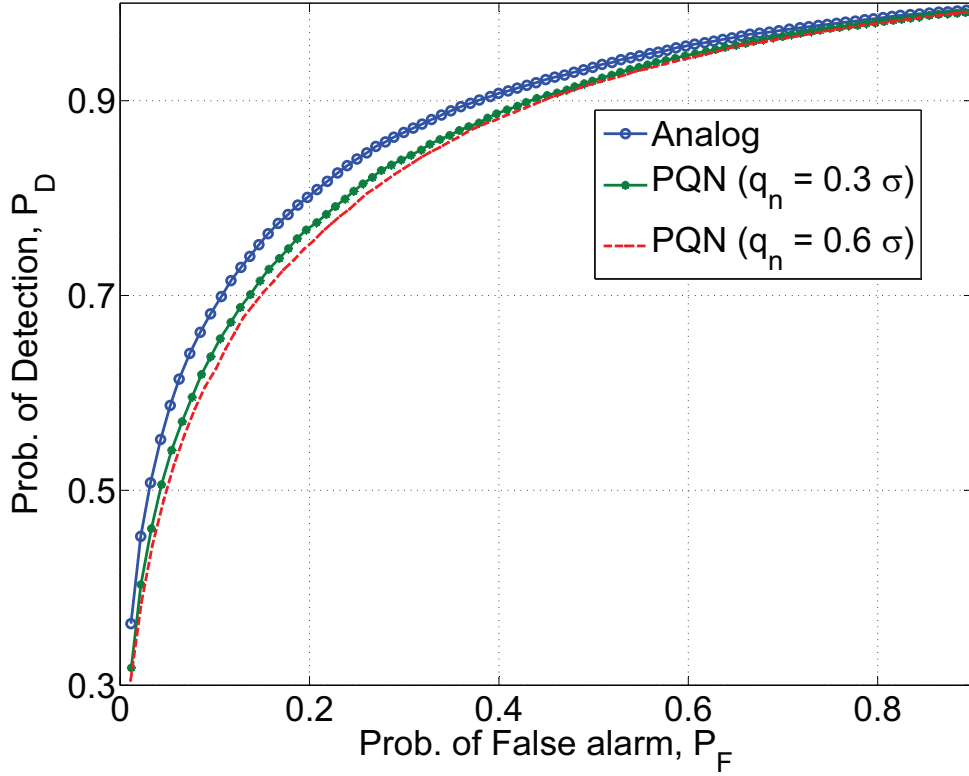


Figure 5.8: Monte Carlo based Receiver Operating Characteristics: Performance of the PQN model based fusion rule is very close to the analog transmission case. The PQN model based fusion rule is, thus, a good and computationally efficient approximation to the exact fusion rule derived in Section 5.2.

5.5 Summary

In this chapter, the problem of fusing correlated sensor decisions for the detection of a random event was considered. Sensor observations (or features extracted thereof) were first quantized using uniform multilevel quantizers before their transmission to the fusion center. Inter-modal dependence was assumed to be unknown and was approximated using copula functions. A GLRT based decision fusion algorithm that can fuse both hard and soft local decisions was derived. The important problem of selecting the best copula was embedded in the GLRT formulation. It was noted that the derived copula-based fusion algorithm becomes computationally expensive as the

number of sensors and/or number of quantization levels increases. A novel approach based on Widrow's additive quantization noise model was developed which requires deliberate injection of an external noise at the receiver before fusion. The addition of external noise at the fusion center effectively 'filters' the baseband CFs by rejecting the repetitive CF lobes that arise due to quantization. Since this process is analogous to low pass filtering (LPF) in signal processing, we term this noise, the LPF-noise.

As an illustrative example, a Gauss-Gauss detection problem was presented. Gaussian noise sources were used to generate LPF-noise at the fusion center, and results for two different quantization step sizes were obtained. Our results show that the PQN model based fusion can be considerably accurate provided the CF of the input signals are bandlimited and Widrow's first quantization theorem is satisfied. The key to the success of the PQN model based fusion is the choice of the external noise source used for filtering the baseband CF. Design of a noise source that introduces minimal distortion while filtering is a topic of future research.

Chapter 6

Quantification of Neural Synchrony Using Copulas

A parametric copula-based framework for general hypothesis testing problems was developed in the previous chapters. Our focus, in these chapters, was to exploit the spatial dependence structure between the heterogeneous sensor observations while assuming temporal independence. In this chapter, we relax the assumption of statistical independence in time. As a specific application, we study the phenomenon of synchronization between multiple simultaneously recorded electroencephalographic (EEG) signals. Synchrony between the multiple EEG channels is known to be a key feature suggestive of many neurophysiological diseases. For example, a decrease in synchrony among the different channels is indicative of Alzheimer's disease (AD), while an enhanced synchronization is known to be related to epileptic seizures [26,80]. In this chapter, we consider the problem of detecting the onset of AD.

AD is the most common form of dementia, which causes problems with memory, thinking and behavior. It is a degenerative disease that slowly worsens over time. It is the sixth leading cause of death in the United States as reported by Alzheimer's Association [1]. Deaths due to AD have increased by as much as 47.1% between

the years 2000-2006. While there are no current treatments that can entirely stop Alzheimers from progressing, medications are available that could slow the worsening of dementia thus improving the quality of life for those with the disease and also their caregivers. Medication for AD is most effective when applied at an early stage. It is, therefore, important to develop methods for early diagnosis of AD.

Many studies have reported that AD causes a decreased inter-channel dependence when compared to their age-matched control subjects. EEG signals from patients suffering from AD exhibit reduced coherence (See [55, 102] for an in-depth review). Several measures have been proposed for EEG synchrony quantification ([14, 26, 60, 78, 93, 102] provide recent reviews on EEG synchrony measures). However, reliable detection of AD remains to be a difficult problem especially for patients in the pre-symptomatic phase (also known as Mild Cognitive Impairment (MCI)). This poses severe limitations in the diagnosis of early stage AD. While some measures such as correlation coefficient, corr-entropy, coefficient, coh-entropy and mutual information are bivariate, and thus applicable to only a pair of channels, measures such as Granger causality, although multivariate, fail to account for nonlinear interdependencies. Thus, the problem of quantifying dependence or synchrony between multiple time series remains difficult especially for biomedical signals such as EEG that are known to exhibit nonlinear and non-Gaussian statistics.

In this chapter, a novel copula based approach is proposed based on copula theory for quantifying synchrony between multiple EEG channels. As discussed in Chapter 2, copula functions are good candidates for modeling complex statistical dependencies and play an important role in characterizing multivariate statistical dependence. The copula derived synchrony measures are then used as *features* to distinguish MCI patients from control subjects. Following the same approach as in [26], we evaluate the classification performance using the leave-one-out cross validation procedure using

the dataset collected at the RIKEN Brain Science Institute, Japan. This is the same dataset that was used in [26].

The rest of the chapter is structured as follows. In Section 6.1, we provide a brief review of two multivariate synchrony measures, the Granger causality and stochastic event synchrony (SES), that are used in conjunction with the copula derived features for classification. Application of copula theory for neural synchrony quantification is discussed in Section 6.2. Two commonly used copula models, the multivariate Gaussian and the Student's t copula functions are analyzed. In Section 6.3, the synchrony measures are applied to the EEG dataset for the purpose of diagnosing MCI; we describe the dataset, and present the classification results. The chapter is summarized in Section 6.4.

6.1 Synchrony Measures: A Review

As discussed above, a considerable amount of research has been recently devoted to detecting fluctuations in EEG synchrony, and a number of measures have been proposed for this purpose. A detailed study of these measures for early diagnosis of AD was recently undertaken by Dauwels et al. [26]. The authors report that, out of the several dependence measures such as, (a) correlation coefficient and its analogues in frequency and time-frequency domain, (b) several Granger causality measures, (c) phase synchrony, (d) state space based synchrony, (e) information theoretic interdependence measures, and (f) stochastic event synchrony (SES), only two measures, the Granger causality (specifically, the full frequency directed transfer function discussed below) and SES, showed statistically significant differences between the MCI and the control subjects. Using discriminant analysis (DA) with Granger causality and SES measures as features, the authors obtained a classification accuracy of about 83% through the leave-one-out cross validation procedure. Our goal, in this chapter, is

to see if the classification accuracy can be improved further by augmenting the two measures with the copula derived synchrony measures.

We now briefly review the Granger causality and SES measures.

