

Rowan University

Rowan Digital Works

Theses and Dissertations

12-31-2008

Use of linear predictive features and pattern recognition techniques to develop a vector quantization based blind SNR estimation system

Russell Paul Ondusko III
Rowan University

Follow this and additional works at: <https://rdw.rowan.edu/etd>



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

Recommended Citation

Ondusko, Russell Paul III, "Use of linear predictive features and pattern recognition techniques to develop a vector quantization based blind SNR estimation system" (2008). *Theses and Dissertations*. 708.
<https://rdw.rowan.edu/etd/708>

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact graduateresearch@rowan.edu.

USE OF LINEAR PREDICTIVE FEATURES AND PATTERN
RECOGNITION TECHNIQUES TO DEVELOP A VECTOR
QUANTIZATION BASED BLIND SNR ESTIMATION SYSTEM

by
Russell Paul Ondusko III

A Thesis Submitted to the
Graduate Faculty in Partial Fulfillment
of the Requirements for the Degree of
MASTER OF SCIENCE

Department: Electrical and Computer Engineering
Major: Engineering (Electrical Engineering)

Approved:

Members of the Committee

In Charge of Major Work

For the Major Department

For the College

Rowan University
Glassboro, New Jersey
2008
© Russell Paul Ondusko III

ABSTRACT

Russell Paul Ondusko III

USE OF LINEAR PREDICTIVE FEATURES AND PATTERN RECOGNITION TECHNIQUES
TO DEVELOP A VECTOR QUANTIZATION BASED BLIND SNR ESTIMATION SYSTEM

2007/08

Dr. Ravi Ramachandran

Master of Science in Electrical and Computer Engineering

Signal-to-noise ratio is defined as the ratio of a given transmitted signal to the background noise of the transmission medium. Signal-to-noise ratio (SNR) is a common concept found in all forms of electrical communications. The easiest way to measure the signal-to-noise ratio is through intrusive means in which a corrupt signal is compared to its original signal. This technique is inefficient and impractical because it requires the original signal and can only be used to theoretically test the noise properties of a channel rather than estimate the SNR of a communicated signal. Characteristics of speech signals can be used to develop non-intrusive methods for estimating the SNR of a signal. These methods do not require knowledge of the original speech signal for analysis.

In this thesis a Vector Quantization (VQ) based pattern recognition system approach is applied to estimate the SNR of speech signals. Features for the VQ system are derived from the speech signals through linear predictive analysis. The system is trained and tested on a range of 0 to 30 dB SNR in which codebook size, codebook spacing, training sets, and decision methods are studied to determine the best system architecture for a robust SNR estimation system for speech signals. The optimal feature for estimating the SNR of any speech signal regardless of the spectrum of the background noise is determined through analysis of testing results. An ensemble of classifiers approach is used to perform both decision level and distance level fusion using various combination rules to determine the best feature combination and fusion technique for a robust SNR estimating system for speech signals.

ACKNOWLEDGMENTS

I would like to thank my parents, entire family, and friends for their love and support through my college years, as well as my entire educational experience. The work of my family helped me to accomplish the goals I set for myself in attending higher education.

I would like to thank my peers in the Graduate Student Office for their help and determination through the completion of this research. In particular, Matthew Marbach for laying the groundwork for this research and Andrew McClellan for his help in the research of this system.

I would like to thank the staff of the Electrical and Computer Engineering Department of Rowan University, for imparting the knowledge and skills necessary to complete my research and to further my general knowledge in engineering. I would like to thank Dr. Hristescu for her assistance as a member of my thesis defense committee. I would also like to thank my graduate advisors Dr. Ravi Ramachandran and Dr. Linda Head for their guidance, help, and assistance throughout my entire college career.

This work was supported by the U.S. Air Force Research Laboratory, Rome NY, under contracts FA8750-05-C-0029 and F30602-03-C-0067.

TABLE OF CONTENTS

ABSTRACT.....	i
ACKNOWLEDGMENTS.....	ii
LIST OF FIGURES.....	vii
LIST OF TABLES.....	xii
CHAPTER 1 – INTRODUCTION.....	1
1.1 Measures of Speech Quality.....	1
1.2 Objectives of Thesis.....	2
1.3 Expected Contribution.....	3
1.4 Focus and Organization.....	3
CHAPTER 2 – BACKGROUND.....	5
2.1 SNR Estimation Techniques.....	6
2.2 Linear Predictive Features.....	10
2.2.1 Levinson Durbin Algorithm.....	11
2.2.2 LP Cepstrum.....	12
2.2.3 Reflection Coefficients.....	13
2.2.4 Log Area Ratio.....	13
2.2.5 Adaptive Component Weighted Cepstrum.....	13
2.2.6 Postfilter Cepstrum.....	14
2.2.7 Line Spectral Frequencies.....	15

2.3 Vector Quantization.....	15
2.3.1 Linde-Buzo-Gray Algorithm.....	16
2.4 Genetic Algorithms.....	16
2.5 MCRA Benchmark.....	20
2.5.1 Signal Energy Calculation.....	20
2.5.2 Signal Presence Probability.....	21
2.5.3 Noise Spectrum Estimation.....	23
2.5.4 A Posteriori Signal to Noise Estimation.....	24
CHAPTER 3 – APPROACH.....	25
3.1 Feature Extraction.....	25
3.2 Performance Scoring.....	26
3.2.1 VQ Classification.....	26
3.2.2 VQ Estimation.....	26
3.2.3 Absolute Error.....	27
3.3 Codebook Variables.....	28
3.3.1 Codebook Size.....	28
3.3.2 Codebook Spacing.....	29
3.4 Codebook Architecture.....	30
3.4.1 Individual Noise Spectrum Systems.....	30
3.4.2 Robust Multi-Spectrum System.....	30
3.4.3 Feature Fusion.....	31
3.4.3.1 Decision Level Fusion.....	31

3.4.3.1.1 Optimization.....	32
3.4.3.2 Distance Level Fusion.....	35
3.4.4 MCRA Benchmark.....	37
CHAPTER 4 – IMPLEMENTATION AND RESULTS.....	39
4.1 Feasibility Study.....	39
4.1.1 Initial AWGN Performance Results.....	39
4.1.2 Codebook Size.....	40
4.1.3 Codebook Spacing.....	47
4.1.4 VQ Estimation.....	52
4.1.4.1 Motivation.....	52
4.1.4.2 AWGN Results.....	60
4.2 Noise Spectrum Robustness.....	66
4.2.1 Pink Noise.....	66
4.2.2 CPV Noise.....	71
4.2.3 Cross Spectrum Testing.....	75
4.2.4 Multiple Spectrum System.....	82
4.2.4.1 Robust Spectrum Results.....	82
4.2.4.2 Beyond Range Codebook System.....	88
4.3 Feature Fusion.....	94
4.3.1 Unweighted Fusion.....	94
4.3.2 Weighted Fusion.....	105
4.4 MCRA Benchmark.....	110

CHAPTER 5 – CONCLUSIONS.....113

 5.1 Synopsis of Thesis.....113

 5.2 Summary of Accomplishments.....113

 5.3 Final Recommendations.....115

REFERENCES.....117

LIST OF FIGURES

Figure 1. Block Diagram of Feature Extraction Process.....	26
Figure 2. Block Diagram of Decision Level Fusion.....	32
Figure 3. Block Diagram of Distance Level Fusion.....	37
Figure 4. Comparison of LSF Resolution Classification AAE.....	49
Figure 5. Comparison of CEP Resolution Classification AAE.....	50
Figure 6. Comparison of REFL Resolution Classification AAE.....	50
Figure 7. Comparison of LAR Resolution Classification AAE.....	51
Figure 8. Comparison of ACW Resolution Classification AAE.....	51
Figure 9. Comparison of PFL Resolution Classification AAE.....	52
Figure 10. Error Analysis Graphs for Line Spectral Frequency Feature.....	54
Figure 11. Error Analysis Graphs for LP Cepstrum Feature.....	55
Figure 12. Error Analysis Graphs for Reflection Coefficient Feature.....	56
Figure 13. Error Analysis Graphs for Log Area Ratio Feature.....	57
Figure 14. Error Analysis Graphs for ACW Cepstrum Feature.....	58
Figure 15. Error Analysis Graphs for PFL Cepstrum Feature.....	59
Figure 16. Comparison of Classification and Estimation AAE for LSF with 5 dB Spaced Codebooks.....	62
Figure 17. Comparison of Classification and Estimation AAE for LSF with 3 dB Spaced Codebooks.....	62

Figure 18. Comparison of Classification and Estimation AAE for LSF with 1 dB Spaced Codebooks.....63

Figure 19. Comparison of Classification and Estimation AAE for CEP with 1 dB Spaced Codebooks.....63

Figure 20. Comparison of Classification and Estimation AAE for REFL with 1 dB Spaced Codebooks.....64

Figure 21. Comparison of Classification and Estimation AAE for LAR with 1 dB Spaced Codebooks.....64

Figure 22. Comparison of Classification and Estimation AAE for ACW with 1 dB Spaced Codebooks.....65

Figure 23. Comparison of Classification and Estimation AAE for PFL with 1 dB Spaced Codebooks.....65

Figure 24. Comparison of Classification and Estimation AAE for LSF in the Pink System with 1 dB Spaced Codebooks.....68

Figure 25. Comparison of Classification and Estimation AAE for CEP in the Pink System with 1 dB Spaced Codebooks.....68

Figure 26. Comparison of Classification and Estimation AAE for REFL in the Pink System with 1 dB Spaced Codebooks.....69

Figure 27. Comparison of Classification and Estimation AAE for LAR in the Pink System with 1 dB Spaced Codebooks.....69

Figure 28. Comparison of Classification and Estimation AAE for ACW in the Pink System with 1 dB Spaced Codebooks.....70

Figure 29. Comparison of Classification and Estimation AAE for PFL in the Pink System with 1 dB Spaced Codebooks.....	70
Figure 30. Comparison of Classification and Estimation AAE for LSF in the CPV System with 1 dB Spaced Codebooks.....	72
Figure 31. Comparison of Classification and Estimation AAE for CEP in the CPV System with 1 dB Spaced Codebooks.....	73
Figure 32. Comparison of Classification and Estimation AAE for REFL in the CPV System with 1 dB Spaced Codebooks.....	73
Figure 33. Comparison of Classification and Estimation AAE for LAR in the CPV System with 1 dB Spaced Codebooks.....	74
Figure 34. Comparison of Classification and Estimation AAE for ACW in the CPV System with 1 dB Spaced Codebooks.....	74
Figure 35. Comparison of Classification and Estimation AAE for PFL in the CPV System with 1 dB Spaced Codebooks.....	75
Figure 36. AAE for All Features when Trained on AWGN and Tested on Pink Noise with 1 dB Resolution.....	79
Figure 37. AAE for All Features when Trained on AWGN and Tested on CPV Noise with 1 dB Resolution.....	79
Figure 38. AAE for All Features when Trained on Pink Noise and Tested on AWGN with 1 dB Resolution.....	80
Figure 39. AAE for All Features when Trained on Pink Noise and Tested on CPV Noise with 1 dB Resolution.....	80

Figure 40. AAE for All Features when Trained on CPV Noise and Tested on AWGN with 1 dB Resolution.....	81
Figure 41. AAE for All Features when Trained on CPV Noise and Tested on Pink Noise with 1 dB Resolution.....	81
Figure 42. Comparison of Robust System LSF Feature AAE for All Tested Noise Types	85
Figure 43. Comparison of Robust System CEP Feature AAE for All Tested Noise Types.....	86
Figure 44. Comparison of Robust System REFL Feature AAE for All Tested Noise Types.....	86
Figure 45. Comparison of Robust System LAR Feature AAE for All Tested Noise Types.....	87
Figure 46. Comparison of Robust System ACW Feature AAE for All Tested Noise Types.....	87
Figure 47. Comparison of Robust System PFL Feature AAE for All Tested Noise Types	88
Figure 48. Comparison of 33 Codebook Robust System LSF Feature AAE for All Tested Noise Types.....	91
Figure 49. Comparison of 33 Codebook Robust System CEP Feature AAE for All Tested Noise Types.....	91
Figure 50. Comparison of 33 Codebook Robust System REFL Feature AAE for All Tested Noise Types.....	92
Figure 51. Comparison of 33 Codebook Robust System LAR Feature AAE for All Tested Noise Types.....	92

Figure 52. Comparison of 33 Codebook Robust System ACW Feature AAE for All Tested Noise Types.....93

Figure 53. Comparison of 33 Codebook Robust System PFL Feature AAE for All Tested Noise Types.....93

Figure 54. AAE For all Tested Noise Types Under Best Unweighted Fusion Combination104

Figure 55. AAE For all Tested Noise Types Under Second Best Unweighted Fusion Combination.....105

Figure 56. AAE For all Tested Noise Types Under Best Weighted Fusion Combination109

Figure 57. AAE For all Tested Noise Types Under Second Best Weighted Fusion Combination.....110

Figure 58. AAE For all Tested Noise Types Under Best MCRA Decision Method.....112

LIST OF TABLES

Table 1: Cepstrum Codebook Size Performance.....	42
Table 2: Line Spectral Frequency Codebook Size Performance.....	43
Table 3: Reflection Coefficient Codebook Size Performance.....	44
Table 4: Log Area Ratio Codebook Size Performance.....	45
Table 5: ACW Cepstrum Codebook Size Performance.....	46
Table 6: PFL Cepstrum Codebook Size Performance.....	47
Table 7: AWGN Results Comparing Codebook Resolution Through OAAE in dB.....	49
Table 8: AWGN Results Comparing VQ Classification and VQ Estimation OAAE in dB	61
Table 9: OAAE Results in db for System Designed with Pink Noise.....	67
Table 10: OAAE Results in dB for System Designed with CPV Noise.....	72
Table 11: OAAE in dB for System Trained on AWGN and Tested on Pink Noise.....	76
Table 12: OAAE in dB for System Trained on AWGN and Tested on CPV Noise.....	77
Table 13: OAAE in dB for System Trained on Pink Noise and Tested on AWGN.....	77
Table 14: OAAE in dB for System Trained on Pink Noise and Tested on CPV Noise.....	77
Table 15: OAAE in dB for System Trained on CPV Noise and Tested on AWGN.....	78
Table 16: OAAE in dB for System Trained on CPV Noise and Tested on Pink Noise.....	78
Table 17: AWGN OAAE in dB for Robust System.....	83
Table 18: Pink Noise OAAE in dB for Robust System.....	83
Table 19: CPV Noise OAAE in dB for Robust System.....	83

Table 20: CMV Noise OAAE in dB for Robust System.....	84
Table 21: AWGN OAAE in dB for Robust 33 Codebook System.....	90
Table 22: Pink Noise OAAE in dB for Robust 33 Codebook System.....	90
Table 23:CPV Noise OAAE in dB for Robust 33 Codebook System.....	90
Table 24: CMV Noise OAAE in dB for Robust 33 Codebook System.....	90
Table 25: Decision Level Mean Fusion Combination OAAE Results in dB.....	97
Table 26: Decision Level Median Fusion Combination OAAE Results in dB.....	98
Table 27: Decision Level Trimmed Mean Fusion Combination OAAE Results in dB....	99
Table 28: Distance Level Minimum Fusion Combination OAAE Results in dB.....	100
Table 29: Distance Level Mean Fusion Combination OAAE Results in dB.....	101
Table 30: Distance Level Median Fusion Combination OAAE Results in dB.....	102
Table 31: Distance Level Trimmed Mean Fusion Combination OAAE Results in dB...	103
Table 32: Four Best Fusion Combinations OAAE in dB.....	104
Table 33: Decision Level Weighted Fusion OAAE Results in dB.....	108
Table 34: Three Best Weighted Fusion Combinations OAAE in dB.....	109
Table 35: MCRA Benchmark OAAE in dB.....	111

CHAPTER 1 – INTRODUCTION

1.1 Measures of Speech Quality

Speech signals, like any signal, are subject to additive noise from many sources.

Transmission of an analog speech signal is subject to noise corruption from its channel, wired or wireless. The channel the signal passes through has a large bearing on the type of noise in the signal. Quantization noise will occur during the process of digitizing a speech signal. Applications of speech signals, including speaker identification and speech detection, are highly dependent on the amount of noise in a signal. Confidence metrics have been developed for such applications based on the quality of the incoming speech signal [1][2]. These confidence metrics have many applications but could be most important in security based applications. Two methods that exist to judge speech quality are the Mean Opinion Score and having knowledge of a signal before it has been degraded by noise. The mean opinion score technique requires multiple votes on the quality of speech signal from human listeners. Generally this is done by giving the signal a score from 1 to 5 with 1 being highly corrupt and 5 being clean. This technique is time consuming, lacks the detail and resolution of a sophisticated analysis technique, and is highly susceptible to error since it is solely based on human opinion. Having knowledge of a signal before corruption is an accurate way to judge signal quality, however, the process is intrusive and not always feasible. The goal of this thesis is to develop an efficient, accurate, non-intrusive, method for determining speech signal quality. The

signal to noise ratio of a speech signal has been identified as a good measure of speech signal quality. A pattern recognition system will be devised to use linear predictive feature data to train and test a VQ Classifier based system to provide an estimate of the signal to noise ratio, or SNR, of a speech signal regardless of the spectrum of the additive noise.

1.2 Objectives of Thesis

The main objectives of this thesis are:

- 1. To investigate six linear predictive based speech signal features identified from use in speaker recognition systems and their contribution to a signal to noise ratio estimating system.*
- 2. To implement a pattern recognition approach to signal to noise ratio estimation using vector quantization classifiers.*
- 3. To investigate classifier parameters when designing the VQ system and their effect on signal to noise ratio estimation.*
- 4. To study the ability of a VQ based pattern recognition system to estimate signal to noise ratio of multiple noise types including additive white Gaussian noise, pink noise, Continental Poor Voice (CPV) noise, and Continental Mid Voice (CMV) noise.*
- 5. To train a system robust to all trained and untrained noise spectrum types.*
- 6. To study feature fusion techniques to identify the best feature combination and fusion scheme for a robust system.*

1.3 Expected Contribution

This thesis details the research performed to construct a robust, non-intrusive pattern recognition based signal to noise ratio estimating system to provide speech signal quality estimates. The goals are to create a system which performs better than a tested benchmark system and to achieve less than a 3 dB average SNR estimation error within the range of 0 to 30 dB SNR.

1.4 Focus and Organization

This thesis studies the creation and performance of a VQ Classifier based pattern recognition system which uses linear predictive features derived from speech signals corrupted to known SNR levels to accurately estimate the signal to noise ratio of the entire speech signal. The thesis is divided into the following chapters:

Chapter 1 is an introduction to the motive and approach of developing a pattern recognition based SNR estimating system.

Chapter 2 is a literature review of methods used to estimate signal to noise ratio and noise estimation in general and a background on algorithms employed in the system presented. This chapter provides information on six linear predictive features, vector quantization, the genetic algorithm as an optimization technique, and the minima controlled recursive averaging algorithm which is used as a benchmark.

Chapter 3 defines the approach used for the specific application presented here and explains how all algorithms previously explained are used to generate this specific system. This includes a step by step analysis of the system design and research direction.

The contribution of each training parameter, each feature, and each post processing step is analyzed through their effect on an error based performance measure.

Chapter 4 presents and discusses the results obtained through experimentation at each step of the research process.

Chapter 5 summarizes the results and draws conclusions as to the final design of a robust signal to noise ratio estimator for the purpose of providing information on speech quality for the purpose of developing confidence metrics.

CHAPTER 2 – BACKGROUND

Noise is present in almost any form of electrical signal. Communication systems are notoriously subject to channel noise. The increased popularity of speech signal applications creates a need to investigate noise and its effect on speech signals, a common communication signal. Applications such as speaker detection and speech detection are popular areas of research and whose performance is highly dependent on noise in the speech signals. Many systems exist to study and remove noise from speech signals including noise estimation and speech enhancement. Various techniques are employed in these systems, spectral estimation being one of them.

Noise estimation had been a thoroughly researched signal processing tool, employing many methods to improve its accuracy. Spectrum estimation is a widely adopted technique with many forms of implementation in which signal properties are used to estimate the spectrum of the signal noise. Pattern recognition is a less popular approach to noise estimation in a system. Our application involves estimating the SNR of a speech signal for input into other systems and not for direct signal processing. This provides a unique opportunity to apply a pattern recognition approach to estimating the noise in a speech signal, a signal type which already has many pattern recognition based applications such as speech and speaker identification.

This chapter provides a literature review on noise estimation techniques for speech signals and an explanation of the terminology and algorithms employed to study a pattern recognition approach to the problem of noise estimation in a speech signal.

2.1 SNR Estimation Techniques

Noise estimation in speech signals is important for many applications. Noise estimation is usually the first step in a larger process such as speech enhancement. Many techniques have been researched to estimate noise in speech signals.

Pattern recognition has been used for the purpose of SNR estimation in speech signals for an application attempting noise suppression [3]. This technique focused on determining the SNR in different frequency bands for separate frames, as opposed to the SNR of an entire sentence. The SNR range was from -10 to 20 dB. The speech based feature studied in this application was amplitude modulation spectrograms (AMS), which are based on neurophysiologic findings on amplitude modulation processing in mammals. The AMS feature was calculated for 32 millisecond frames with 16 millisecond overlap and a 16 kHz sampling rate and results in a 15 x 15, two-dimensional complex AMS feature for each frame, with one temporal dimension and one dimension related to frequency bands. This pattern recognition approach was trained using forty-two different types of non-artificial noise mostly from traffic, machinery, and social environments, and testing was performed on fifty-four types of noise including those trained on. The pattern recognition technique employs a single standard feed-forward neural network [4]. This neural network has an input layer of 255 neurons, one for each location in the 15 x 15 AMS matrix; 160 neurons in the hidden layer; and 15 neurons in the output layer, one to estimate the SNR of each frequency band. The output of this neural network at each band is a value between 0.5 and .95 within which there is a linear trend corresponding from -10 to 20 dB. The AMS feature based system provides frame by frame SNR estimation for

fifteen frequency bands with an average error across all frequency bands of 5.6 dB.

Though this pattern recognition method differs in output and application from our own, information from this method could be used in further investigation. It was found that SNR estimation was more difficult in higher frequency bands.

