Automatic Modulation Classification Using Grey Relational Analysis

Matthew J. Price

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical Engineering

Jeffrey H. Reed, Chair
Tamal Bose
Ashwin E. Amanna

April 25, 2011
Blacksburg, Virginia

Keywords: Automatic Modulation Classification, Grey Relational Analysis,
Cyclostationary Analysis, Cumulants, Haar Wavelet Transform

Automatic Modulation Classification Using Grey Relational Analysis

Matthew J. Price

ABSTRACT

One component of wireless communications of increasing necessity in both civilian and military applications is the process of automatic modulation classification. Modulation of a detected signal of unknown origin requiring interpretation must first be determined before the signal can be demodulated. This thesis presents a novel architecture for a modulation classifier that determines the most likely modulation using Grey Relational Analysis with the extraction and combination of multiple signal features. An evaluation of data preprocessing methods is conducted and performance of the classifier is investigated with the addition of each new signal feature used for classification.

Grant Information

The research presented in this investigation was supported by the Federal Railroad Administration, Office of Research and Development, FRA Grant No. DTFR53-09-H-00021. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the view of the Federal Railroad Administration and/or U.S. DOT.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Quick and accurate signal detection and classification of radio signals is an increasing necessity. Radio spectrum is a scarce commodity, not because every radio band is constantly occupied, but because most bands have already been licensed and assigned to specific services [1]. Licensing of the radio spectrum is an important tool for minimizing interference between users as each licensed user has exclusive rights to its spectrum band. Some spectrum bands, particularly bands allocated to TV broadcast service in the United States, are underutilized and the Federal Communication Commission (FCC) is therefore considering a new spectrum allocation policy that would permit the use of unlicensed radios in these bands [2].

The new IEEE 802.22 WRAN standard aims to increase the efficiency of spectrum use in the TV broadcast spectrum in order to provide wireless broadband access to rural and remote areas with performance comparable to existing broadband access technologies [3]. For this policy to work, an unlicensed user must detect and identify incumbent signals. The secondary user also must not interfere with the primary user services and be able to detect

primary user signals quickly in low signal-to-noise environments [4]. Once an incumbent is recognized, the unlicensed radio must immediately switch to a different frequency band that has been identified through Dynamic Spectrum Access. Dynamic Spectrum Access (DSA) is the process of identifying available unused regions of spectrum and regions used by secondary users [5]. The almost instantaneous decision making required for DSA is best accomplished by radio artificial intelligence [6]. Cognitive Radio (CR), first introduced by Joseph Mitola [7], has the ability to sense itself, the surrounding radio environment and users, and to make decisions on how to use available spectrum resources.

In order to distinguish between primary and secondary user signals and establish communications with other secondary users, an effective method for classification of the observed signal modulation is required and it is predicted that incumbent signals will need to be detected in -15dB SNR environment [8, 9]. Some detectors proposed for Wireless Regional Area Networks (WRAN) observe the energy occupying a spectrum band only and do not rely on statistical signal features for signal classification [10]. They are therefore not appropriate for weak signals or signals overwhelmed by RF noise since the detector is not able to discriminate between the energy level of a signal corrupted by noise and noise only. The CR must therefore use a more sophisticated classification method.

Automatic modulation classification (AMC) is the identification of an underlying symbol constellation from observed noisy measurements and is an intermediate step between digital signal detection and demodulation [11]. The AMC performed in this work uses a feature-based (FB) approach that extracts statistical properties which are then compared to a reference library of statistical information for specific modulations. This method is more robust in noisy environments than energy detection since additional noise and interference in the observed channel will have different statistical properties than the signal to be detected [12].

AMC is a necessary tool for secondary users operating in spectrum white space and also aids civilian authorities in the monitoring of unauthorized transmitters. The need for modulation classification extends into the military realm of electronic warfare as well. Signals must be

classified for interception and surveillance of enemy communications, target acquisition and homing, and the identification and jamming of enemy communications and navigational systems.

1.2 Contributions

The new decision engine for automatic modulation classification that is presented in this work uses Grey Relational Analysis and the creation of comparison vectors from the fusion of unrelated signal features. The GRA-AMC performance is tested in both low- and high-noise environments.

The GRA-AMC decision engine is a single-step process (as opposed to hierarchical approaches [11,13]). It also does not require training of the radio like classifiers that implement Hidden Markov Models [14] or Artificial Neural Networks [8, 15–17], which are sensitive to disparity between the training and test data, such as changes in noise distribution or variance [11].

The generic architecture of the GRA-AMC is illustrated by Figure 1.1. Each signal feature is extracted separately from the incoming signal. These independent signal features are then inserted into the Φ_{test} Signal Feature Vector (SFV) to form a representation of the incoming signal. A corresponding set of signal features is stored for each modulation in the radio reference library and all modulations are stored in the Φ_{comp} matrix for comparison. The Φ_{test} SFV is then compared to each row of Φ_{comp} representing each modulation choice and each modulation is given a grey relational grade. This grade is a quantification of similarity between 0 and 1, where 1 is the most similar.

The SFV can also be occupied by a single feature vector that contains many data points, such as a cyclostationary profile. The classification process is then defined as a Multiple Attribute Decision Making (MADM) problem, using the combination of signal features instead of combining classification techniques, as in [17]. The SFV reduced data set is then compared

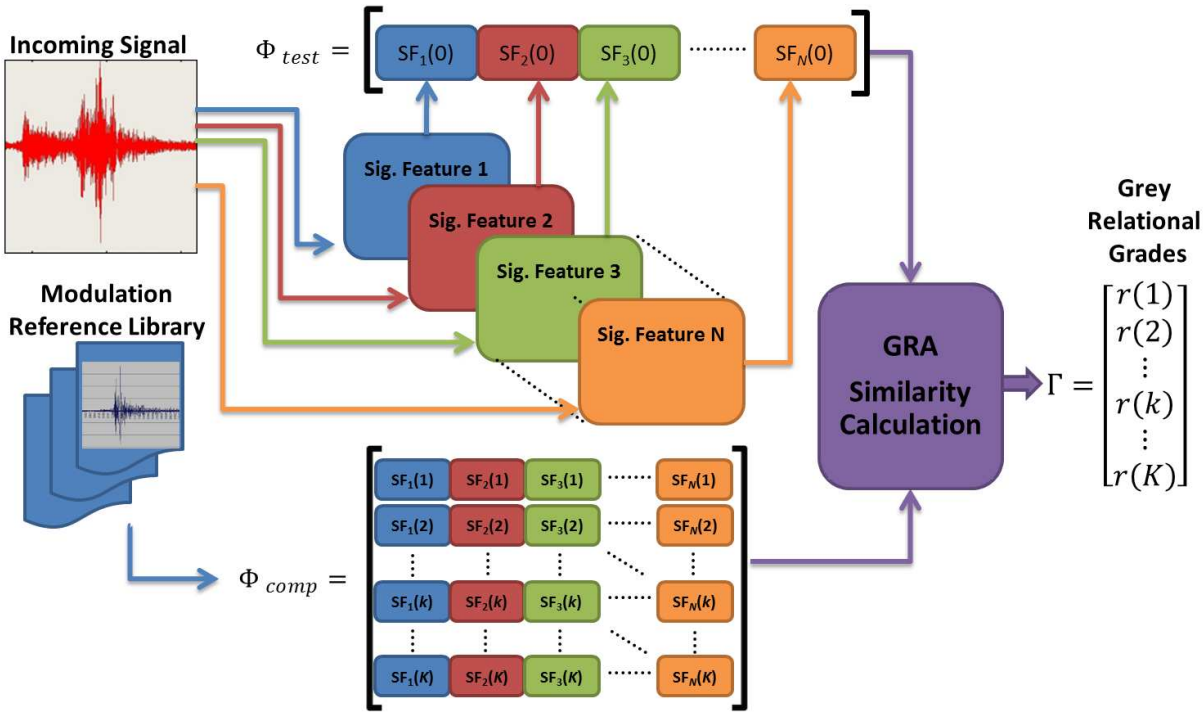


Figure 1.1: The GRA-AMC Architecture

to a library of reference vectors for different modulations using the GRA which is designed to compare vectors of information, regardless of the units involved. Each element in a comparison vector can have a completely different unit (or unitless score) as long as all compared vectors share the same signal attributes in the same order [18]. The GRA solves the MADM problem by combining the entire range of signal feature metrics for every alternative SFV into a single ranking value, yielding a quick SFV comparison [17]. The signal is classified as the modulation associated with the SFV that is decided by the GRA to be the most likely (highest ranked) match.

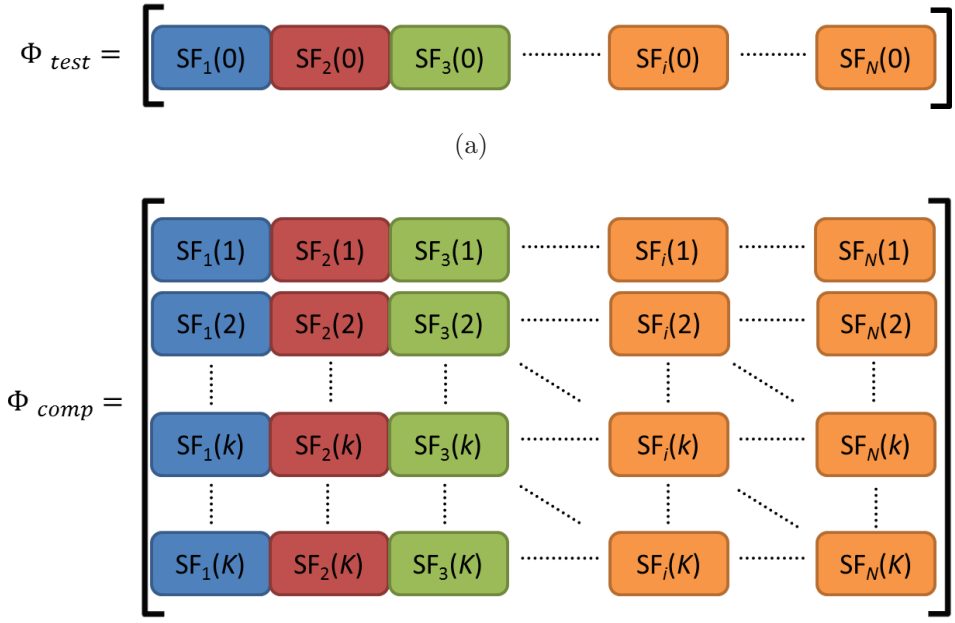


Figure 1.2: Signal Feature Combination: (a) Incoming Signal SFV; (b) Modulation Comparison Matrix.

The GRA-AMC developed for this work uses a set reference SFVs for four digital modulation types: Binary Phase-Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), Frequency-Shift Keying (FSK) and Minimum-Shift Keying (MSK). Each reference SFV will be occupied by one or a combination of three signal features: cyclostationary profiles, cumulants, and wavelet transform variance.

A summary of the contributions presented in this work is as follows:

- Application of Grey Relational Analysis to signal feature comparison in Automatic Modulation Classification
- Quantitative analysis of data preprocessing techniques for cyclostationary analysis
- Quantitative analysis of the redefining of the distance metric used by the GRA algorithm
- Quantitative analysis of the combination of multiple signal features into single signal feature vector for classification
- Investigation into the true amount of unique distinguishing information added by combining signal features

The first two contributions listed above have already been published in [19] and are also currently under review for publication for the Springer Journal of Analog Integrated Circuits and Signal Processing (AISCP).

1.3 Computer Simulations

Simulations performed to test the GRA-AMC assumed a zero-mean AWGN channel with no fading, no frequency or phase error, perfect synchronization and rectangular pulse shaping with *a priori* knowledge of baud rate and carrier frequency. Each signal consisted of 10,000 data samples corresponding to 2,000 sent symbols. Each simulation consisted of 1,200 Monte-Carlo trials (300 for each modulation) for an SNR range of -20dB to +15dB in 5dB increments. A confusion matrix was then produced for each SNR. The received signal, for which the modulation is to be classified, is modeled as (1.1) where $n(t)$ is additive white Gaussian noise with power σ_N^2 as in [20]. The signal-to-noise ratio (SNR) is then defined by (1.2), where E_S is the signal energy.

$$r(t) = s(t) + n(t) \tag{1.1}$$

$$SNR = 10\log_{10}\left(E_S/\sigma_N^2\right) \tag{1.2}$$

In real-world environment, coarse estimates of the signal characteristics such as carrier frequency and signal bandwidth can be determined from the estimated power spectral density. Baud rate can be estimated via a tracking loop, and symbol timing can be recovered with fractional sampling schemes [11].

Each modulation was assigned a probability of correct classification (P_{CC}) for each SNR, which is the percentage of trials that resulted in a correct classification of that modulation. The *average* probability of classification ($P_{CC,\ Avg}$) was taken over all four modulation types to summarize the performance of the classifier. This can be observed in the simulation confusion matrix as the ratio of the sum of the values in the main diagonal to the total number of signals sent. For an overall numerical comparison between techniques, the overall probability of classification ($P_{CC,\ Total}$) was defined as the average of all $P_{CC,\ Avg}$ over the SNR range of -20dB to 0dB. Much work has already been conducted in the modulation classification in signal-to-noise environments above 0dB [21]. Therefore the focus of this thesis is in the negative SNR range.

To obtain a complete picture of the performance of the classifier, the probability of a false positive should also be tabulated. Since the GRA-AMC must classify the observed spectrum as one of four modulations, presenting the overall probability of a false positive will not be useful as it will simply be $P_{fp,\ Total}$=1-$P_{CC,\ Total}$. Even though a decrease of $P_{fp,Total}$ is important, it is very desirable that every modulation not be classified as one particular modulation when the noise floor is raised. Therefore we will instead present the average of the maximum P_{fp} for each modulation, $\overline{P}_{fp,max}$, out of all tested SNR scenarios. These performance measures of the proposed GRA-AMC are all extracted from the tabulated confusion matrices and Receiver Operating Characteristic (ROC) curves.

1.4 Key Findings

The investigations conducted in the following chapters resulted in key findings including:

- A significant increase in classifier performance results from the correct combination of data preprocessing techniques.
- Redefining the distance metric used by the GRA greatly reduces bias towards one modulation in low SNR.
- Combining signal features for classification with the GRA will only compound the effects of noise if each individual signal feature performs poorly in low SNR.
- Signal feature data fusion requires that each added signal feature add unique statistical information to the signal feature vector.

Chapter 2

Background

This chapter presents examples of work done for modulation classification using methods other than Grey Relational Analysis. An overview of the performance of the individual signal features that are used in this thesis is presented as a summary of selected literature. Origins and past applications of Grey Relational Analysis are presented along with a detailed look at the GRA algorithm itself.

2.1 Literature Review of Automatic Modulation Classification

This section explores commonly-used methods for automatic modulation classification with a focus on feature-based approaches. Feature-based (FB) approaches to AMC have been implemented using a wide variety of signal features such as instantaneous amplitude, phase, and frequency, zero-crossing intervals, wavelet transforms, amplitude and phase histograms, constellation shapes, and others [21, 22]. Likelihood-based (LB) approaches have also been used for AMC, but these require a particularly laborious and complex process that attempts to derive a mathematical model for the signals being classified. LB methods are also sensitive

to modeling errors, and have a complexity that creates a potential for high latency that is inadequate, especially during a military engagement. Although the FB method may not be optimal, it is simpler to implement, and can have comparable performance [11].

2.1.1 Artificial Neural Networks

One common decision method for FB automatic modulation classification is the use of artificial neural networks. The Artificial Neural Network (ANN) is a computational model that is inspired by biological neural networks. The ANN model adapts and changes shape during the "learning" or "training" phase and can subsequently make decisions based off the training information [15]. This method has had much success in negative SNR using the cyclostationary profile signal feature [8, 15, 17]. However, this high performance is also directly related to the fine cyclic resolution that is required to identify distinguishing features of the cyclostationary profile in high noise [23]. The ANN model has also been used in combination with cumulants [16], but the performance is reduced since cumulant estimation is easily corruptible in noise.

2.1.2 Hidden Markov Models

A discrete sequence of data can be described as a Markov process if the future state of the process, given the present, is independent of the past. The probability of arriving at the next state in the sequence can be determined by adding the probabilities of every way of arriving at that state. The model is referred to as a Hidden Markov Model (HMM) if the underlying state is unknown, as in the case of a signal awaiting identification [24]. The HMM has the same role as a decision engine in modulation classification as the ANN or GRA and has been successfully implemented for classification with the cyclostationary profile [14, 24].

2.1.3 Hierarchical Decision Making

Hierarchical decision trees (sometimes referred to as the "divide and conquer" approach) are the preferred method for modulation classification when the signal features being estimated and interpreted are cumulants or wavelet variance [11,24]. Modulations can first be classified by class (i.e. PSK, PAM, QAM, FSK) followed by a intra-class decision between different modulation orders (for PSK: BPSK, QPSK, 8-PSK, etc.). The simplest design uses a hard-decision threshold method to decide between branches of the decision tree. Modifications to the hierarchical approach have also been made with the substitution of an ANN at each decision point in the hierarchical tree instead of the hard-decision threshold [5, 13].