6.1.1 Granger Causality

Granger causality refers to a family of synchrony measures that are derived from linear stochastic models of time series. Let $z_1(l), z_2(l), \dots, z_N(l)$ denote measurements from N different channels where l is the time index. Assume, without loss of generality, that the signals have zero mean and unit variance, and consider the following multivariate autoregressive (MVAR) model,

$$\bar{z}(l) = \sum_{m=1}^p \mathbf{A}(j) \bar{z}(l-m) + \bar{e}(l), \quad (6.1)$$

where $\bar{z}(l) = (z_1(l), \dots, z_N(l))^T$, p is the model order, the model coefficients $\mathbf{A}(j)$ are $N \times N$ matrices and $e(l)$ is a zero-mean Gaussian random vector of size N . Thus, each signal $z_n(l)$ is assumed to linearly depend on its own ‘ p ’ past values, and the ‘ p ’ past values of the other signals, $z_j(l)$, $j = 1, \dots, N$ and $j \neq n$. The above model, Eq. (6.1), can also be cast as

$$\bar{e}(l) = \sum_{m=0}^p \tilde{\mathbf{A}}(j) \bar{z}(l-m), \quad (6.2)$$

where $\tilde{\mathbf{A}}(0) = \mathbf{I}$ (identity matrix) and $\tilde{\mathbf{A}}(j) = -\mathbf{A}(j)$ for $j > 0$. Transforming Eq. (6.2) into the frequency domain, one obtains

$$Z(f) = \tilde{\mathbf{A}}^{-1}(f) E(f) = \mathbf{H}(f) E(f), \quad (6.3)$$

where ‘f’ is the frequency variable. The power spectrum matrix of the signal $\bar{z}(l)$ is then

$$\mathbf{S}(\mathbf{f}) = \mathbb{E} [Z(\mathbf{f})Z^\dagger(\mathbf{f})] = \mathbf{H}(\mathbf{f})\mathbf{V}\mathbf{H}^\dagger(\mathbf{f}), \quad (6.4)$$

where \mathbf{V} is the covariance matrix of $\bar{z}(l)$ and $Z^\dagger(\mathbf{f})$ is the hermitian conjugate of $Z(\mathbf{f})$. Granger causality measures such as partial coherence (PC), directed transfer function (DTF), full frequency directed transfer function (ffDTF), partial directed coherence (PDC) and others can be defined in terms of the matrices, \mathbf{A} , \mathbf{H} and \mathbf{S} . In this work, we have used the ffDTF, $F_{ij}^2(\mathbf{f})$, where

$$F_{ij}^2(\mathbf{f}) = \frac{|H_{ij}(\mathbf{f})|^2}{\sum_{\mathbf{f}} \sum_{k=1}^N |H_{ij}(\mathbf{f})|^2} \in [0, 1] \quad (6.5)$$

quantifies the fraction of inflow to channel i stemming from channel j .

6.1.2 Stochastic Event Synchrony

Stochastic event synchrony refers to a recently developed family of measures that quantifies similarity between point processes [27–29]. Time-frequency transforms of each signal are first approximated as a sum of half-ellipsoid basis functions referred to as “bumps”. For example, Fig. 6.1 shows the bump model corresponding to an EEG signal. Each bump may be considered as an event on the time-frequency plane, and the resulting bump model may be considered as a two-dimensional point process (event sequence). Only those “bumps” with energy larger than a preset threshold are retained in the bump models of each signal, which are then aligned to quantify the coincidence between the signals. For example, Fig. 6.2 shows a schematic for two bump models, (say) E and E' . The more the overlap between the bump models, ‘stronger’ is the dependence between the two signals.

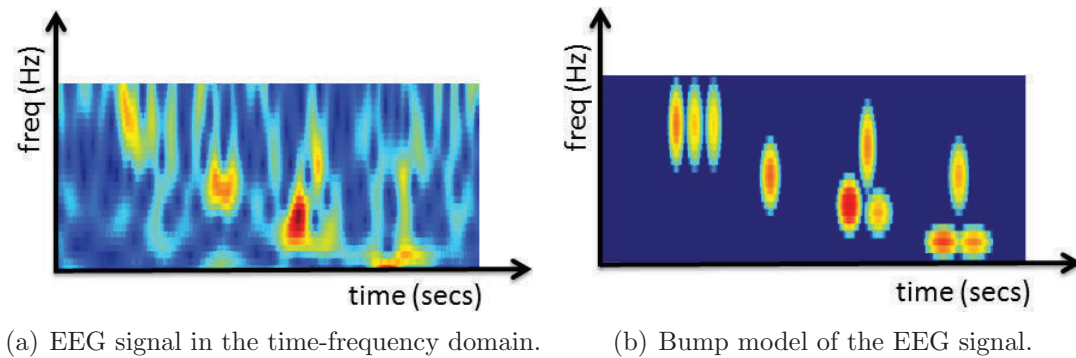


Figure 6.1: Bump modeling

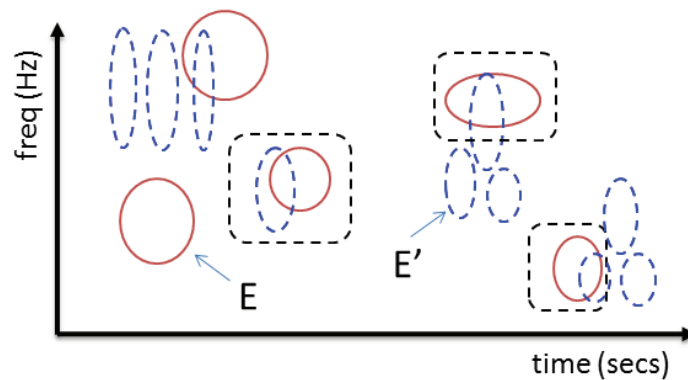


Figure 6.2: Bump matching of two event sequences, E and E' . The pairs in the dashed square box show coincident bumps.

The alignment of the bump models is posed as a statistical inference problem. The goal is to estimate the parameter $\theta = (\rho_{SES}, \delta_t, \delta_f, s_t, s_f)$, where

- ρ_{SES} : fraction of non-coincident bumps
- δ_t and δ_f : average time and frequency offsets between coincident bumps
- s_t and s_f : variance of the time and frequency offsets.

Details of the underlying generative model $p(E, E', \theta)$, and an iterative algorithm that results in the maximum a posteriori (MAP) estimate, $\hat{\theta}$, can be found in [28, 29].

6.2 Copula based Synchrony Measure

In the context of time series modeling, copulas have been utilized to characterize both, (a) the temporal dependence, *i.e.*, by considering the conditional PDF $f(z_n(l)|z_n(l-1), \dots, z_n(l-p))$ (p is the model order), and (b) the multivariate dependence across multiple time series $\{z_1(l), z_2(l), \dots, z_N(l)\}$ conditioned on some past information set $\mathcal{F}_{(l-1)}$ [21, 77]. They have been used intensively in modeling financial time series data (*e.g.*, [42, 83]).

In this paper, we consider the application of copulas to measure EEG synchrony. Specifically, we use copulas to estimate the d -dimensional joint PDF $f(\{z_1(l)\}_{(l-p)}^l, \dots, \{z_N(l)\}_{(l-p)}^l)$ where $d = N \cdot (p+1)$. Now, as discussed in Chapter 2, Sklar's theorem proves the existence of a copula density $c(\cdot)$ for any multivariate PDF with continuous marginals so that it can be expressed as in Eq. (2.6). However, identifying the true underlying copula density is difficult in most practical scenarios. We, therefore, select a copula density (say $k(\cdot)$) *a priori* and fit the desired marginals (if known). In this chapter, we explore the use of the Gaussian and Student's t copula functions.

The d -dimensional Gaussian copula density is given as

$$k_g(u_1, \dots, u_d) = \frac{\phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))}{\phi(\Phi^{-1}(u_1)) \dots \phi(\Phi^{-1}(u_d))}, \quad (6.6)$$

where the index R refers to the correlation matrix, and $u_i = F_i(z_i) \sim \mathcal{U}(0, 1)$ (uniform distribution). Φ and ϕ are standard Gaussian CDF and PDF respectively. Like other copulas, it provides the flexibility of having disparate marginals.