Spectrum estimation is a common noise estimation technique in which the actual spectrum of the noise signal is estimated. Spectrum estimation has many applications in noise including spectral subtraction for signal enhancement. Standard spectral estimation is typically performed for an entire signal, which does not take into account non-stationary noise. Non-stationary noise is noise whose spectrum changes in time. One advancement in noise spectrum estimation in speech signals involves the use of low frequency regions of the signal to track spectral amplitude [5]. This technique improves on Boll's method [6], a well known spectral subtraction method which obtains the noise spectrum from non-speech segments of a signal. Using spectral subtraction to remove noise from a speech signal often adds musical noise to the signal. Musical noise consists of tones added to a speech signal from isolated patches of noise which have not been removed and is generally combated by adjusting the noise spectrum being subtracted by a constant weighting factor. A constant weighting factor, however, will not perform well in the presence of non-stationary noise. The amplitude of the noise can be tracked easily during non-speech segments of a speech signal. In [5] the tendency of speech data to fall within 50 Hz to 3.5 kHz is utilized to develop a noise amplitude tracking method. Using the high frequency end of the signal spectrum would result in the need for high sampling rates to properly apply this technique, so the frequency range of 0 to 50 Hz was studied.

The technique compares the low frequency spectrum of any segment of a speech signal to the low frequency spectrum of identified non-speech portions of the speech signal to provide a more variable noise spectrum amplitude. This application is rooted heavily in spectral subtraction for speech enhancement and is useful for obtaining estimates of noise levels in segments of a speech signal as opposed to an overall estimate for the total speech signal.

Higher order statistics have also been a focus of research for noise estimation and determining signal to noise ratio in speech signals [7]. Higher order statistics focuses on separating signal and noise energies. In work done by E. Nemer, et al. [7], segments of a speech signal are broken up into frequency sub-bands. The kurtosis and energy of each subband for each segment is calculated. It is assumed that the noise has a normal Gaussian distribution so the kurtosis of noise is zero. Assuming the kurtosis of noise is zero the calculated energy is both the noise and speech signal energy together, while the kurtosis is just the speech signal kurtosis. The energy of the speech signal can be estimated from the kurtosis, and the noise energy is obtained by subtracting the speech signal energy from the total calculated energy from the subband. This process can provide an SNR estimate for each subband in each segment of a speech signal. Using this data a total signal SNR could be estimated, however only in the presence of normally distributed noise.

A Minima Controlled Recursive Averaging (MCRA) algorithm has been applied to noise spectrum estimation for the purpose of speech enhancement [8]. This algorithm estimates noise based on past spectral power values and signal presence probability. The

algorithm uses the Short Time Fourier Transform to obtain both time and frequency data in the signal. The STFT breaks the signal into frames and frequency sub-bands for analysis. The algorithm estimates speech presence probability in a frequency subband of a signal frame from the local energy of the noisy speech signal and a local minimum. A signal presence probability is generated for each frame to perform noise estimation. A local energy matrix is obtained by smoothing the magnitude squared of the STFT in both time and frequency. Next a local minimum matrix is obtained through the recursive process of taking the minimum between the local energy in the local energy matrix and the minimum calculated for the prior temporal point. To obtain a local minimum, as opposed to a signal minimum, a temporary variable is used to hold the minimum of the previous temporal point, which is reinitialized after a set number of sub-bands. The ratio between the minimum local energy matrix and the local energy matrix is used to estimate signal presence probability by assuming signal presence in a specific frame at a specific subband if the ratio at that point is above a set threshold. Since this algorithm was implemented for the purpose of speech enhancement, the threshold is set to be more sensitive to signal presence. This is so it is less likely to remove actual speech data during spectral subtraction. The noise spectrum in a frame of the speech signal is now obtained by averaging past spectral powers using a smoothing parameter which varies with signal presence probability. The MCRA algorithm is employed as a benchmark for the method discussed in this paper.

There are many techniques for estimating noise in a speech signal. The technique developed in this thesis estimates the signal to noise ratio of an entire speech signal. Most

techniques estimate a frame based SNR which may be further split into various sub-bands, which is required for their specific applications. Obtaining full sentence SNR estimates would require a proper technique for combining multiple SNR estimates in a speech signal.

2.2 Linear Predictive Features

One of the most important factors in the success of a pattern recognition system is the feature data used. The feature data must differ between classes significantly enough to provide a means of classification. Without the ability to extract meaningful, usable information from a signal, pattern recognition could not occur. A set of features was chosen for speech signal SNR estimation based on their popularity in other pattern recognition based speech applications including speech and speaker recognition. These features are derived from a linear predictive analysis of a speech signal, and display a change in behavior as noise in a signal increases. The Levinson Durbin Algorithm [1] is used to derive a linear predictive filter for each frame of a speech signal being analyzed. This filter vector is then used to derive all features used in this application. The features of interest are the LP Cepstrum (CEP), ACW Cepstrum, PFL Cepstrum, Reflection Coefficients (REFL), Log Area Ratios (LAR), and Line Spectral Frequencies (LSF). The behavior of these features has been observed to be dependent on the amount of noise in a signal. The magnitudes of the LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum all increase as noise in a signal increases. The angles which comprise the Line Spectral Frequencies become more evenly spaced as the noise in a signal increases. The ACW Cepstrum and PFL Cepstrum are designed to be more

robust to noise, and are expected to classify the amount of noise in a speech signal less accurately than the other features.

2.2.1 Levinson Durbin Algorithm

The Levinson-Durbin recursion algorithm [9] is used to calculate the predictor coefficients for a linear predictive filter. First, the p order autocorrelation matrix for the signal frame being analyzed is calculated, where p is the desired order of the linear predictive filter. The diagonal of the autocorrelation matrix is set as the variable R , where the first value in R is position 0 and the last value is position p . Second, all variables used in the algorithm are initialized to zero. These variables include the filter coefficients, a temporary filter variable, and the reflection coefficients:

$$\begin{aligned} a(i) &= 0 \\ a_{prev}(i) &= 0 \\ refl(i) &= 0 \\ i &= 1 \rightarrow p \end{aligned}$$

where a is the linear predictive filter, a_{prev} is a temporary variable used in the algorithm, $refl$ is the reflection coefficient vector, and p is the desired filter order. A final variable denoted Energy is initialized as the energy of the frame, or the first value, $R(0)$ of the autocorrelation matrix. Finally the algorithm runs a loop to recursively update the filter coefficients according to the following psuedo-code:

```

for i=1 → p
    Acc =  $\sum_{j=1}^{i-1} a_{prev}(j) * R(i-j)$ 
    refl(i) =  $\frac{R(i) - ACC}{Energy}$ 
    a(i) = refl(i)
    for j=1 → i-1
        a(j) = a_prev(j) - refl(i) * a_prev(i-j)
    end
    Energy = Energy * [1 - refl(i) * refl(i)]
    for j=1 → p
        a_prev(i) = a(j)
    end
end
end

```

2.2.2 LP Cepstrum

The cepstrum is defined as the inverse of the power spectrum of a signal which is the inverse transform of the logarithm of a signal's z transform. The cepstrum $cx[n]$ corresponding to the signal $x[n]$ has a z transform:

$$Cx(z) = \log(X(z))$$

The LP Cepstrum [10] can be computed from the linear predictive filter coefficients calculated from the Levinson-Durbin algorithm:

$$C_{LP}(n) = a_n + \sum_{i=1}^{n-1} \frac{i}{n} C_{LP}(i) a_{n-i}$$

2.2.3 Reflection Coefficients

The Reflection Coefficients are calculated within the Levinson-Durbin Algorithm as the refl vector. The polynomial z-transform of the Linear Predictive filter,

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k}, \text{ can be obtained by the recursion [10]:}$$

$$A^{(0)}(z) = 0$$

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-1} A^{(i-1)}(z^{-1})$$

$$A(z) = A^{(p)}(z)$$

where k_i are called the PARCOR reflection coefficients.

2.2.4 Log Area Ratio

Log Area Ratios [11] are obtained directly from the reflection coefficients of a signal:

$$g_i = \log\left[\frac{A_{i+1}}{A_i}\right] = \log\left[\frac{1-k_i}{1+k_i}\right] \quad \text{where } 1 \leq i \leq p$$

where g_i are the log area ratios, A is the predictor coefficient polynomial, and k_i are the reflection coefficients.

2.2.5 Adaptive Component Weighted Cepstrum

The ACW Cepstrum [10] was developed as a feature more robust to noise for the purpose of speech signal recognition applications. Due to its robustness to noise it is not expected to perform well as a pattern recognition feature for determining noise level in a speech signal. However, it was added to the feature set to determine its ability in estimating speech signal SNR and to see if it could contribute in the later stages of the system. The

ACW Cepstrum is derived from the partial fraction expansion of the reciprocal of the linear predictive filter:

$$\frac{1}{A(z)} = \sum_{k=1}^p \frac{\lim_{z \rightarrow f_k} [(1 - f_k z^{-1}) / A(z)]}{1 - f_k z^{-1}} = \sum_{k=1}^p \frac{r_k}{1 - f_k z^{-1}}$$

It has been observed that the residues r_k deviate as the signal becomes more corrupt.

These residues are set to one, removing their deviation due to noise to create a feature more robust to noise. After setting the r_k values to one there is an altered transfer function:

$$\frac{N(z)}{A(z)} = \sum_{k=1}^p \frac{1}{1 - f_k z^{-1}} = \frac{1}{A(z)} \sum_{k=1}^p \prod_{i=1, i \neq k}^p (1 - f_i z^{-1})$$

$$\frac{N(z)}{A(z)} = p \frac{1 - \sum_{k=1}^{p-1} b_k z^{-k}}{1 - \sum_{k=1}^{p-1} a_k z^{-k}}$$

The ACW Cepstrum is then calculated from the polynomial $N(z)/A(z)$ in the same manner the LP Cepstrum is calculated from $A(z)$.

2.2.6 Postfilter Cepstrum

The Postfilter Cepstrum [10] is a weighted form of the LP cepstrum originally designed to enhance noisy speech. This feature was also designed to be robust to noise in a signal, so it is not expected to perform well for the classification of a noise level in a speech signal. The postfilter cepstrum can be calculated using a technique based on the predictor coefficients:

$$H_{pf}(z) = \frac{A(z/\beta)}{A(z/\alpha)} \quad \text{where } 0 < \beta < \alpha \leq 1$$

or from a technique using the LP Cepstrum:

$$PFL_n = CEP_n(\alpha - \beta^n)$$

$$\alpha = 1.0$$

$$\beta = 0.9$$

2.2.7 Line Spectral Frequencies

Line Spectral Frequencies [12] are calculated from the Linear predictive filter $A(z)$. $A(z)$ is an all pole filter which can be broken into symmetric and anti-symmetric polynomials $P(z)$ and $Q(z)$:

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1})$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1})$$

where p is again the order of prediction. The LSF feature is computed as the angles between 0 degrees and 180 degrees not inclusive of the polynomials $P(z)$ and $Q(z)$.

2.3 Vector Quantization

Vector Quantization involves compressing a data set of vectors into a smaller data set to represent the whole. Clustering algorithms are used to compress the vector data to retain the best overall representation of the original data. Vector quantization can be used for the simple application of data compression, but can also be used for signal coding and use in pattern recognition systems. Vector quantization for pattern recognition involves compressing sets of data representing different classes into separate VQ codebooks. The codebook refers to a set of compressed data representing a single class. Each vector in a codebook is referred to as a codeword. Several algorithms can be used to perform vector quantization, and the Linde-Buzo-Gray algorithm was used in this application.

2.3.1 Linde-Buzo-Gray Algorithm

The Linde-Buzo-Gray algorithm [13] is a clustering algorithm which was designed for vector quantization. The process requires a data set, codebook size, and a distance measure. The data set is the set of vectors which will be compressed into the codebook. The codebook size must be a power of two due to the manner in which this particular algorithm clusters. The distance measure is the method with which distances will be measured between data points in the algorithm and is generally squared euclidean distance. The algorithm begins by creating a codebook of size one by taking the mean of the data. The mean is taken separately for each vector dimension. The second step involves developing a codebook of size two. The data point is “perturbed” to make a second data point. Perturbation occurs by changing one vector dimension slightly in the original data point. Voronoi regions are established for each of the data points encompassing all original data points closest to the compressed and perturbed data points. The codebook is expanded to two vectors by finding the mean vector for each Voronoi region. The average distance from each data vector to its assigned codeword is calculated, and if the total distance for each region is below a set threshold the two new codewords are kept. If the threshold is not met the algorithm returns to a codebook of size one and attempts to expand using a different perturbation. Each codebook expansion after occurs in the same manner in which each codeword is perturbed, Voronoi regions are established, new codewords are created, and distance thresholds are checked.

2.4 Genetic Algorithms

Genetic Algorithms are search algorithms which mimic the observed Darwinian

evolutionary behavior [14],[15]. Darwinian evolution operates on the principle that organisms more fit to survive are more likely to breed fit offspring. This observed behavior can be adapted for many applications and is often used for optimization. The main variables for use of a genetic algorithm include population variables, genetic operators, and stopping criteria. Stopping criteria is common to all optimization and search techniques to assure the algorithm achieves a desired result if a desired result is attainable within the constraints of the search.

Population variables include population size, chromosome phenotype, and chromosome genotype. Population size refers to how many chromosomes will be in the population, which can change for each iteration of the population. Chromosome phenotype refers to the physical structure of the chromosome. Chromosome length is a common phenotype for single dimensional chromosomes. The type of data at each position in a chromosome is another phenotype. It is difficult to breed offspring from chromosomes of different phenotype. Chromosome genotype is the data that is used to generate the chromosome. For example, if a chromosome were representing a filter the genotype would be the filter coefficients.

Genetic operators include the general operations included in genetics and how they are applied to a specific application. The most common genetic operators are fitness, selection, crossover, and mutation. Fitness refers to an organism's ability to survive in its environment. This concept can be applied to other applications by determining a chromosome's behavior in a system as compared to the desired result. System error is an example of a fitness measure that could be applied to a chromosome. Selection refers to

the process of deciding which chromosomes from a population should be mated to create a chromosome for the next population. This process is based on the fitness associated with each chromosome in the population. Selection criteria must be established for a specific application including number of parent chromosomes used to generate a single offspring and how fitness is used to determine parents. In many applications two chromosomes are selected as parents, and chromosome fitness is used to generate a probability distribution from which to randomly choose each parent. Two factors in deciding the number of parents a chromosome may have are deciding whether a chromosome can reproduce asexually with itself and whether a chromosome can reproduce multiple times in a single selection process. For example, once a chromosome has been picked as a parent of an offspring should it be taken from the pool of candidates when selecting the next parent for that same offspring or any other offspring in the next population. Crossover refers to the method in which the parents of an offspring will be combined. Methods for crossover depend heavily on the type of data contained within the chromosomes. One common technique involves selecting each data point in the child chromosome randomly from one of its parents. For example consider two chromosomes of length ten holding alphanumeric data are being bred into an offspring:

$$\begin{aligned}
 C_1 &= a\ 1\ f\ 4\ d\ d\ 3\ e\ 1\ 1 \left\{ \begin{array}{l} 1 \geq pc_i \geq .5 \\ .5 > pc_i \geq 0 \end{array} \right. \\
 C_2 &= v\ 2\ 2\ 3\ e\ f\ 1\ s\ 2\ 4 \\
 pc &= .10\ .23\ .56\ .64\ .12\ .78\ .95\ .51\ .36\ .74 \\
 C_{child} &= v\ 2\ f\ 4\ e\ d\ 3\ e\ 2\ 1
 \end{aligned}$$

where C1 is the first parent, C2 is the second parent, pc is a random crossover probability vector with values between 0 and 1 inclusive, i is a position in a chromosome or

probability vector, C_{child} is the child chromosome, and the crossover rule builds the child chromosome from randomly selecting a parent for each position in the chromosome. For numerical data the crossover technique could involve operations such as taking the mean of its parents. Mutation in a genetic algorithm refers to data in the child chromosome which is not inherited from a parent chromosome. Mutation in Darwinian evolution provides the opportunity for the advancement or change of a species if the change benefits the organism and helps prevent population saturation. In a genetic algorithm mutation plays the same role. Mutation allows a change in the search pattern of the algorithm that would not be found simply by parent crossover, and it prevents the entire population from being duplicate copies of one specific chromosome. Rules are designed for mutation to specifically fit the application of the algorithm. These rules could include mutation probability, or the likelihood a piece of data in a child chromosome will mutate from its original value obtained from crossover, and mutation type, or how that data can change. High mutation probability effectively removes the search behavior of the algorithm by creating child chromosomes with no relation to the parent chromosomes, while low mutation probability promotes population convergence.

Some genetic algorithms use a rule called elitism [15]. Elitism exploits the fitness of a good solution in a population by allowing it to copy itself into the child population. Elitism could promote convergence, but is useful in assuring that good solutions are not lost by creating new populations. Elitism can be used to copy good performing chromosomes into the child population by any number of rules which suit the application. Often, the best performing chromosome is saved in the next population iteration, however

other rules could be developed such as saving a certain percent of chromosomes or all chromosomes which fall in a desired fitness range.

2.5 MCRA Benchmark

A benchmark test is needed to compare the results of the method presented in this thesis with another SNR estimation technique. The a posteriori SNR estimation from the MCRA algorithm was selected as the benchmark algorithm [8]. Most of the noise estimation calculation is done in the power spectral domain. The signal is divided into spectral frames and the energy of those frames is used to determine the signal presence probability by comparing the ratio between the energy of the noisy speech and the minimum energy in a particular frame. The principle behind this algorithm is that there is more energy in the regions where there are both speech and noise. The signal presence probability is used to adjust a smoothing parameter that averages past spectral power values. Through almost every step of the estimation process smoothing is performed to smooth the transition between regions with speech and silent portions of the signal. The smoothing takes into account the strong correlation of speech in consecutive frames. The smoothing parameters are also defined to preserve speech assigning a higher probability of speech presence in particular frames of the signal. The average of past spectral values produces an estimate of the power spectrum of the noise.

2.5.1 Signal Energy Calculation

A corrupted speech signal, denoted by $y(n)$, is divided into overlapping frames through a windowing function. The signal is then analyzed using the Short Time Fourier Transform. The original signal is composed of both clean speech represented by $x(n)$ and $d(n)$ the

noise accompanying the clean speech, where n represents the discrete time index.

$$Y(k,l) = \sum_{n=0}^{N-1} y(n+lM)h(n)e^{-j(2\pi/N)nk}$$

where k is the frequency bin index, l is the time frame index, h is an N sized analysis window, and M is the time frame update size. $Y(k,l)$, $X(k,l)$, and $D(k,l)$ represent the STFT of observed signal, the clean signal and the noise signal respectively. The goal is to estimate variance of the noise in each subband represented by $\lambda_d(k,l) = E[|D(k,l)|^2]$. The variance will then be used to estimate a posterior SNR, by comparing the variance of the noise in each subband to the energy of the original signal in each subband.

2.5.2 Signal Presence Probability

To decide if speech is present or absent in a subband of a particular frame the ratio between the local energy of a frame and the minimum energy of the signal during that time frame is used in a decision rule. To obtain the local energy of the speech signal the magnitude squared of the STFT of the noisy signal is smoothed in both time and frequency. In frequency the windowing function is applied to the energy, and in our case the window is a local average of two bins near point of interest and that point.

$$S_f(k,l) = \sum_{i=w}^w b(i) |Y(k-i,l)|^2$$

A first order recursive averaging is performed on the signal in time to smooth the signal further and produce the spectral power of the signal with respect to time.

$$S(k,l) = \alpha_s S(k,l-1) + (1 - \alpha_s) S_f(k,l)$$

The α_s is the smoothing parameter and is a value between 0 and 1. The smoothed version

of the calculated power spectrum is then used to calculate the minimum energy of the signal in a given frame. Initially the minimum local energy, $S_{\min}(k,l)$, and a temporary variable matrix are initialized to the indexed energy in the signal, $S(k,l)$. The minimum computation is completed by completing a sample wise comparison of the local minimum energy and the minimum of the previous frame. The lower value of the two is selected as the minimum for that that particular frame. Since the minimum energy, $S_{\min}(k,l)$ is updated each time, the minimum of the previous frame is the minimum of all the frames before it within some resolution defined by L . When L frames have been read the minimum is taken as the lesser of the local energy and the previous frame of the temporary matrix. The ratio between the local energy of the original signal and the calculated minimum is then calculated.

$$S_r(k,l) = S(k,l) / S_{\min}(k,l)$$

A Bayes minimum cost decision rule is then used to decide if there speech is absent or present at a particular location in the signal. A conversion of the decision rule, created by Cohen,

$$S_r(k,l) \begin{matrix} > \\ < \end{matrix} \delta \begin{matrix} H_1 \\ H_0 \end{matrix}$$

was employed to create an indicator matrix to denote speech absence or presence. The indicator matrix $I(k,l)$ is assigned a value of 1 when $S_r(k,l) > \delta$, and a value of 0 when $S_r(k,l) < \delta$. The indicator matrix is used to estimate the speech presence probability.

$$p(k,l) = \alpha_p p(k,l-1) + (1 - \alpha_p) I(k,l)$$

The α_p is another smoothing parameter between 0 and 1. The estimator has three features that make it robust to different types of noise. The δ is not sensitive to either the type of level of noise in a particular signal. The estimator also used α_p to take into account the increased probability of speech presence in frames near each other. When $S_r < \delta$ falsely deciding speech absence is small, due to the ratio of spectral power of the signal to the estimated spectral power of the noise.