2.1.4 Selected Signal Features

The signal features used by the GRA-AMC in this thesis (cyclostationary profiles, cumulants, and wavelets) have all been used by the other classification techniques as discussed in the previous sections. Table 2.1 summarizes the performance of modulation classifiers in selected literature that use each of these signal features. Previous modulation classification work done using the cyclostationary profile has had very favorable results in negative SNR. However, these results required a large observational window and fine cyclic resolution. Both of these caveats require more processing time that this thesis assumes to be unavailable. Classifiers that use cumulants and wavelets also perform well, but the SNR operational range must stay above 0dB.

2.2 Grey Relational Analysis: Origins and Past Use

Grey Systems Theory (GST) is a set of predictive and comparative concepts and algorithms first developed by Professor Deng Julong in 1982. If a system where all relevant information is known is described as white and a system where all relevant information is unknown is

Table 2.1: Selected Signal Feature Performance in Literature

Author(s)	Signal Feature	Performance Notes
Fehske *et al.* [15]	Cyclostationary Profile	Almost 100% P_{CC} when above: -9dB for ANN trained on various SNR 4dB for ANN trained on 6dB
Kim *et al.* [14]	Cyclostationary Profile	P_{CC} above 90% when above -10dB for 200 averaging blocks using HMM classifier
Like *et al.* [5]	Cyclostationary Profile	P_{CC} above 90% when above 2.5dB for 1-antenna proposed system.
Liu *et al.* [17]	Cyclostationary Profile	P_{CC} above 90% when above -12dB. (Fine cyclic resolution)
Ramkumar [24]	Cyclostationary Profile	P_{CC} above 95% when above 2dB.
Yuan *et al.* [25]	Cyclostationary Profile	P_{CC} above 90% when above -6dB. (Frequency smoothing window L=25)
Sun [26]	Stationary Cumulants	P_{CC} above 80% when above 3dB.
Swami, Sadler [11]	Stationary Cumulants	Almost 100% P_{CC} when above 5dB.
Ho *et al.* [27, 28]	Wavelets	Almost 100% P_{CC} for: BPSK above 2dB, QPSK/8-PSK above 8dB BFSK/QFSK above 4dB, 8FSK above 18dB
Hong, Ho [29]	Wavelets	P_{CC} above 97% when above 5dB.
Prakasam, Madheswaran [30]	Wavelets	Almost 100% P_{CC} when above 5dB, with low P_{fp}.
Wu *et al.* [16]	Wavelets	P_{CC} above 90% when above 8dB.

black, the grey system will lie in the middle. The term "grey" refers to the incompleteness of known information. GST was developed from a need to analyze systems with extremely limited information where the requirements of traditional statistical methods could not be satisfied and has been proven to be useful for analyzing poor, incomplete, and uncertain data [18, 31].

One practical algorithm that has risen from GST is Grey Relational Analysis (GRA), which is used for determining the best solution from extremely limited raw data [32]. GRA maps the Euclidean distance between two vectors to a normalized measurement of similarity and produces a grey relational grade between 0 and 1 [33]. A relational grade of 1 represents an exact match. This method of vector comparison also has the advantage of not requiring each vector element to have the same units. Instead, vectors can consist of any combination of relevant system information with original units - whether they are physical quantities or unitless numerical scores [18]. GRA has already been applied to manufacturing [34], project management [35], the location estimation of mobile phones [36], economics, marketing, agriculture, and road traffic safety research [18].

One such application was conducted in the city of Shanghai in 1989 [32]. Grey relational analysis was used to evaluate the choices for traffic reconstruction projects. A set of ideal criteria was formed (Fig. 2.1a) to describe what were considered the ideal resulting attributes of the construction project. Some attribute metrics are common metrics that one would associate with such a project (i.e. cost, traffic volume). Many others were unitless scores that were determined by other means to describe the effect of the construction (i.e. functionality, safety, feasibility). Corresponding scores and estimates were determined for six different construction project choices (Fig. 2.1b). The array of ideal metric values was then compared to the matrix of choices, using GRA. A subway project was consequently selected.

The goal of this work is to apply GRA to modulation classification, taking advantage of GRA's ability to compare vectors occupied by an assortment of completely different metrics. Different signal features will be collected and fused into a single vector to be interpreted by

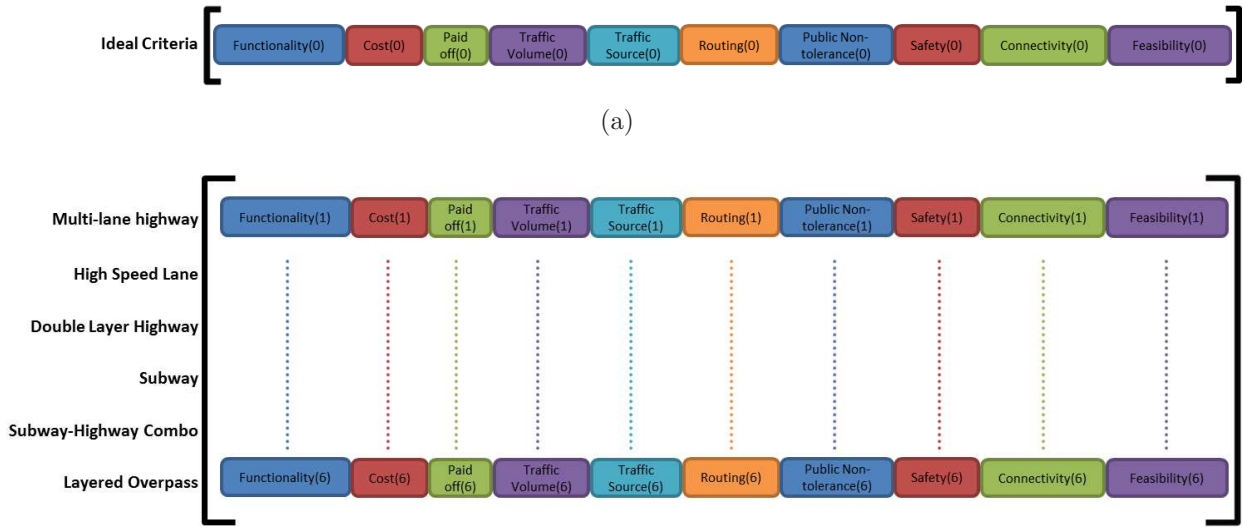


Figure 2.1: Shanghai Project Feature Vectors: (a) Ideal attribute values; (b) Project choices with corresponding attribute values.

the GRA. Grey Relational Analysis has the unique ability to apply a relational element that is entirely dependent on the set of reference vectors that define the grey relational space.

2.3 The GRA Algorithm

This section describes the generic GRA algorithm. GRA quantifies the similarity between a "reference" (or observed) data series and a matrix of "compared" data series, called an evaluation matrix. All K number of "compared" data series in the evaluation matrix define the grey relational space (GRS) [18]. Each of these alternative data series will have n number of attributes. In the context of this work the evaluation matrix will consist of K reference modulation SFVs, each with n extracted signal features.

First, let the observed test sequence (from the received signal) consist of n sample points $\varphi_i(0)$, $i = 1, 2, 3, \ldots, n$ where each sample point is an attribute of the signal. Each reference vector of the GRS matrix is defined as $\Phi(k) = [\varphi_1(k), \varphi_2(k), \cdots, \varphi_n(k)]$, $k = 1, 2, 3, \ldots, K$.

Equations (2.1) and (2.2) define Φ_{test} as the observed vector and Φ_{comp} as the comparative set of reference vectors.

$$\Phi_{test} = [\varphi_1(0), \varphi_2(0), \cdots \varphi_i(0), \cdots, \varphi_n(0)] \tag{2.1}$$

$$\Phi_{comp} = \begin{bmatrix} \Phi(1) \\ \Phi(2) \\ \vdots \\ \Phi(k) \\ \vdots \\ \Phi(K) \end{bmatrix} = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) & \cdots & \varphi_i(1) & \cdots & \varphi_n(1) \\ \varphi_1(2) & \varphi_2(2) & \cdots & \varphi_i(2) & \cdots & \varphi_n(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \varphi_1(k) & \varphi_2(k) & \cdots & \varphi_i(k) & \cdots & \varphi_n(k) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \varphi_1(K) & \varphi_2(K) & \cdots & \varphi_i(K) & \cdots & \varphi_n(K) \end{bmatrix} \tag{2.2}$$

The absolute value of the difference between the test sequence Φ_{test} and the k^{th} reference vector $\Phi_{comp}(k)$ is determined using (2.3), creating the deviation matrix $\Delta\Phi$ defined in (2.4). The Euclidean distance between the corresponding attributes of the test vector and each comparable vector is subsequently calculated, as in (2.5). Euclidean distance is the standard method for quantifying distance, but the use of other methods will be explored later in this work.

$$\Delta\varphi_i(k) = \mid \varphi_i(0) - \varphi_i(k) \mid \tag{2.3}$$

$$\Delta\Phi = \begin{bmatrix} \Delta\varphi_1(1) & \Delta\varphi_2(1) & \cdots & \Delta\varphi_i(1) & \cdots & \Delta\varphi_n(1) \\ \Delta\varphi_1(2) & \Delta\varphi_2(2) & \cdots & \Delta\varphi_i(2) & \cdots & \Delta\varphi_n(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \Delta\varphi_1(k) & \Delta\varphi_2(k) & \cdots & \Delta\varphi_i(k) & \cdots & \Delta\varphi_n(k) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \Delta\varphi_1(K) & \Delta\varphi_2(K) & \cdots & \Delta\varphi_i(K) & \cdots & \Delta\varphi_n(K) \end{bmatrix} \tag{2.4}$$

$$ED(k) = \sqrt{\sum_{i=1}^{n} (\Delta\varphi_i(k))^2} \tag{2.5}$$

Next, the minimum and maximum values of matrix $\Delta\Phi$ ($\Delta\varphi_{max}$ and $\Delta\varphi_{min}$) must both be identified as shown in (2.6) and (2.7). The final step is to assign each comparative vector a Grey Relational Grade, as illustrated by (2.8), where the recognition coefficient $\xi \in [0, 5]$. The recognition coefficient ξ is used to minimize or weaken the effect of $\Delta\varphi_{max}$ as it grows large. It affects only the 'magnitude' of the grey relational grade, but does not change the relative relationships between the comparative sequences [33]. Since the selected choice will still have the highest grey relational grade regardless of the value of ξ, the ξ value is a nonessential component of the GRA for this work and was set to 1, as in the case of the cardiac arrhythmia identification conducted in [37].

$$\Delta\varphi_{min} = \min_{\forall k}[\min_{\forall i} \Delta\varphi_i(k)] \tag{2.6}$$

$$\Delta\varphi_{max} = \max_{\forall k}[\max_{\forall i} \Delta\varphi_i(k)] \tag{2.7}$$

$$r(k) = \exp\left[\xi \left(\frac{ED(k)}{\Delta\varphi_{max} - \Delta\varphi_{min}}\right)^2\right] \tag{2.8}$$

The Grey Relational Grades, $r(k)$, are placed into a matrix, Γ, for final comparison as shown in (2.9).

$$\Gamma = [r(1), r(2), \ldots, r(k), \ldots, r(K)] \tag{2.9}$$

The relational grades provide a quantification of similarity between the test vector and the set of reference vectors. As the relational grade increases, so does similarity and the closer the test vector is to the reference vector in terms of distance. However due to the relational nature of the GRA, $\sum_{k=1}^{K} r\,(k) = 1$ and therefore no reference vector will ever have an $r(k)$ value of exactly 1.

Chapter 3

Single Feature GRA-AMC

This chapter explores modulation classification using only one signal feature: the cyclostationary profile (or α-profile). The α-profile is not a single numerical metric, but an entire vector of data points whose pattern is then interpreted and matched to the known α-profile of a particular modulation. The equation for Φ_{test} in these simulations is defined as

$$\Phi_{test} = \Big[\alpha\text{-profile}\Big] \tag{3.1}$$

A cyclostationary process is defined as a random process with statistical parameters, such as the autocorrelation function, that vary periodically with time [38, 39]. Most man-made signals in communications, telemetry, radar, and sonar, will have statistical parameters that exhibit periodicity and therefore can be modeled as cyclostationary [40]. Because of the periodic behavior of the modulation process its statistical description varies with time periodically [39]. Cyclic spectral analysis algorithms are used to estimate the correlation between spectral components of these signals [41] and create an α-profile that has distinguishing features caused by data symbol changes. The spectral correlation function (SCF) is one such algorithm [42] that is the tool used to create the α-profile.

Cyclostationary (CS) detection is an increasingly favorable alternative to energy detectors. Using energy detection to identify signals is extremely challenging if the observed channel

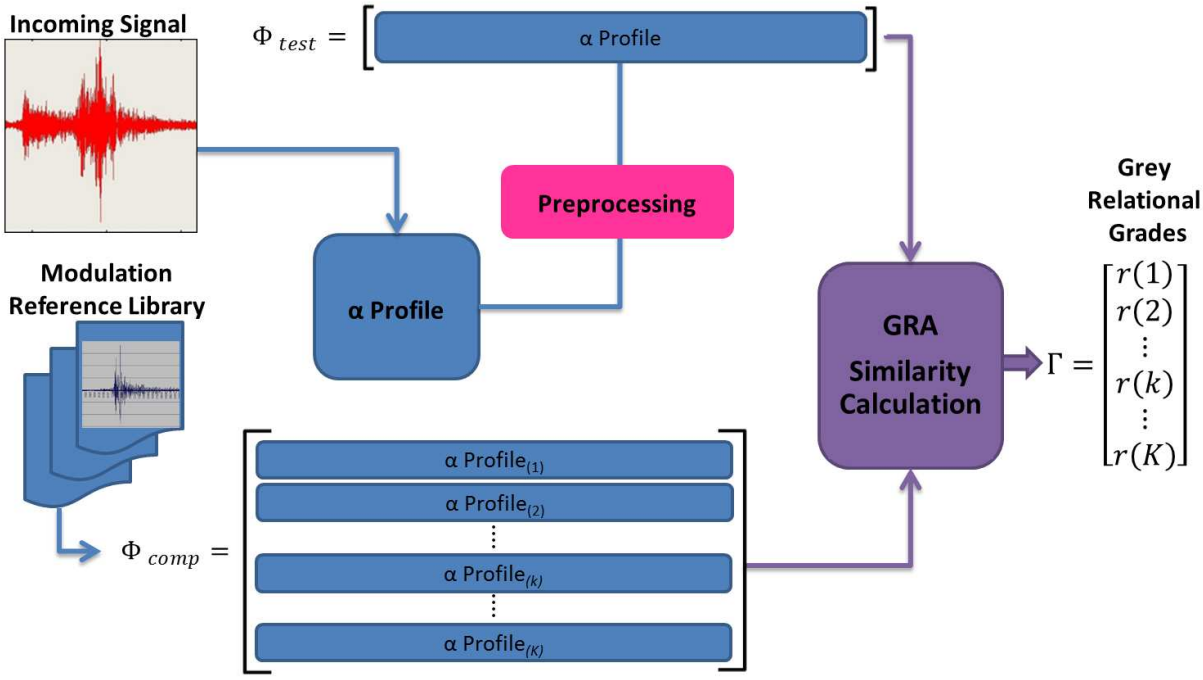


Figure 3.1: Single Feature (α-Profile) GRA-AMC Architecture

contains fluctuating or high background noise. In contrast, cyclostationary detection exploits the hidden periodicities that are unique to man-made signals and has the potential to detect and classify modulation type in the same process [4, 43]. CS detectors have also shown favorable performance in low SNR conditions [15, 25]. A distinction can be made between noise which is wideband, stationary, and non-correlated and the modulated signal's spectral correlation and periodicity [17].

CS spectral analysis has been applied to time difference of arrival estimation, signal detection and classification and even used for the study of interstellar medium in Astronomy [44, 45]. This work was inspired by the use of GRA for the identification of cardiac arrhythmias from the pattern recognition of electrocardiogram (ECG) signals [37, 46]. An ECG signal represents the changes in electrical potential during the heartbeat as recorded with non-invasive electrodes on the limbs and chest. The result is a unique signal profile that can be used for identification.

3.1 The Spectral Coherence Profile

The α-profile, or spectral coherence profile, is defined as the maximum of the magnitude of the Spectral Coherence Function (SCF). The magnitude is used to account for the unknown phase of the SCF [5].

$$\alpha\text{-profile} = \max_{\forall f} |C_X^\alpha(f)| \tag{3.2}$$

The SCF, $C_X^\alpha(f)$, of the function $x(t)$ is defined as

$$C_X^\alpha(f) = \frac{S_X^\alpha(f)}{[S_X^{\alpha\,*}\left(f+\frac{\alpha}{2}\right) S_X^\alpha\left(f-\frac{\alpha}{2}\right)]^{1/2}} \tag{3.3}$$

and $S_X^\alpha(f)$ is the spectral correlation density (SCD) of $x(t)$:

$$S_X^\alpha(f) = \lim_{T\to\infty}\lim_{\Delta t\to\infty}\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2}\frac{1}{T}X_T\left(t, f+\frac{\alpha}{2}\right)X_T^*\left(t, f-\frac{\alpha}{2}\right)dt \tag{3.4}$$

where $X_T(t, f)$ is the Fourier transform of $x(t)$ and α is a cycle frequency [5].