The t -copula density is given as,

$$k_t(u_1, \dots, u_d) = \frac{f_{\nu, R}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))}{\prod_{i=1}^d f_\nu(t_\nu^{-1}(u_i))}, \quad (6.7)$$

where $f_{\nu,R}(\cdot)$ and $t_{\nu}(\cdot)$ are the standard t -joint density and CDF with ν degrees of freedom respectively. Thus, in addition to the correlation matrix R , the t -copula is also parameterized by ν - the degrees of freedom in Eq. (6.7), which is directly related to the so called tail dependence or the joint probability of extreme events [30]. For example, in the case of a bivariate random vector (X_1, X_2) , the tail dependence measures the nature of $c(F_1(x_1), F_2(x_2))$ in the upper-right and lower-left quadrants of \mathbb{I}^2 - the unit square [70]. It is quantified by the upper and lower tail dependence coefficients λ_u and λ_l defined as

$$\lambda_u = \lim_{q \rightarrow 1} P(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q)) \quad (6.8)$$

$$\lambda_l = \lim_{q \rightarrow 0} P(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q)) \quad (6.9)$$

provided the limits exist.

For the t -copula density, the upper and lower tail dependence coefficients are equal and can be derived as [70],

$$\begin{aligned} \lambda_u = \lambda_l &= 2t_{\nu+1} \left(-\sqrt{(\nu+1)} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) \\ &= \lambda. \end{aligned} \quad (6.10)$$

Thus, ν is inversely proportional to the tail dependence; the number of degrees of freedom ν decreases with increasing tail dependency λ . Further, as $\nu \rightarrow \infty$, the t -copula approaches the Gaussian copula.

Now, given the measurements, we use the canonical maximum likelihood (CML) approach to infer the parameters of the copula density (ν for the t -copula and R for both the t and Gaussian copula densities); the approach allows us to estimate the copula parameters without any assumptions on the parametric forms of the marginal distributions (see Section 2.2.2). The entries of the correlation matrix R quantify the coupling between different signals and can be used to define synchrony measures.

Moreover, we suggest the use of multi-information [94] as a global synchrony measure to summarize the dependence structure in a single number.

Multi-information ‘ $\mathcal{I}(\cdot)$ ’ between d random variables X_1, \dots, X_d is given as

$$\begin{aligned} \mathcal{I}(X_1; \dots; X_d) &= \int f(x_1, \dots, x_d) \log \frac{f(x_1, \dots, x_d)}{\prod_{i=1}^d f(x_i)} dx_1 \dots dx_d \\ &= \int f(x_1, \dots, x_d) \log \frac{f_p(\mathbf{x}) c(F_1(x_1), \dots, F_d(x_d))}{f_p(\mathbf{x})} dx_1 \dots dx_d \end{aligned} \quad (6.11)$$

$$= \mathbb{E}_f \log c(F_1(x_1), \dots, F_d(x_d)), \quad (6.12)$$

where $\mathcal{I}(\cdot) \geq 0$. It is zero when variables are statistically independent. However, note that it is impossible to compute Eq. (6.12) since the true joint PDF $f(\cdot)$ and hence the copula $c(\cdot)$ is unknown. We approximate the integral over $f(\cdot)$ in Eqs. (6.11) and (6.12) by a sample expectation, and replace the true copula $c(\cdot)$ by a trial copula $k(\cdot)$ chosen *a priori*, e.g., Gaussian or t-copula, leading to the following expression:

$$\mathcal{I}_k(\cdot) = \frac{1}{L} \sum_l \log k(\hat{F}_1(x_{1l}), \dots, \hat{F}_d(x_{dl})), \quad (6.13)$$

where

$$\hat{F}_n(a) = \frac{1}{L} \sum_{l=1}^L \mathbb{I}(z_{nl} \leq a). \quad (6.14)$$

$\mathbb{I}(\mathcal{E})$ in (6.14) indicates the occurrence of event \mathcal{E} , and L denotes the sample size.

Thus, for $N = 2$ and the model order $p = 2$, multi-information is estimated as

$$\mathcal{I}_k(\cdot) = \frac{1}{L} \sum_l \log k\left(\hat{F}_1(z_1(l)), \hat{F}_1(z_1(l-1)), \hat{F}_2(z_2(l)), \hat{F}_2(z_2(l-1))\right). \quad (6.15)$$

As discussed in the preceding chapters, application of an arbitrary copula density $k(\cdot)$ ($\neq c(\cdot)$) introduces model mismatch errors, and the performance of a copula based algorithm is thus highly dependent on the choice of $k(\cdot)$. This error can be quantified using the KL divergence, $D(f, f_k)$ as discussed in Chapter 3. As shown in Section 3.2.1, higher the value of $\mathcal{I}_k(\cdot)$, smaller is the mismatch error (in terms of KL divergence).

6.3 Early Diagnosis of Alzheimer’s Disease

In this section, we describe the EEG data in brief and also address the issue of selecting the copula that *best* fits the data.

6.3.1 EEG Data

Silver (Ag/AgCl) electrodes (disks of diameter 8mm) were placed on 21 sites according to 10-20 international system, with the reference electrode on the right ear-lobe. EEG were obtained with Biotop 6R12 (NEC San-ei, Tokyo, Japan) at the rate of 200 samples per second. The acquired data was then band pass filtered (4 Hz - 30 Hz) using a third-order Butterworth filter. A common reference (right ear-lobe) was used for data analysis.

The subjects comprised two study groups. The first was a group with 25 patients diagnosed as suffering from MCI when the EEG recordings were carried out. Later on, they all developed mild AD. The criteria for inclusion into the MCI group were a mini mental state exam (MMSE) score = 24, though the average score in the MCI group was 26 (standard deviation 1.8). The second group known as the control set consisted of 56 age-matched, healthy subjects who had no memory or other cognitive impairments. The average MMSE of this control group was 28.5 (standard deviation 1.6). The ages of the two groups were 71.9 ± 10.2 and 71.7 ± 8.3 , respectively.

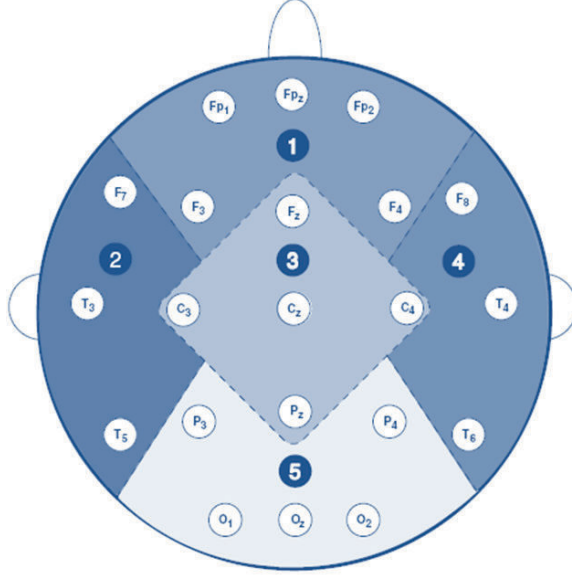


Figure 6.3: Clustering of the 21 EEG channels into 5 zones indicated by the colors and dashed lines, (1) frontal, (2) left temporal, (3) central, (4) right temporal, and (5) occipital

All recording sessions were conducted with the subjects in an awake but resting state with eyes closed. A 5-minute long EEG was recorded. Only those subjects were retained in the analysis whose EEG recordings showed maximum artifact-free data. Based on this requirement, the number of subjects in the two groups described above was further reduced to 22 and 38, respectively. From each subject, one artifact-free EEG segment of 20s was analyzed (for each of the 21 channels).

6.3.2 Computation of EEG Synchrony

In order to reduce the dimensionality of the problem, the 21 EEG channels are clustered into five groups and the signals in each group are averaged, thus obtaining five spatially averaged EEG signals per subject. The five groups correspond to frontal, left temporal, central, right temporal and occipital regions (see Fig.6.3). The copula based synchrony measures (R, ν and $\mathcal{I}_k(\cdot)$) are obtained using the spatially averaged signals (*i.e.*, $N = 5$).