2.5.3 Noise Spectrum Estimation

When estimating the noise spectrum, techniques are employed to assign a greater probability to speech presence. The original implementation of this technique was used in the creation of a speech enhancement system. In such a system predicting speech absence when it is present can severely distort the enhanced signal. A temporal recursive smoothing is performed on the noisy speech signal during periods of speech absence.

$$\begin{aligned} \lambda_d(k, l+1) = & \lambda_d(k, l)p(k, l) \\ & + [\alpha_d \lambda_d(k, l) + (1 - \alpha_d) |Y(k, l)|^2] \\ & \times (1 - p(k, l)) \end{aligned}$$

The α_d is another smoothing parameter between 0 and 1. The smoothing parameter is adjusted in time by the signal presence probability $p(k, l)$. The noise power spectrum is an average of past spectral power values.

$$\lambda_d(k, l+1) = E \left[|D(k, l)|^2 \right]$$

2.5.4 A Posteriori Signal to Noise Estimation

The posterior SNR is then estimated by comparing the power of the STFT of the noisy signal to the estimated variance of the noise in a particular subband. The posterior SNR is defined thusly because it is calculated from past spectral power values.

$$SegSNR(l) = 10 \log \frac{\sum_{k=1}^N |Y(k,l)|^2 - \sum_{k=1}^N \lambda_d(k,l)}{\sum_{k=1}^N \lambda_d(k,l)}$$

CHAPTER 3 – APPROACH

3.1 Feature Extraction

Sentences from the TIMIT database are used to collect feature information from speech signals [16]. The New England Dialect portion of the database is used, which includes 38 speakers uttering 10 sentences, totaling 380 sentences. This data set is broken into two smaller sets for training and testing. Each experiment is run twice using one data set for just training the VQ codebooks and the other set for just testing the system. Both sets are used for training and testing to obtain more data on system error. The first step in feature extraction, figure 1, is adding noise to a speech signal. Next, the Levinson Durbin algorithm was used to perform linear predictive (LP) analysis of the sentence. This step broke the sentence into frames of 240 data points with 160 overlapping points and generated a 12th order LP filter for each frame. Two processes, Feature Extraction and Energy Calculations, run in parallel. Feature extraction uses the LP based filters to calculate the respective 12th dimensional features for each frame. Energy calculations observe each frame to determine which signal frames may have originally corresponded to silent portions in the speech. Energy thresholding removes all features corresponding to frames whose energy falls below a constant energy threshold to remove the silent portions of the signal. Finally, the feature vectors for that signal are passed to the VQ system for training or testing.

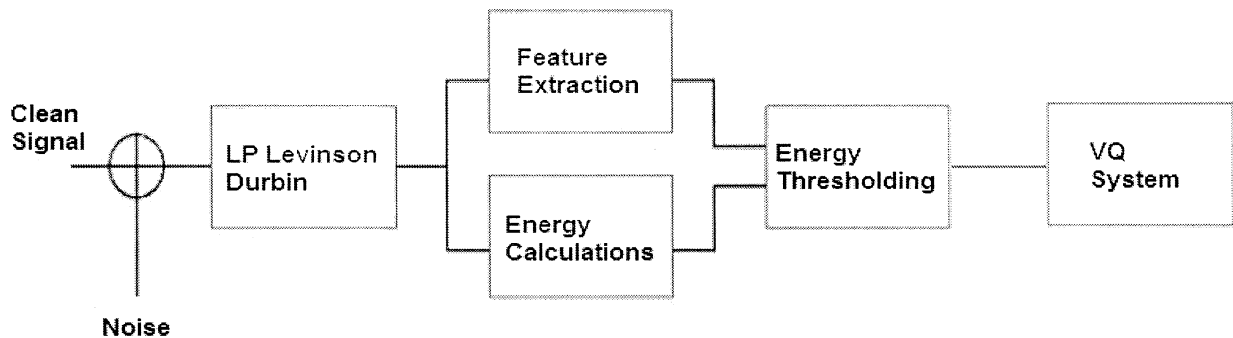


Figure 1. Block Diagram of Feature Extraction Process

3.2 Performance Scoring

3.2.1 VQ Classification

VQ Classification is a pattern recognition technique in which the VQ system identifies which class input data falls under and provides this class as an output. Using VQ classification in this context would refer to assigning to the sentence the SNR value of the codebook with the smallest Euclidean distance from all feature vectors of a sentence. This is referred to in our experimentation as a hard decision. This approach to obtaining speech signal SNR with a VQ system was used mostly to determine the feasibility of using the features selected for implementation in the VQ system.

3.2.2 VQ Estimation

VQ Estimation was an alternative approach to VQ classification which is referred to as a soft decision. VQ estimation exploits the fact that the classes being identified are actually on a linear scale. This allows the use of more than one VQ codebook to estimate the value being identified, as opposed to designating a specific class. In this case the SNR is estimated through a combination of the codebooks with the smallest Euclidean distance

from the feature vectors of the speech signal. The signal SNR is estimated by converting the distances of the three closest codebooks into probabilities, assigning the highest probability to the smallest distance:

$$p_i = \frac{(s - d_i)}{s}$$

where p_i is the probability assigned to one of the three smallest distances, s is the sum of the three smallest distances, and d_i is the distance being transformed; and summing the SNR codebook values multiplied by their respective probability:

$$SNR = \sum_{i=1}^N SNR_i \cdot p_i$$

where SNR is the final sentence SNR estimate, SNR_i is the SNR value associated with codebook i , p_i is probability i in order from smallest probability to largest, and N is the number of estimations used in the decision which will equal 3. It is expected that the soft decision estimation technique would reduce the error in obtaining the SNR of a sentence as compared to the hard decision classification method.

3.2.3 Absolute Error

The nature of this pattern recognition application allows the development of using SNR error as a performance measure. The output of the VQ system for an input speech signal is the classified or estimated SNR of the signal. During testing the actual SNR of the input signal is known. Since the output of the system is not a class but a value within a linear range the absolute error of the signal is easily obtained for use as a performance measure. The AE, Absolute Error, is simply the absolute value of the actual speech signal test SNR subtracted from the estimated speech signal SNR. This provides a performance

measure for a single sentence. During testing each SNR level is tested using all test speech signals. The system performance is tested at a single SNR level by taking the AAE, Average Absolute Error, for the SNR level. The AAE is obtained by taking the mean of all speech signals with the same SNR when input into the VQ system:

$$AAE_i = \sum |SNR_i - i|$$

where AAE_i is the absolute average error at SNR level i , SNR_i is the group of individual SNR estimates for sentences corrupted at SNR level i , and i is the SNR of the input test sentences. This performance measure shows system performance at one point within the SNR test range. Total system performance is obtained by calculating the OAAE, Overall Average Absolute Error. OAAE is found by taking the mean of all AAEs calculated for the system:

$$OAAE = \sum_{i=0}^{30} \left(\frac{AAE_i}{31} \right)$$

where $OAAE$ is the total system performance, and AAE_i is the AAE calculated at each test SNR level i . The OAAE is the performance of the system over the entire test range.

3.3 Codebook Variables

3.3.1 Codebook Size

VQ Codebooks are created by compressing data using the Linde-Buzo-Gray Algorithm. The algorithm compresses data to a power of two codebook vectors. The most effective codebook size should be found for proper performance. A codebook compressed too small may not retain the information needed to distinguish the data for that SNR level, while a codebook made too large may also cause overlap between SNR levels, as well as

increase training and testing time. The first step in experimentation is to determine a proper codebook size for the selected features. Training and testing was performed with codebook sizes of 16 code vectors to 256 code vectors. An increase in performance as the codebook size increases until reaching a peak where performance would begin to drop as the classes began to overlap is expected, but it is not expected that the trend will perform the same for individual features.

3.3.2 Codebook Spacing

Due to the unique nature of this pattern recognition application there are not a set number of classes to recognize. The classes of this pattern recognition system are the SNR levels of the signals used to train the VQ codebooks. The input to the system could come from a sentence of any SNR level, most often within the specified system range. VQ Codebooks could be created using sentences from any SNR level. It was decided codebooks will be created at evenly spaced intervals within the system's range of 0 dB SNR to 30 dB SNR. First, codebooks will be created at 5 dB spacing, creating a system of seven total codebooks from 0 to 30 dB. This system will be used to quickly identify the feasibility of the selected features and determine a proper codebook size for each feature. Codebook spacing will be reduced to 3 dB increments, creating a system of 11 total codebooks and 1 dB increments, creating a system of 31 codebooks. These systems will be tested to determine the effect on reducing codebook size on the performance of the system. Reducing the codebook spacing is expected to reduce system error by closing the gaps in classification.

3.4 Codebook Architecture

3.4.1 Individual Noise Spectrum Systems

The VQ based pattern recognition approach will first be tested on systems designed for signals corrupted by individual noise types. The systems are trained on feature data extracted from sentences corrupted by additive noise generated with the same frequency spectrum. Testing is performed by comparing feature data extracted from speech signals not used in the training set against the codebooks to obtain a vector containing the distances from the speech signal's feature data as compared to each of the codebooks. Testing is performed separately for each of the six LP features. Initial results will be obtained from testing systems with only feature data extracted from signals corrupted with the same noise spectrum type. To test system robustness to noise spectrum type, each individual spectrum system will be tested using feature data extracted from signals corrupted by different noise spectrum types. The ability to estimate speech signal SNR is expected to be present when a system is tested using feature data extracted from signals corrupted with the same noise spectrum type but not from signals corrupted by different noise spectrum types. The three noise spectrum types to be used in this experiment will be AWGN (additive white Gaussian noise), Pink Noise, and CPV (continental poor voice) Noise.

3.4.2 Robust Multi-Spectrum System

An SNR estimating system cannot be built to work with only a specific noise spectrum. In a real world application the system must work with all noise spectra, or any noise spectra which are likely to appear under the conditions of its use. A robust system can be

designed by extracting training data from signals corrupted by AWGN, pink noise, and CPV noise individually. Feature data will be extracted from each sentence after being corrupted by just AWGN at the set SNR level, just Pink noise at the set SNR level, and just CPV noise at the set SNR level. This will increase the data being compressed into each codebook by a factor of three. The system must also be tested against an untrained noise spectrum, so the system testing will include AWGN, Pink noise, CPV noise, and an untrained CMV noise. It is expected that the system will perform well for all trained noise spectra. By blending the three noise spectra in the training data the system is expected to perform well for the untrained noise spectrum.

3.4.3 Feature Fusion

Feature fusion is an ensemble of classifiers technique [17] which combines information from two or more features in an attempt to use complementary data from the features to reduce the VQ system estimation error. Feature fusion could reduce error in this application if a combination of features and fusion method is found which generally brings the SNR estimate closer to the actual speech signal SNR.

3.4.3.1 Decision Level Fusion

Decision level fusion, figure 2, combines the SNR estimates of a combination of the six features into one final estimate. This can be done by averaging the SNR estimates, taking the median of the estimates, or taking a trimmed mean of the estimates. The trimmed mean method uses the average of all but the highest and lowest estimates being used. These decision rules can be used due to the nature of this pattern recognition application. Since the purpose of the application is to estimate an SNR value within a specific range

and not to classify an input from multiple dissimilar classes, fusion of multiple feature outputs is possible. Decision level fusion will reduce system error if a combination of features has complementary data or if the collection of decisions brings the SNR estimate closer to the speech signal's SNR. Complementary data using just two features, for example, would occur if one feature generally estimates a higher SNR than the actual SNR while another feature estimates a lower SNR than the actual SNR.

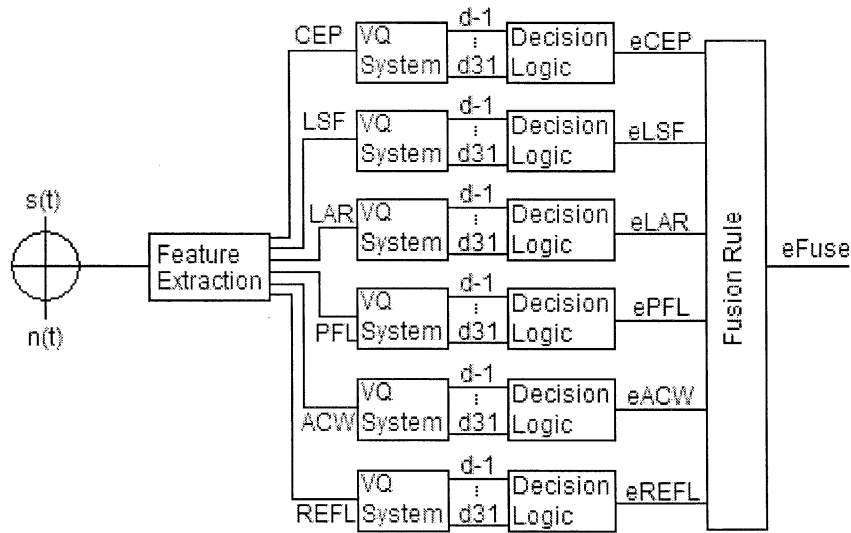


Figure 2. Block Diagram of Decision Level Fusion
The process of determining the individual feature SNR estimates, eCEP through eREFL, remains unchanged. The estimates are then combined into one final SNR estimate, eFuse.

3.4.3.1.1 Optimization

Decision level fusion could potentially be improved by weighting the features used. Unweighted decision level fusion gives equal importance to all features in the set being used to estimate the speech signal SNR. Only the mean decision level technique can be easily adapted to weighting. All features in a set will be given a weight and the total of

the weights should sum to one. The weights are multiplied by their respective feature's SNR estimate and the weighted estimates are summed into the final estimate. Each feature's weight then corresponds to its fraction of the final speech signal SNR estimate.

For this application using a Genetic Algorithm optimization technique is used to obtain the weights. The algorithm is initialized with a population of random weights for a feature set. Each set of weights is a chromosome in the population. Each set of random weights in the population sums to one to properly perform weighted decision level fusion. The OAAE is calculated separately for each chromosome and is then used as the chromosome's fitness. The next generation of the population is then formed. The next generation population has been kept at the same size as the previous population, and for this experiment the population consists of 10 sets of weights. Using elitism the best performing set of weights is saved as the first chromosome in the new population. Child chromosome formation consists of combination and mutation. Combination consists of blending the weights of two parent chromosomes into one child. First two parent sets must be selected. Each chromosome in the current population is mapped to a distribution from zero to one giving each weight set a portion of the distribution associated with its fitness as compared to the other sets fitnesses. The size of the distribution for a single weight set, or probability it will be chosen as a parent, is determined in the same way soft decision is performed. The probability is obtained by subtracting the sets fitness from the sum of all fitnesses and dividing this value by the sum of all fitnesses. The weights are now mapped to the selection distribution in specific non-overlapping ranges corresponding to their probability of parent selection. Two random values between zero

and one are generated, and their position on the range selects which two parents are selected. Using this method, the same weight set can be selected as both parents. Combination of the two chosen parents was then done by randomly selecting which parent to choose each weight. A random value was generated between zero and one to determine each specific weight's parent. If the value was greater than 0.5 it was taken from the first parent, otherwise it was taken from the second parent. The process is completed for nine children in the new population, making ten children including the elite chromosome. Mutation is the next step in creating a child population. Since the weights are randomly selected from either parent, each weight set will not necessarily sum to one as required. Mutation was completed simply by dividing each weight by the sum of its weight set so the sets sum to one. This algorithm is run on a single combination of features until one of three stopping criteria are reached. The three criteria include reaching the maximum population, falling below an error threshold, or when the population reaches a saturation state. The maximum population threshold occurs after the creation of the 1000th population. The error threshold occurs if the OAAE of a weighted combination set has fallen below 1 dB. The saturation threshold occurs when all sets in a child population are the same, and is the most likely stopping criteria to occur. If this threshold is reached the best set of weights has been found and each subsequent child population would only produce the same weight sets.

This experiment was used to determine the best set of weights and best features for use in decision fusion. The experiment was run on each combination of features, a total of 100 times for each combination to acquire statistics about the algorithm and its

decisions on the best weight sets. Weight sets were obtained by using the Genetic Algorithm to optimize the performance of classifying the training data with the VQ system. The training data includes all sentences used to create the codebooks at all codebook SNR levels after corruption by AWGN, pink noise, and CPV noise separately. Once each weight set is obtained using the training data it is used to determine the effect of those weights when combining the testing data. The testing data includes all sentences not used to create the codebooks corrupted by AWGN, pink noise, CPV noise, and CMV noise. The output of this algorithm provides the weights selected for each feature for all trials of all combinations, the number of populations created before a stopping criteria has been reached, and the OAAE of the test signals for AWGN, pink noise, CPV noise, and CMV noise using each specific weight set. The weight and OAAE information can be used to select the best weights to use in the final application as well as give information on the performance of this optimization technique. The number of trials combined with the OAAE data received from the algorithm is useful for reporting the efficiency and performance of this algorithm for this application. It is expected that using the Genetic Algorithm optimization technique to find the best combination and weight set will reduce error in the final system. One benefit of this approach is that after the experiment is run and a weight set and feature combination is decided on there is little added complexity to the VQ system.

3.4.3.2 Distance Level Fusion

Distance level fusion functions by combining the distances of each feature at each codebook to then estimate the SNR based on the combination of distances, as opposed to

the combination of feature estimates. Distance level fusion, figure 3, requires that the values which represent the distances from each codebook be compatible for all features used in the fusion. Since each feature's distance values are representative of the typical component values for each feature, the distances must be transformed to be properly combined. First, the distances are transformed for each individual feature by dividing them by the sum of all distances for that feature for that particular sentence. This technically provides an inverse probability, however the SNR estimate is made on the smallest distances which correlate to the lowest probabilities. After the distances for each feature have been transformed using this method, they can be combined using several techniques. The rules used to combine distances of multiple features include the mean distance, minimum distance, median distance, and trimmed mean distance. The mean distance averages the transformed distances from a specific codebook level for all features used in the fusion. The minimum distance approach takes the smallest transformed distance from a specific codebook level of all the features involved in the fusion. The median distance approach uses the median of the transformed distances from a specific codebook level of all the features involved in the fusion. Finally, the trimmed mean approach takes the average of the transformed distances from a specific codebook level, leaving out the largest and smallest transformed distance.

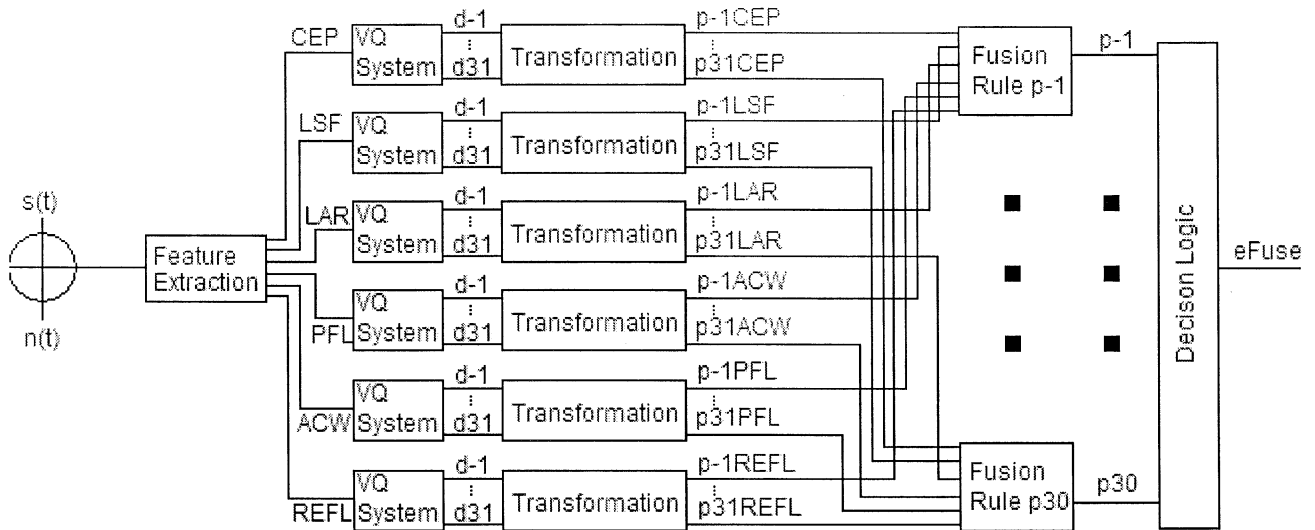


Figure 3. Block Diagram of Distance Level Fusion

After the distances are calculated for each feature from their specific codebook set they are transformed, $p-1$ CEP and $p31$ CEP refer to the transformed CEP distance. The transformed distances are combined into an individual distance for that SNR level, $p-1$ through $p31$, and normal decision logic is used to obtain the final SNR estimate, eFuse.

3.4.4 MCRA Benchmark

The Minima Controlled Recursive Averaging algorithm was used as a benchmark to compare the SNR estimation technique presented in this thesis to an existing noise estimation algorithm. The MCRA algorithm was designed to estimate the noise spectrum in frames of a speech signal for the purpose of speech enhancement. As a benchmark for an SNR estimation system, the estimated noise spectrum was used to calculate an SNR level for the signal. The MCRA algorithm uses the Short Time Fourier Transform to break a signal into frames to perform noise spectrum estimation in each frame. The STFT breaks the signal into frames of 240 data points with 80 overlapping points. The parameters for the algorithm were selected based on the original implementation by Cohen, et al. these parameters included $\alpha_d=0.9$, $\alpha_s=0.8$, $\alpha_p=0.18$, $\delta=4.8$, a rectangular

window of size 3 was used to smooth the signal energy in spectral domain, and a hamming window is used in the calculation of short-time Fourier transform. The algorithm estimates the noise signal for each frame to provide an SNR estimate for each frame of the speech signal. The VQ based pattern recognition approach to SNR estimation presented in this thesis estimates the SNR level of the whole speech signal, as opposed to the frame SNR estimation provided by the MCRA algorithm. The frame SNR estimates of the MCRA algorithm are combined into one SNR estimate for the entire speech signal. In work done by Cui et al. [18] an algorithm is used which provides frame based SNR estimates while a single SNR estimate is needed for the entire signal. To correct this problem a floor is set on frame based SNR estimates at 0 dB, and all estimates below 0 dB are set to 0 dB. The total signal SNR is obtained by averaging all non-zero frame SNR estimates. The floor is set at 0 dB to remove valleys in the frame estimates caused by pauses in the speech signal. For the MCRA benchmark test, three techniques are implemented to obtain the single SNR estimate of the signal from the multiple frame SNR estimates. The first estimate is obtained by averaging all frame SNR estimates. The second estimate is obtained by averaging all frame SNR estimates above 0 dB, not inclusive. The third estimate is obtained by averaging all frame SNR estimates above a percentage of the maximum frame SNR estimate. This technique sets the floor at 0.65 multiplied by the maximum frame SNR estimate.