Figure 3.2 shows the ideal reference α-profiles for the four modulation types used for this work. These modulation schemes were simulated using MATLAB with 2000 symbols at a baud rate of 100Hz, sampled at 500Hz. The α-profiles were calculated with 100 discrete cycle frequency values and a carrier frequency to sampling frequency ratio of F_C/F_S=0.2.

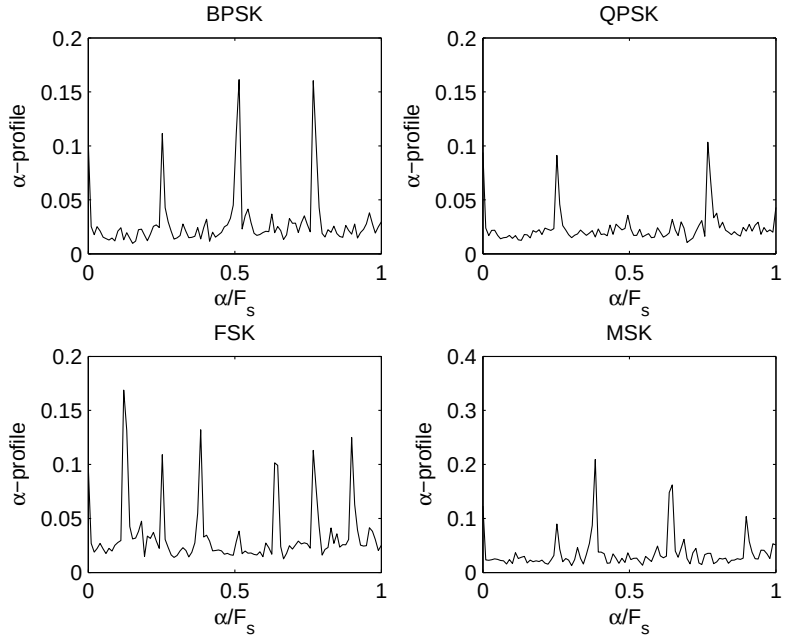


Figure 3.2: α-Profiles for BPSK, QPSK, FSK and MSK

The cyclic resolution of the α-profiles produced by the MATLAB code from [24] used in this work is very coarse (high $\Delta\alpha$) and a likely improvement in performance should immediately occur if this resolution is increased [14]. In the $S_X^\alpha(f)$ limit equation (3.4), Δt is the amount of time over which the spectral components are correlated. This limits the temporal resolution of the signal to Δt and thusly produces a cyclic resolution of approximately $\Delta\alpha = 1/\Delta t$ [39]. A frequency smoothing factor of M=100, as defined in [47],was used in the creation of the α-profile.

3.2 Data Preprocessing

Any pattern recognition modulation classifier will require three steps: data acquisition and preprocessing, feature extraction and representation, and decision-making [48]. Data preprocessing can be defined as the transferring of an original sequence of data into a comparable sequence for the purpose of equal comparison by the identical scaling and weighting of the data [34]. The application of GRA to wireless communications automatically creates the need for data preprocessing, as no wireless signal can be transmitted without being affected by the environment it travels through. Each known modulation type produces a different arrangement of peaks in its SCD plot that are then exhibited by the α-profile [8]. However, in low SNR conditions, the performance of the SCF will become worse, since the spectrum lines associated to important parameters such as carrier frequency are obscured by those associated with noise [49]. Therefore, data preprocessing must be conducted to convert the original data into a comparable vector of data so everything is equally scaled, the effects of noise are diminished and the distinguishing peaks stand out [21, 34].

As frequency offsets and additive noise occur, these effects are transferred to the α-profile. The effect of frequency offset will be investigated in Section 3.4. An entire reference library of signals in every type of noisy environment is not feasible and to even attempt to create such a set of reference sequences is impractical. A large reference library that accounts for as many frequency offsets and noise conditions as possible would be enormous and the latency associated with comparing the incoming signal to the large reference library would defeat the purpose of using the fast-deciding GRA. The received signal must instead be compared to the reference vectors that contain the signal features of the theoretical ideal (zero noise) modulations. To counter the effects of noise on the received signal (and consequentially any extracted signal features) both normalization and weighting are implemented to determine the best combination of preprocessing techniques for this application. This preprocessing will attempt to overcome the effects of noise to more closely match the incoming modulation to one in the reference set.

3.2.1 Normalization

A noisy RF environment can affect these SFV metrics greatly, so the first step is to normalize the received data. Three different normalization methods will be investigated later in this chapter and are outlined in Table 3.1 for the original sequence $\varphi_i^0(k)$.

Table 3.1: Normalization Methods

Maximum Value (MV) [34]	$\varphi_i(k) = \frac{\varphi_i^0(k)}{\max\limits_{\forall i} \varphi_i^0(k)}$
"Higher is Better" (HiB) [34]	$\varphi_i(k) = \frac{\varphi_i^0(k) - \min\limits_{\forall i} \varphi_i^0(k)}{\max\limits_{\forall i} \varphi_i^0(k) - \min\limits_{\forall i} \varphi_i^0(k)}$
Statistical [50]	$\varphi_i(k) = \frac{\varphi_i^0(k) - \mu}{\sigma}$

3.2.2 Weighting

In the case of the α-profile, the features that distinguish between modulation types are the location and magnitude of the peaks. Weighting assigns a quantified value of significance to each data point in the α-profile as a function of the magnitude (i.e. a higher profile magnitude will be assigned a higher significance). The same weighting function can be used for all modulation types and the received signal, if all are normalized by the same method (see Table 3.1).

After Maximum Value (MV) normalization, the distinguishing peaks of all four modulations in low noise were above a magnitude of 0.25. Using this as a threshold, the weighting coefficient for each vector element $\varphi_i(k)$ was originally defined by (3.5) as inspired by [51] and illustrated in Figure 3.3.

$$W(\varphi_i(k)) = \frac{1}{2}\left\{\tanh((|\varphi_i(k)| - 0.25) \cdot 60 - 1) + 1\right\} \tag{3.5}$$

However, if the normalization method is changed, or noise in the channel induces a greater

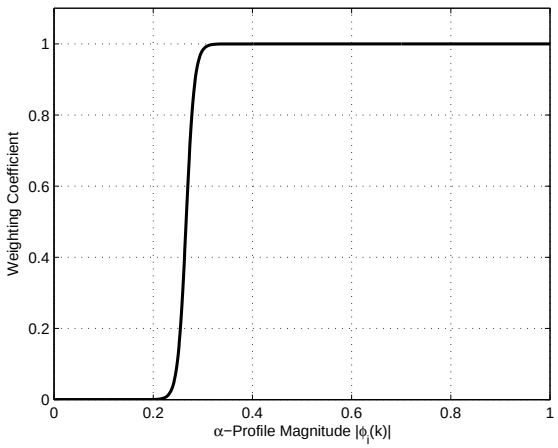


Figure 3.3: Weighting Function Example

number of peaks in the α-profile, the weighting threshold will also need to be changed. A more adaptable weighting method should have a statistical component. Figure 3.3 illustrates that the tanh weighting function is practically a hard-decision threshold, where peaks above 0.25 are weighted to 1 or almost 1 and the noisy data around the peaks is cast away.

In order to have an adaptable weighting scheme, a statistical hard-decision threshold (3.6) was used. The self-adjusting threshold value is the mean of the α-profile vector plus one standard deviation. This method has the flexibility to be applied to any vector, no matter how it is normalized.

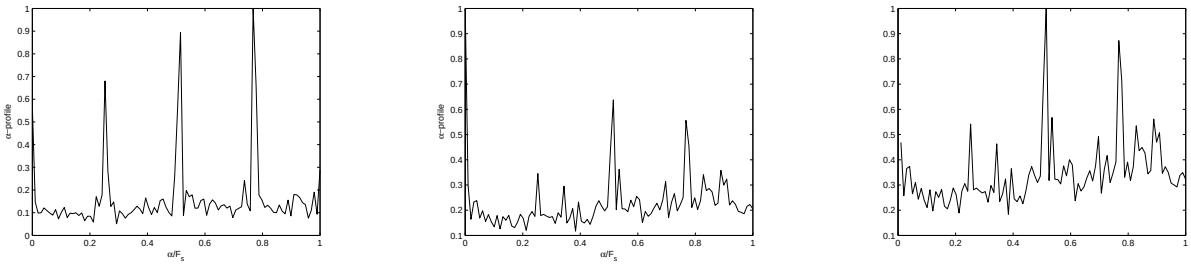
$$W(\varphi_i(k)) = \begin{cases} 1; & \varphi_i(k) \geq \mathrm{E}\left[\Phi(k)\right] + \sigma_{\Phi(k)} \\ 0; & \varphi_i(k) < \mathrm{E}\left[\Phi(k)\right] + \sigma_{\Phi(k)} \end{cases} \tag{3.6}$$

As baud rate and carrier frequency change, the cyclic resolution of the α-profile will also change. Therefore expecting the exact-numbered data point where the distinguishing peaks will occur in the α-profile vector requires observing the exact same frequency band and perfect synchronization with the signal every time. If this can be achieved, then the most ideal weighting scheme would be to weight those specific data points containing distinguishing

peaks to 1 and all else to 0. However, for an adaptive receiver and classifier this method is simply not practical as it is dependent on not only an exact knowledge of how the α-profile will appear, but is also dependent on the resolution of the Fast Fourier transforms (FFTs) used in the calculation of the α-profile.

3.2.3 α-Profile Abridgment

Grey Relational Analysis is a tool designed to focus on the differences between two or more vectors of information. To take full advantage of the GRA, any traits that are common to all four modulations are useless for distinguishing one modulation from another. The performance of the classifier is dependent upon the uniqueness of the information contained in each SFV. Therefore, since the SCF defined by (3.3) has a magnitude constrained to be within $[0, 1]$ with $C_X^0(f) = 1$ for all f [15], the first element of each α-profile was discarded to create an abridged profile. The resulting abridged α-profiles have more distinguishing characteristics after normalization (especially when received in a noisy environment) when the rest of the data points were no longer scaled relative to the always-largest $C_X^\alpha(f)$ value at $\alpha = 0$.



(a) (b) (c)

Figure 3.4: Abridgment Effect on BPSK α-Profile: (a) Theoretical (SNR=∞), MV Normalized; (b) SNR=-5dB, MV Normalized; and, (c) SNR=-5dB, Abridged and MV Normalized.

Figure 3.4 illustrates the effect of abridgment with MV normalization on a BPSK α-profile. With no abridgment, the $C_X^0(f)$ element dominates, but when the profile is abridged first,

the three distinguishing peaks are much closer in magnitude to those of the ideal α-profile.

3.3 Single Feature Classification Results

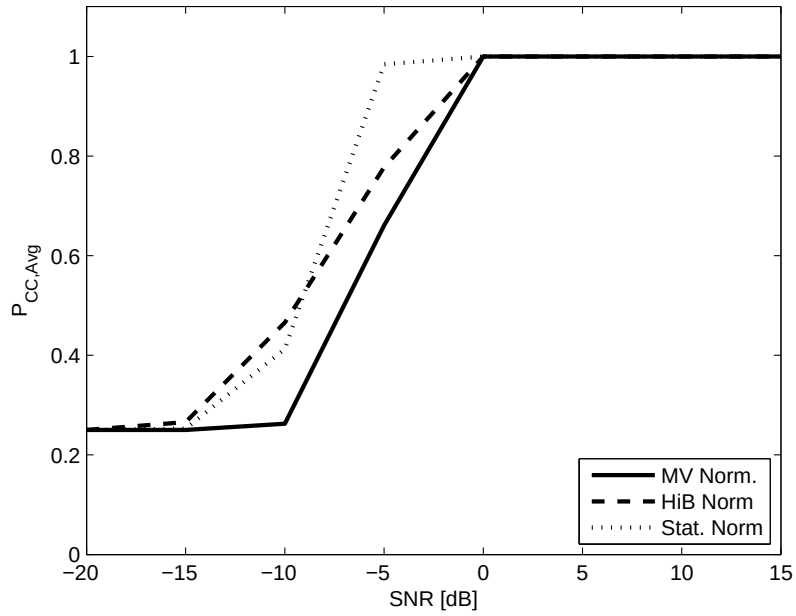
This section summarizes the results of combining all of the previously-discussed data preprocessing methods in an effort to determine and select the combination that yields the best GRA-AMC performance. The top-performing data preprocessing combination will then be used in subsequent chapters by the GRA-AMC as additional features are added to the SFV.

The overall probability of classification did increase with hard-decision weighting for MV normalization, and "Higher is Better" (HiB) normalization with an abridged profile. However, the most dramatic difference was found when the abridged profile was implemented instead of the complete profile. Taking away the data point that all four modulations had in common increased the effect of the distinguishing peaks, despite any noise that was also in the α-profiles. Figure 3.5 and Fig. 3.6 illustrate the resulting effects between all three normalization methods.

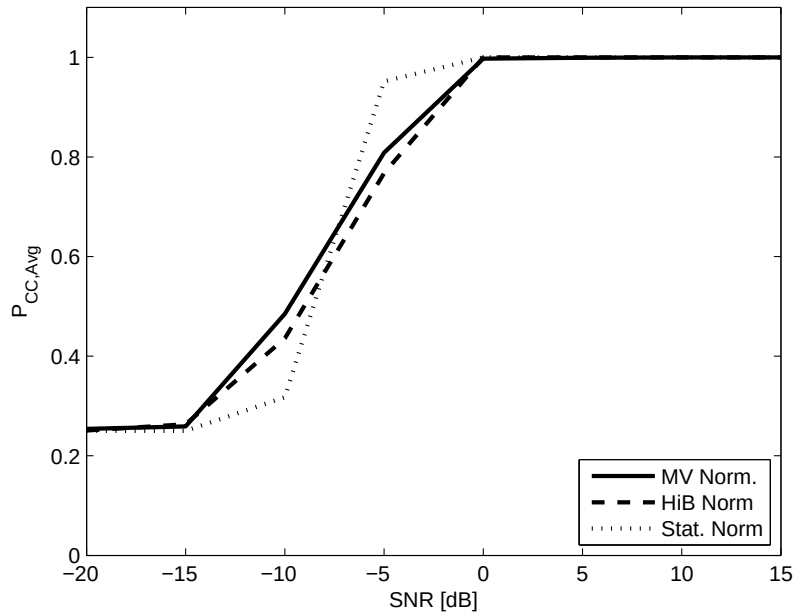
While profile abridgment had a positive impact on $P_{CC,\ Total}$, weighting had a major impact with only the simplest normalization method: MV normalization. This effect is tabulated in Tables 3.2 and 3.3. The overall best combination was consistently a statistical normalization of the α-profile with no weighting. This combination yielded both the highest $P_{CC,\ Total}$ and the lowest $\overline{P}_{fp,max}$. This is the trend that must be continued to show improvement with the addition of other signal features in the comparison vectors.

Figure 3.8 illustrates the most favorable data preprocessing method results. Figure 3.8a shows the probability of classification for each of the four modulations as SNR is incremented, and Fig. 3.8b is an illustration of the Receiver Operating Characteristics (ROC) curve.