We follow the procedure outlined in [26] for computing the Granger causality (ffDTF) and the SES measures. An average ffDTF measure, \bar{F} , is obtained by first averaging over all (integer) frequencies within the frequency band of 4 to 30 Hz, and then across the 10 pairs of spatially averaged channels. That is,

$$\bar{F} = \sum_{i=1}^5 \sum_{j>i}^5 \left(\frac{1}{27} \sum_{f=4}^{30} F_{ij}(f) \right) \quad (6.16)$$

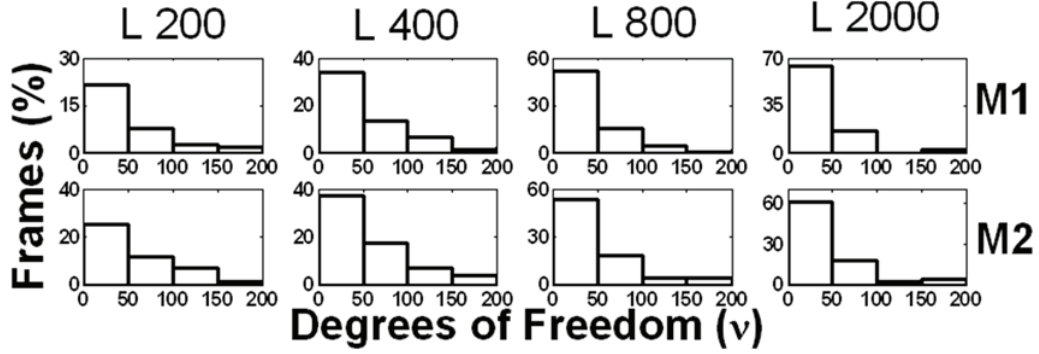
In order to obtain the SES measure, bump models for all the 21 EEG channels are first obtained. A bump model for each of the five regions are then obtained by means of the aggregation algorithm described in [107]. SES parameters for all pairs of regions are then computed, which are finally averaged over all the region pairs. Thus, one set of average SES parameters are obtained per subject.

Further, since spontaneous EEG are known to be highly non-stationary, it may not be meaningful to estimate the copula synchrony and Granger causality measures using the entire duration of EEG signals (20 seconds). Therefore, we divide the EEG signals into small segments or frames, each of length L samples. We consider several frame lengths in our analysis (L = 1 sec, 2 sec, 4 sec, 10 s). The Granger causality and the copula synchrony measures are obtained by averaging over all the frames. Computation of the SES parameter does not require such division of EEG signals into smaller segments and subsequent averaging of the computed values since it is applicable to non-stationary signals [26].

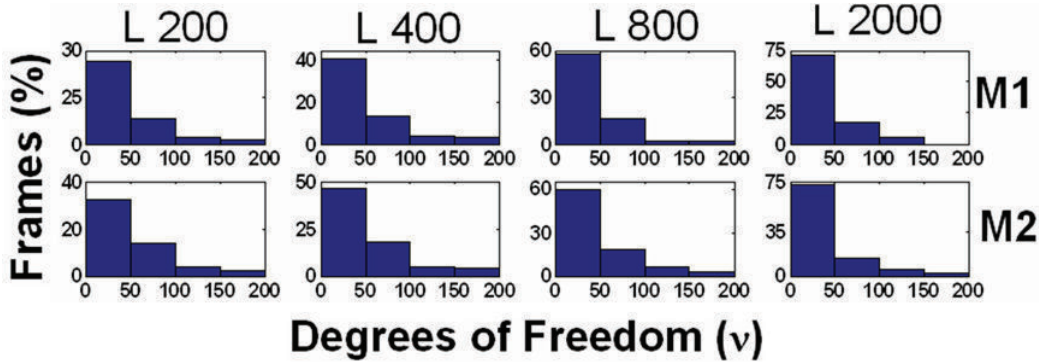
Next, we discuss the important issue of selecting a copula function.

6.3.3 Copula Selection

As discussed earlier, the performance of a copula based algorithm is highly dependent on the choice of $k(\cdot)$ used to model the dependence structure between multiple signals.



(a) Normalized histogram of ν estimated using data from subjects with mild cognitive impairment (MCI)



(b) Normalized histogram of ν estimated using data from age-matched *control* subjects

Figure 6.4: Normalized histogram of estimates of degrees of freedom (ν) computed for different model orders and frame lengths (L)

Our analysis, in this section, indicates that the t-copula is more suitable than the Gaussian copula for EEG synchrony quantification. We provide the details below.

Figure 6.4 shows the normalized histogram of ν , estimated from the spatially averaged EEG data for model orders one and two (M1 and M2) and for different frame lengths (L). Histograms for both, the *control* subjects and the patients with MCI, are shown. (See [18] for extension of the definition and analysis of ν to the multivariate t-copula).

It is evident from Fig. 6.4 that an appreciable number of frames exhibit tail dependence, i.e., they have smaller values of ν (See Eq. (6.10)). For example, about 75 % of the total frames from the control subjects have $\nu < 50$ when model order, p , and the frame length, L , are set to 2 and 2000 respectively (M2 L2000). The

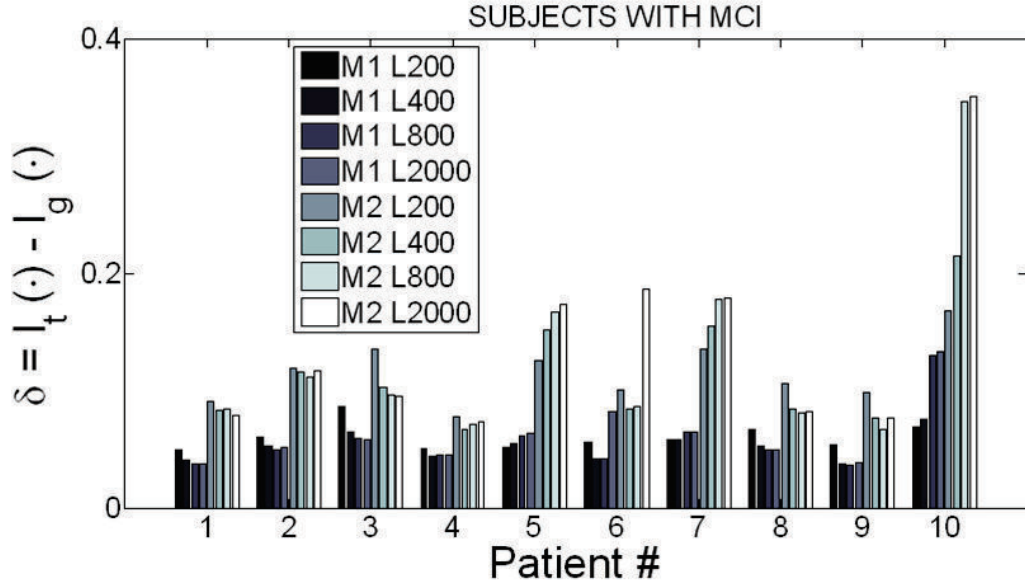


Figure 6.5: Difference of multi-information estimates based on the t and Gaussian copula functions.

Gaussian copula with zero upper and lower tail dependence will be unable to *capture* this characteristic of the EEG data. It is important to note that the t -copula will also be able to characterize the *Gaussian-like* behavior of certain frames with larger values of ν as it *neats* the Gaussian copula (i.e., it approaches the Gaussian copula as $\nu \rightarrow \infty$).

Next, we estimate multi-information for each frame by approximating $\mathbb{E}_f \log k(\cdot)$ by Eq. (6.13). We then average over all frames to obtain estimates of $\mathcal{I}(\cdot)$ using both the Gaussian copula ($\mathcal{I}_g(\cdot)$) and the t -copula ($\mathcal{I}_t(\cdot)$) functions. We plot in Fig. 6.5, the difference, δ , between the two estimates, $\mathcal{I}_t(\cdot)$ and $\mathcal{I}_g(\cdot)$, of multi-information for ten patients. It is clear from Fig. 6.5, that δ is always positive, i.e., $\mathcal{I}_t(\cdot) \geq \mathcal{I}_g(\cdot)$ for any choice of the model order and frame length. Thus, the t -copula provides a better estimate (in the KL divergence sense) of the true joint PDF than the Gaussian copula density (see Section 3.2.1). Though results for only ten patients are presented here, we observed that the non-negativity of δ was true for all patients. In fact the

t-copula was a better fit for every frame as well (*i.e.*, $\delta \geq 0$ for every frame). Similar results were obtained for all patients in the control set.

6.3.4 Classification Results

In the above sections, we discussed the copula approach to jointly model multiple EEG signals. Having learnt the copula parameters from data, we now proceed to the task of distinguishing MCI patients from the age-matched control subjects. In order to compare our results with [26], we use the same classification method - the linear and quadratic discriminant analysis with leave-one-out cross validation as used in [26].