CHAPTER 4 – IMPLEMENTATION AND RESULTS

4.1 Feasibility Study

The feasibility of developing a signal to noise ratio estimating system using a pattern recognition approach which utilizes linear predictive features was completed by studying the system's performance and defining system parameters with just AWGN. The two main parameters of the system included VQ codebook size and the spacing between VQ codebooks. VQ codebook size indicates how many codewords, or vectors of data, there are in a codebook after the compression process. Larger codebooks take longer to train, however the training process is only completed once before system implementation. Larger codebook sizes also require the comparison of test features to more codewords during the testing process, adding additional computational complexity to the system. Codebook spacing indicates what SNR levels are trained into codebooks. This allows better resolution during the testing process. Decreasing codebook spacing also increases training and testing time by creating more codebooks to train during training and more codewords to compare during testing. During this feasibility study the motivation for implementing a softer SNR decision method, or an estimating method, was developed based on observing the error distribution of initial systems. Implementing this approach adds little computational complexity to the system.

4.1.1 Initial AWGN Performance Results

Additive white Gaussian noise was utilized for the initial system tests. Codebooks of 32

to 256 codewords in steps of powers of two were created to test system performance to select a codebook size for each linear predictive feature. After selecting codebook sizes for each of the linear predictive features the codebook spacing, or resolution, was tested. Codebooks were spaced with 5 dB SNR resolution initially. Later 3 dB SNR resolution and 1 dB SNR resolution were tested. Since the SNR spectrum being studied was 0 to 30 dB, using a 5 dB resolution meant having at least a 2.5 dB error built into the system for signals corrupted at SNR levels in the center of two codebooks assuming a correct classification of one of the adjacent codebooks. 5 dB resolution also means an error of a multiple of 5 dB with incorrect classification. Using a 3 dB resolution reduced the built in error to 1.5 dB for signals corrupted at a level between codebooks and decreased the incorrect classification error to a multiple of 3 dB. A resolution of 1 dB reduces these sources of error again. System testing was only completed on integer SNR levels between 0 dB and 30 dB SNR, so the error between 1 dB spaced codebooks is not seen in these results.

4.1.2 Codebook Size

The first experiment performed to determine the feasibility of using a VQ Classifier based pattern recognition approach to providing an SNR estimate for a speech signal was to test different codebook sizes for each of the six features on a simple system. The system tested was trained in 5 dB increments from 0 to 30 dB SNR with codebooks sizes of 32 codewords, 64 codewords, 128 codewords, and 256 codewords. The systems were tested based on their classification performance and not average error. Confidence intervals were calculated to give an understanding of how these systems could be

expected to perform in a real application as opposed to the small trial size used. Tables 1 through 6 show the performance results for the systems to compare which size codebook to choose for each feature. Performance results in this experiment are shown in percentage of signals correctly classified. For all features it was observed that at each level of testing the largest codebook size, 256 codewords, performed overall the best. This codebook size was picked to be the standard size for each feature for all further experiments. Observing each features behavior in their 256 codebook systems show promising performance for further research using all six linear predictive features. The AWC Cepstrum and PFL Cepstrum, however, have the steepest drop in performance as the SNR of the test signals increases indicating that they will perform the poorest for AWGN. The ACW Cepstrum and PFL Cepstrum were kept for further testing to determine if they will contribute to performance in more complex systems. Another observed behavior from each of the codebook sizes is the drop in performance as the SNR of the test signals increases with a peak at the highest tested SNR. It can be inferred that more highly corrupt sentences, with lower SNR, have a greater difference in feature vector behavior than cleaner signals, and the peak at the clean extrema of the test range likely occurs because the system tends to classify a signal as a neighboring class and the 30 dB codebook only has one neighbor.

**Table 1: Cepstrum Codebook Size Performance
Measured in Percent Properly Classified**

Cepstrum		VQ Size			
	95% Confidence	32	64	128	256
30	Upper Bound	67.5	79.96	82.86	89.41
	Calculated Mean	60.53	73.68	76.84	84.21
	Lower Bound	53.56	67.41	70.83	79.01
25	Upper Bound	63.9	63.9	59.75	63.9
	Calculated Mean	56.84	56.84	52.63	56.84
	Lower Bound	49.78	49.78	45.51	49.78
20	Upper Bound	61.31	68	75.05	69.53
	Calculated Mean	54.21	61.05	68.42	62.63
	Lower Bound	47.11	54.1	61.79	55.73
15	Upper Bound	70.54	79.48	75.05	78.5
	Calculated Mean	63.68	73.16	68.42	72.11
	Lower Bound	56.83	66.84	61.79	65.71
10	Upper Bound	81.41	80.93	86.17	84.76
	Calculated Mean	75.26	74.74	80.53	78.95
	Lower Bound	69.11	68.54	74.88	73.14
5	Upper Bound	87.11	88.03	85.23	90.77
	Calculated Mean	81.58	82.63	79.47	85.79
	Lower Bound	76.05	77.23	73.72	80.81
0	Upper Bound	93.42	94.28	95.54	97.15
	Calculated Mean	88.95	90	91.58	93.68
	Lower Bound	84.48	85.72	87.62	90.22

Table 2: Line Spectral Frequency Codebook Size Performance Measured in Percent Properly Classified

Line Spectral Frequency			VQ Size		
	95% Confidence	32	64	128	256
30	Upper Bound	67.5	73.05	71.05	70.04
	Calculated Mean	60.53	66.32	64.21	63.16
	Lower Bound	53.56	59.58	57.38	56.28
25	Upper Bound	48.07	57.65	61.83	70.54
	Calculated Mean	41.05	50.53	54.74	63.68
	Lower Bound	34.04	43.4	47.64	56.83
20	Upper Bound	77.52	72.05	75.54	77.03
	Calculated Mean	71.05	65.26	68.95	70.53
	Lower Bound	64.59	58.47	62.35	64.03
15	Upper Bound	78.01	83.33	82.38	83.81
	Calculated Mean	71.58	77.37	76.32	77.89
	Lower Bound	65.15	71.4	70.25	71.98
10	Upper Bound	87.57	86.17	91.22	90.77
	Calculated Mean	82.11	80.53	86.32	85.79
	Lower Bound	76.64	74.88	81.42	80.81
5	Upper Bound	94.28	94.7	98.65	97.92
	Calculated Mean	90	90.53	95.79	94.74
	Lower Bound	85.72	86.35	92.93	91.55
0	Upper Bound	97.54	98.65	99	99.34
	Calculated Mean	94.21	95.79	96.32	96.84
	Lower Bound	90.88	92.93	93.63	94.35

Table 3: Reflection Coefficient Codebook Size Performance Measured in Percent Properly Classified

Reflection Coefficient			VQ Size		
SNR (dB)	95% Confidence	32	64	128	256
30	Upper Bound	72.05	76.53	77.52	76.04
	Calculated Mean	65.26	70	71.05	69.47
	Lower Bound	58.47	63.47	64.59	62.91
25	Upper Bound	53.42	53.42	58.18	71.89
	Calculated Mean	46.32	46.32	51.05	65.08
	Lower Bound	39.21	39.21	43.93	58.26
20	Upper Bound	70.54	77.52	69.53	72.05
	Calculated Mean	63.68	71.05	62.63	65.26
	Lower Bound	56.83	64.59	55.73	58.47
15	Upper Bound	70.54	76.53	81.9	82.86
	Calculated Mean	63.68	70	75.79	76.84
	Lower Bound	56.83	63.47	69.68	70.83
10	Upper Bound	82.38	82.86	82.86	86.64
	Calculated Mean	76.32	76.84	76.84	81.05
	Lower Bound	70.52	70.83	70.83	75.47
5	Upper Bound	84.76	89.86	90.32	88.95
	Calculated Mean	78.95	84.74	85.26	83.68
	Lower Bound	73.14	79.61	80.21	78.42
0	Upper Bound	99	97.92	97.92	99.34
	Calculated Mean	96.32	94.74	94.74	96.84
	Lower Bound	93.63	91.55	91.55	94.35

**Table 4: Log Area Ratio Codebook Size Performance
Measured in Percent Properly Classified**

Log Area Ratio			VQ Size		
SNR (dB)	95% Confidence	32	64	128	256
30	Upper Bound	65.96	75.05	75.54	79.48
	Calculated Mean	58.95	68.42	68.95	73.16
	Lower Bound	51.93	61.79	62.35	66.84
25	Upper Bound	51.83	57.65	62.35	63.18
	Calculated Mean	44.74	50.53	55.26	56.08
	Lower Bound	37.65	43.4	48.17	48.99
20	Upper Bound	60.79	69.02	71.05	72.55
	Calculated Mean	53.68	62.11	64.21	65.79
	Lower Bound	46.58	55.19	57.38	59.03
15	Upper Bound	75.05	74.05	74.05	79.96
	Calculated Mean	68.42	67.37	67.37	73.68
	Lower Bound	61.79	60.68	60.68	67.41
10	Upper Bound	88.03	82.38	86.17	83.33
	Calculated Mean	82.63	76.32	80.53	77.37
	Lower Bound	77.23	70.25	74.88	71.4
5	Upper Bound	88.95	91.22	90.32	91.22
	Calculated Mean	83.68	86.32	85.26	86.32
	Lower Bound	78.42	81.42	80.21	81.42
0	Upper Bound	97.92	97.54	97.92	99.34
	Calculated Mean	94.74	94.21	94.74	96.84
	Lower Bound	91.55	90.88	91.55	94.35

**Table 5: ACW Cepstrum Codebook Size Performance
Measured in Percent Properly Classified**

ACW Cepstrum			VQ Size		
	95% Confidence	32	64	128	256
30	Upper Bound	60.27	74.55	78.01	72.05
	Calculated Mean	53.16	67.89	71.58	65.26
	Lower Bound	46.04	61.24	65.15	58.47
25	Upper Bound	49.14	46.44	52.36	58.47
	Calculated Mean	42.11	39.47	45.26	51.32
	Lower Bound	35.07	32.5	38.17	44.18
20	Upper Bound	53.42	68	64.42	58.18
	Calculated Mean	46.32	61.05	57.37	51.05
	Lower Bound	39.21	54.1	50.32	43.93
15	Upper Bound	56.07	55.55	61.31	75.05
	Calculated Mean	48.95	48.42	54.21	68.42
	Lower Bound	41.82	41.3	47.11	61.79
10	Upper Bound	66.47	65.45	74.55	76.53
	Calculated Mean	59.47	58.42	67.89	70
	Lower Bound	52.47	51.39	61.24	63.47
5	Upper Bound	76.04	75.54	80.93	79.96
	Calculated Mean	69.47	68.95	74.74	73.68
	Lower Bound	62.91	62.35	68.54	67.41
0	Upper Bound	92.98	97.15	96.36	97.54
	Calculated Mean	88.42	93.68	92.63	94.21
	Lower Bound	83.86	90.22	88.91	90.88

**Table 6: PFL Cepstrum Codebook Size Performance
Measured in Percent Properly Classified**

PFL Cepstrum			VQ Size		
SNR (dB)	95% Confidence	32	64	128	256
	30	Upper Bound	65.45	66.98	77.03
	Calculated Mean	58.42	60	70.53	71.05
	Lower Bound	51.39	53.02	64.03	64.59
25	Upper Bound	46.44	57.65	55.02	57.41
	Calculated Mean	39.47	50.53	47.89	50.26
	Lower Bound	32.5	43.4	40.77	43.12
20	Upper Bound	56.07	53.96	53.96	65.45
	Calculated Mean	48.95	46.84	46.84	58.42
	Lower Bound	41.82	39.73	39.73	51.39
15	Upper Bound	66.47	72.05	77.52	76.04
	Calculated Mean	59.47	65.26	71.05	69.47
	Lower Bound	52.47	58.47	64.59	62.91
10	Upper Bound	70.54	74.05	70.54	76.04
	Calculated Mean	63.68	67.37	63.68	69.47
	Lower Bound	56.83	60.68	56.83	62.91
5	Upper Bound	79.48	83.81	79.48	86.64
	Calculated Mean	73.16	77.89	73.16	81.05
	Lower Bound	66.84	71.98	66.84	75.47
0	Upper Bound	90.32	94.7	97.92	97.92
	Calculated Mean	85.26	90.53	94.74	94.74
	Lower Bound	80.21	86.35	91.55	91.55

4.1.3 Codebook Spacing

Since the classes being classified by this pattern recognition approach are actually SNR levels which fall on a linear range and training data is obtained by corrupting a set of speech signals to the required levels of corruption, it is possible to increase the resolution of the classification system by creating codebooks of data corrupted to levels closer on the linear range. Three systems were created by spacing codebooks at 5 dB

intervals, 3 dB intervals, and 1 dB intervals. Table 7 shows the classification overall absolute average error for each of these three systems for comparison. It can be seen that as the system resolution increases, or the spacing between codebooks decreases, the OAAE decreases. This can be attributed to the fact that when an error is made at a higher resolution it is usually on a lower magnitude. Also, less spacing between codebooks provides smaller ranges that are not covered for direct classification which requires these ranges to be rounded to a local trained SNR level. In table 7 it can be seen that the OAAE for each of the linear predictive features reduces between at least .23 dB for the Reflection Coefficients and at most .29 dB for the PFL Cepstrum. Figures 4 through 9 compare the average absolute error at integer test SNR levels between 0 dB SNR and 30 dB SNR for the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum features respectively. Each graph shows the same general behavior when comparing the use of different SNR resolutions. When codebooks are spaced 5 dB apart, the trained SNR levels have the lowest error while the error increases as the distance from a codebook increases. The same behavior can be seen for 3 dB spaced codebooks, but the error at trained levels is higher and the error at untrained levels is lower than with 5 dB spacing. The system with 1 dB resolution has been trained at all tested levels. It has higher error at its trained levels than the 5 dB and 3 dB systems have at their trained levels, however it has much lower error in the areas which the 5 dB and 3 dB systems peak between trained levels. The general behavior observed is that as the resolution increases, or the spacing decreases, the behavior of the AAE smooths which results in a lower OAAE, providing better overall system

performance.

Table 7: AWGN Results Comparing Codebook Resolution Through OAAE in dB

Feature	Codebook Space	Classification OAAE
LSF	1 dB	1.76
LSF	3 dB	1.87
LSF	5 dB	2.02
CEP	1 dB	1.85
CEP	3 dB	1.96
CEP	5 dB	2.09
REFL	1 dB	1.84
REFL	3 dB	1.92
REFL	5 dB	2.07
LAR	1 dB	1.83
LAR	3 dB	1.94
LAR	5 dB	2.08
ACW	1 dB	2.09
ACW	3 dB	2.23
ACW	5 dB	2.35
PFL	1 dB	2.01
PFL	3 dB	2.16
PFL	5 dB	2.30