The ROC curve plots P_{fp} against P_{CC}. The point (0,1) represents perfect classification of the modulation with no false positives while the point (1,1) indicates always classifying

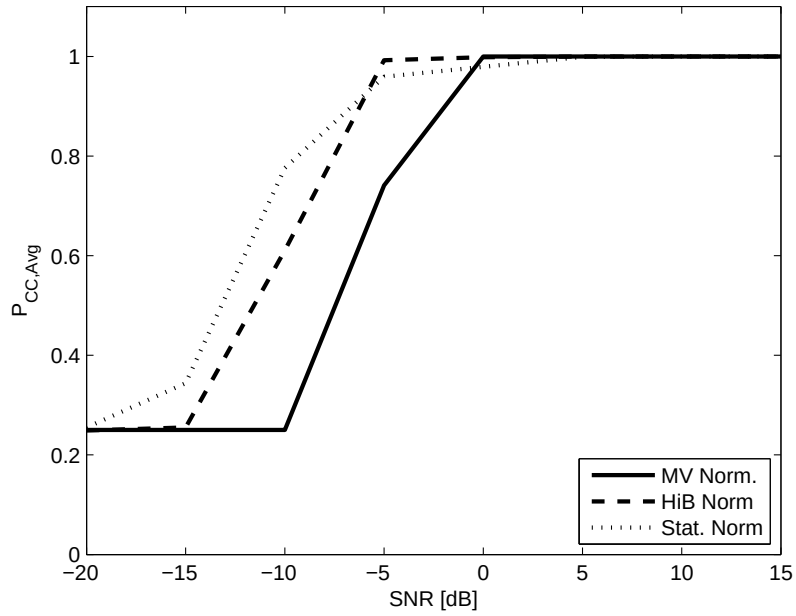


(a)

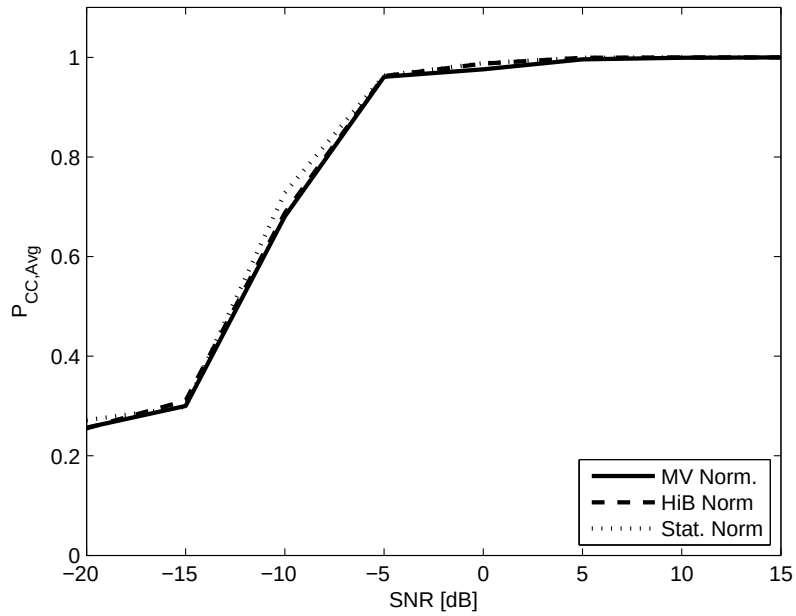


(b)

Figure 3.5: Single Feature Performance Results: (a) Complete α-Profile; (b) Complete α-Profile + HD.



(a)



(b)

Figure 3.6: Single Feature Performance Results: (a) Abridged α-Profile; (b) Abridged α-Profile + HD.

Table 3.2: $P_{CC,\ Total}$: Complete(C) or Abridged(A) Profile and/or Hard-Decision Weighting.

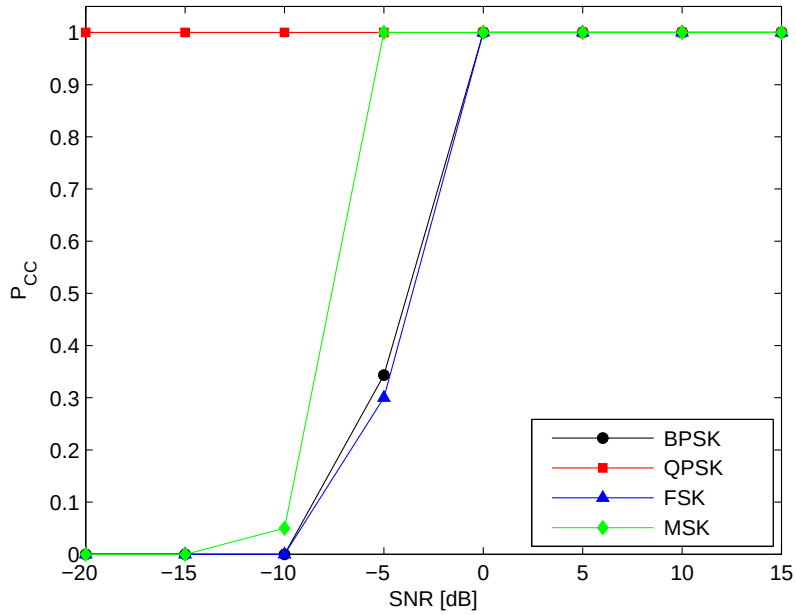
Normalization	C. Profile	C. Profile+HD	**A. Profile**	A. Profile+HD
Maximum Value	48.47%	56.08%	49.82%	63.48%
"Higher is Better"	55.18%	54.35%	62.12%	64.07%
Statistical	57.98%	55.38%	**66.27%**	65.00%

Table 3.3: $\overline{P}_{fp,max}$: Complete(C) or Abridged(A) Profile and/or Hard-Decision Weighting.

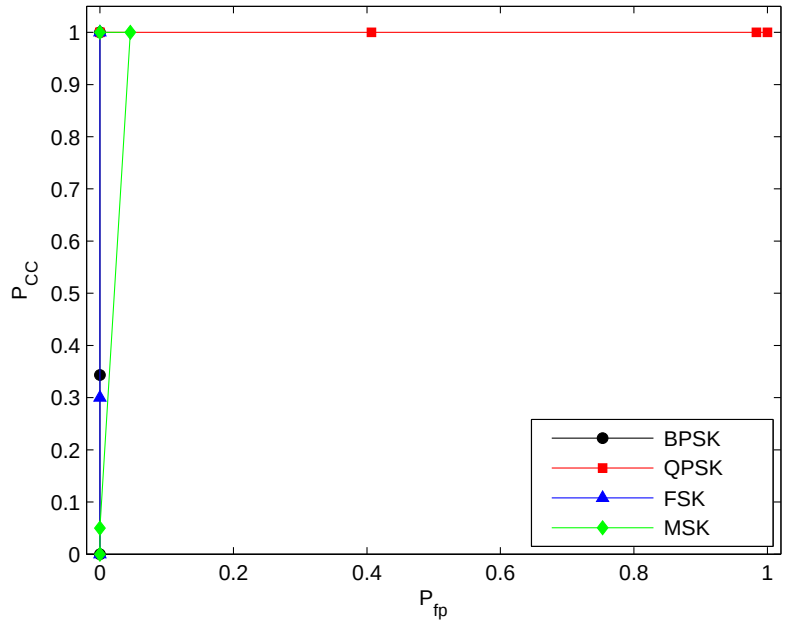
Normalization	C. Profile	C. Profile+HD	**A. Profile**	A. Profile+HD
Maximum Value	26.14%	28.22%	25.00%	25.00%
"Higher is Better"	31.75%	31.94%	25.47%	25.03%
Statistical	25.53%	26.39%	**24.83**%	25.19%

the received signal as one particular modulation, whether or not it is correct [52]. The performance of the GRA-AMC is better if the ROC curve falls faster. ROC plots often have a curve that rises steeply and moves towards P_{fp}=1 and P_{CC}=1 [30]. However, this is characteristic of a receiver that indicates the presence or absence of a signal. Since the proposed GRA-AMC always assumes that there is a signal present, it will always select one of the four possible modulation types, creating a false positive for one of the other three. So, as the P_{CC} of one modulation decreases, P_{fp} of one or all other modulations will increase. The goal is to increase $P_{CC,Total}$ while keeping the maximum P_{fp} of all four modulations as low as possible.

To illustrate non-favorable results, Fig. 3.7 shows the performance results of using MV normalization only, with no weighting. For this data preprocessing combination, the probability of classification for BPSK, FSK, and MSK drops off quickly after 0dB SNR. However, QPSK P_{CC}=100% for all SNR. This is only because the GRA determines QPSK to be the most likely choice of the four modulations. QPSK has the fewest number of distinguishing peaks

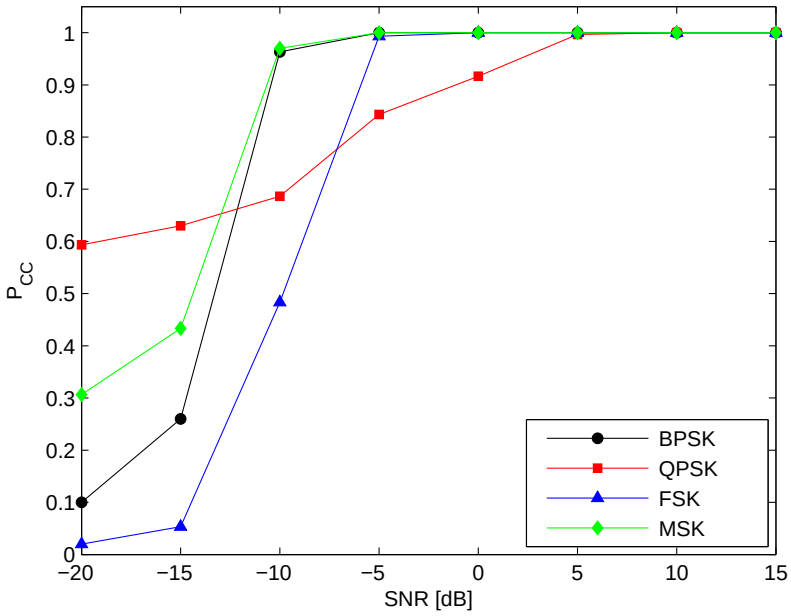


(a) P_{CC} vs. SNR

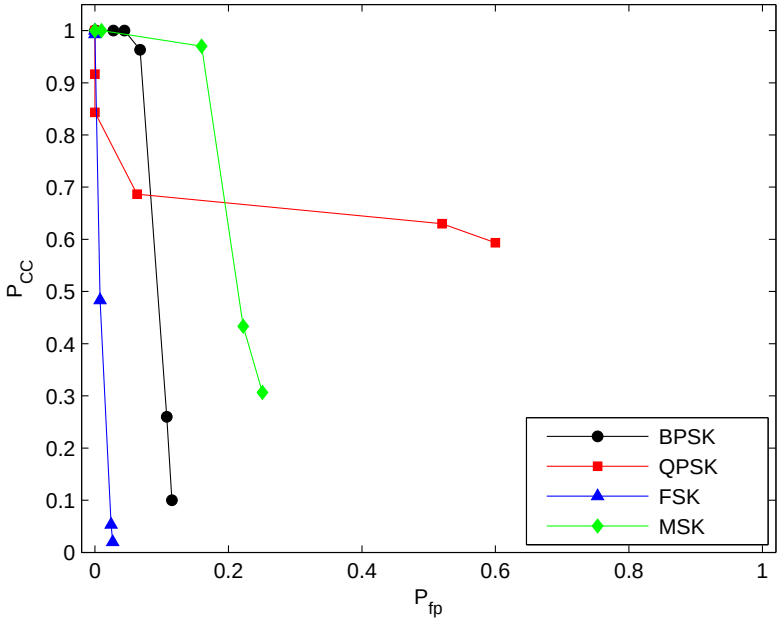


(b) ROC Curve

Figure 3.7: Single Feature Performance Results: MV Normalization with Complete α-Profile



(a) P_{CC} vs. SNR



(b) ROC Curve

Figure 3.8: Single Feature Performance Results: Statistical Normalization with Abridged α-Profile

Table 3.4: MV Norm., Complete α-Profile Confusion Matrix: -20dB to 5dB

		Estimated			
		BPSK	QPSK	FSK	MSK
True, 5dB	BPSK	300	-	-	-
	QPSK	-	300	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, 0dB	BPSK	300	-	-	-
	QPSK	-	300	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -5dB	BPSK	103	197	-	-
	QPSK	-	300	-	-
	FSK	-	169	90	41
	MSK	-	-	-	300
True, -10dB	BPSK	-	300	-	-
	QPSK	-	300	-	-
	FSK	-	300	-	-
	MSK	-	285	-	15
True, -15dB	BPSK	-	300	-	-
	QPSK	-	300	-	-
	FSK	-	300	-	-
	MSK	-	300	-	-
True, -20dB	BPSK	-	300	-	-
	QPSK	-	300	-	-
	FSK	-	300	-	-
	MSK	-	300	-	-

Table 3.5: Stat. Norm., Abridged α-Profile Confusion Matrix: -20dB to 5dB

		Estimated			
		BPSK	QPSK	FSK	MSK
True, 5dB	BPSK	300	-	-	-
	QPSK	1	299	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, 0dB	BPSK	300	-	-	-
	QPSK	25	275	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -5dB	BPSK	300	-	-	-
	QPSK	40	253	-	7
	FSK	-	-	298	2
	MSK	-	-	-	300
True, -10dB	BPSK	289	8	1	2
	QPSK	42	206	5	47
	FSK	16	44	145	95
	MSK	3	5	1	291
True, -15dB	BPSK	78	162	5	55
	QPSK	30	189	8	73
	FSK	32	180	16	72
	MSK	35	126	9	130
True, -20dB	BPSK	30	192	9	69
	QPSK	31	178	9	82
	FSK	40	179	6	75
	MSK	33	169	6	92

in its α-profile and therefore will have the smallest amount of total distance between the ideal QPSK profile and random noise with no distinguishing peaks. These P_{CC} results are misleading and this is the reason why Fig. 3.7b must be shown in juxtaposition. Although it appears that QPSK is always being classified correctly in Fig. 3.7a, the adjacent Fig. 3.7b ROC curve indicates the probability of a QPSK false positive approching 100% at the same time.

In contrast, Fig. 3.8a indicates worse QPSK classification at low SNR. However, not only do the other modulations have a higher probability of classification, but the P_{fp} values in Fig. 3.8b are not dominated by a single modulation as in 3.7b.

The corresponding confusion matrices for the results shown in Fig. 3.7 and Fig. 3.8 are tabulated in Table 3.4 and Table 3.5respectively. The confusion matrix is a simple tool that can quickly show insight into the classifier behavior. For an ideal classifier all results, each 100%, are contained in the diagonal matrix elements [30]. The confusion matrices were tabulated for SNR ranging between -20dB and 0dB.

3.4 Effect of Frequency Offset

As mentioned in Section 3.2, factors such as frequency offset have an influence over the calculated cyclostationary profile. This section addresses the deterioration of classifier performance that results from too much frequency offset.

The results in the other sections of this thesis assume perfect synchronization and a known carrier frequency and baud rate. This is usually not the case in real-world scenarios and there is almost always some degree of frequency offset at the receiver. This section investigates frequency offset using the model defined in [53], where normalized carrier frequency offset is defined as $\epsilon = \Delta f T_s$, where T_s is the timing between consecutive samples and Δf represents the carrier frequency offset. The value of ϵ was set to be 0.01, creating an offset 5% of the carrier frequency for these simulations.

The effect of frequency offset is compounded by the coarse cyclic resolution of the α-profiles created for this work by MATLAB code from [24], as discussed in Section 3.1. The effect is best illustrated by the frequency-offset BPSK α-profile in Fig. 3.9.

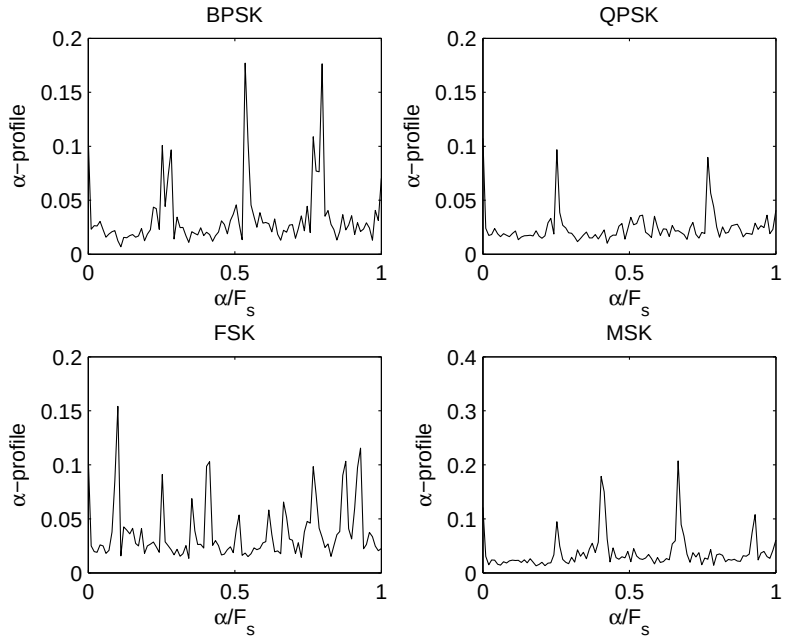


Figure 3.9: α-Profiles for BPSK, QPSK, FSK and MSK at 5% Carrier Frequency Offset

Since GRA determines the total distance between the calculated cyclostationary profile and each stored reference profile, a shift in the location of peaks should make little difference. However, this shift can inadvertently creates two peaks from one, or cause a peak that was previously represented by two data points, to become a peak with a single data point width. As there are only a few large-valued data points that are significant, the deletion or addition of a peak-valued data point has a large impact.

The cyclic resolution must be increased to show clear and distinct peaks, even as the observed frequency changes. Unfortunately, time constraints did not make this possible. Figure 3.10 shows the extreme performance decrease when cyclic resolution remains coarse and frequency offset occurs.