For each frame of length ‘L’ samples, we first compute μ_l , the mean of the absolute values of all entries of the correlation matrix R estimated using CML. Two features, $\bar{\mu}_\rho$ and $\bar{\sigma}_\rho$, are then obtained by computing the average and standard deviation of μ_l over all frames. Estimates for the multi-information $\hat{\mathcal{I}}_h$ and the number of degrees of freedom ν are also obtained in the similar way (*i.e.*, by averaging over all the frames).

Table 6.1 shows the leave-one-out classification rate for different features obtained with optimal parameter settings (model order and frame length). Classification rates obtained using the Gaussian copula are also included. Results in rows 1, 2 and 3 of Table 6.1 have been reported in [26] and correspond to the stochastic event synchrony measure (ρ_{SES}), the full frequency directed transfer function (ffDTF) - a Granger causality measure [60], and the combination of both respectively. It can be seen that using copula measures (both Gaussian and Student’s t) in conjunction with ρ_{SES} and ffDTF improves the classification rate to 85%. Several existing synchrony measures were considered in [26] and the authors observed no improvement beyond 83.3%.

Next, we compare the Gaussian copula and the t-copula models (*e.g.*, classification rates obtained using $(\rho_{SES}, \text{ffDTF}, \mathcal{I}_g(\cdot))$ vs. that using $(\rho_{SES}, \text{ffDTF}, \mathcal{I}_t(\cdot))$). It can

Table 6.1: Classification rates using linear and quadratic discriminant analysis with leave-one-out cross validation

| Features | Linear | Quadratic | Copula Features |
|--|--------|-----------|------------------------------|
| ρ_{SES} | 70.0% | 70.0% | - |
| ffDTF | 68.3% | 75.0% | - |
| ρ_{SES}, ffDTF | 83.3% | 83.3% | - |
| $\rho_{SES}, \text{ffDTF}, h_g(\cdot)$ | 85.0% | 80.0% | $\bar{\sigma}_\rho$ |
| | 85.0% | 81.7% | $\bar{\mu}_\rho$ |
| | 83.3% | 80.0% | $\hat{\mathcal{L}}_g(\cdot)$ |
| $\rho_{SES}, \text{ffDTF}, h_t(\cdot)$ | 85.0% | 81.67% | $\bar{\sigma}_\rho$ |
| | 85.0% | 83.30% | $\bar{\mu}_\rho$ |
| | 83.3% | 81.67% | $\hat{\mathcal{L}}_t(\cdot)$ |
| | 85.0% | 85.0% | ν |

be seen that the performance of the t-copula feature set is never lower than that of the Gaussian copula density. This is probably because the t-copula nests the Gaussian copula as discussed in Section 6.3.3.

6.4 Summary

In this chapter, the problem of quantifying synchrony between multiple EEG channels was considered. Several measures stemming from various fields such as physics, information theory, statistics and signal processing have been proposed in the past for this purpose. While some of these measures are bivariate, the remaining ones, although multivariate, fail to account for nonlinear interdependencies. As we have shown in this chapter, the approach based on copula theory addresses both these limitations. Two commonly used copula models, the multivariate Gaussian and the

Student's t copula functions were analyzed, and the important issue of selecting the copula that *best* fits the data was also discussed. Both these copula functions provide the flexibility of having disparate distributions for the individual EEG channels. We estimate the marginals using the empirical distribution functions, thus avoiding the need for making any assumptions regarding their parametric forms.

Application of the copula derived synchrony measures for distinguishing MCI patients from the age-matched control subjects has also been investigated. Our results show that the copula based features when used in conjunction with other synchrony measures help enhance the detection of AD onset. Our results prove the feasibility of the copula approach. More general multivariate copula functions not limited to the correlation matrix for dependence characterization will be investigated in the future.

Chapter 7

Conclusion

In this dissertation, a general parametric framework for processing signals of disparate and statistically dependent modalities has been developed. Modeling heterogeneous signals is a complex problem. For example, while it is intuitive that human speech is correlated with the lip movements, it is not immediately clear how these signals, that belong to completely different ‘domains’, could be modeled. The lack of adequate models for heterogeneous signals is, in fact, one of the main reasons limiting the development of multimodal signal processing algorithms. The modeling issues and challenges involved have been discussed in detail in Chapter 2, and a solution based on the statistical theory of copulas has been proposed. We show how copula functions possess all the *ingredients* necessary for modeling the joint statistics of heterogeneous data. The important problem of selecting the best copula for binary hypothesis testing and classification problems has been addressed. Copula selection methods for both, (a) the training-testing paradigm, when one has access to a training set so that models could be learned prior to system deployment, and, (b) when models have to be learned in real-time, have been developed. An important advantage of the proposed copula based approach is that no assumptions on the source of heterogeneity are necessary; the same machinery holds for fusion of multiple modalities, samples, algorithms or

multiple classifiers. The biometrics example presented in Chapter 3 is an illustration of multi-algorithm fusion, while the footstep detection example considered in Chapter 4 fuses acoustic and seismic features derived from signals of disparate modalities.

The framework has also been extended to the case when the local heterogeneous measurements are quantized before their transmission to the fusion center. The local quantizers were assumed to be uniform multibit quantizers and a GLRT based rule for fusing correlated soft decisions was derived. A novel suboptimal but computationally efficient fusion rule has also been developed that involves deliberately contaminating the quantized observations with *controlled* noise at the fusion center. Addition of noise at the fusion center helps to *linearize* the highly nonlinear quantization process thus resulting in significant computational savings. It has been demonstrated that the performance of the suboptimal fusion rule is very close to that of the optimal test when the characteristic functions of the input signals is bandlimited.

The ability of copulas to characterize nonlinear dependencies was further exploited to address problems in medical diagnosis. Specifically, a novel copula based method for quantifying neural synchrony has been developed. It has been demonstrated that the copula derived synchrony measures when used in conjunction with Granger causality and stochastic event synchrony helps in the early diagnosis of Alzheimer's disease.

7.1 Future Research Directions

We now present some future research directions.

- **Dynamic Dependence:** The role of copulas in heterogeneous signal processing was investigated in this dissertation. However, it was assumed that the signals were stationary. Application of copulas for dynamic systems where the signal

statistics as well as the spatio-temporal dependence between sensor observations evolves with time needs to be investigated.

- Combination of multiple copula densities: Different copula functions exhibit different behavior and a combination of multiple copula functions may better characterize dependence between several modalities than just using a single copula function. It would be interesting to explore this multi-model approach in detail.

Several interesting extensions are possible on the basis of the results obtained in Chapter 5.

- Non-uniform Quantization: A copula-based test statistic to fuse correlated multibit local sensors' decisions was derived in this chapter assuming that the local sensor observations underwent uniform quantization before their transmission to the fusion center. It would be interesting to extend this formulation to the case when the quantizers are non-uniform. The implications of non-uniform quantization on the design of the PQN model-based suboptimal fusion rule could also be explored.
- Design of LPF-noise: It is necessary to use a noise source that causes minimal distortion while filtering out the repetitive CF lobes. A formal procedure needs to be developed for this purpose.
- Stochastic Resonance and LPF-noise effect: It has been shown that the detection performance of a distributed detection system can be enhanced by the addition of controlled noise at the local sensors. It may be interesting to explore the effect of LPF-noise in conjunction with the SR phenomenon. *Could this result in performance enhancement in addition to reducing the system computational complexity?* This question needs to be investigated.

Our results in Chapter 6 proved the feasibility of the copula approach for quantifying synchrony between multiple biomedical signals. However, our study was limited to the Gaussian and the Students t copula functions. It would be interesting to consider copula models belonging to more generalized families. The vine copulas [65] are known to be flexible graphical models of dependence, and can be built up using a cascade of bivariate copulas. Their use for early diagnosis of neurophysiological disorders such as Alzheimers disease needs to be investigated.

Appendix A

Optimal Likelihood Ratio Test For Example 3.1

We first derive $f(z_1, z_2, z_3)$, the joint PDF of the observations under hypothesis H_1 .