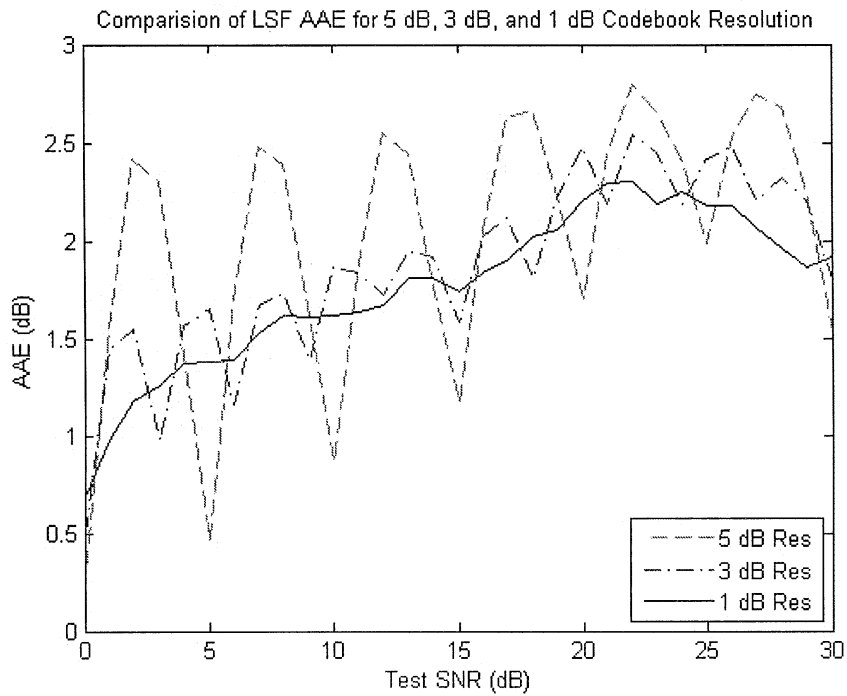


Figure 4. Comparison of LSF Resolution Classification AAE

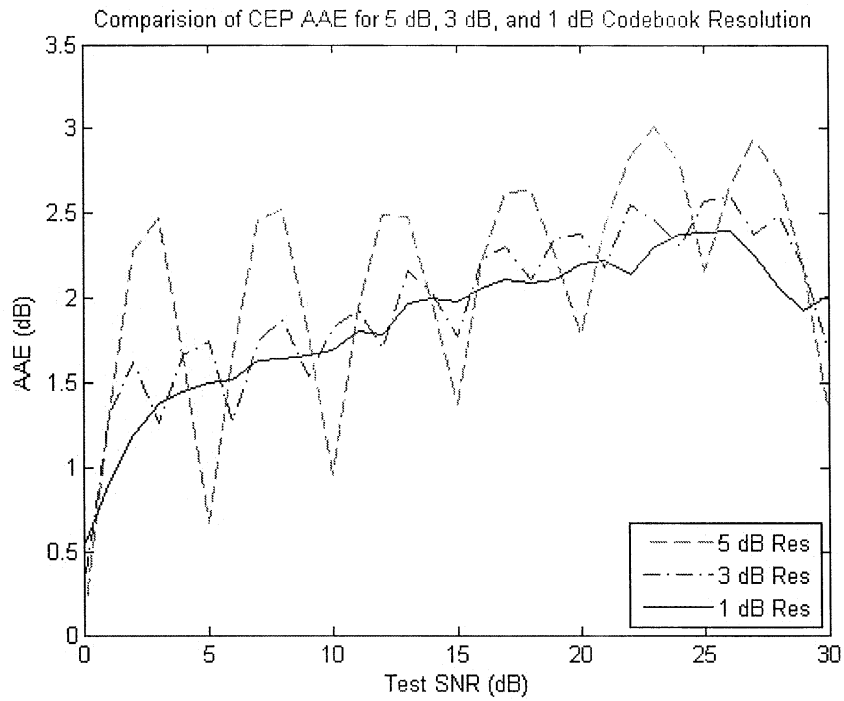


Figure 5. Comparison of CEP Resolution Classification AAE

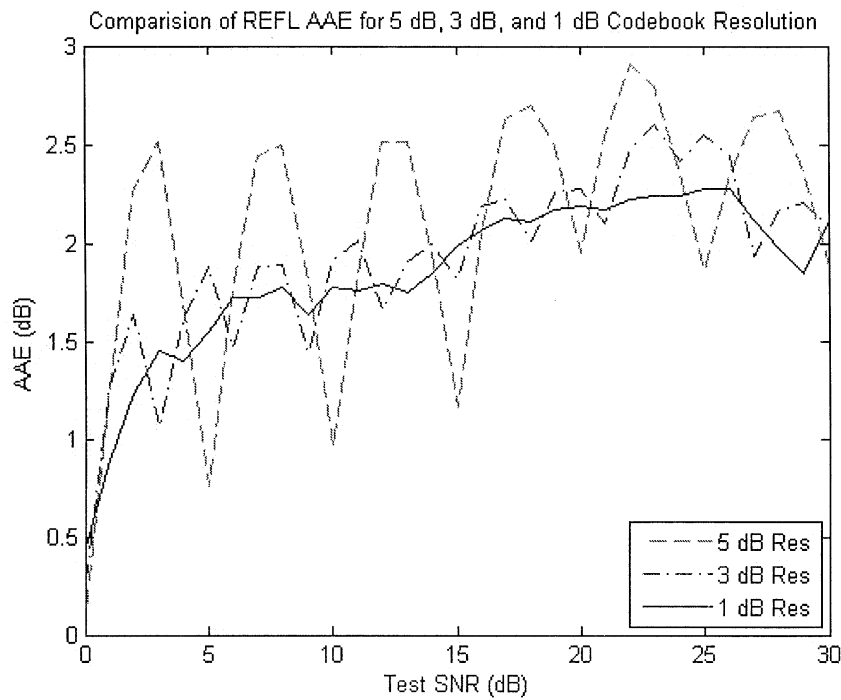


Figure 6. Comparison of REFL Resolution Classification AAE

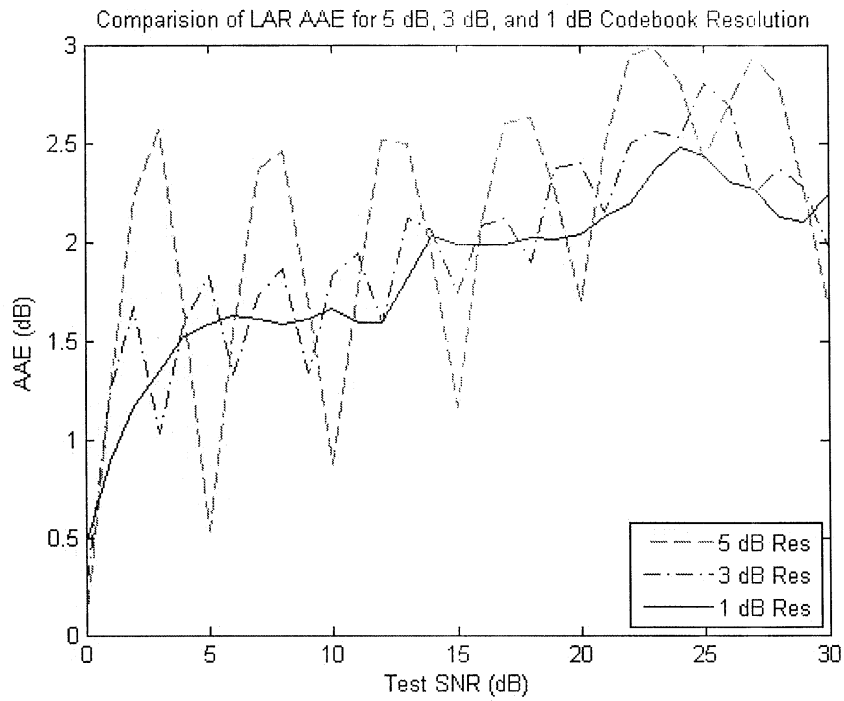


Figure 7. Comparison of LAR Resolution Classification AAE

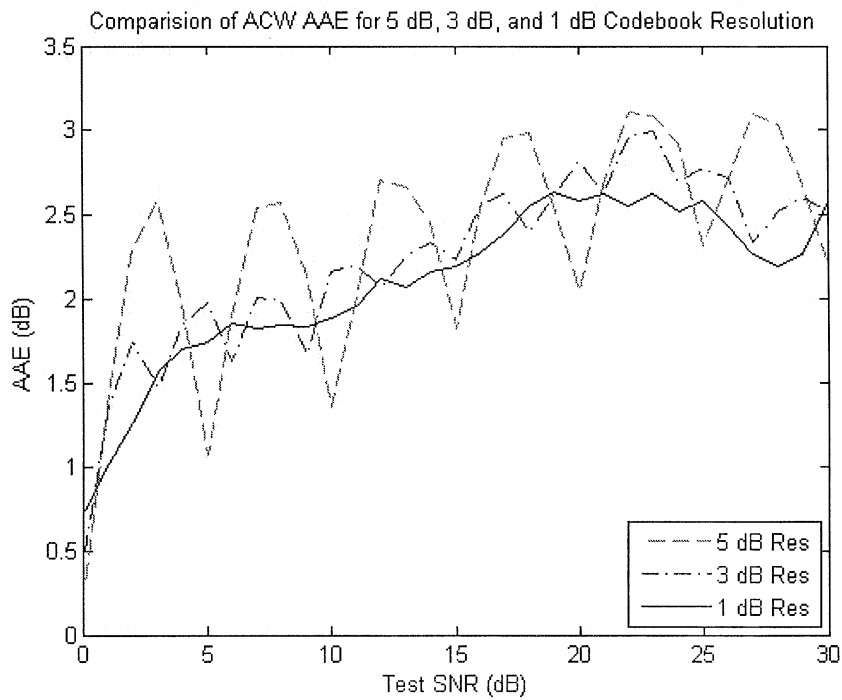


Figure 8. Comparison of ACW Resolution Classification AAE

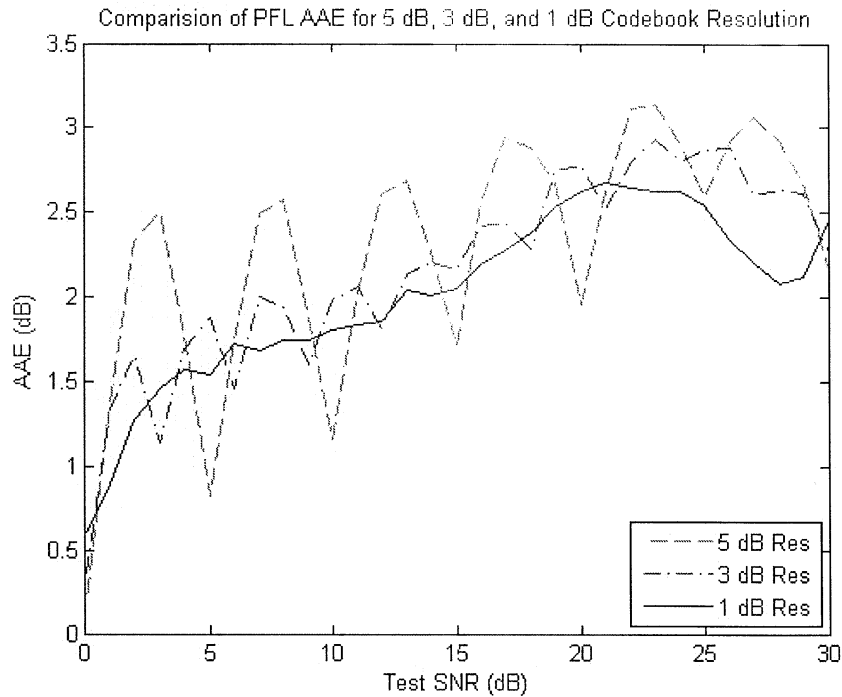


Figure 9. Comparison of PFL Resolution Classification AAE

4.1.4 VQ Estimation

4.1.4.1 Motivation

Based on the results of the histograms in figures 10 through 15, corresponding to the error distribution of the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum respectively, a soft decision scoring method was developed to estimate the SNR of a sentence based on the distances to the VQ codebooks of each SNR level. This method normalized each distance into a probability:

$$p_i = \frac{s - d_i}{s * (dn - 1)}$$

where p_i is the probability calculated for a specific SNR codebook i , s is the sum of the

calculated distances obtained for each codebook. d_i is the distance calculated for the specific SNR codebook i , and d_n is the number of total SNR levels. This equation assigns a higher probability to smaller distances. These probabilities add up to 1 and can be used to estimate the sentence's SNR:

$$SNR = \sum_1^{d_n} p_i * SNR_i$$

where SNR_i is the SNR the probability p_i was calculated for. By definition, a hard decision is a decision or classification on the SNR based on the smallest distance only. A soft decision or estimation is obtained from a combination of the data gained from the pattern recognition process. It was observed that using the soft scoring method presented on all codebook distances, the estimated SNR would always be close to the middle SNR value of the test. In the case of this experiment, testing from 0 to 30 dB at intervals of 5 dB, the SNR was always estimated to be approximately 15 dB. This was due to the nature of the distances being used to calculate the probabilities. Though the distances are smaller for the SNR levels close to the correct level and larger for those SNR levels farther from the correct SNR level, the distances are relatively close. Use of the three smallest distances as opposed to every distance was motivated since incorrect classifications are observed to classify the SNR of a sentence as a neighboring SNR level.