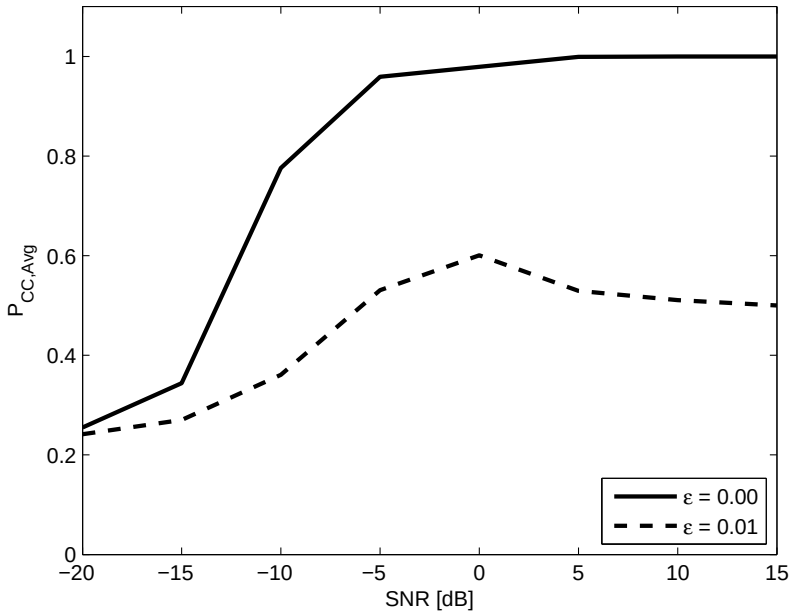


Figure 3.10: Frequency Offset Effect

3.5 Redefining GRA Distance

Different distance metrics have found use in quantifying similarity for case-based reasoning [54] and determining image quality [55, 56]. The GRA-AMC simulations to this point have used Euclidean distance to measure the difference between the observed signal feature vector, Φ_{test}, and the matrix of reference signal features, Φ_{comp}. This section briefly explores the effect of redefining the distance metric used by the GRA algorithm to compare α-profiles. To accomplish this, (2.8) was redefined with the generic distance metric $D(k)$ and the squared exponent of the inner term was replaced with p, to create the alternate equation (3.7).

$$r(k) = \exp\left[\xi\left(\frac{D(k)}{\Delta\varphi_{max} - \Delta\varphi_{min}}\right)^{p}\right] \tag{3.7}$$

Minkowski, Chebyshev, and Canberra distance metrics were all tried. The commonly-used Manhattan (or city block) distance and Euclidean distance functions are variants of the

Minkowski r-distance function [57]. All distance metrics are summarized in Table 3.6.

Table 3.6: Distance Metrics

Manhattan	$D(k) = \sum_{i=1}^{n} \lvert\varphi_i(0) - \varphi_i(k)\rvert$
Euclidean	$D(k) = \sqrt{\sum_{i=1}^{n} \lvert\varphi_i(0) - \varphi_i(k)\rvert^2}$
Minkowski	$D(k) = \left(\sum_{i=1}^{n} \lvert\varphi_i(0) - \varphi_i(k)\rvert^r\right)^{1/r}$
Chebyshev	$D(k) = \max_{\forall i} \lvert\varphi_i(0) - \varphi_i(k)\rvert$
Canberra	$D(k) = \sum_{i=1}^{n} \frac{\lvert\varphi_i(0)-\varphi_i(k)\rvert}{\lvert\varphi_i(0)+\varphi_i(k)\rvert}$

Both r and p were incremented from 1 to 3 and all possible combinations were attempted for Minkowski distance. The r value does not apply to Chebyshev and Canberra but p was incremented for these as well. The performance results of the GRA-AMC using all tested distance metrics are listed in Table 3.7.

The Minkowski $r = 3$ distance yielded slightly better results for both $\mathrm{P_{CC}}$ and $\overline{\mathrm{P}}_{\mathrm{fp,max}}$, but the true improvement in performance cannot be observed from these metrics alone. Figure 3.11 shows that $\mathrm{P_{CC,Avg}}$ is almost identical for Minkowski $r = 2$ and Minkowski $r = 3$ distance. The true effect can be observed in more detail in Fig. 3.12 and Table 3.8. Instead of three modulations performing well at the expense of the fourth (QPSK) when using Euclidean distance, the use of Minkowski $r = 3$ distance averages out the performance of all four modulation curves. Each of the four ROC curves in Fig. 3.12b is steep, illustrating an overall good performance. This is a tremendous improvement over the Euclidean distance results (Fig. 3.8) which show a high false positive rate for QPSK as the noise floor is increased.

The GRA-AMC is not designed to be a signal detector. It assumes that there is always a modulation present in the data that it collects for input. This means that one of the four

Table 3.7: Single Feature Performance Results with Redefined GRA Distance

Distance Metric	r	p	$P_{CC,\ Total}$	$\overline{P}_{fp,max}$
Minkowski	1	1	59.77%	26.33%
	1	2	59.87%	27.00%
	1	3	39.68%	25.00%
	2	1	66.27%	24.83%
	2	2	66.27%	24.83%
	2	3	65.97%	24.78%
	3	1	66.47%	24.56%
	3	2	67.52%	24.56%
	3	3	67.37%	24.81%
Chebyshev	-	1	64.02%	24.69%
	-	2	62.90%	25.92%
Canberra	-	1	30.20%	26.22%
	-	2	27.72%	28.17%

Table 3.8: Minkowski r=3 Distance Confusion Matrix: -20dB to 5dB

		Estimated			
		BPSK	QPSK	FSK	MSK
True, 5dB	BPSK	300	-	-	-
	QPSK	1	299	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, 0dB	BPSK	300	-	-	-
	QPSK	10	290	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -5dB	BPSK	300	-	-	-
	QPSK	36	257	2	5
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -10dB	BPSK	296	2	-	2
	QPSK	57	146	62	35
	FSK	12	18	232	38
	MSK	4	-	1	295
True, -15dB	BPSK	133	55	62	50
	QPSK	62	85	89	64
	FSK	66	76	104	54
	MSK	65	57	81	97
True, -20dB	BPSK	75	61	103	61
	QPSK	73	68	92	67
	FSK	68	68	110	54
	MSK	72	66	99	63

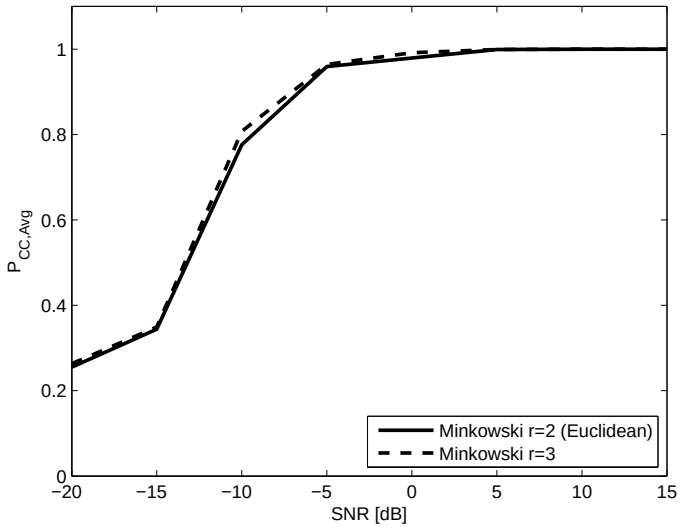
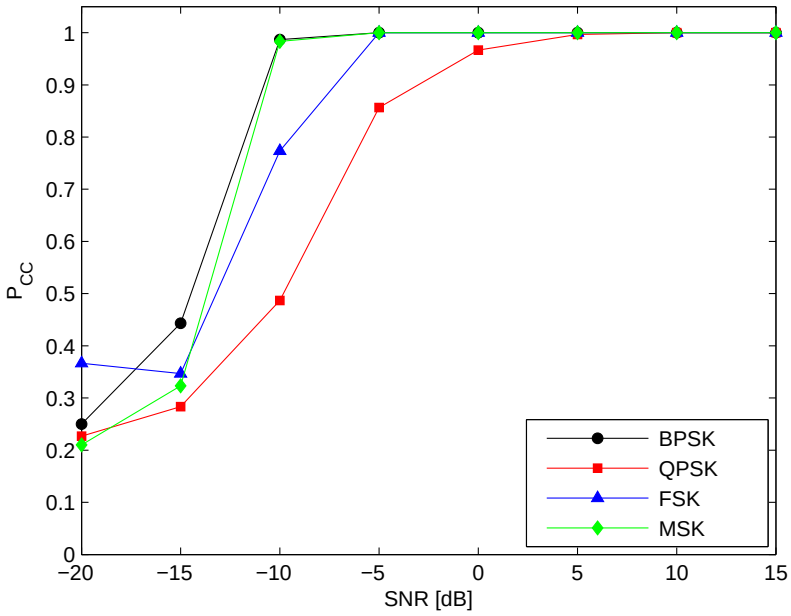


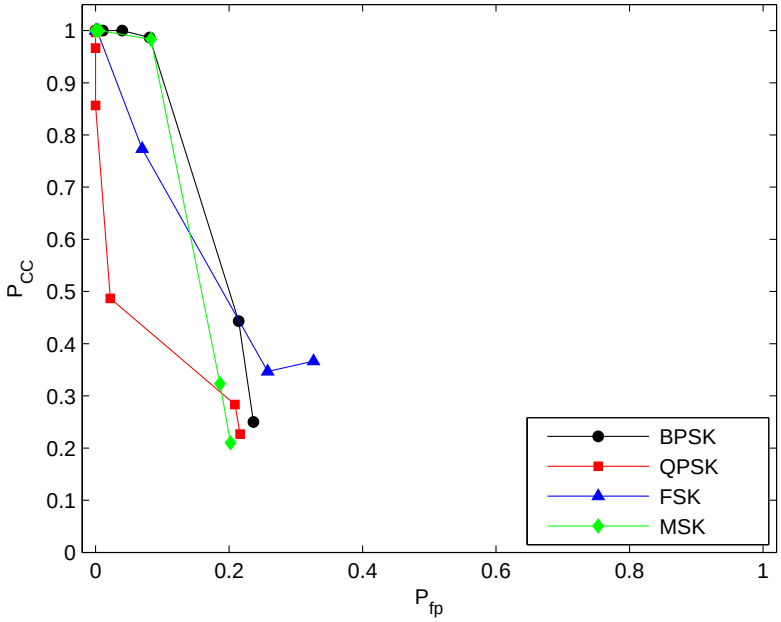
Figure 3.11: $P_{CC,Avg}$: Minkowski r=2 Distance vs. Minkowski r=3 Distance

modulation choices will be selected and it is only a matter of which modulation is the most likely choice. The goal of using different distance metrics for GRA was to eliminate any overwhelming bias for or against one particular modulation (in this case, a bias for QPSK above all other modulations when in high noise). Minkowski $r = 3$ distance successfully dampens this bias, causing all four modulations to have much closer correct-classification and false-positive rates.

Although there was significant performance improvement as a result of using Minkowski $r = 3$ distance, Chapter 4 continues to use Euclidean distance for the GRA so the performance resulting from the addition of signal features to the SFV can be investigated without bias.



(a) P_{CC} vs. SNR



(b) ROC Curve

Figure 3.12: Single Feature Performance Results: Statistical Normalization with Abridged α-Profile and Minkowski r=3 Distance

Chapter 4

Signal Feature Data Fusion

This chapter investigates the effect of combining two completely different signal feature metrics into a single vector for comparison by the GRA algorithm. Different signal features are fused into the Signal Feature Vector for the purpose of improving classifier performance by adding more distinguishing signal attributes. Both wavelet magnitude variance and stationary cumulants will be combined with the cyclostationary profile to form new SFVs for comparison by the GRA.

4.1 Cyclostationary Profile, Wavelet Magnitude Variance Data Fusion

The wavelet transform is widely used in many fields of signal processing, especially image compression, speech processing, and modulation classification [16]. Applying a wavelet transform to a digital modulation creates distinct patterns [27] from the transient information of the signal that can then be used for modulation classification, just as distinct patters are created from the formation of the α-profile in Chapter 3. The variance of the magnitude of the received signal Haar continuous wavelet transform (CWT) for two different scales

($a = 2, 3$) was used and added as a second type of signal feature for the SFV, redefined in (4.1). Larger scales represent coarser and coarser resolutions as a is increased.

In order for the added wavelet metrics to notably improve performance of the GRA-AMC, the α-profile is first condensed into a single-valued metric. Simply adding two wavelet metrics to an α-profile vector that already contains 99 noise-corrupted data points will have a minimal effect. So, the total distance between the calculated α-profile and each reference α-profile is determined and placed side-by-side with the additional wavelet metrics so that each new metric can have a greater impact on the classifier performance.

Because $\sum \Delta\,(\alpha\text{-profile})$ is different for each reference vector, $\Phi_{comp}\,(k)$, the element in Φ_{test} that is assigned $\sum \Delta\,(\alpha\text{-profile})$ is given a null value so the $\sum \Delta\,(\alpha\text{-profile})$ closest to zero is determined to be the most similar. The new Φ_{test} and $\Phi_{comp}\,(k)$ are defined by (4.1a) and (4.1b) respectively.

$$\Phi_{test} = \Big[0, \quad \mathrm{Var}\left(|CWT|_{a=2}\right), \quad \mathrm{Var}\left(|CWT|_{a=3}\right)\Big] \tag{4.1a}$$

$$\Phi_{comp}\,(k) = \Big[\sum \Delta\,(\alpha\text{-profile}), \quad \mathrm{E}\left[\mathrm{Var}\left(|CWT|_{a=2}\right)\right], \quad \mathrm{E}\left[\mathrm{Var}\left(|CWT|_{a=3}\right)\right]\Big] \tag{4.1b}$$

The variance of the magnitude of the CWT was determined after peak removal. Peaks occur on symbol transitions and a median filter is necessary to suppress these peaks and keep them from affecting the variance calculation [21]. Peaks were removed through the use of a median filter with a width of 10 to improve results as recommended by [28]. Wavelets have the advantage of being calculated with fast algorithms [58] just like cumulants in Section 4.2.

In the published literature, there have been examples of signal classification using only one wavelet scale [28] and many wavelet scales [16]. But when one scale has been used, the conclusion from [28] was that "combining the results from several scales will improve the identification accuracy". Taking this advice, two different scales of the Haar CWT were

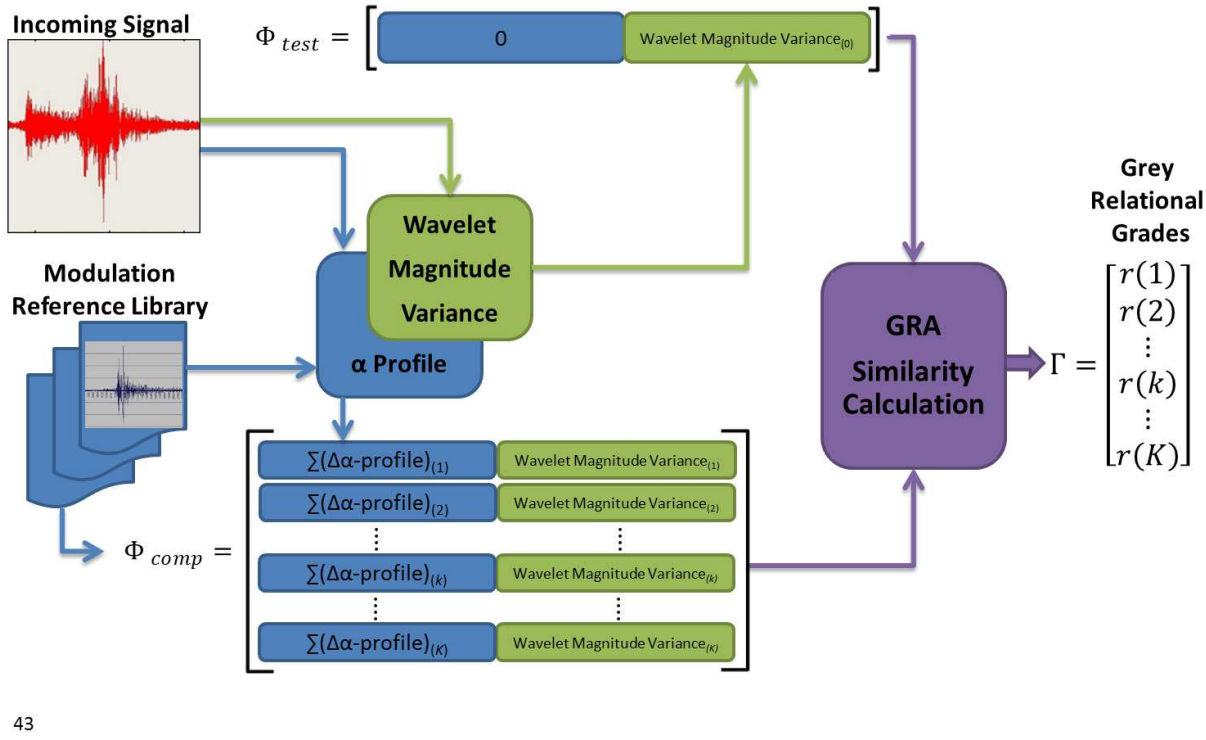


Figure 4.1: Double Feature GRA-AMC Architecture: α-Profile and Wavelet Magnitude Variance

calculated and added to the SFV.