The PDFs of the auxiliary random variables, W_1 and W_2 , are

$$p_{W_1}(w_1) = \frac{1}{\sqrt{\pi\lambda_1}} \exp\left(-\frac{1}{\lambda_1}w_1^2\right), \quad -\infty < w_1 < \infty, \quad (\text{A.1})$$

and

$$p_{W_2}(w_2) = \frac{1}{\lambda_1^{a_1}\Gamma(a_1)} w_2^{a_1-1} \exp\left(-\frac{w_2}{a_1}\right), \quad 0 < w_2 < \infty, \quad (\text{A.2})$$

respectively, where $\lambda_1 = 2\sigma_1^2$ denotes the twice of the variance of W_1 . We define another random variable, W , which follows a Gaussian distribution with zero mean and variance σ_1^2 ($W \sim \mathcal{N}(0, \sigma_1^2)$), and consider the following functions,

$$z_1 = g_1(w, w_1, w_2) = w, \quad -\infty < z_1 < \infty, \quad (\text{A.3})$$

$$z_2 = g_2(w, w_1, w_2) = w^2 + w_1^2, \quad 0 < z_1^2 < z_2 < \infty, \quad (\text{A.4})$$

$$z_3 = g_3(w, w_1, w_2) = \frac{w_2}{w_2 + w^2 + w_1^2}, \quad 0 < z_3 < 1. \quad (\text{A.5})$$

It is easy to verify that the variables, Z_1 , Z_2 and Z_3 , follow Gaussian, exponential and Beta distribution. Our goal, here, is to derive the joint PDF, $f(z_1, z_2, z_3) = p_{Z_1, Z_2, Z_3}(z_1, z_2, z_3; H_1)$.

Using the *Jacobian method of transformation* [16, p. 185], the joint PDF of the transformed variables is given as,

$$f(z_1, z_2, z_3) = p_{W, W_1, W_2}(h_1(z_1, z_2, z_3), h_2(z_1, z_2, z_3), h_3(z_1, z_2, z_3)) |J| \quad (\text{A.6})$$

$$= p_W(h_1(z_1, z_2, z_3)) p_{W_1}(h_2(z_1, z_2, z_3)) p_{W_2}(h_3(z_1, z_2, z_3)) |J|, \quad (\text{A.7})$$

where, the inverse mappings are

$$w = h_1(z_1, z_2, z_3) = z_1 \quad (\text{A.8})$$

$$w_1 = h_2(z_1, z_2, z_3) = \sqrt{z_2 - z_1^2} \quad (\text{A.9})$$

$$w_2 = h_3(z_1, z_2, z_3) = \frac{z_2 z_3}{1 - z_3} \quad (\text{A.10})$$

and, the Jacobian, J , is

$$J = \begin{bmatrix} \frac{\partial w}{\partial z_1} & \frac{\partial w}{\partial z_2} & \frac{\partial w}{\partial z_3} \\ \frac{\partial w_1}{\partial z_1} & \frac{\partial w_1}{\partial z_2} & \frac{\partial w_1}{\partial z_3} \\ \frac{\partial w_2}{\partial z_1} & \frac{\partial w_2}{\partial z_2} & \frac{\partial w_2}{\partial z_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\partial w_1}{\partial z_1} & \frac{1}{2}(z_2 - z_1^2)^{-\frac{1}{2}} & 0 \\ \frac{\partial w_2}{\partial z_1} & \frac{\partial w_2}{\partial z_2} & \frac{z_2}{(1 - z_3)^2} \end{bmatrix} \quad (\text{A.11})$$

$$= \frac{1}{2} \frac{z_2}{\sqrt{z_2 - z_1^2} (1 - z_3)^2}. \quad (\text{A.12})$$

Using Eqs. (A.8), (A.9), (A.10) and (A.12), in Eq. (A.6), and after some algebraic steps, we have,

$$f(z_1, z_2, z_3) = \frac{1}{2\pi\Gamma(a_1)\lambda_1^{1+a_1}} \frac{z_2^{a_1} z_3^{a_1-1}}{\sqrt{z_2 - z_1^2} (1 - z_3)^{1+a_1}} \exp\left(-\frac{z_2}{\lambda_1(1 - z_3)}\right), \quad (\text{A.13})$$

$$-\infty < z_1 < \infty, \quad 0 < z_1^2 < z_2 < \infty \quad \text{and} \quad 0 < z_3 < 1.$$

Observations, z_1 , z_2 and z_3 , have the following distributions under H_0 ,

$$Z_1 \sim \mathcal{N}(0, \sigma_0^2), \quad Z_2 \sim \text{exponential}(\lambda_0), \quad Z_3 \sim \text{beta}(a_0, b_0 = 1). \quad (\text{A.14})$$

Further, they are statistically independent. Therefore, their joint PDF is

$$g(z_1, z_2, z_3) = \prod_{n=1}^3 g(z_n) \quad (\text{A.15})$$

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{z_1^2}{2\sigma_0^2}\right) \cdot \frac{1}{\lambda_0} \exp\left(-\frac{z_2}{\lambda_0}\right) \cdot \\ &\quad \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} z_3^{a_0-1} (1 - z_3)^{b_0-1} \\ &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \frac{a_0}{\lambda_0} z_3^{a_0-1} \exp\left(-\left(\frac{z_1^2}{2\sigma_0^2} + \frac{z_2}{\lambda_0}\right)\right), \end{aligned} \quad (\text{A.16})$$

$$-\infty < z_1 < \infty, \quad 0 < z_2 < \infty, \quad \text{and} \quad 0 < z_3 < 1.$$

With (A.13) and (A.15), the LRT statistic can be obtained as,

$$\begin{aligned}
\mathcal{T}_{LR} &= \frac{\prod_{l=1}^L f(z_{1l}, z_{2l}, z_{3l})}{\prod_{l=1}^L g(z_{1l}, z_{2l}, z_{3l})} \\
&= \frac{\prod_{l=1}^L \frac{z_{2l}^{a_1} z_{3l}^{a_1-1} (1-z_{3l})^{-(1+a_1)}}{2\pi\Gamma(a_1)\lambda_1^{1+a_1}\sqrt{z_{2l}-z_{1l}^2}} \exp\left(-\frac{z_{2l}}{\lambda_1(1-z_{3l})}\right)}{\prod_{l=1}^L \frac{1}{\sqrt{2\pi\sigma_0^2}} \frac{a_0}{\lambda_0} z_{3l}^{a_0-1} \exp\left(-\left(\frac{z_{1l}^2}{2\sigma_0^2} + \frac{z_{2l}}{\lambda_0}\right)\right)} \mathbb{I}(z_2 > z_1^2), \quad (\text{A.17})
\end{aligned}$$

where, $\mathbb{I}(\mathcal{E})$ is an indicator of event \mathcal{E} , included to ensure the support, $0 < z_1^2 < z_2 < \infty$, for the variable Z_2 under H_1 .

Bibliography

- [1] 2009 Alzheimer’s Disease, Facts and Figures, Executive Summary. http://www.alz.org/alzheimers_disease_facts_figures.asp.
- [2] Nist biometric scores set, 2004. <http://www.itl.nist.gov/iad/894.03/biometricscores/>.
- [3] V. Aalo and R. Viswanathan. On distributed detection with correlated sensors: two examples. *IEEE Transactions on Aerospace and Electronic Systems*, 25(3):414–421, 1989.
- [4] B. Aiazzi, L. Alparone, and S. Baronti. Estimation based on entropy matching for generalized Gaussian pdf modeling. *IEEE Signal Processing Letters*, 6(6):138–140, 1999.
- [5] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society*, 28(1):131–142, 1966.
- [6] P. Atrey, M. Hossain, A. El Saddik, and M. Kankanhalli. Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems*, 16(6):345–379, 2010.
- [7] J.-F. Aujol, G. Aubert, and L. Blanc-Feraud. Wavelet-based level set evolution for classification of textured images. *IEEE Transactions on Image Processing*, 12(12):1634–1641, 2003.
- [8] D. Bamber. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12, 1975.
- [9] M. J. Beal, N. Jojic, and H. Attias. A graphical model for audiovisual object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):828–836, 2003.
- [10] D. Berg. Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15:675–701, 2009.
- [11] C. M. Bishop and J. Lasserre. Generative vs. discriminative? Getting the best of both worlds. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,