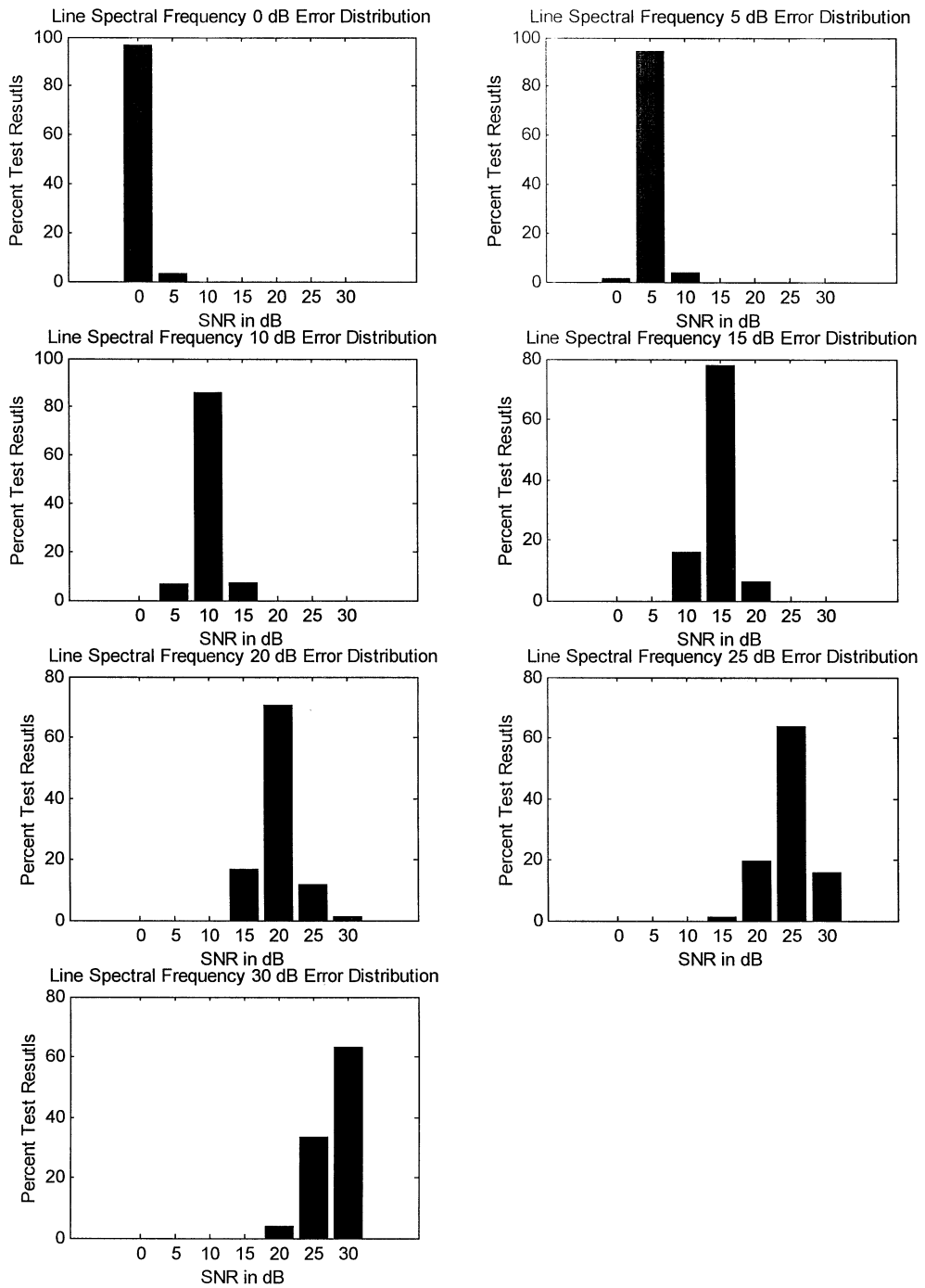


Figure 10. Error Analysis Graphs for Line Spectral Frequency Feature

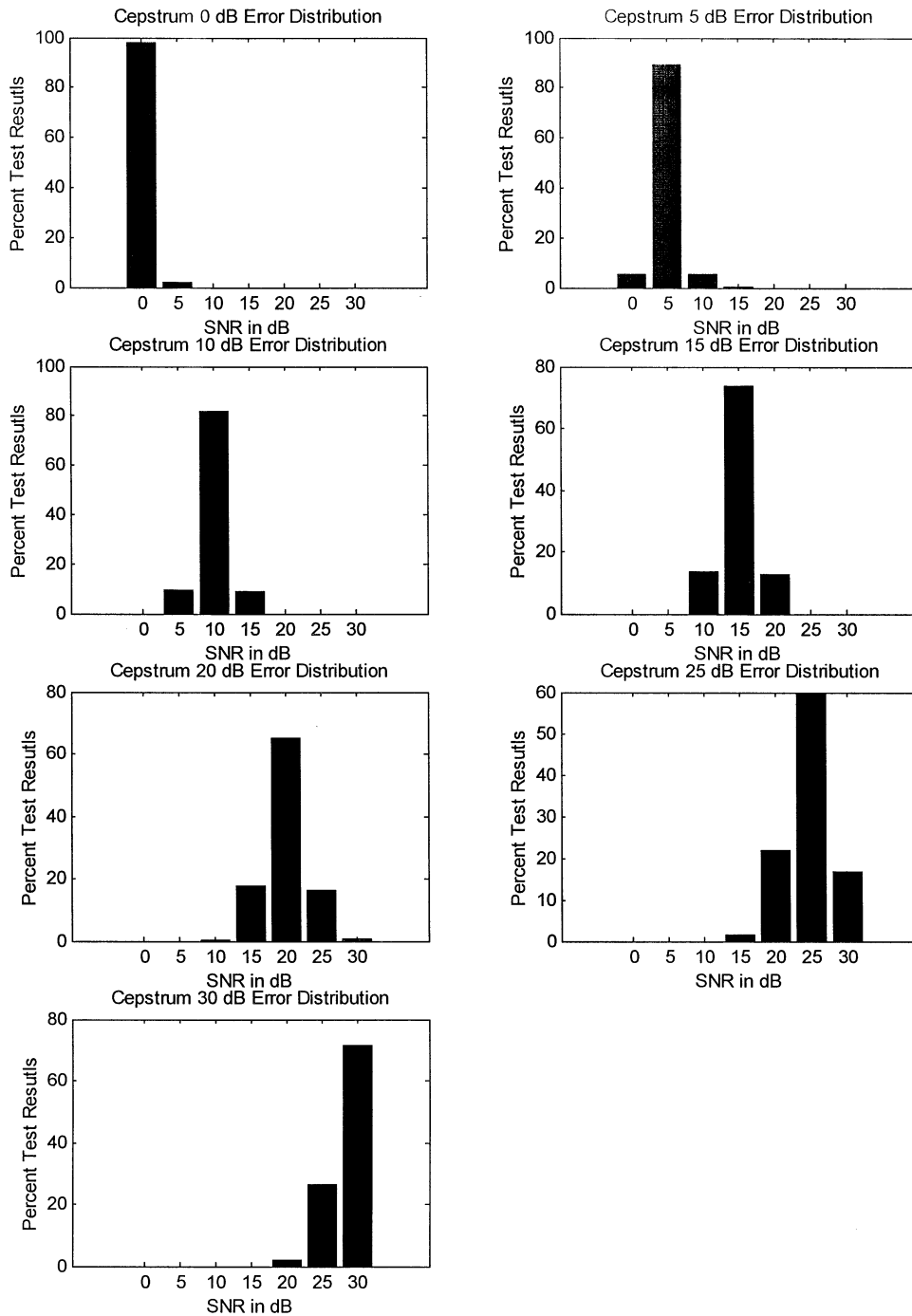


Figure 11. Error Analysis Graphs for LP Cepstrum Feature

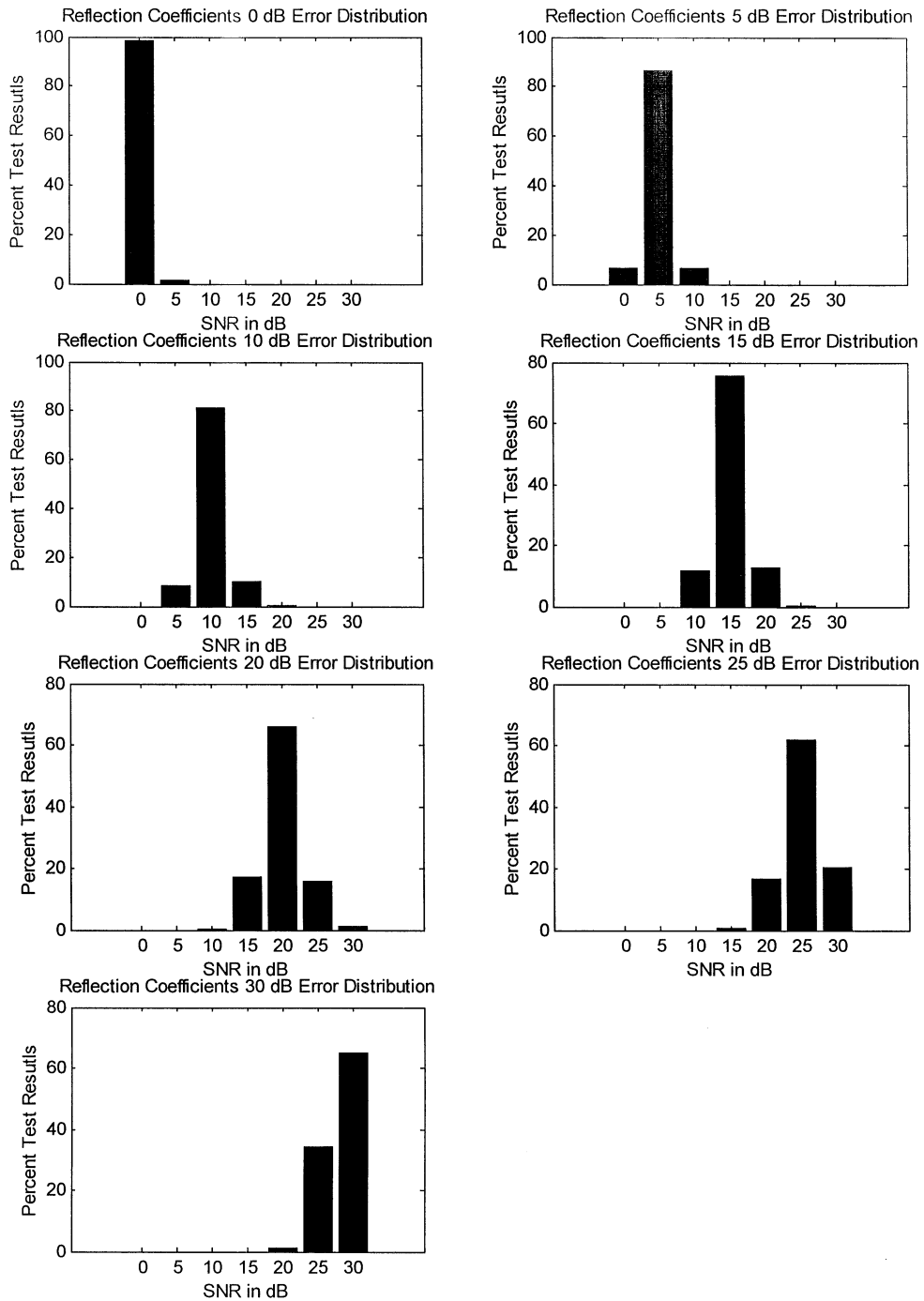


Figure 12. Error Analysis Graphs for Reflection Coefficient Feature

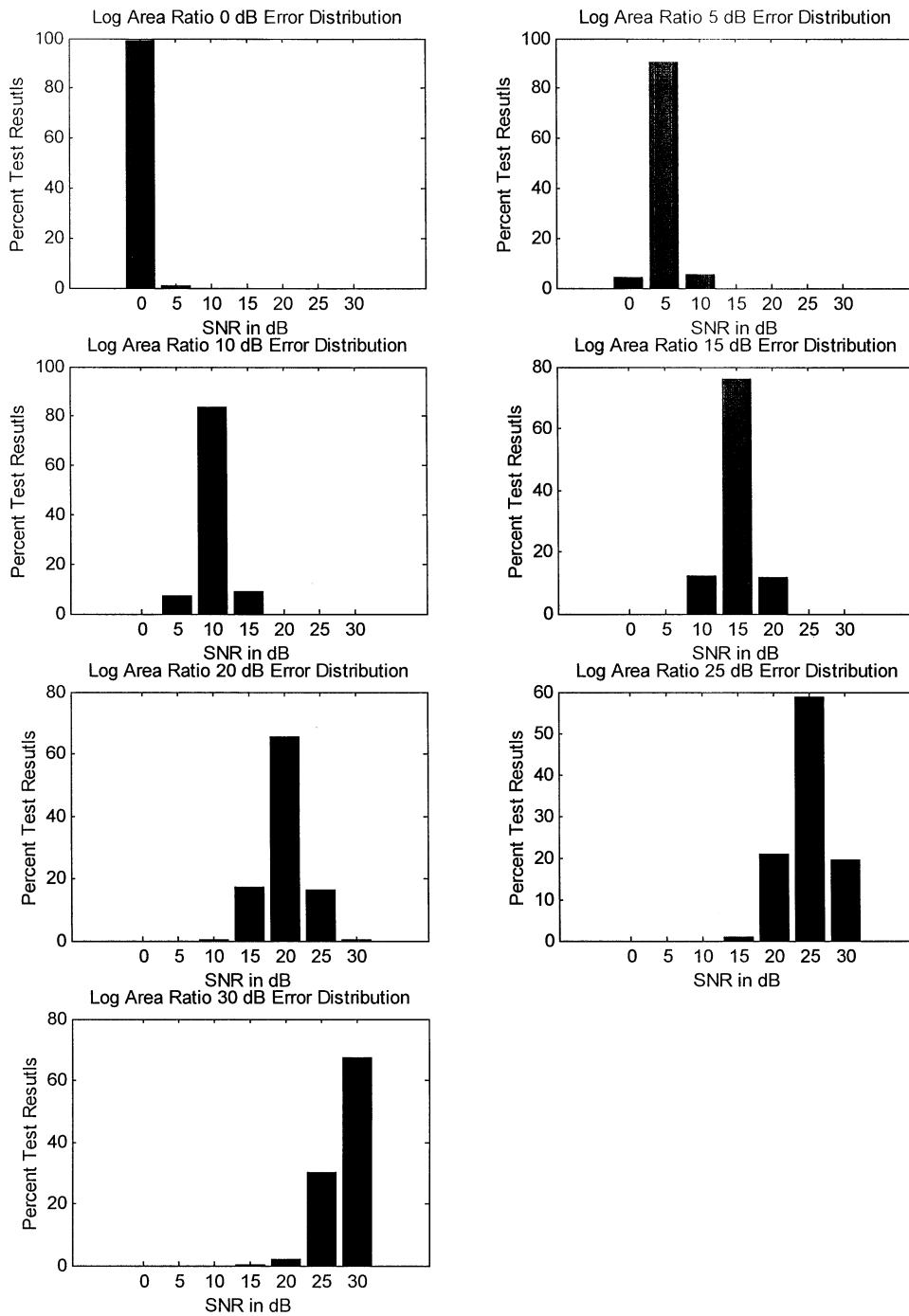


Figure 13. Error Analysis Graphs for Log Area Ratio Feature

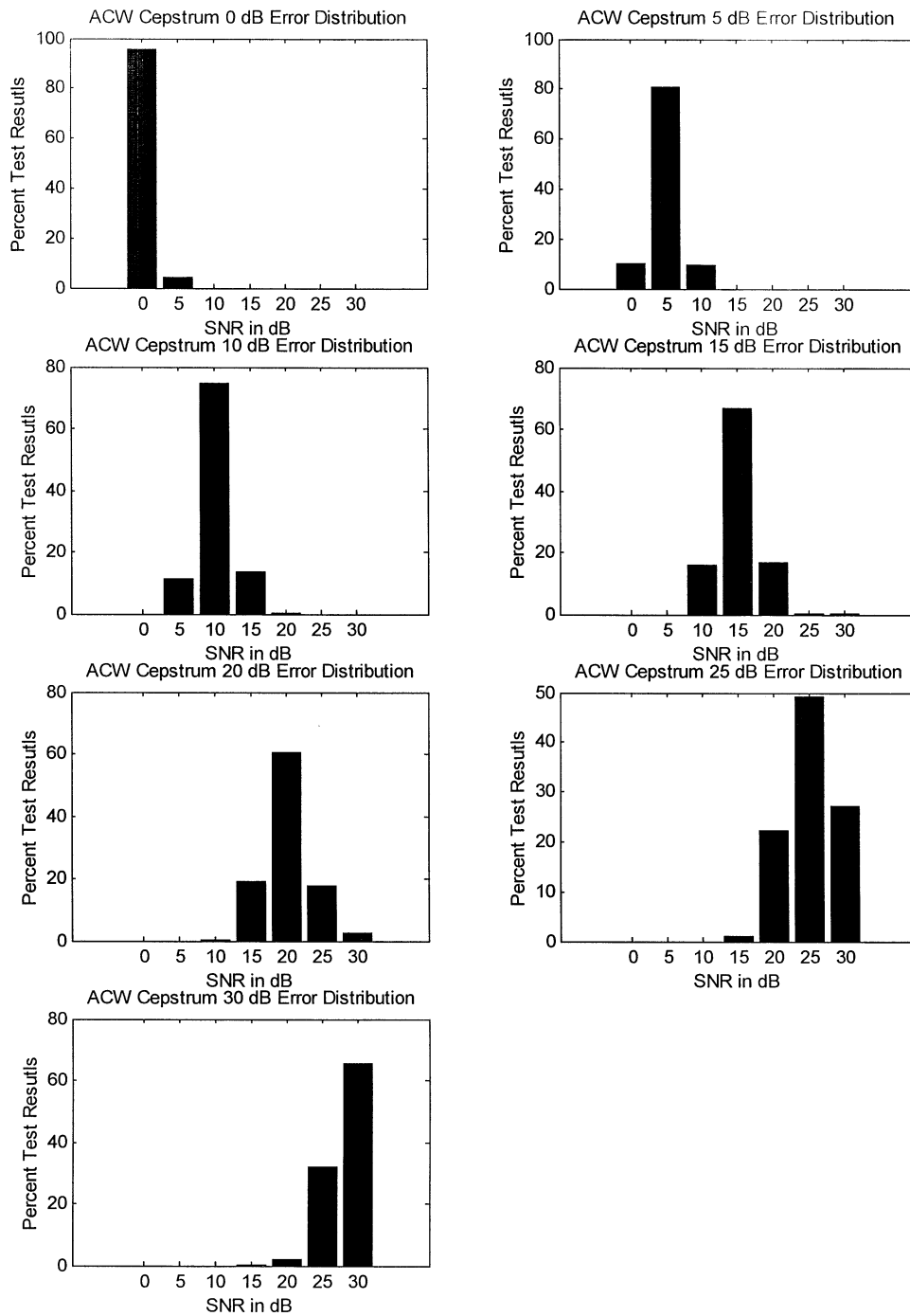


Figure 14. Error Analysis Graphs for ACW Cepstrum Feature

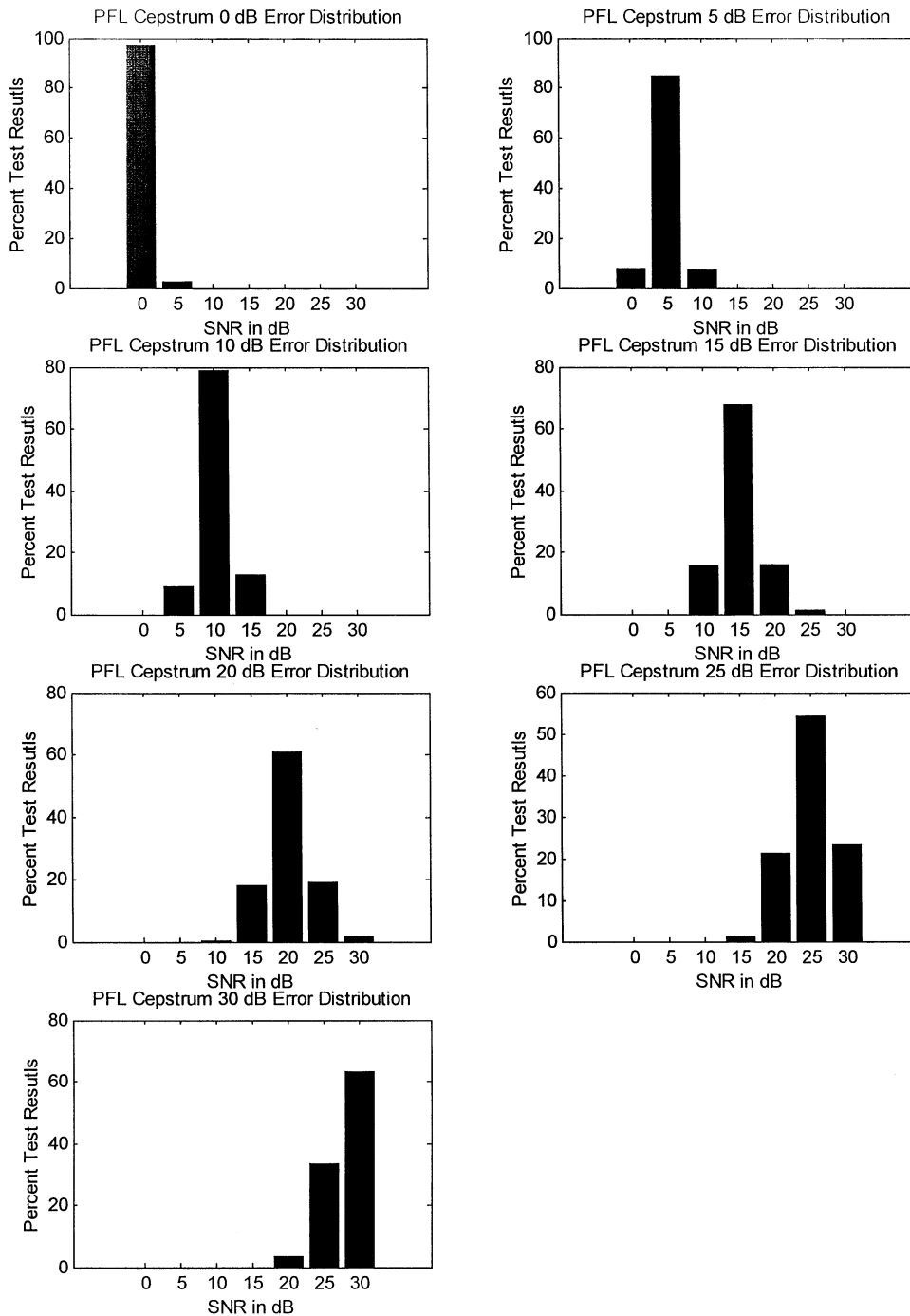


Figure 15. Error Analysis Graphs for PFL Cepstrum Feature

4.1.4.2 AWGN Results

Implementing the softer decision estimation method it was expected to see error reduce. Table 8 compares the OAAE obtained for each of the three resolutions when a soft decision is used which provides an estimation based on a weighted average of the three smallest codebook decisions. It is observed that the soft decision estimation technique effectively reduces error for the 3 dB and 1 dB codebook space resolutions for all linear predictive features. When codebooks are spaced 5 dB apart the OAAE of the system increases in all cases. Figures 16 through 18 compare the hard decision classification method to the soft decision estimation method for the Line Spectral Frequency feature for the systems built with 5 dB spacing, 3 dB spacing, and 1 dB spacing respectively. All features show similar behavior, so the LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum only compare the system with 1 dB spacing in figures 19 through 23 respectively. When comparing the hard decision classification method to the soft decision estimation method using the LSF system with 5 dB spacing, the same general behavior is observed for both methods, however the estimation method performs more poorly at lower SNR levels and peaks to approximately 5 dB at either extrema of the tested range. This peak is attributed to the weighted averaging of three SNR levels denoted by the codebooks associated with the lowest distances. This averaging, though weighted, generally averages the three closest trained SNR levels evenly causing any signal below 5 dB to be estimated close to 5 dB and any signal above 25 dB to be estimated close to 25dB. The same behavior is observed at the extrema for 3 dB spacing and 1 dB spacing, however to a lesser degree due to the

resolution of the system. It is also observed for the 3 dB spaced system and all 1 dB spaced systems that the soft decision estimation method generally provides SNR estimates of corrupted signals with less error, especially in the cleaner region associated with higher SNR levels. The behavior of the estimation method peaking at the extrema and having the gap widen when compared to classification error as signals become less corrupt is observed in all six features. This leads to the conclusion of using a final system which employs the soft decision estimation technique with codebooks created with 1 dB resolution. Table 8 shows that the best performing feature for the AWGN system is the Log Area Ratio with an OAAE of 1.59 dB for the soft decision estimation method at 1 dB resolution.

Table 8: AWGN Results Comparing VQ Classification and VQ Estimation OAAE in dB

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	1.76	1.61
LSF	3 dB	1.87	1.78
LSF	5 dB	2.02	2.35
CEP	1 dB	1.85	1.68
CEP	3 dB	1.96	1.82
CEP	5 dB	2.09	2.33
REFL	1 dB	1.84	1.62
REFL	3 dB	1.92	1.72
REFL	5 dB	2.07	2.21
LAR	1 dB	1.83	1.59
LAR	3 dB	1.94	1.71
LAR	5 dB	2.08	2.23
ACW	1 dB	2.09	1.85
ACW	3 dB	2.23	1.94
ACW	5 dB	2.35	2.33
PFL	1 dB	2.01	1.78
PFL	3 dB	2.16	1.89
PFL	5 dB	2.30	2.32

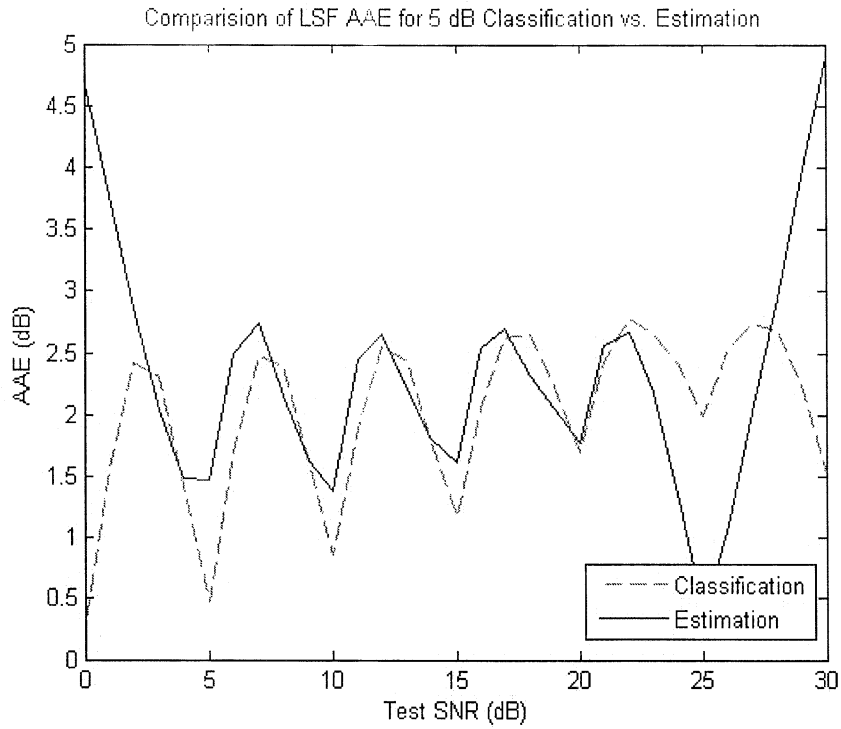


Figure 16. Comparison of Classification and Estimation AAE for LSF with 5 dB Spaced Codebooks

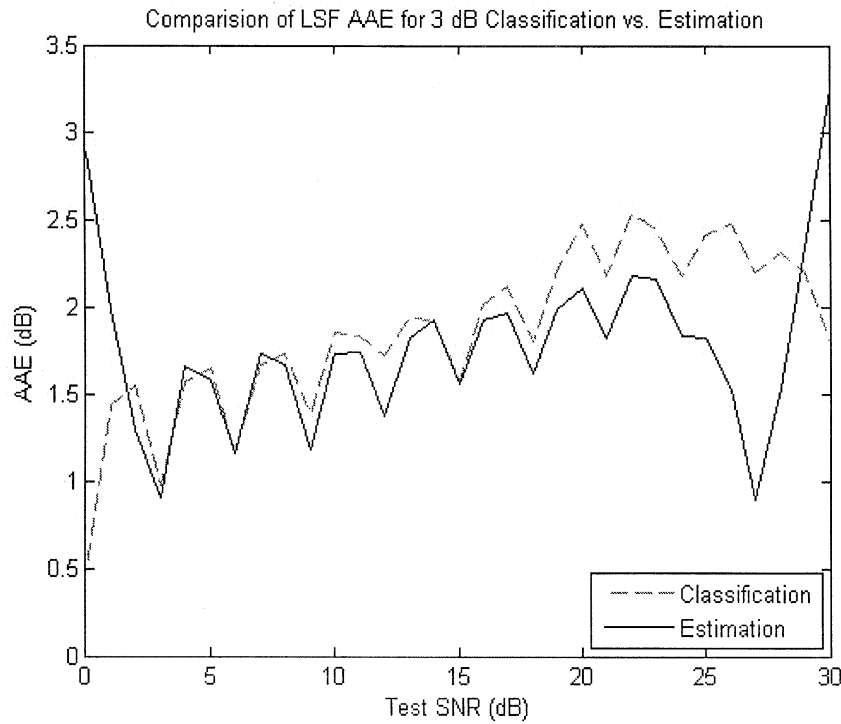


Figure 17. Comparison of Classification and Estimation AAE for LSF with 3 dB Spaced Codebooks

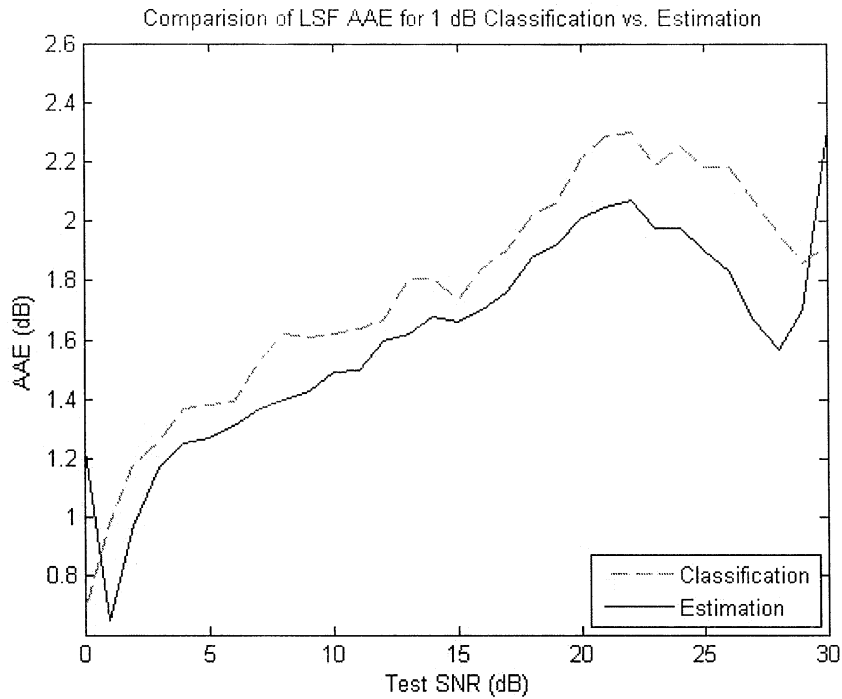


Figure 18. Comparison of Classification and Estimation AAE for LSF with 1 dB Spaced Codebooks

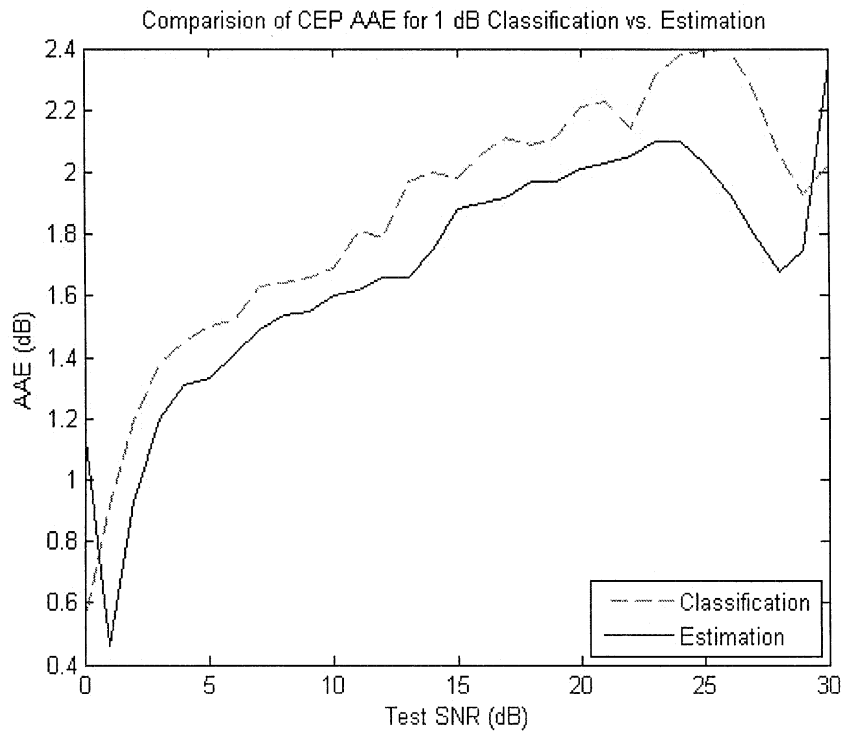


Figure 19. Comparison of Classification and Estimation AAE for CEP with 1 dB Spaced Codebooks

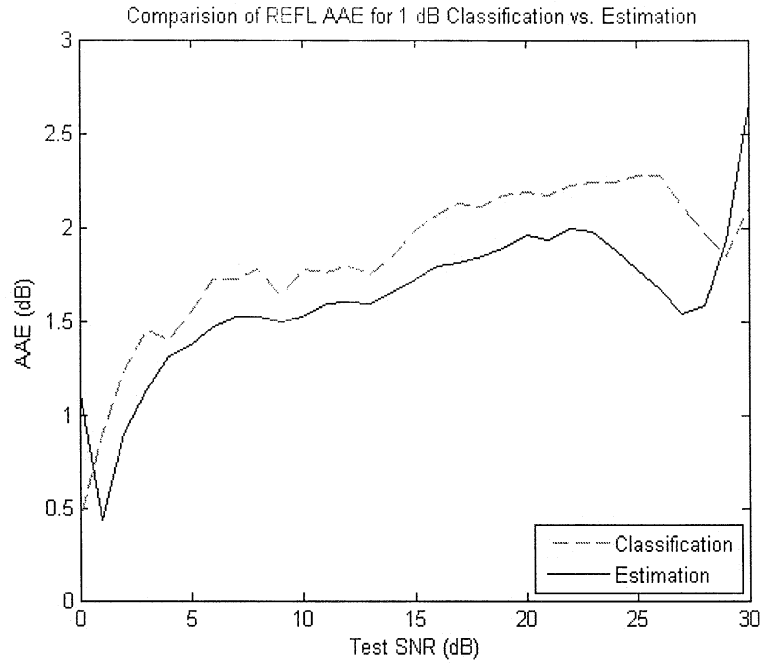


Figure 20. Comparison of Classification and Estimation AAE for REFL with 1 dB Spaced Codebooks

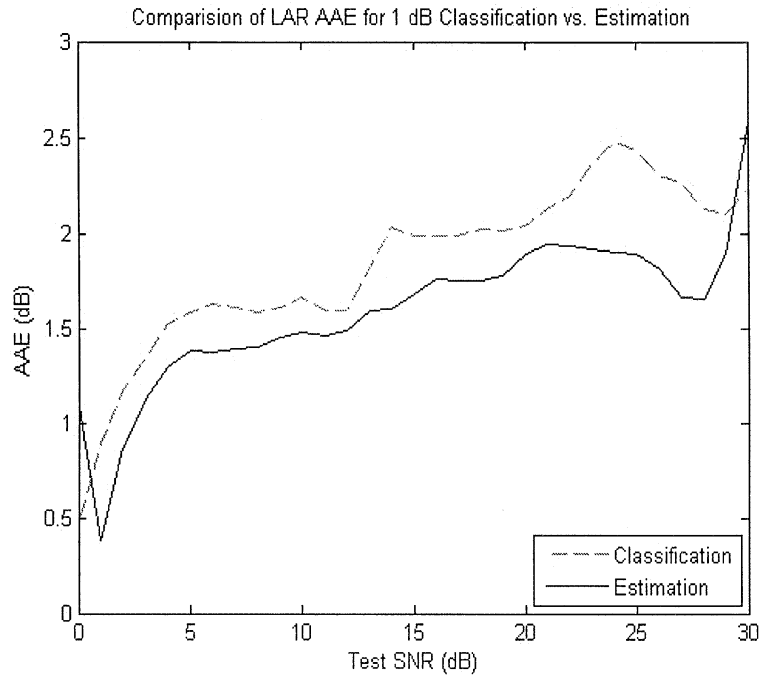


Figure 21. Comparison of Classification and Estimation AAE for LAR with 1 dB Spaced Codebooks

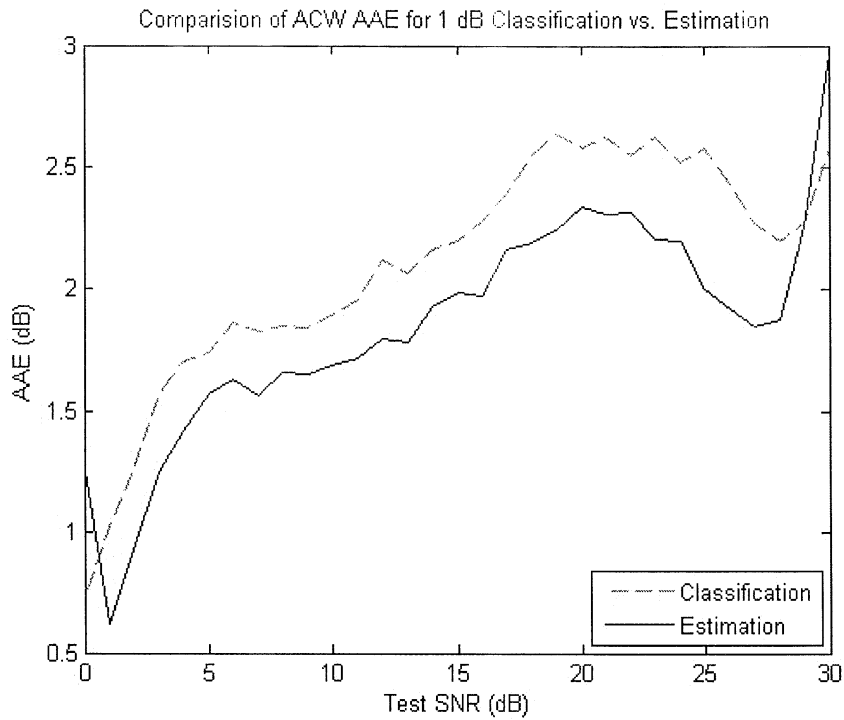


Figure 22. Comparison of Classification and Estimation AAE for ACW with 1 dB Spaced Codebooks