4.1.1 The Haar Wavelet Transform

Let $s(t) = \widetilde{s}(t)\, e^{j(\omega_c t+\theta_c)}$ represent the transmitted complex signal, where ω_c is carrier frequency and θ_c is carrier phase. In (4.2) S is the signal power, N is the number of observed symbols, T is the symbol period and $u_T(t)$ is the standard unit pulse of duration T.

$$\widetilde{s}_{PSK}(t) = \sqrt{S}\sum_{i=1}^{N} e^{j\varphi_i} u_T(t-iT), \varphi_i \in \left\{\frac{2\pi}{M}(m-1), m = 1, 2, \ldots, M\right\} \tag{4.2a}$$

$$\widetilde{s}_{FSK}(t) = \sqrt{S}\sum_{i=1}^{N} e^{j(\omega_i t+\theta_i)} u_T(t-iT), \omega_i \in \{\omega_1, \omega_2, \ldots, \omega_M\}, \theta_i \in (0, 2\pi) \tag{4.2b}$$

The continuous Haar wavelet transform (CWT) of the signal $s(t)$ is defined [29] to be

$$CWT(a,\tau) = \frac{1}{\sqrt{a}}\int s(t)\,\Psi^*\left(\frac{t-\tau}{a}\right)dt \tag{4.3}$$

where a is the baby wavelet scale. The Haar wavelet [27] function was chosen for simplicity and the mother wavelet is defined as (4.4).

$$\Psi(t) = \begin{cases} 1; & t \in [0, \frac{1}{2}) \\ -1; & t \in [\frac{1}{2}, 1) \\ 0; & t \notin [0, 1) \end{cases} \tag{4.4}$$

Different modulations have different unique CWTs. When the Haar wavlet is within a symbol time:

$$|CWT_{PSK}(a,\tau)| = \frac{4\sqrt{S}}{\sqrt{a}\omega_c}\sin^2\left(\frac{\omega_c a}{4}\right) \tag{4.5a}$$

$$|CWT_{FSK}(a,\tau)| = \frac{4\sqrt{S}}{\sqrt{a}(\omega_c+\omega_i)}\sin^2\left[\frac{(\omega_c+\omega_i)a}{4}\right] \tag{4.5b}$$

where S_i and ω_i are the power and frequency of symbol i.

Amplitude normalization of signal $\widetilde{s}(t)$was conducted before the CWT was calculated:

$$\widetilde{s}(t) = \frac{\widetilde{s}(t)}{|\widetilde{s}(t)|} \tag{4.6}$$

and from (4.2a) and (4.2b) the amplitude normalized modulated signals [29] become:

$$\widetilde{s}_{PSK}(t) = \sum_{i=1}^{N} e^{j\varphi_i} u_T(t - iT) \tag{4.7a}$$

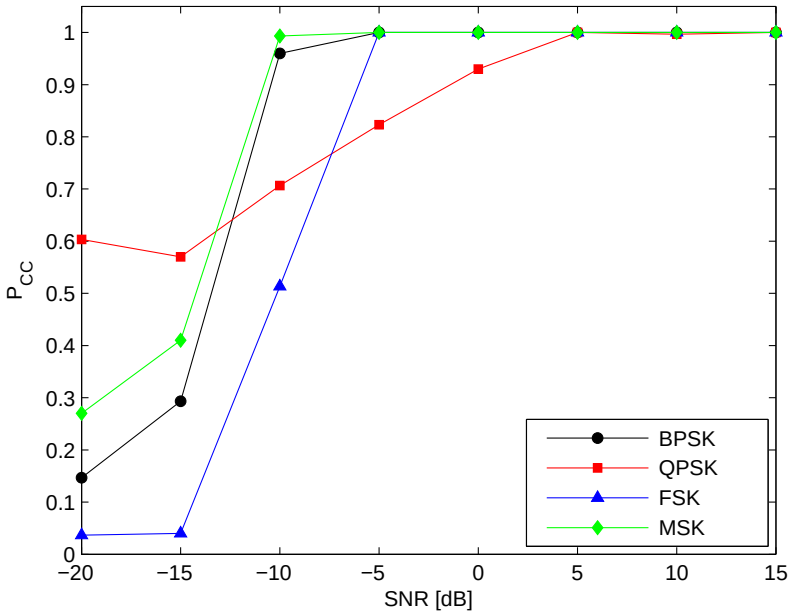
$$\widetilde{s}_{FSK}(t) = \sum_{i=1}^{N} e^{j(\omega_i t + \theta_i)} u_T(t - iT) \tag{4.7b}$$

4.1.2 Classification Results

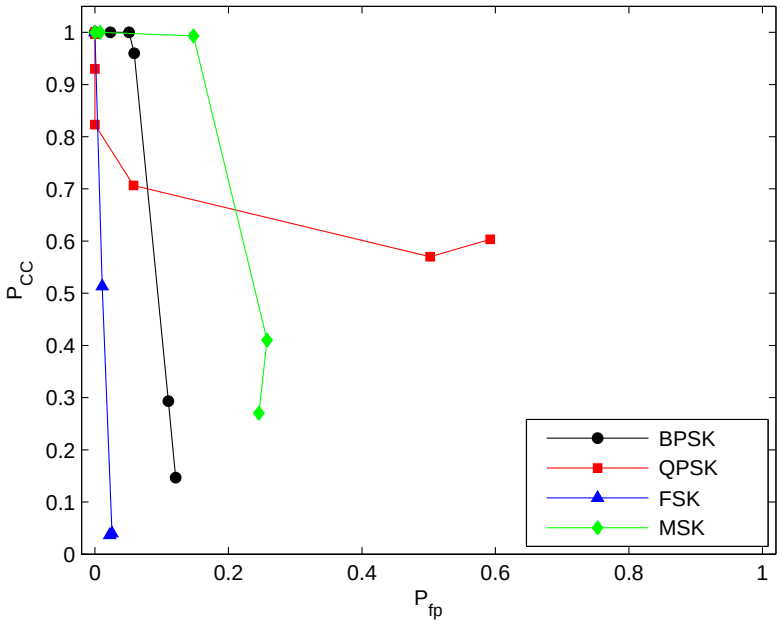
Table 4.1: Data Fusion Comparison: α-Profile, Wavelet Magnitude Variance

Features	$P_{CC,\ Total}$	$\overline{P}_{fp,max}$
α-Profile	66.27%	24.83%
α-Profile, Wavelet Mag. Variance	66.48%	24.92%

Despite the favorability of wavelet transforms for signal classification, a classifier that only uses wavelets will break down below 5dB [16, 27–30]. The lack of statistically-significant improvement in $P_{CC,Total}$ as illustrated in Table 4.1 was therefore expected and predicted in the literature. Instead, the goal of adding wavelet magnitude variance as the second signal feature in the SFV was to cause a performance improvement similar to that caused by the substitution of Euclidean distance for $r = 3$ Minkowski distance in Section 3.5. However, there was no such improvement in performance either.



(a) P_{CC} vs. SNR



(b) ROC Curve

Figure 4.2: Classification Performance: α-Profile, Wavelet Magnitude Variance

Table 4.2: α-Profile, Wavelet Magnitude Variance Fusion Confusion Matrix: -20dB to 5dB

		Estimated			
		BPSK	QPSK	FSK	MSK
True, 5dB	BPSK	300	-	-	-
	QPSK	-	300	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, 0dB	BPSK	300	-	-	-
	QPSK	21	279	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -5dB	BPSK	300	-	-	-
	QPSK	46	247	-	7
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -10dB	BPSK	288	9	-	3
	QPSK	42	212	9	37
	FSK	11	42	154	93
	MSK	-	1	1	298
True, -15dB	BPSK	88	150	5	57
	QPSK	43	171	9	77
	FSK	22	168	12	98
	MSK	34	134	9	123
True, -20dB	BPSK	44	173	6	77
	QPSK	40	181	8	71
	FSK	31	185	11	73
	MSK	38	175	6	81

4.2 Cyclostationary Profile, Stationary Cumulants Data Fusion

This section explores the addition of estimated stationary cumulants as the second signal feature to be combined with the cyclostationary profile to form the SFV. Cumulants of the received signal are calculated in addition to the α-profile and all are combined for comparison. Higher order statistics, such as cumulants, characterize the shape of the distribution of the noisy baseband samples and have been found effective for discriminating between modulation subclasses (i.e. PSK, PAM, QAM) [11]. Just as in Section 4.1, the α-profile is condensed into a single metric before being inserted into the SFV. The new Φ_{test} and $\Phi_{comp}(k)$ are defined by (4.8a) and (4.8b) respectively.

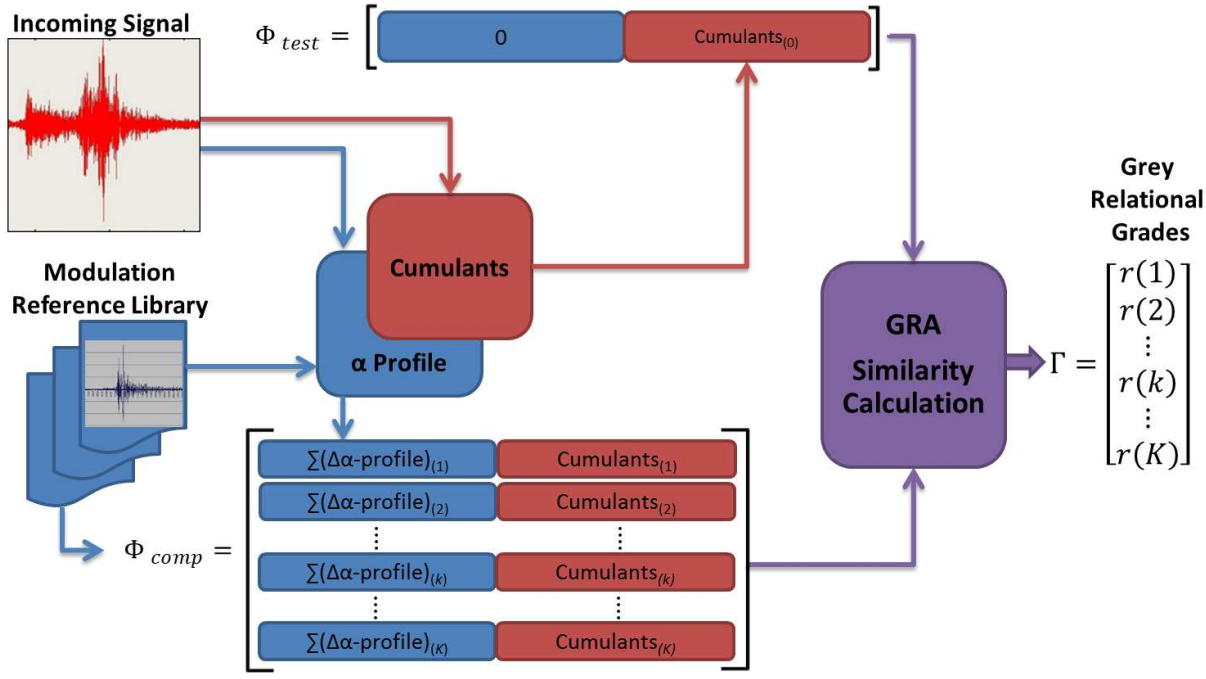


Figure 4.3: Double Feature GRA-AMC Architecture: α-Profile and Cumulants

$$\Phi_{test} = \left[0, \quad \hat{C}_{20}, \quad \hat{C}_{21}, \quad \hat{C}_{40}, \quad \hat{C}_{41}, \quad \hat{C}_{42}, \quad \frac{|\hat{C}_{40}|}{|\hat{C}_{42}|}\right] \tag{4.8a}$$

$$\Phi_{comp}(k) = \left[\sum \Delta(\alpha\text{-profile}), \quad C_{20}, \quad C_{21}, \quad C_{40}, \quad C_{41}, \quad C_{42}, \quad \frac{|C_{40}|}{|C_{42}|}\right] \tag{4.8b}$$

Both stationary [24, 26] and cyclic [59–62] cumulants have been used for modulation classification. A reasonably good estimate of cyclic parameters requires many more samples than for stationary parameters [20] so stationary cumulants were chosen for this work. Sixth-order cumulants were also considered for addition to the signal feature vector, however when SNR is lower than 0dB, the performance of the classifier breaks down when sixth-order cumulants are used. Since the estimation of high-order cumulants amplifies the Gaussian noise power, the influence of Gaussian noise increases as the cumulant order increases [26, 61]. Cyclic cumulants as an individual classifier have better performance in high noise than stationary cumulants, and future work should explore the use of cyclic cumulants in combination with the cyclostationary profile for classification.

4.2.1 Cumulants

For a complex-valued stationary random process $y(n)$ assuming independent and identically distributed (i.i.d.) equiprobable symbols, second-order and fourth-order cumulants can be defined by (4.9) and (4.10).

$$C_{20} = E\left[y^2(n)\right] \tag{4.9a}$$

$$C_{21} = E\left[|y(n)|^2\right] \tag{4.9b}$$

$$C_{40} = \text{cum}\left[y(n), y(n), y(n), y(n)\right] \tag{4.10a}$$

$$C_{41} = \text{cum}\left[y(n), y(n), y(n), y^*(n)\right] \tag{4.10b}$$

$$C_{42} = \text{cum}\left[y(n), y(n), y^*(n), y^*(n)\right] \tag{4.10c}$$

$$\text{cum}\left(w, x, y, z\right) = E\left(wxyz\right) - E\left(wx\right)E\left(yz\right) - E\left(wy\right)E\left(xz\right) - E\left(wz\right)E\left(xy\right) \tag{4.11}$$

For the purposes of testing, estimates of the cumulants in (4.9) and (4.10) must be calculated by averaging over a finite number of samples, assuming that $y(n)$ is zero-mean [11]. These estimates are presented as (4.12) and (4.13).

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} y^2(n) \tag{4.12a}$$

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^2 \tag{4.12b}$$

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} y^4(n) - 3\hat{C}_{20}^2 \tag{4.13a}$$

$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} y^3(n)y^*(n) - 3\hat{C}_{20}\hat{C}_{21} \tag{4.13b}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2 \tag{4.13c}$$

To account for the presence of noise, the estimated normalized cumulants are determined using (4.14) where σ_N^2 is the noise power.

$$\hat{C}_{4k} = \frac{\hat{C}_{4k}}{\hat{C}_{21}^2 - \sigma_N^2}, k = 0, 1, 2 \tag{4.14}$$

Assuming an i.i.d. signal with average power normalized to 1, the associated ideal high-order cumulants can be expressed by Table 4.3. The $\frac{|C_{40}|}{|C_{42}|}$ ratio is an additional metric that has been used to distinguish between PSK and FSK modulations [62] and was therefore also chosen to be part of the SFV.

Table 4.3: Theoretical Values for Cumulants

Constellation	C_{20}	C_{21}	C_{40}	C_{41}	C_{42}	$\frac{\lvert C_{40}\rvert}{\lvert C_{42}\rvert}$
BPSK	1	1	-2	-2	-2	1
QPSK	0	1	1	0	-1	1
FSK	0	1	0	0	-1	0
MSK	0	1	0	0	-1	0

4.2.2 Classification Results

Table 4.4: Data Fusion Comparison: α-Profile, Cumulants

Features	$P_{CC,\ Total}$	$\overline{P}_{fp,max}$
α-Profile	66.27%	24.83%
α-Profile, Cumulants	71.45%	20.31%

The addition of cumulants to the SFV created and improvement in the performance of the GRA-AMC that is summarized by Table 4.4. The confusion matrix (Table 4.5) illustrates that the main improvement came from a greater ability to distinguish between PSK and FSK signals. It is important to remember that the GRA-AMC is comparing SFVs with only seven distinct metrics and is still able to have near-perfect P_{CC} for all four modulations down to -5dB SNR. As SNR goes below -15dB, the P_{CC} for both BPSK and QPSK *increased*. As the noise in the channel becomes overwhelming, the GRA is biased towards choosing a PSK modulation. This is a logical bias since the theoretical cumulant values for FSK and MSK are identical (Table 4.3). Since most of the information contained in the SFV highlights the difference between PSK and FSK, this suggests a useful adaptation by combining the GRA with hierarchical AMC techniques which traditionally use hard thresholds that can be deemed moot in a high noise environment. Artificial Neural Networks have already been combined with hierarchical decision making as discussed in Section 2.1.3 and the application

of the GRA in this way could improve performance with its quick decision algorithm and no need to train the radio in advance.

The true determinant of performance is the type of metrics that are chosen by the designer for comparison. If the chosen signal features are distinguishable at a low SNR, then there is no need for neural network training [15, 16] or hierarchical decision making [11, 13, 24] with the GRA-AMC.