- D. Heckerman, A. F. M. Smith, and M. West, editors, *Bayesian Statistics*, volume 8, pages 3–24. Oxford University Press, 2007.
- [12] R. E. Bland. Acoustic and seismic signal processing for footstep detection. Master’s thesis, Massachusetts Institute of Technology, Cambridge, 2006.
- [13] A. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.
- [14] M. Breakspear. Dynamic connectivity in neural systems: theoretical and empirical considerations. *Neuroinformatics*, 2, 2004.
- [15] T. Butz and J.-P. Thiran. From error probability to information theoretic (multi-modal) signal processing. *Signal Processing*, 85(5):875–902, 2005.
- [16] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, 2002.
- [17] M. T. Chan. Hmm-based audio-visual speech recognition integrating geometric and appearance-based visual features. In *Proc. IEEE Fourth Workshop Multimedia Signal Processing*, pages 9–14, 2001.
- [18] Yin Chan and Haijun Li. Tail dependence for multivariate t-copulas and its monotonicity. *Insurance: Mathematics and Economics*, 42:763–770, 2008.
- [19] Biao Chen, Lang Tong, and P. K. Varshney. Channel-aware distributed detection in wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):16–26, 2006.
- [20] Hao Chen, P. K. Varshney, S. M. Kay, and J. H. Michels. Theory of the stochastic resonance effect in signal detection: Part i - fixed detectors. *IEEE Transactions on Signal Processing*, 55(7):3172–3184, 2007.
- [21] X. Chen and Y. Fan. Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130:307–335, 2006a.
- [22] Simon R. Cherry. Multimodality imaging: Beyond PET/CT and SPECT/CT. *Seminars in nuclear medicine*, 39(5):348–353, September 2009.
- [23] C. Cortes and M. Mohri. AUC optimization vs. error rate minimization. In *Advances in Neural Information Processing Systems*. MIT Press, 2003.
- [24] R. Cutler and L. Davis. Look who’s talking: speaker detection using video and audio correlation. In *Proc. IEEE Int. Conf. Multimedia and Expo ICME 2000*, volume 3, pages 1589–1592, 2000.
- [25] S. C. Dass, K. Nandakumar, and A. K. Jain. A principled approach to score level fusion in multimodal biometric systems. In *Audio and Video based Biometric Person Authentication*, pages 1049–1058, 2005.

- [26] J. Dauwels, F. Vialatte, and A. Cichocki. A comparative study of synchrony measures for the early diagnosis of Alzheimer’s disease based on EEG. *NeuroImage (in press)*. <http://www.dauwels.com/files/NeuroImage2009.pdf>.
- [27] J. Dauwels, F. Vialatte, T. Rutkowski, and A. Cichocki. *Advances in Neural Information Processing Systems*, volume 20, chapter Measuring neural synchrony by message passing, pages 361–368. MIT Press, Cambridge, MA, 2008.
- [28] J. Dauwels, F. Vialatte, T. Weber, and A. Cichocki. Quantifying statistical interdependence by message passing on graphs, part 1: One-dimensional point processes. *Neural Computation*, 21(8):2152–2202, 2009.
- [29] J. Dauwels, F. Vialatte, T. Weber, T. Musha, and A. Cichocki. Quantifying statistical interdependence by message passing on graphs, part ii: Multidimensional point processes. *Neural Computation*, 21(8):2203–2268, 2009.
- [30] Stefano Demarta and Alexander J. McNeil. The t copula and related copulas. *International Statistical Review*, 73:111–129, 2005.
- [31] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.
- [32] A. A. Dilbazar, H. O. Park, and T. W. Berger. The application of dynamic synapse neural networks on footstep and vehicle recognition. In *International Joint Conference on Neural Networks*, pages 1842–1846, August 2007.
- [33] M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, 11(2):146–158, 2002.
- [34] E. Drakopoulos and C.-C. Lee. Optimum multisensor fusion of correlated local decisions. *IEEE Transactions on Aerospace and Electronic Systems*, 27(4):593–606, 1991.
- [35] S. Eguchi and J. Copas. Interpreting kullback-leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, 97(9):2034–2040, October 2006.
- [36] A. Ekimov and J. M. Sabatier. Vibration and sound signatures of human footsteps in buildings. *Journal of Acoustical Society of America*, 120(2):762–768, 2006.
- [37] A. Ekimov and J. M. Sabatier. Passive ultrasonic method for human footstep detection. In *SPIE Unattended Ground, Sea and Air Sensor Technologies and Applications IX*, volume 6562, May 2007.
- [38] M. A. F. Figueiredo and A. K. Jain. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):381–396, 2002.

- [39] J. W. Fisher III and T. Darrell. Speaker association with signal-level audiovisual fusion. *IEEE Transactions on Multimedia*, 6(3):406–413, 2004.
- [40] J. W. Fisher III, T. Darrell, W. T. Freeman, and P. Viola. Learning joint statistical models for audio-visual fusion and segregation. In *Neural Information Processing Systems (NIPS)*, volume 13, pages 772–778, 2000.
- [41] Tieyan Fu, Xiao Xing Liu, Lu Hong Liang, Xiaobo Pi, and A. V. Nefian. Audio-visual speaker identification using coupled hidden markov models. In *Proc. Int. Conf. Image Processing ICIP 2003*, volume 3, 2003.
- [42] C. Genest, M. Gendron, and Michael Bourdeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 00:1–10, 2009.
- [43] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance:Mathematics and Economics*, 44:199–213, 2009.
- [44] R. Gittins. *Canonical Analysis: A Review with Applications in Ecology*. Springer, 1985.
- [45] R. M. Gray and Jr. Stockham, T. G. Dithered quantizers. *IEEE Transactions on Information Theory*, 39(3):805–812, 1993.
- [46] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics*. Wiley, 1966.
- [47] R. Gupta and A. O. Hero. High-rate vector quantization for detection. *IEEE Transactions on Information Theory*, 49(8):1951–1969, 2003.
- [48] Fredrik Gustafsson and Richard Karlsson. Statistical results for system identification based on quantized observations. *Automatica*, 45(12):2794–2801, 2009.
- [49] J. Hershey and J. Movellan. Audio-vision: Using audio-visual synchrony to locate sounds. In *Advances in Neural Information Processing Systems (NIPS)*, volume 12, pages 813–819, 1999.
- [50] I. Y. Hoballah and P. K. Varshney. Distributed bayesian signal detection. *IEEE Transactions on Information Theory*, 35(5):995–1000, 1989.
- [51] I. Y. Hoballah and P. K. Varshney. An information theoretic approach to the distributed detection problem. *IEEE Transactions on Information Theory*, 35(5):988–994, 1989.
- [52] K. M. Houston and D. P. McGaffigan. Spectrum analysis techniques for personnel detection using seismic sensors. In E. M. Carapezza, editor, *Unattended Ground Sensor Technologies and Applications V*, volume 5090, pages 162–173. SPIE, 2003.

- [53] J. E. Hurtado. *Structural Reliability: Statistical Learning Perspectives (Lecture Notes in Applied and Computational Mechanics)*. Springer, 2004.
- [54] A. J. Izenman. *Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning*. Springer, 2008.
- [55] J. Jeong. EEG dynamics in patients with Alzheimer’s disease. *Clinical Neurophysiology*, 115:1490–1505, 2004.
- [56] H. Joe. *Multivariate Dependence and Related Concepts*. Chapman and Hall, 1997.
- [57] H. Joe and J. J. Xu. The estimation method of inference functions for margins for multivariate models. Technical report, University of British Columbia, 1996.
- [58] R. L. Joshi and T. R. Fischer. Comparison of generalized Gaussian and Laplacian modeling in dct image coding. *IEEE Signal Processing Letters*, 2(5):81–82, 1995.
- [59] M. Kam, Q. Zhu, and W. S. Gray. Optimal data fusion of correlated local decisions in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, 28(3):916–920, 1992.
- [60] M. Kaminski and H. Liang. Causal influence: Advances in neurosignal analysis. *Critical review in biomedical engineering*, 33:347–430, 2005.
- [61] S. M. Kay. *Fundamentals of Statistical Signal Processing: Vol. II Detection Theory*. Prentice Hall PTR, 1993.
- [62] E. Kidron, Y. Y. Schechner, and M. Elad. Pixels that sound. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2005.
- [63] E. Kidron, Y. Y. Schechner, and M. Elad. Cross-modal localization via sparsity. *IEEE Transactions on Signal Processing*, 55(4):1390–1404, 2007.
- [64] K. Kokkinakis and A. K. Nandi. Speech modelling based on generalized Gaussian probability density functions. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP ’05)*, volume 1, pages 381–384, 2005.
- [65] D. Kurowicka and R. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modeling*. John Wiley and Sons Ltd., 2006.
- [66] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer, 3 edition, 2008.
- [67] R. Leonardi, P. Migliorati, and M. Prandini. Semantic indexing of soccer audiovisual sequences: a multimodal approach based on controlled markov chains. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(5):634–643, 2004.