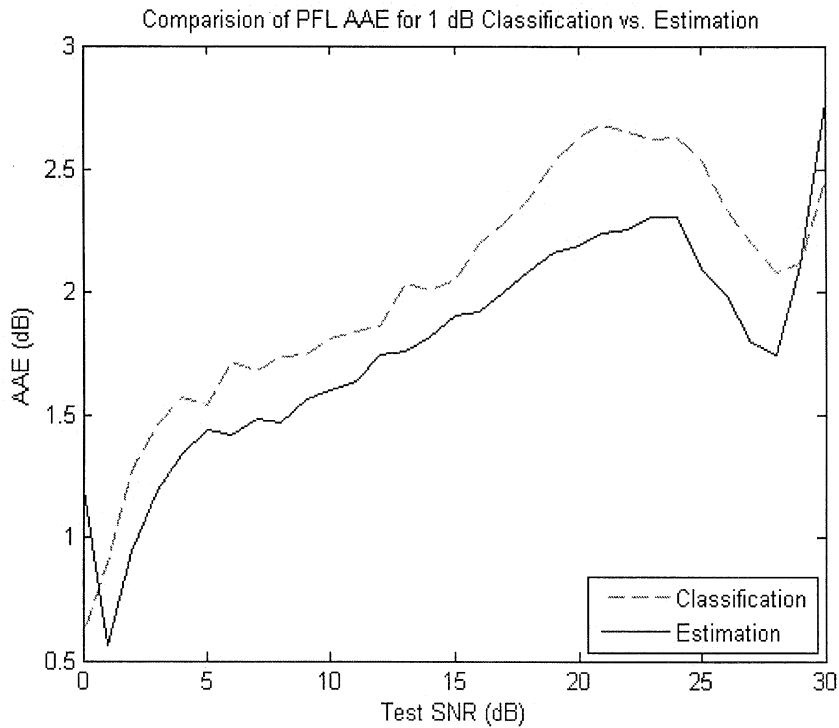


Figure 23. Comparison of Classification and Estimation AAE for PFL with 1 dB Spaced Codebooks

4.2 Noise Spectrum Robustness

Low OAAE errors for all linear predictive features on a system designed specifically for AWGN has been achieved. A system designed for AWGN only takes into account one spectrum of noise. To develop a SNR estimation system robust to the spectrum of the additive noise in the signal, other types of noise must be tested and further made into a single system which will perform well regardless of the spectrum of the noise corrupting a signal. First, a VQ system was designed specifically for Pink noise to test its ability to estimate the SNR of a signal corrupted with Pink noise. Second, a system was designed specifically for CPV noise. After testing these systems on the noise spectrum for which they were trained, each of three systems designed for Soft Decision specific noise spectrum; including AWGN, Pink noise, and CPV noise; were tested on the two spectra not included in their training. Finally, a system was developed by using all three noise spectra for training to determine the system's ability to act as a speech signal SNR estimator that is robust to the unspecified noise spectrum corrupting the input signal. To further test the robustness of this system a fourth untrained noise spectrum, CMV noise, was tested against this robust system.

4.2.1 Pink Noise

The system designed for Pink noise performed with the same OAAE and AAE behavior as the system for AWGN when comparing their separate results. The OAAE for both the hard decision classification method and the soft decision estimation method at the three tested resolutions can be seen in table 9. One notable difference is that the soft decision estimation method does improve the system designed with 5 dB SNR resolution

for all features. The OAAE for each feature increases for the Pink system when compared to the AWGN system indicating that Pink noise SNR is harder to classify than AWGN SNR. The best feature for the Pink noise system, as seen in table 9, is the Reflection coefficients with an OAAE for the soft decision estimation method at 1 dB resolution of 1.85 dB. Figures 24 through 29 show a comparison of the hard decision classification method and the soft decision estimation method for the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratio, ACW Cepstrum, and PFL Cepstrum features respectively. The same overall behavior is observed for each of these features when regarding Pink noise as with AWGN. These results show that the system will work when estimating the SNR of a signal corrupted by Pink noise if Pink noise is used to train the system.

Table 9: OAAE Results in db for System Designed with Pink Noise

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	2.18	1.94
LSF	3 dB	2.25	1.98
LSF	5 dB	2.35	2.32
CEP	1 dB	2.24	2.01
CEP	3 dB	2.34	2.04
CEP	5 dB	2.43	2.36
REFL	1 dB	2.17	1.85
REFL	3 dB	2.30	1.94
REFL	5 dB	2.41	2.32
LAR	1 dB	2.22	1.89
LAR	3 dB	2.35	1.96
LAR	5 dB	2.46	2.34
ACW	1 dB	2.49	2.14
ACW	3 dB	2.63	2.24
ACW	5 dB	2.73	2.54
PFL	1 dB	2.44	2.09
PFL	3 dB	2.56	2.18
PFL	5 dB	2.69	2.51

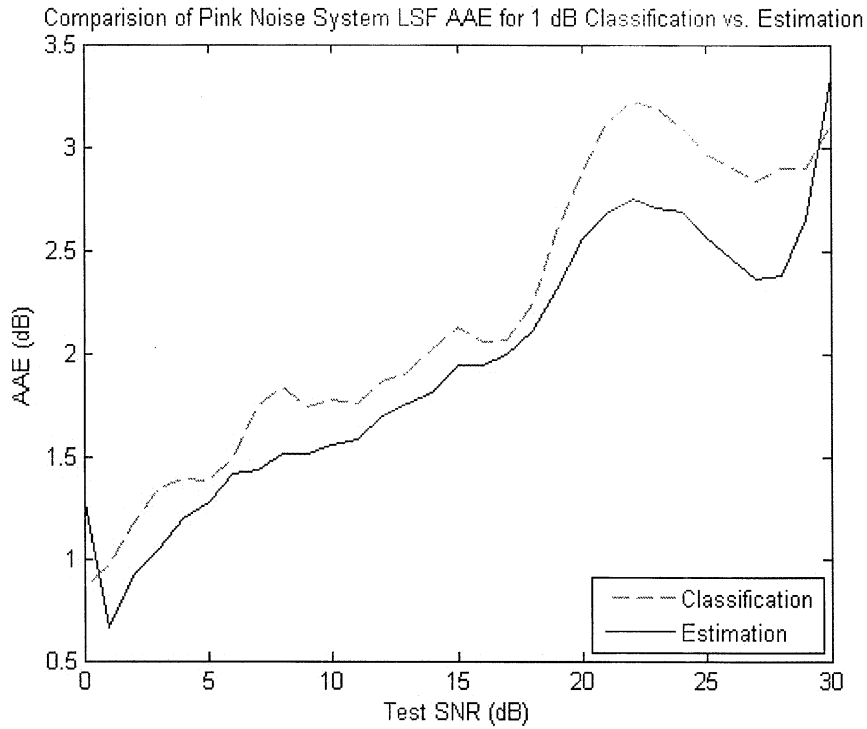


Figure 24. Comparison of Classification and Estimation AAE for LSF in the Pink System with 1 dB Spaced Codebooks

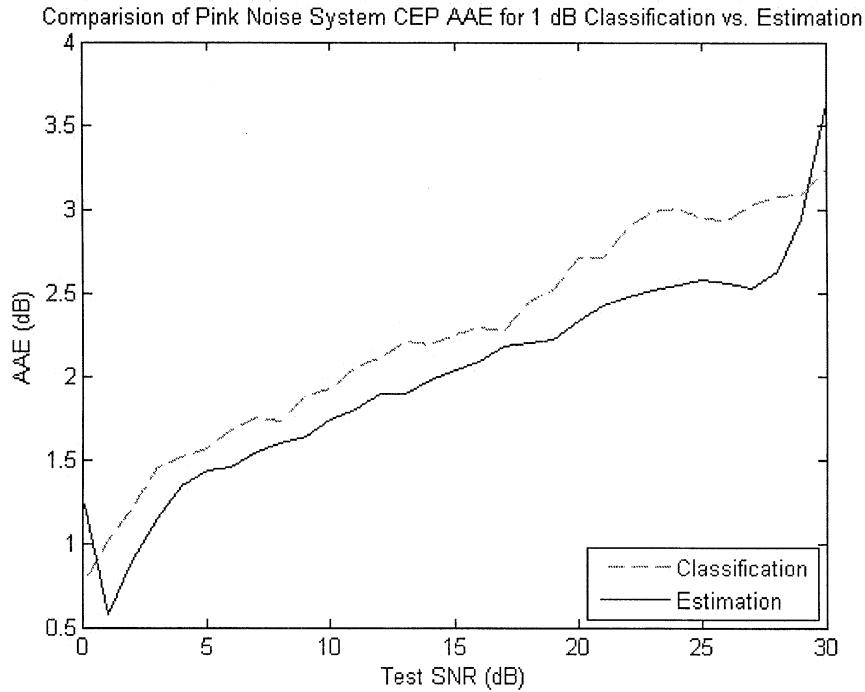


Figure 25. Comparison of Classification and Estimation AAE for CEP in the Pink System with 1 dB Spaced Codebooks

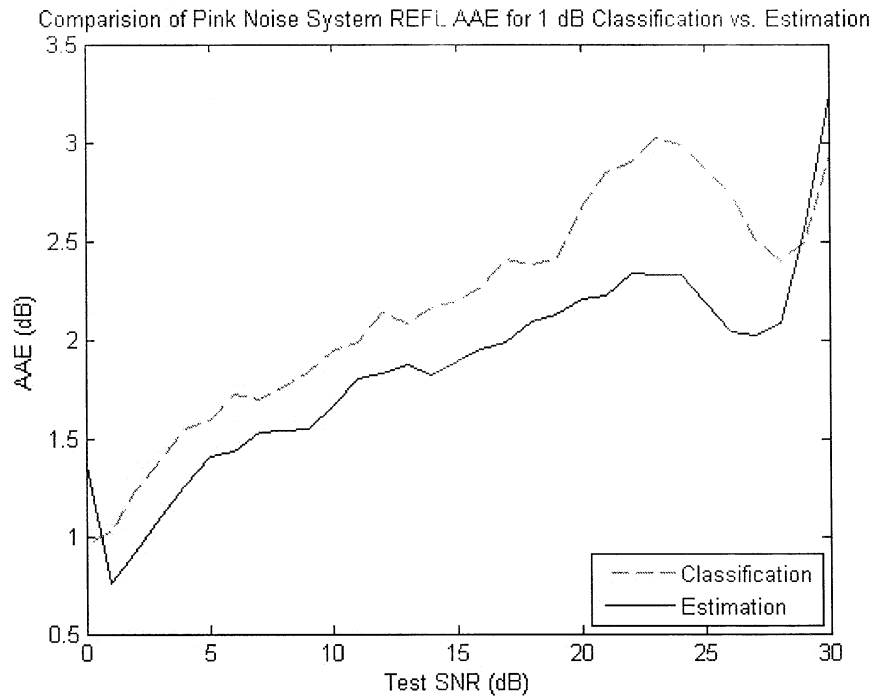


Figure 26. Comparison of Classification and Estimation AAE for REFL in the Pink System with 1 dB Spaced Codebooks

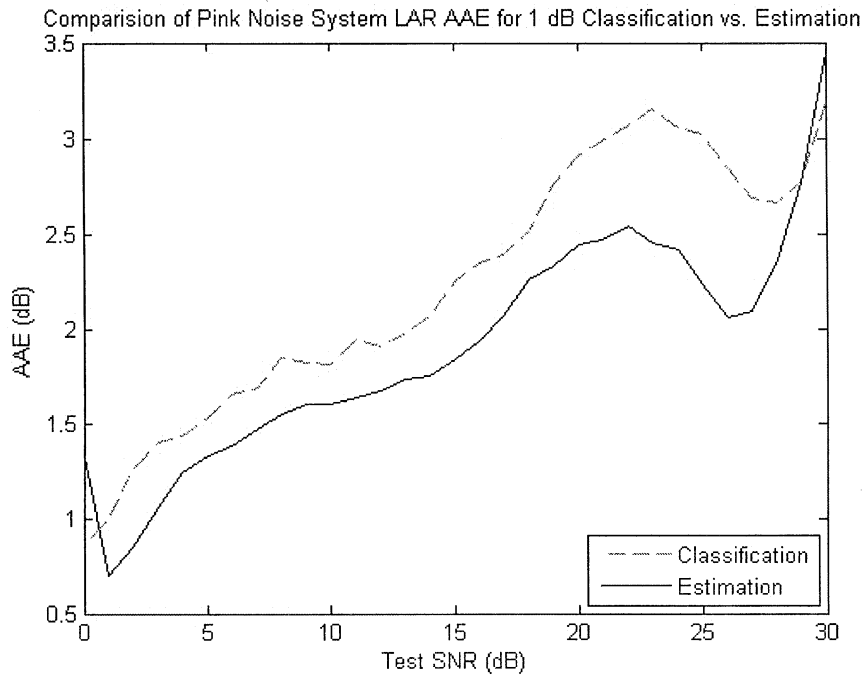


Figure 27. Comparison of Classification and Estimation AAE for LAR in the Pink System with 1 dB Spaced Codebooks

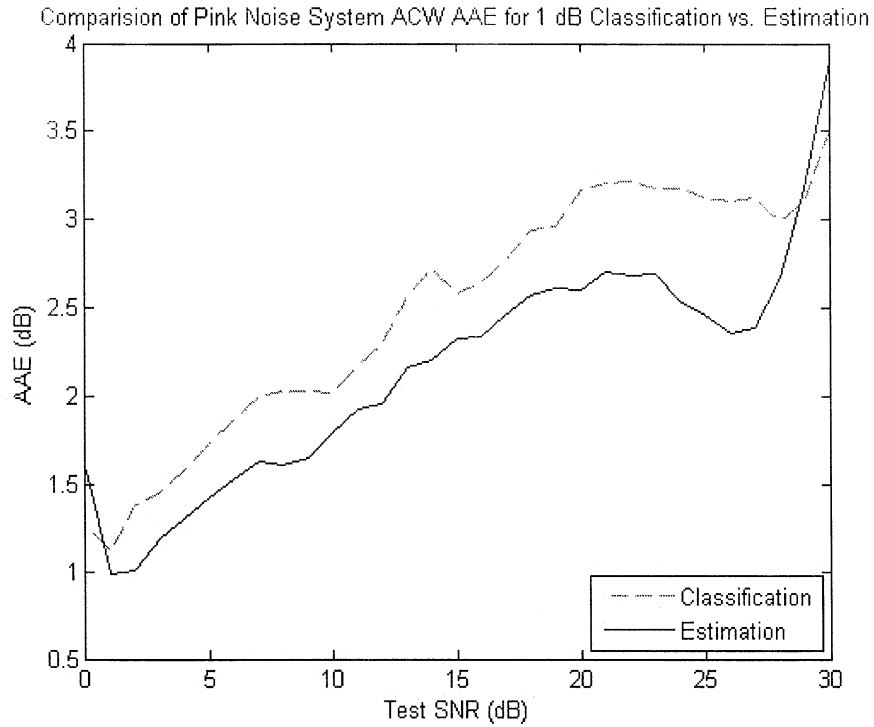


Figure 28. Comparison of Classification and Estimation AAE for ACW in the Pink System with 1 dB Spaced Codebooks

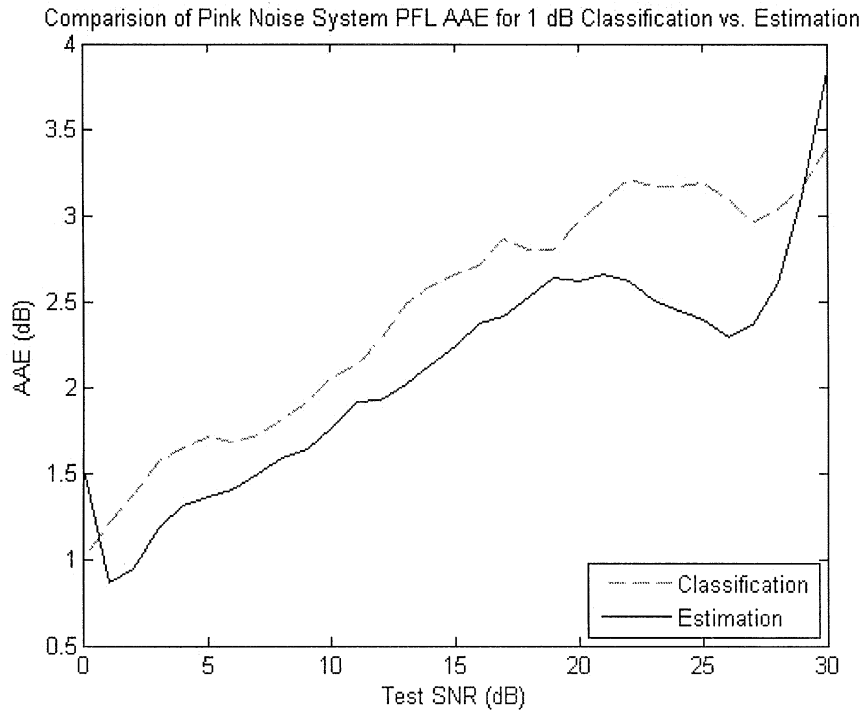


Figure 29. Comparison of Classification and Estimation AAE for PFL in the Pink System with 1 dB Spaced Codebooks

4.2.2 CPV Noise

The system designed for CPV noise performed with the same OAAE and AAE behavior as the system for AWGN and Pink Noise. The OAAE for both the hard decision classification method and the soft decision estimation method at the three tested resolutions can be seen in table 10. Again the soft decision estimation method does improve the system designed with 5 dB SNR resolution for all features, which is not observed for AWGN. The OAAE for each feature increases for the CPV system when compared to the AWGN and Pink noise systems indicating that CPV noise SNR is harder to classify than either other types of noise corruption. The best feature for the CPV noise system, as seen in table 10, is the LP Cepstrum with an OAAE for the soft decision estimation method at 1 dB resolution of 2.06 dB. Figures 30 through 35 show a comparison of the hard decision classification method and the soft decision estimation method for the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratio, ACW Cepstrum, and PFL Cepstrum features respectively. The same overall behavior is observed for each of these features when regarding CPV noise as with AWGN and Pink noise. Again, these results show that the system will work when estimating the SNR of a signal corrupted by CPV noise if CPV noise is used to train the system.