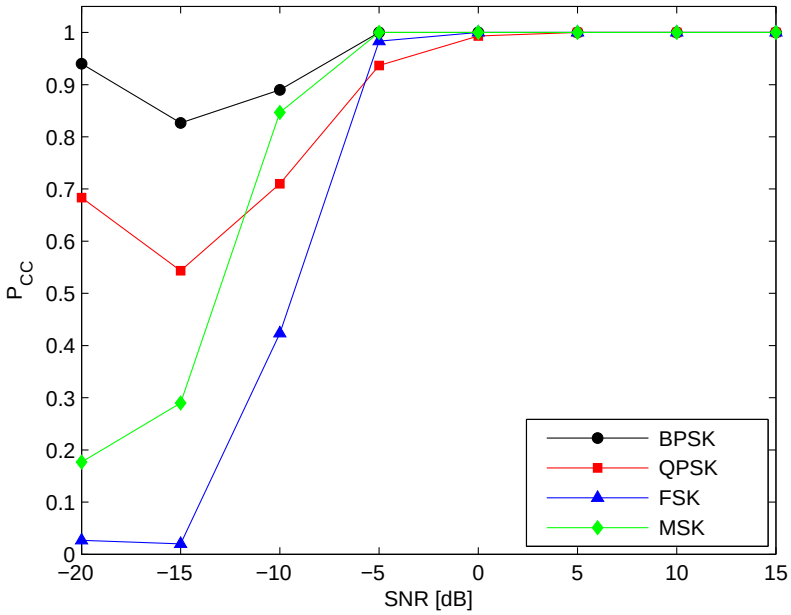
Table 4.6 and Fig. 4.5 compare the classifier performances for the data fusion techniques explored in this chapter to the Chapter 3 single feature classifier performance.

4.2.3 Different Methods for Condensing the α-Profile

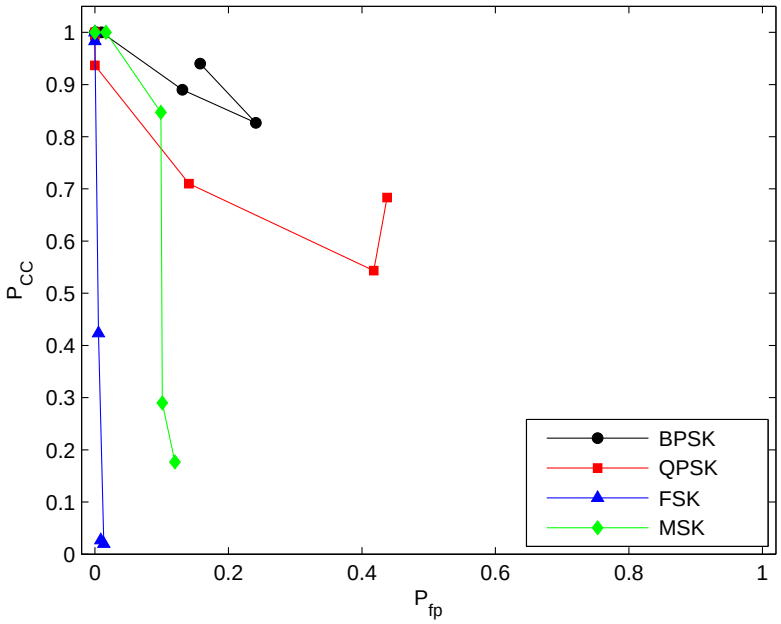
The condensation of the α-profile to create a single-valued metric for the first element of the SFV brings into question what the best method is for such an operation. Additional simulations were conducted to investigate using different quantifications of similarity for the α-profile besides the total Euclidean distance to create the condensed α-profile metric. In $\Phi_{comp}(k)$, the $\sum \Delta(\alpha\text{-profile})$ value was replaced by either cross-correlation or cosine similarity; both of which have a desired ideal value of 1. The Φ_{test} array was redefined as (4.15) and the results are tabulated in Table 4.7.

$$\Phi_{test} = \left[1, \quad \hat{C}_{20}, \quad \hat{C}_{21}, \quad \hat{C}_{40}, \quad \hat{C}_{41}, \quad \hat{C}_{42}, \quad \frac{|\hat{C}_{40}|}{|\hat{C}_{42}|}\right] \tag{4.15}$$

Even with the addition of cumulants to the SFV, other comparative methods for condensing α-profile to a single-valued metric decreased performance. Euclidean distance was found to be the most effective method of the three.



(a) P_{CC} vs. SNR



(b) ROC Curve

Figure 4.4: Classification Performance: α-Profile, Cumulants

Table 4.5: α-Profile, Cumulants Confusion Matrix: -20dB to 5dB

		Estimated			
		BPSK	QPSK	FSK	MSK
True, 5dB	BPSK	300	-	-	-
	QPSK	-	300	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, 0dB	BPSK	300	-	-	-
	QPSK	-	298	-	-
	FSK	-	-	300	-
	MSK	-	-	-	300
True, -5dB	BPSK	300	-	-	-
	QPSK	9	281	-	10
	FSK	-	-	295	5
	MSK	-	-	-	300
True, -10dB	BPSK	267	32	-	1
	QPSK	53	213	5	29
	FSK	38	76	127	59
	MSK	27	19	-	254
True, -15dB	BPSK	248	49	-	3
	QPSK	82	163	8	47
	FSK	79	174	6	41
	MSK	56	153	4	87
True, -20dB	BPSK	282	18	-	-
	QPSK	42	205	4	49
	FSK	45	188	8	59
	MSK	55	188	4	53

Table 4.6: Data Fusion Performance Summary

Features	$P_{CC,\ Total}$	$\overline{P}_{fp,max}$
α-Profile	66.27%	24.83%
α-Profile, Wavelet Mag. Variance	66.48%	24.92%
α-Profile, Cumulants	71.45%	20.31%

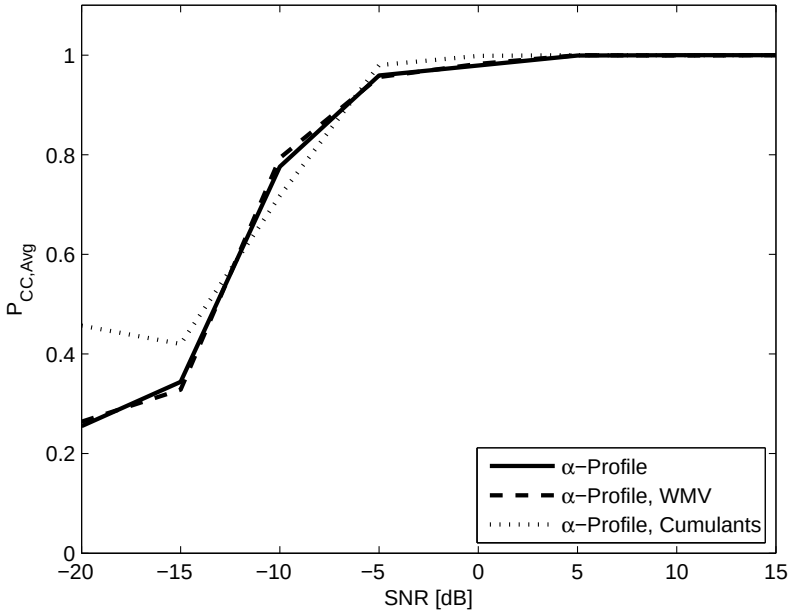


Figure 4.5: Data Fusion $P_{CC,Avg}$ Comparison

Table 4.7: Double Feature Performance Results: α-Profile Similarity Metric Comparison

Similarity Metric	$P_{CC,\ Total}$	$\overline{P}_{fp,max}$
Euclidean Distance	71.45%	20.31%
Cosine Similarity	66.10%	21.36%
Cross-Correlation	64.73%	21.22%

4.2.4 GRA-AMC Extension to Cochannel Signals

An additional advantage to using cumulants is that they help determine if there is more than one occupied channel in the observed spectrum and if so, classify the modulation types of both. This is not particularly feasible with a metric such as the α-profile since all distinguishing peaks would be overlapping, creating a pattern that would be recognizable. Classification of cochannel signals has been successfully conducted with cumulants, but mostly to distinguish between intra-class PSK signals and usually with a high SNR (above 10dB) when the number of observed symbols is low [59, 60]. The GRA algorithm is only designed to identify the one case that is most similar, but the GRA-AMC was modified to indicate the two most likely modulations in the channel in an effort to explore the extent to which GRA could be applied to common communications scenarios. The GRA-AMC performance results are shown in Table 4.8 for signal-to-noise ratio ranging from 0dB to 15dB.

Two signals of the same modulation type were never sent, as the GRA-AMC can only rank the four modulation choices in the reference library by which is most likely. So to accomplish cochannel classification, the top two ranked modulation types were tabulated. The classifier had perfect performance when BPSK was combined with any other modulation. These results were only matched by the FSK/MSK combination. The poorest results were from the QPSK/FSK combination which had an average P_{CC} across the 0dB to 15dB SNR range of approximately 50%.

4.3 Investigating the Uniqueness of Added Signal Feature Information

The performance of GRA-AMC in high noise is entirely dependent upon the performance of the individual performance of each signal feature that is utilized in signal feature data fusion. The addition of wavelet magnitude variance had no effect on overall performance.

Table 4.8: Co-Channel Classification 0dB to 15dB

		Estimated					
		BPSK/QPSK	BPSK/FSK	BPSK/MSK	QPSK/FSK	QPSK/MSK	FSK/MSK
True, 15dB	BPSK/QPSK	300	-	-	-	-	-
	BPSK/FSK	-	300	-	-	-	-
	BPSK/MSK	-	-	300	-	-	-
	QPSK/FSK	-	-	-	108	-	192
	QPSK/MSK	1	-	-	4	258	37
	FSK/MSK	-	-	-	-	-	600
True, 10dB	BPSK/QPSK	300	-	-	-	-	-
	BPSK/FSK	-	300	-	-	-	-
	BPSK/MSK	-	-	300	-	-	-
	QPSK/FSK	-	-	-	167	-	133
	QPSK/MSK	-	-	-	-	240	60
	FSK/MSK	-	-	-	-	-	300
True, 5dB	BPSK/QPSK	-	300	-	-	-	-
	BPSK/FSK	-	300	-	-	-	-
	BPSK/MSK	-	-	300	-	-	-
	QPSK/FSK	127	-	-	173	-	-
	QPSK/MSK	143	-	-	-	157	-
	FSK/MSK	103	140	57	-	-	-
True, 0dB	BPSK/QPSK	300	-	-	-	-	-
	BPSK/FSK	-	300	-	-	-	-
	BPSK/MSK	-	-	283	-	-	17
	QPSK/FSK	1	5	-	150	-	144
	QPSK/MSK	1	-	7	-	38	255
	FSK/MSK	-	-	-	-	-	300

The addition of stationary cumulants caused an increase in $P_{CC,Total}$, mostly because of the resulting tendency to classify all signals as a type of PSK modulation in high noise.

Even when adding signal features that have better individual performance in high noise, a further determination must be made: does the additional signal feature add any unique distinguishing statistical information that the other signal features do not possess? For example, the cyclostationary profiles for QPSK and 8-PSK modulations are identical. This is an instance where the addition of other signal features is essential. The periodicity of the received signal can be interpreted using several different signal features.

Principal Component Analysis (PCA) is a method by which the interdependence of different categories of data within a matrix can be observed [63]. All three signal features that have been used in this work were combined into a single comparison matrix, as illustrated in Fig. 4.6. The purpose of this analysis is to determine which signal features contribute unique characteristics to the comparison matrix and consequentially, the ability to distinguish between modulation types.

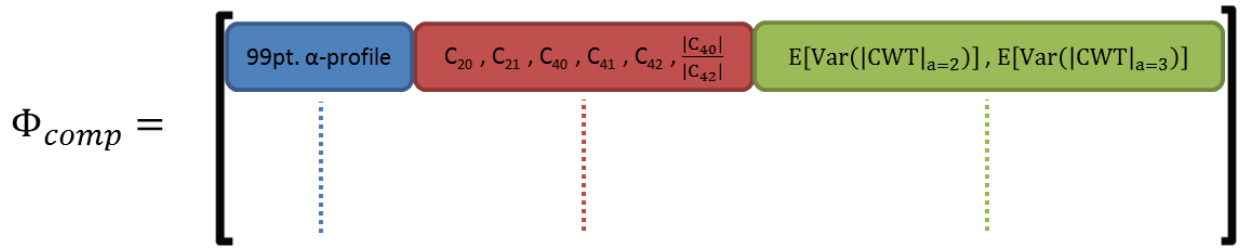


Figure 4.6: Comparison Matrix for Principal Component Analysis

Principal Component Analysis transforms a set of values that could possibly be interrelated (correlated) and projects them onto a plane whose axes are orthogonal and uncorrelated. Distinguishing features are then easily highlighted. The farther the data point of a specific attribute is from the origin, the more variability it brings to the comparison matrix. This variability translates to the ability to distinguish between modulation types.

All distinguishing features have been labeled in the MATLAB-created PCA plot (Fig. 4.7).

Each point whose label begins with 'A' is one of the 99 data points representing the α-profile. Figure 4.8 highlights each of the distinguishing points for the α-profile that are labeled in Fig. 4.7. Each point or set of points is associated with a distinguishing peak. The points labeled with a 'C' are correspondingly associated with the stationary cumulants. These appear to bring the most unique and distinguishing information to the matrix. But as previously discussed in Section 4.2.2, this information yields more improvement when distinguishing between inter-class modulations (PSK vs. FSK).

The two values for wavelet magnitude variance ('WV') are very close to the origin of Fig. 4.7 and barely add any unique information useful for classification. In fact, adding signal features to the SFV that do not bring with them unique distinguishing information have the potential for reducing or degrading classifier performance when classifying in high noise. Noise will affect every signal feature, but a signal feature that barely adds any new information to the GRA-AMC is not worth compounding the effect of noise on the final decision.

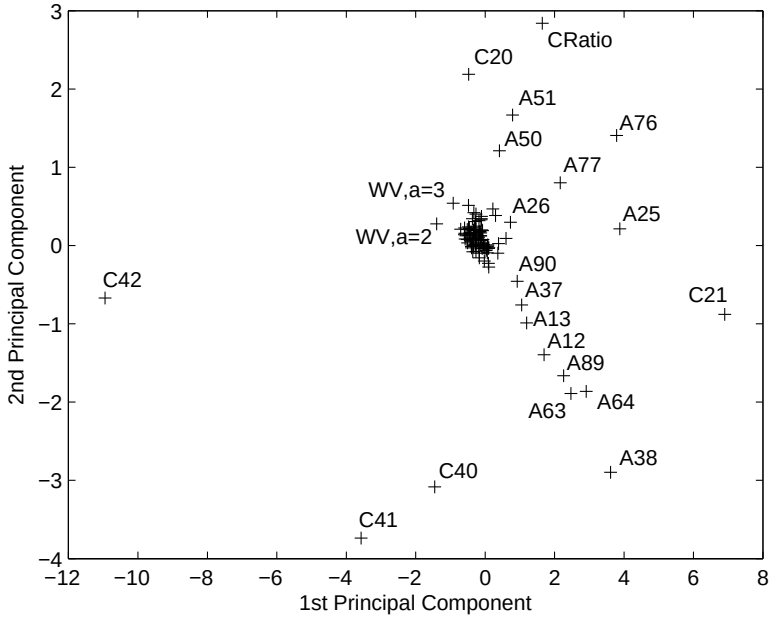


Figure 4.7: Principal Component Analysis Plot

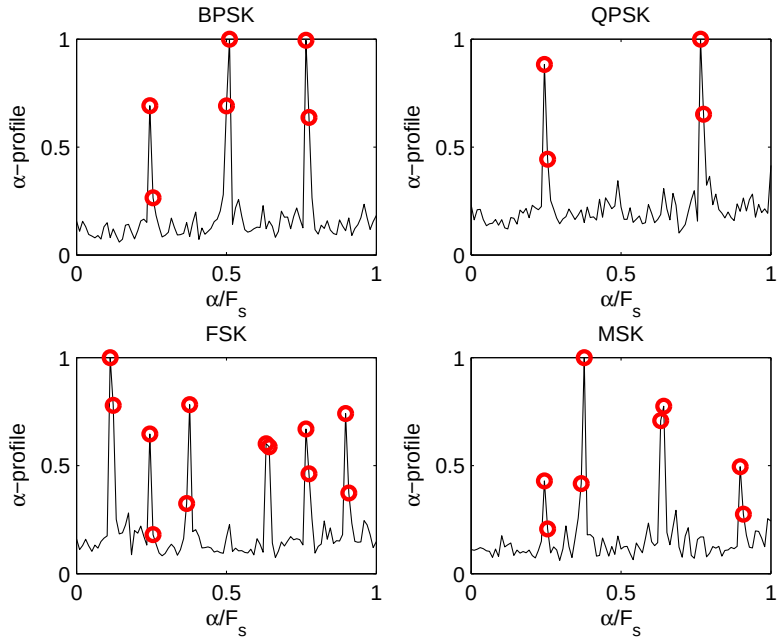


Figure 4.8: Distinguishing α-Profile Data Points from PCA Plot

Chapter 5

Summary, Conclusions, and Suggestions for Future Work

5.1 Summary and Concluding Remarks

Automatic modulation classification has become an essential component of civilian radio policy and electronic warfare. A novel modulation classifier architecture was presented in this thesis that uses Grey Relational Analysis as a decision engine that compares extracted signal features to choose the most likely modulation. The classifier framework presented here attempts to improve efficiency by taking advantage of the multiple statistical features that can be extracted from man-made signals while using only a single similarity algorithm to determine the most likely choice.