- [68] D. Mari and S. Kotz. *Correlation and Dependence*. Imperial College Press, London, 2001.
- [69] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.
- [70] R. B. Nelsen. *An Introduction to Copulas*. Springer-Verlag, 1999.
- [71] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. In *Neural Information Processing Systems (NIPS)*, volume 14, pages 841–848, 2001.
- [72] L. O’Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.
- [73] C. H. Papadimitriou and J. N. Tsitsiklis. Intractable problems in control theory. In *Proc. 24th IEEE Conf. Decision and Control*, volume 24, pages 1099–1103, 1985.
- [74] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3 edition, 1991.
- [75] L. Pardo. *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC, 2006.
- [76] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- [77] Andrew J Patton. *Handbook of Financial Time Series*, chapter Copula-Based Models for Financial Time Series, pages 767–785. Springer, 2009.
- [78] E. Pereda, R. Q. Quiroga, and J. Bhattacharya. Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, 77:1–37, 2005.
- [79] D. Povey and P. C. Woodland. Minimum phone error and i-smoothing for improved discriminative training. In *Proc. IEEE Int Acoustics, Speech, and Signal Processing (ICASSP) Conf*, volume 1, 2002.
- [80] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Physical Rev. E*, 65, 2002.
- [81] L. Rabiner and B.-H. Juang. *Fundamentals of speech recognition*. Prentice Hall, Englewood Cliffs, New Jersey, 1993.
- [82] L. Roberts. Picture coding using pseudo-random noise. *IRE Transactions on Information Theory*, 8(2):145–154, 1962.

- [83] Georg Michael Rockinger and Eric Jondeau. Conditional dependency of financial series: An application of copulas. *SSRN eLibrary*, 2001. <http://ssrn.com/paper=262642>.
- [84] M. E. Sargin, E. Erzin, Y. Yemez, and A. M. Tekalp. Lip feature extraction based on audio-visual correlation. In *Proc. of European Signal Processing*, 2005.
- [85] L. L. Scharf and C. T. Mullis. Canonical coordinates and the geometry of inference, rate, and capacity. *IEEE Transactions on Signal Processing*, 48(3):824–831, 2000.
- [86] L. Schuchman. Dither signals and their effect on quantization noise. *IEEE Transactions on Communication Technology*, 12(4):162–165, 1964.
- [87] J. H. Shapiro. Bounds on the area under the ROC curve. *Journal of the Optical Society of America A*, 16(1):53–57, 1999.
- [88] K. Sharifi and A. Leon-Garcia. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(1):52–56, 1995.
- [89] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.
- [90] A. J. Simpson and M. J. Fitter. What is the best index of detectability? *Psychological Bulletin*, 80:481–488, 1973.
- [91] M. Slaney and M. Covell. FaceSync: A linear operator for measuring synchronization of video facial images and audio tracks. In *Advances in Neural Information Processing Systems (NIPS)*, volume 13, pages 814–820, 2000.
- [92] G. E. Sleaf, M. Ladd, T. S. McDonald, and G. J. Elbring. Acoustic and seismic modalities for unattended ground sensors. In *SPIE Unattended Ground, Sea and Air Sensor Technologies and Applications*, volume 3713, pages 2–9, July 1999.
- [93] C. J. Stam. Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clinical Neurophysiology*, 116:2266–2301, 2005.
- [94] M. Studený and J. Vejnarová. The multiinformation function as a tool for measuring stochastic dependence. In *Learning in Graphical Models*, pages 261–298. Kluwer Academic Publishers, 1998.
- [95] A. Subramaniam, K. G. Mehrotra, C. K. Mohan, P. K. Varshney, and T. Damarla. Analysis of acoustic and seismic sensor data in indoor environments. In *MSS BAMS*, 2007.
- [96] G. Succi, D. Clapp, R. Gampert, and G. Prado. Footstep detection and tracking. In *SPIE Unattended Ground Sensor Technologies and Applications III*, volume 4393, pages 22–29, Sep. 2001.

- [97] G. Succi, G. Prado, R. Gampert, T. Pedersen, and H. Dhaliwal. Problems in seismic detection and tracking. In *SPIE Unattended Ground Sensor Technologies and Applications II*, volume 4040, pages 165–173, April 2000.
- [98] A. Sundaresan, P. K. Varshney, and N. S. V. Rao. Copula-based fusion of correlated decisions. *IEEE Transactions on Aerospace and Electronic Systems*, 47(1):454–471, 2011.
- [99] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulas. Optimal distributed decision fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 25(5):761–765, 1989.
- [100] J. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control*, 30(5):440–446, 1985.
- [101] V. Tyagi. Maximum accept and reject (MARS) training of the HMM-GMM speech recognition systems. In *Interspeech*, Brisbane, 2008.
- [102] P. Uhlhaas and W. Singer. Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. *Neuron*, 52:155–168, 2006.
- [103] G. Van de Wouwer, P. Scheunders, and D. Van Dyck. Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing*, 8(4):592–598, 1999.
- [104] A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- [105] M. K. Varanasi and B. Aazhang. Parametric generalized Gaussian density estimation. *Journal of Acoustical Society of America*, 86:1404–1415, 1989.
- [106] P. K. Varshney. *Distributed Detection and Data Fusion*. Springer-Verlag, 1997.
- [107] F. Vialatte, C. Martin, R. Dubois, J. Haddad, B. Quenet, R. Gervais, and G. Dreyfus. A machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics. *Neural Networks*, 20:194–209, 2007.
- [108] N. Vucic and H. Boche. A tractable method for chance-constrained power control in downlink multiuser miso systems with channel uncertainty. *IEEE Signal Processing Letters*, 16(5):346–349, 2009.
- [109] Yao Wang, Zhu Liu, and Jin-Cheng Huang. Multimedia content analysis-using both audio and visual clues. *IEEE Signal Processing Magazine*, 17(6):12–36, 2000.
- [110] H. White. *Estimation, Inference and Specification Analysis*. Cambridge University Press, 1994.

- [111] B. Widrow. A study of rough amplitude quantization by means of Nyquist sampling theory. *IRE Transactions on Circuit Theory*, 3(4):266–276, 1956.
- [112] B. Widrow and I. Kollar. *Quantization Noise: Roundoff Error in Digital Computation, Signal processing, Control, and Communications*. Cambridge University Press, 2008.
- [113] B. Widrow, I. Kollar, and Ming-Chang Liu. Statistical theory of quantization. *IEEE Transactions on Instrumentation and Measurement*, 45(2):353–361, 1996.
- [114] S. S. Wilks. The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9:60–62, 1938.
- [115] P. Willett, P. F. Swaszek, and R. S. Blum. The good, bad and ugly: distributed detection of a known signal in dependent gaussian noise. *IEEE Transactions on Signal Processing*, 48(12):3266–3279, 2000.
- [116] H. F. Xing, F. Li, and Y.-L. Liu. Wavelet denoising and feature extraction of seismic signal for footstep detection. In *International Conference on Wavelet Analysis and Pattern Recognition*, volume 1, pages 218–223, 2007.
- [117] Z. Zhu and T. S. Huang, editors. *Multimodal Surveillance: Sensors, Algorithms, and Systems*. Artech House, Norwood, MA, 2007.
- [118] A. M. Zoubir and D. R. Iskander. *Bootstrap Techniques for Signal Processing*. Cambridge University Press, 2004.

VITA

AUTHOR Satish Giridhar Iyengar

DATE OF BIRTH September 04, 1980

PLACE OF BIRTH Pondicherry, India

EDUCATION

- Master of Science, 2004, Syracuse University, New York, USA
- Bachelor of Engineering, 2002, University of Mumbai, Mumbai, India

AWARDS

- Travel award, Jan 2005, Workshop on “Mathematics of the Ear and Sound signal processing,” Institute for Pure and Applied Mathematics (IPAM), UCLA
- Best Poster award (EECS), 2010, Nunan Research Day, Syracuse University

PROFESSIONAL EXPERIENCE

- Graduate Assistant, Dept. of EECS, Syracuse University, 2006 to 2011
- Research Intern, Blue Highway Inc., June 2010 to May 2011
- Research Intern, The US Army Research Lab, Adelphi, MD, Aug. 2009 to Oct. 2009
- Graduate Assistant, Dept. of Bio-Engineering and Neuroscience, 2003 to 2005