Chapter 3 explored the use of cyclostationary analysis as the single signal feature used for modulation classification. The decision to use the α-profile as the entire signal feature vector produced the additional need for extensive data preprocessing, yielding a secondary investigation. Normalization methods, weighting schemes and feature abridgment were presented as methods for improving the signal feature before comparison. Every combination of these

techniques was then tested and the combination with the best performance results became the standard preprocessing technique for the α-profile in Chapter 4. The effect of frequency offset was also investigated and it was determined that a cyclostationary profile with a much finer cyclic resolution is a requirement for counteracting any frequency offset that is encountered by the radio.

The largest improvement in classifier performance resulted from substituting the Euclidean distance metric in the GRA itself with Minkowski $r = 3$ distance. Although the $P_{CC,Total}$ from -20dB to 0dB SNR had no significant change, both the P_{CC} and $\overline{P}_{fp,max}$ were balanced out over all four modulation choices. The classifier bias towards choosing QPSK in high noise was eliminated. This was the most significant result of this work in combination with all three data preprocessing techniques. It is suggested that any future work in GRA-AMC should be conducted using Minkowski $r = 3$ distance.

The signal feature vector used by the GRA was redefined in Chapter 4 with the addition of wavelet magnitude variance in Section 4.1 and the condensing of the α-profile into a single-valued metric. This addition appears to have no statistically-significant effect on the classifier performance. Stationary cumulants were also combined with the α-profile in Section 4.2 in an attempt to increase performance in negative SNR. Both $P_{CC,Total}$ and $\overline{P}_{fp,max}$ both improved significantly. However this improvement was likely the result of a bias that was created towards PSK signals in high noise. Cosine similarity and cross-correlation were both investigated as alternatives to Euclidean distance for quantifying cyclostationary profile similarity and condensing that similarity into a single metric. Euclidean distance was still found to be the most reliable comparator.

The result yielded from adding stationary cumulants to the Signal Feature Vector in Section 4.2 suggested that combining the GRA with a hierarchical classification method by replacing hard thresholds with the fast GRA decision engine would increase classifier performance by letting the GRA only consider two possibilities at each hierarchical decision step. Additive channel noise is compounded as a side-effect of stationary cumulant estimation and although

cyclic cumulants have had favorable classification performance for cochannel identification, the cochannel identification was tried with stationary cumulants in Section 4.2.4. Classification was only successful in positive SNR as expected, and the best performance occurred when BPSK was one of the two cochannel signals.

The investigation was completed in Section 4.3 with an analysis of how much unique information was actually added by including additional signal features in the SFV using Principal Component Analysis. This analysis showed that the performance of each signal feature as an individual classifier is not the only characteristic that should be considered when conducting signal feature data fusion. The additional signal feature must also add unique statistical information to aid in distinguishing between modulation types. Wavelet magnitude variance added little in either performance in high noise or distinguishing information. Stationary cumulants were favorable for their distinguishing characteristics. However, individual performance is still low in negative SNR.

5.2 Suggestions for Future Work

The investigations conducted in this thesis have introduced new ways for classification that should be explored further. Additional questions were also raised regarding the favorability of choosing one signal feature over another for classification.

Section 3.4 illustrated the need for a higher cyclic resolution for the cyclostationary profile to counter the effects of frequency offset. General performance of the GRA-AMC would also likely be improved from using an α-profile with a finer cyclic resolution.

The addition of signal features such as cyclic cumulants and phase/amplitude histograms to cyclostationary profile for signal feature data fusion should be investigated. Cyclic cumulants have been used more successfully for classification in high noise than stationary cumulants and the distinguishing characteristics of cyclic cumulants should also satisfy the secondary need to add unique information to the signal feature vector.

The uniqueness of statistical information that was questioned in Section 4.3 also requires further investigation. Many signal features have shared underlying statistical patterns. The relationships between different signal features that are commonly used for modulation classification should be investigated. Combining additional signal features for classification is a tradeoff between processing requirements and the distinguishing characteristics that are added and not shared by other signal features. The underlying relationships between different signal features (such as between cyclostationary profiles and cumulants) can be analyzed to determine combinations of signal features that add the most distinguishing information for the smallest processing requirements.

Bibliography

[1] United states frequency allocation chart. [Online]. Available: http://www.fas.org/spp/military/program/sigint/allochrt.pdf

[2] "FCC 03-322, NPRM - facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies," Federal Communications Commission, Tech. Rep., December 2003.

[3] K. Po and J. Takada, "Signal detection based on cyclic spectrum estimation for cognitive radio in IEEE 802.22 WRAN system," The Institute of Electronics, Information and Communication Engineers, Technical Reportof IEICE.

[4] J. Renard, J. Verlant-Chenet, J.-M. Dricot, P. DeDoncker, and F. Horlin, "Higher-order cyclostationarity detection for spectrum sensing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, 2010.

[5] E. Like, V. D. Chakravarthy, P. Ratazzi, and Z. Wu, "Signal classification in fading channels using cyclic spectral analysis," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, 2009.

[6] T. W. Rondeau, "Application of artificial intelligence to wireless communications," Doctor of Philosophy, Virginia Polytechnic Insitute and State University, Blacksburg, Virginia, September 2007.

[7] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Doctor of Technology, Royal Institute of Technology (KTH), Stockholm, Sweden, May 2000.

[8] T. J. O'Shea, T. C. Clancy, and H. J. Ebeid, "Practical signal detection and classification in GNU radio," in *SDR Forum Technical Conference (SDR 07)*, 2007.

[9] S. Shellhammer, A. Sadek, and W. Zhang, "Technical challenges for cognitive radio in the TV white space spectrum," *2009 Information Theory and Applications (ITA) Workshop*, February 2009.

[10] N. Sai Shankar, "Overview of blind sensing techniques considered in IEEE 802.22 WRANs," in *5th IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2008 (SECON Workshops '08)*, June 2008, pp. 1 – 4.

[11] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 416 – 429, March 2000.

[12] K. Maeda, A. Benjebbour, T. Asai, T. Furuno, and T. Ohya, "Recognition among OFDM-based systems utilizing cyclostationarity-inducing transmission," in *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2007 (DySPAN 2007)*, April 2007, pp. 516 – 523.

[13] N. Kim, N. Kehtarnavaz, M. Yeary, and S. Thornton, "DSP-based hierarchical neural network modulation signal classification," *IEEE Transactions on Neural Networks*, vol. 14, no. 5, pp. 1065 – 1071, September 2003.

[14] K. Kim, I. Akbar, K. Bae, J.-S. Um, C. Spooner, and J. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2007 (DySPAN 2007)*, April 2007, pp. 212 – 215.

[15] A. Fehske, J. Gaeddert, and J. H. Reed, "A new approach to signal classification using spectral correlation and neural networks," in *Proc. First IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks, 2005 (DySPAN 2005)*, 2005, pp. 144 – 150.

[16] Z. Wu, G. Ren, X. Wang, and Y. Zhao, "Automatic digital modulation recognition using wavelet transform and neural networks," in *Advances in Neural Networks ISNN 2004*, ser. Lecture Notes in Computer Science, F. Yin, J. Wang, and C. Guo, Eds. Springer Berlin / Heidelberg, 2004, vol. 3173, pp. 936 – 940.

[17] H. Liu, D. Yu, and X. Kong, "A new approach to improve signal classification in low SNR environment in spectrum sensing," in *3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 2008 (CrownCom 2008)*, May 2008, pp. 1 – 5.

[18] M. Lu and K. Wevers, "Grey system theory and applications: a way forward," *Journal of Grey Systems*, vol. 10, no. 1, pp. 47 – 53, 2007.

[19] A. Amanna, R. Thamvichai, and M. J. Price, "Grey systems theory applications to wireless communications," in *SDR Forum and Technical Conference (SDR'10)*, Washington, D.C., November 2010.

[20] P. Marchand, C. Le Martret, and J.-L. Lacoume, "Classification of linear modulations by a combination of different orders cyclic cumulants," in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, July 1997, pp. 47 – 51.

[21] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET Communications*, vol. 1, no. 2, pp. 137 – 156, April 2007.

[22] S.-Z. Hsue and S. Soliman, "Automatic modulation recognition of digitally modulated signals," in *IEEE Military Communications Conference, 1989 (MILCOM '89) Conference Record. Bridging the Gap. Interoperability, Survivability, Security, 1989*, vol. 3, October 1989, pp. 645 – 649.

[23] G. Yeung and W. Gardner, "Search-efficient methods of detection of cyclostationary signals," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1214 – 1223, May 1996.

[24] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," *IEEE Circuits and Systems Magazine*, vol. 9, no. 2, pp. 27 – 45, 2009.

[25] Q. Yuan, P. Tao, W. Wenbo, and Q. Rongrong, "Cyclostationarity-based spectrum sensing for wideband cognitive radio," in *WRI International Conference on Communications and Mobile Computing, 2009 (CMC '09)*, vol. 1, January 2009, pp. 107 – 111.

[26] G. Sun, "Mpsk signals modulation classification using sixth-order cumulants," in *3rd International Congress on Image and Signal Processing (CISP), 2010*, vol. 9, October 2010, pp. 4404 – 4407.

[27] K. Ho, W. Prokopiw, and Y. Chan, "Modulation identification by the wavelet transform," in *Conference Record, IEEE Military Communications Conference, 1995 (MILCOM 1995)*, vol. 2, November 1995, pp. 886 – 890.

[28] ——, "Modulation identification of digital signals by the wavelet transform," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 169 – 176, August 2000.

[29] L. Hong and K. Ho, "Identification of digital modulation types using the wavelet transform," in *IEEE Military Communications Conference Proceedings, 1999. (MILCOM 1999)*, vol. 1, 1999, pp. 427 – 431.

[30] P. Prakasam and M. Madheswaran, "Digital modulation identification model using wavelet transform and statistical parameters," *J. Comp. Sys., Netw., and Comm.*, vol. 2008, pp. 6:1 – 6:8, January 2008.

[31] Y. Kuo, T. Yang, and G.-W. Huang, "The use of grey relational analysis in solving multiple attribute decision-making problems," *Computers & Industrial Engineering*, vol. 55, no. 1, pp. 80 – 93, 2008.

[32] D. K. W. Ng, "Grey system and grey relational model," *SIGICE Bull.*, vol. 20, no. 2, pp. 2 – 9, 1994.

[33] K.-C. Chang and M.-F. Yeh, "Grey relational analysis based approach for data clustering," *IEE Proceedings -Vision, Image and Signal Processing,*, vol. 152, no. 2, pp. 165 – 172, April 2005.

[34] Y.-m. Xie, H.-p. Yu, J. Chen, and X.-y. Ruan, "Application of grey relational analysis in sheet metal forming for multi-response quality characteristics," *Journal of Zhejiang University - Science A*, vol. 8, pp. 805 – 811, 2007.

[35] S.-J. Huang, N.-H. Chiu, and L.-W. Chen, "Integration of the grey relational analysis with genetic algorithm for software effort estimation," *European Journal of Operational Research*, vol. 188, no. 3, pp. 898 – 909, 2008.

[36] Z. Du and H. Liao, "A novel location estimation algorithm using grey relational analysis," in *Proc. Int. Conf. Communication Technology ICCT '06*, 2006, pp. 1 – 5.

[37] C.-H. Lin, "Classification enhancible grey relational analysis for cardiac arrhythmias discrimination," *Medical and Biological Engineering and Computing*, vol. 44, pp. 311 – 320, 2006.

[38] W. Gardner, "Spectral correlation of modulated signals: Part I analog modulation," *IEEE Transactions on Communications*, vol. 35, no. 6, pp. 584 – 594, June 1987.

[39] W. A. Gardner, *Statistical spectral analysis : a nonprobabilistic theory.* Englewood Cliffs, N.J.: Prentice Hall, 1988.

[40] W. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 14 – 36, April 1991.

[41] R. Roberts, W. Brown, and J. Loomis, H.H., "Computationally efficient algorithms for cyclic spectral analysis," *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 38 – 49, April 1991.

[42] W. Gardner, W. Brown, and C.-K. Chen, "Spectral correlation of modulated signals: Part II digital modulation," *IEEE Transactions on Communications*, vol. 35, no. 6, pp. 595 – 601, June 1987.

[43] W. Gardner and C. Spooner, "Signal interception: performance advantages of cyclic-feature detectors," *IEEE Transactions on Communications*, vol. 40, no. 1, pp. 149 – 159, January 1992.

[44] P. E. Pace, *Detecting and Classifying Low Probability of Intercept Radar*. Boston: Artech House, 2004.

[45] E. W. Rosolowsky, "The spectral correlation function: A new tool for analyzing spectral line maps," July 1998, senior thesis, Swarthmore College.

[46] C.-H. Lin, Y.-C. Du, Y.-F. Chen, and T.-S. Chen, "Multiple ECG beats recognition in the frequency domain using grey relational analysis," in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006 (EMBS '06)*, August 2006, pp. 2154 – 2158.

[47] H.-S. Chen, W. Gao, and D. Daut, "Spectrum sensing using cyclostationary properties and application to IEEE 802.22 WRAN," in *IEEE Global Telecommunications Conference, 2007 (GLOBECOM '07)*, November 2007, pp. 3133 – 3138.

[48] O. Dobre, Y. Bar-Ness, and W. Su, "Higher-order cyclic cumulants for high order modulation classification," in *IEEE Military Communications Conference, 2003 (MILCOM 2003)*, vol. 1, October 2003, pp. 112 – 117.

[49] MaWenjie, Y. ShiMing, R. Wu, X. ZhengHui, and L. WeiMing, "Spectral correlation function in low SNR environment," in *Proc. Asia-Pacific Radio Science Conf*, 2004, pp. 197 – 200.

[50] K. Teknomo. Similarity measurement. [Online]. Available: http://people.revoledu.com/kardi/tutorial/Similarity/

[51] Y. Zhao, J. Gaeddert, L. Morales, K. Bae, J.-S. Um, and J. H. Reed, "Development of radio environment map enabled case- and knowledge-based learning algorithms for IEEE 802.22 WRAN cognitive engines," in *2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 2007 (CrownCom 2007)*, August 2007, pp. 44 – 49.

[52] M. Barreno, A. Cardenas, and J. D. Tygar, "Optimal ROC curve for a combination of classifiers," in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Cambridge, MA: MIT Press, 2008, pp. 57 – 64.

[53] V. Chaithanya and V. Reddy, "Blind modulation classification in the presence of carrier frequency offset," in *International Conference on Signal Processing and Communications, 2010 (SPCOM 2010)*, July 2010, pp. 1 – 5.

[54] H. Nunez, M. Sanchez-Marre, U. Cortes, J. Comas, M. Martinez, I. Rodriguez-Roda, and M. Poch, "A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations," *Environmental Modeling & Software*, vol. 19, no. 9, pp. 809 – 819, 2004, Environmental Sciences and Artificial Intelligence.

[55] E. Silva, K. Panetta, and S. Agaian, "Quantifying image similarity using measure of enhancement by entropy," in *Proceedings: Mobile Multimedia/Image Processing for Military and Security Applications, SPIE Defence and Security Symposium*, vol. 6579, 2007.

[56] H. Li, R. Shi, W. Chen, and I.-F. Shen, "Image tangent space for image retrieval," in *18th International Conference on Pattern Recognition, 2006 (ICPR 2006)*, vol. 2, 2006, pp. 1126 – 1130.

[57] D. Wilson and T. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, pp. 1 – 34, 1997.

[58] M. Unser, A. Aldroubi, and S. Schiff, "Fast implementation of the continuous wavelet transform with integer scales," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3519 – 3523, December 1994.

[59] C. Spooner, "Classification of co-channel communication signals using cyclic cumulants," in *Conference Record of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers, 1995*, vol. 1, 1995, pp. 531 – 536.

[60] C. Spooner, W. Brown, and G. Yeung, "Automatic radio-frequency environment analysis," in *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, 2000.*, vol. 2, 2000, pp. 1181 – 1186.

[61] C. Spooner, "On the utility of sixth-order cyclic cumulants for RF signal classification," in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, 2001.*, vol. 1, November 2001, pp. 890 – 897.

[62] X. Zhou, Y. Wu, and B. Yang, "Signal classification method based on support vector machine and high-order cumulants," *Wireless Sensor Network*, vol. 2, no. 1, pp. 48 – 52, January 2010.

[63] L. I. Smith, *A tutorial on Principal Components Analysis*, University of Ortago Department of Computer Science, New Zealand, February 2002.