Table 10: OAAE Results in dB for System Designed with CPV Noise

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	2.50	2.07
LSF	3 dB	2.62	2.15
LSF	5 dB	2.77	2.58
CEP	1 dB	2.50	2.06
CEP	3 dB	2.63	2.17
CEP	5 dB	2.79	2.64
REFL	1 dB	2.54	2.12
REFL	3 dB	2.72	2.27
REFL	5 dB	2.86	2.71
LAR	1 dB	2.63	2.16
LAR	3 dB	2.76	2.33
LAR	5 dB	2.86	2.71
ACW	1 dB	3.06	2.40
ACW	3 dB	3.27	2.57
ACW	5 dB	3.37	2.92
PFL	1 dB	2.81	2.25
PFL	3 dB	3.01	2.40
PFL	5 dB	3.11	2.70

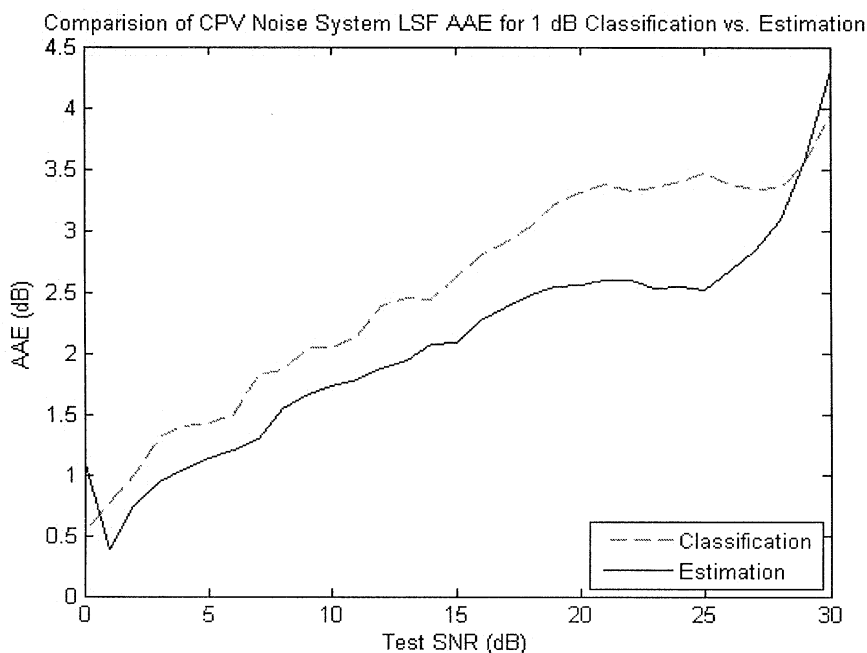


Figure 30. Comparison of Classification and Estimation AAE for LSF in the CPV System with 1 dB Spaced Codebooks

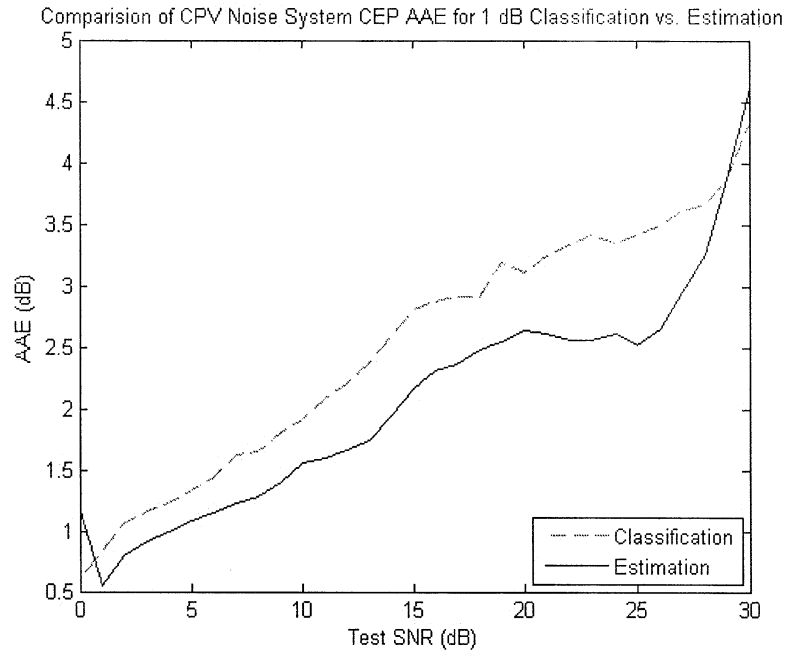


Figure 31. Comparison of Classification and Estimation AAE for CEP in the CPV System with 1 dB Spaced Codebooks

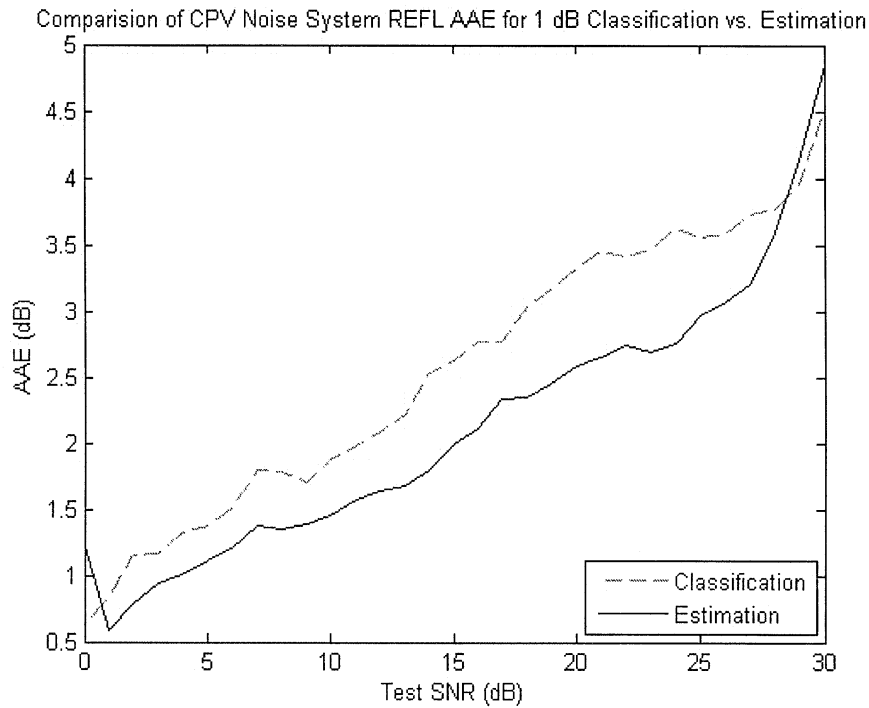


Figure 32. Comparison of Classification and Estimation AAE for REFL in the CPV System with 1 dB Spaced Codebooks

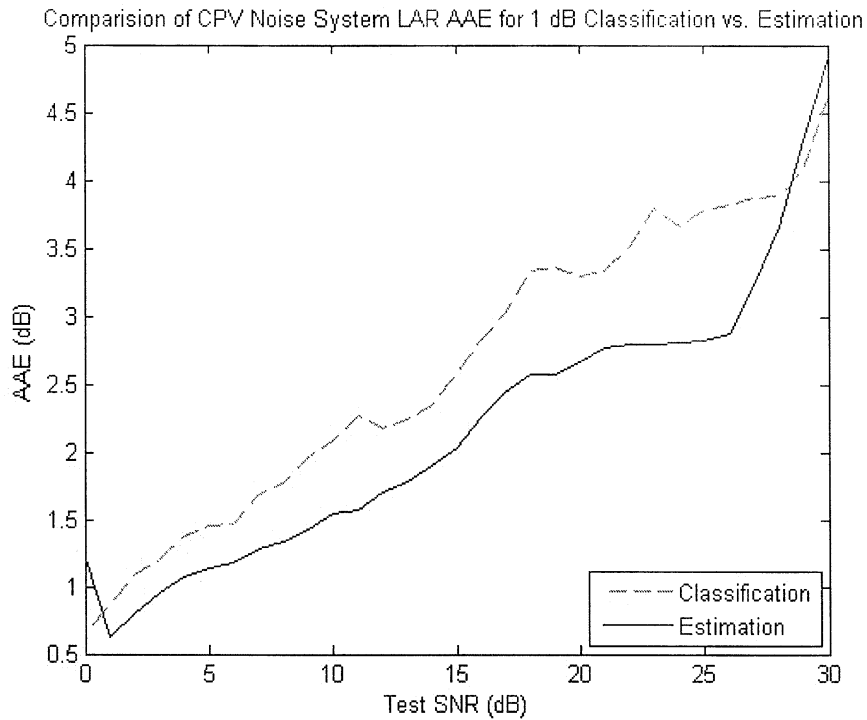


Figure 33. Comparison of Classification and Estimation AAE for LAR in the CPV System with 1 dB Spaced Codebooks

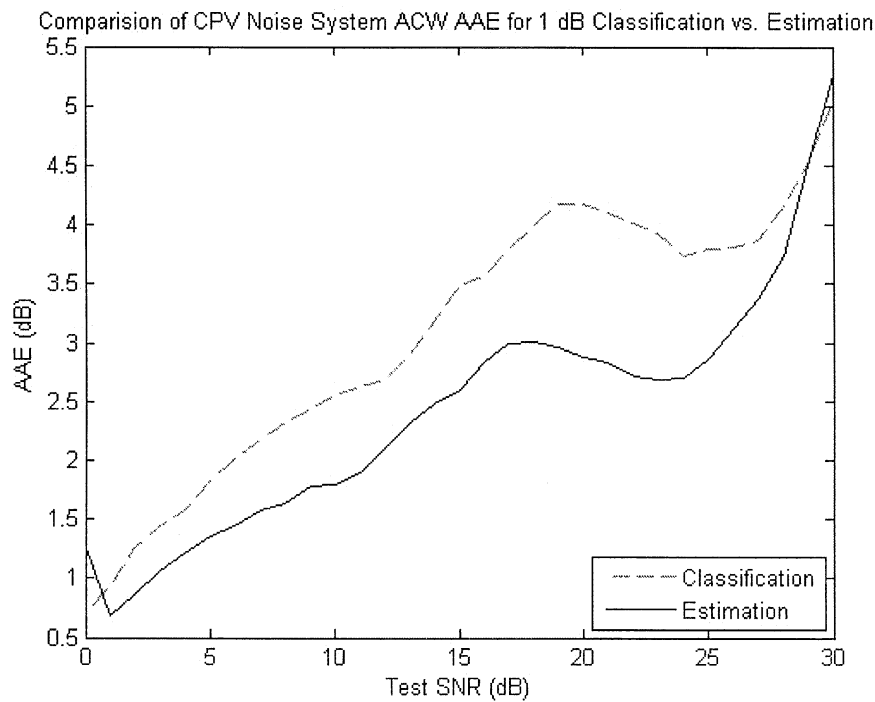


Figure 34. Comparison of Classification and Estimation AAE for ACW in the CPV System with 1 dB Spaced Codebooks

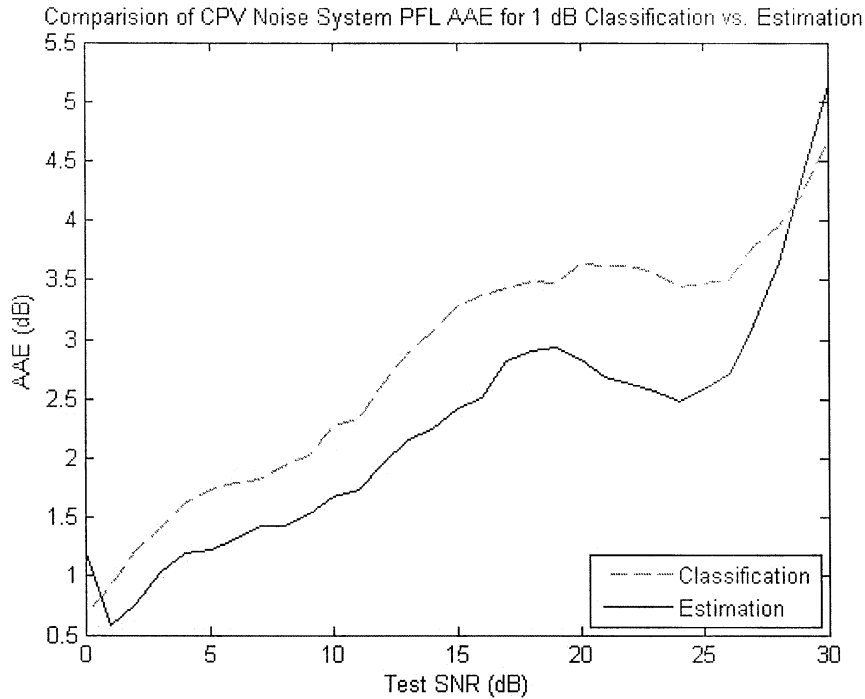


Figure 35. Comparison of Classification and Estimation AAE for PFL in the CPV System with 1 dB Spaced Codebooks

4.2.3 Cross Spectrum Testing

The first test of system robustness was to determine if the systems trained on just AWGN, Pink noise, and CPV noise would perform well when encountering a type of noise they were not trained on. Tables 11 through 16 show the OAAE for the hard decision classification and soft decision estimation with all three system resolutions for a system trained on AWGN and tested on Pink noise, trained on AWGN and tested on CPV noise, trained on Pink noise and tested on AWGN, trained on Pink noise and tested on CPV noise, trained on CPV noise and tested on AWGN, and trained on CPV noise and tested on Pink noise respectively. Figures 36 through 41 show the AAE soft decision estimation method results for each feature at 1 dB resolution on systems trained on AWGN and tested on Pink noise, trained on AWGN and tested on CPV noise, trained on

Pink noise and tested on AWGN, trained on Pink noise and tested on CPV noise, trained on CPV noise and tested on AWGN, and trained on CPV noise and tested on Pink noise respectively. The testing for these systems only included the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, and Log Area Ratios due to the poor performance of these features. The OAAE results show that systems trained on only one type of noise spectrum are not robust to other spectra. The AAE results all follow the same behavior with poor classification at lower SNR levels corresponding to signals that are more heavily corrupt with noise with better classification as the amount of noise in the test signals reduces. This behavior suggests that feature data for cleaner signals is more alike regardless of noise spectrum, which is intuitive. The results of this experiment show that the training process must incorporate multiple noise spectra to create a robust system.

Table 11: OAAE in dB for System Trained on AWGN and Tested on Pink Noise

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	7.41	7.22
LSF	3 dB	7.67	7.38
LSF	5 dB	7.94	7.64
CEP	1 dB	7.08	6.83
CEP	3 dB	7.30	6.96
CEP	5 dB	7.44	7.34
REFL	1 dB	6.83	6.56
REFL	3 dB	7.01	6.81
REFL	5 dB	7.62	7.33
LAR	1 dB	6.70	6.41
LAR	3 dB	6.74	6.54
LAR	5 dB	7.02	6.96

Table 12: OAAE in dB for System Trained on AWGN and Tested on CPV Noise

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	7.95	7.59
LSF	3 dB	8.18	7.16
LSF	5 dB	8.48	6.75
CEP	1 dB	7.93	7.50
CEP	3 dB	8.02	7.17
CEP	5 dB	8.40	6.80
REFL	1 dB	7.73	7.24
REFL	3 dB	8.13	7.07
REFL	5 dB	8.19	6.73
LAR	1 dB	7.91	7.39
LAR	3 dB	8.07	7.08
LAR	5 dB	8.31	6.77

Table 13: OAAE in dB for System Trained on Pink Noise and Tested on AWGN

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	12.01	11.71
LSF	3 dB	12.27	10.97
LSF	5 dB	9.40	10.78
CEP	1 dB	9.65	11.15
CEP	3 dB	11.87	11.32
CEP	5 dB	11.75	10.78
REFL	1 dB	12.72	11.86
REFL	3 dB	12.87	11.76
REFL	5 dB	14.44	10.52
LAR	1 dB	11.35	10.29
LAR	3 dB	10.30	10.87
LAR	5 dB	10.67	10.76

Table 14: OAAE in dB for System Trained on Pink Noise and Tested on CPV Noise

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	12.84	12.35
LSF	3 dB	12.38	11.46
LSF	5 dB	11.23	10.98
CEP	1 dB	11.42	12.55
CEP	3 dB	12.50	11.50
CEP	5 dB	13.85	10.77
REFL	1 dB	12.94	12.67
REFL	3 dB	12.94	12.31
REFL	5 dB	14.97	11.08
LAR	1 dB	11.39	11.58
LAR	3 dB	11.61	11.94
LAR	5 dB	13.05	11.02

Table 15: OAAE in dB for System Trained on CPV Noise and Tested on AWGN

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	7.87	7.74
LSF	3 dB	8.08	7.93
LSF	5 dB	8.51	8.46
CEP	1 dB	8.12	7.94
CEP	3 dB	8.50	8.29
CEP	5 dB	9.15	8.67
REFL	1 dB	8.07	7.78
REFL	3 dB	8.24	8.02
REFL	5 dB	8.56	8.24
LAR	1 dB	7.60	7.38
LAR	3 dB	7.65	7.56
LAR	5 dB	8.31	8.39

Table 16: OAAE in dB for System Trained on CPV Noise and Tested on Pink Noise

Feature	Codebook Space	Hard Decision	Soft Decision
LSF	1 dB	9.26	9.02
LSF	3 dB	9.22	8.91
LSF	5 dB	9.80	9.12
CEP	1 dB	8.83	8.60
CEP	3 dB	9.15	8.75
CEP	5 dB	9.19	8.80
REFL	1 dB	8.78	8.55
REFL	3 dB	8.91	8.49
REFL	5 dB	9.52	8.88
LAR	1 dB	8.68	8.32
LAR	3 dB	8.35	8.15
LAR	5 dB	9.16	8.61

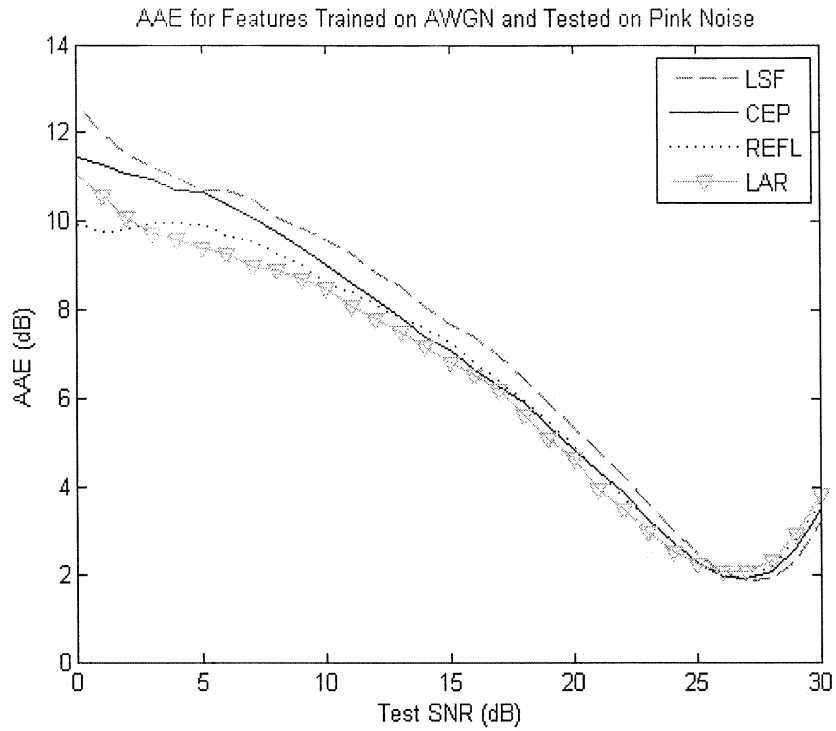


Figure 36. AAE for All Features when Trained on AWGN and Tested on Pink Noise with 1 dB Resolution

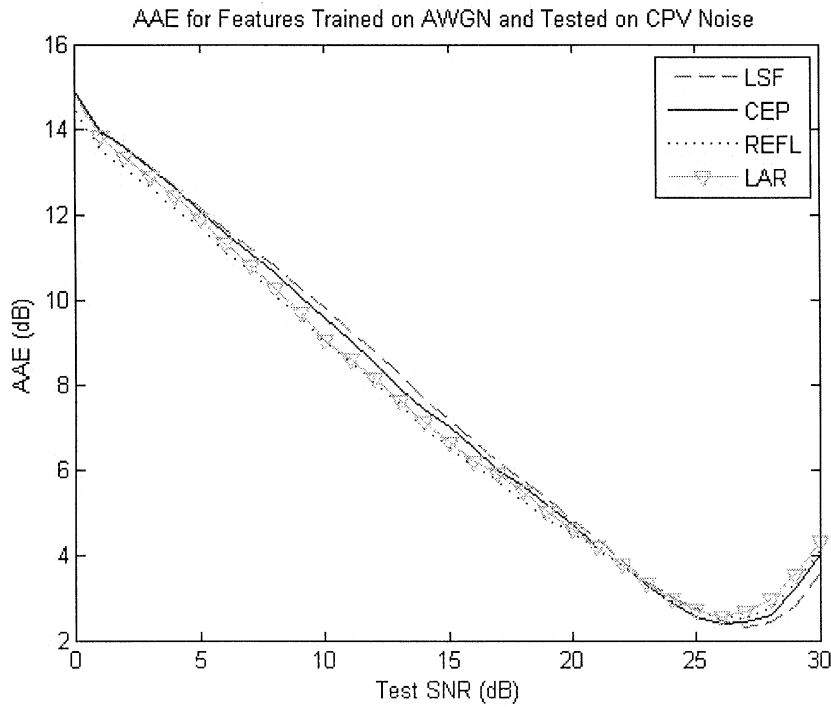


Figure 37. AAE for All Features when Trained on AWGN and Tested on CPV Noise with 1 dB Resolution

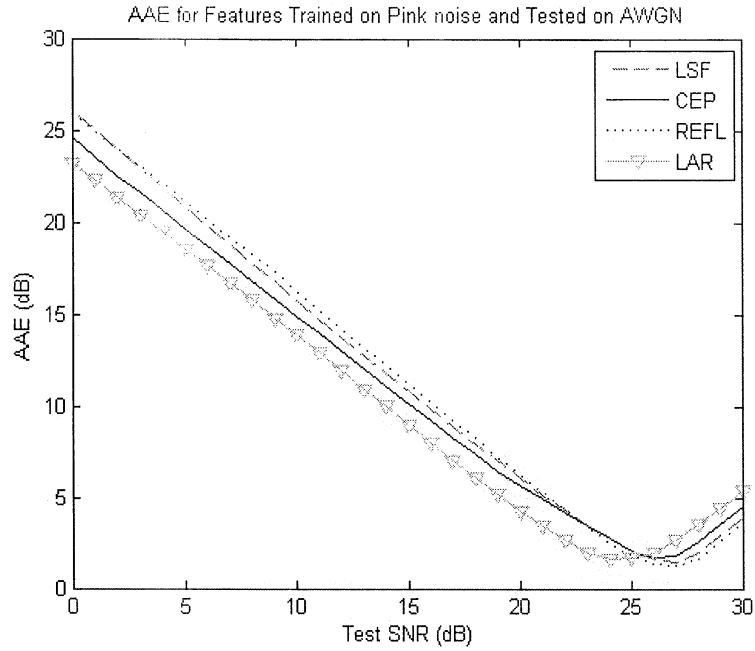


Figure 38. AAE for All Features when Trained on Pink Noise and Tested on AWGN with 1 dB Resolution

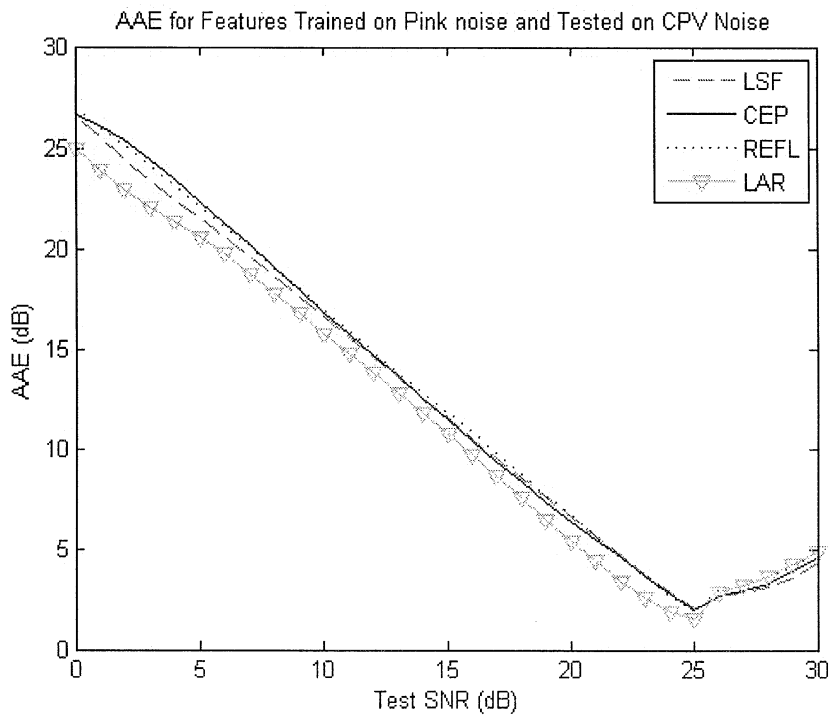


Figure 39. AAE for All Features when Trained on Pink Noise and Tested on CPV Noise with 1 dB Resolution

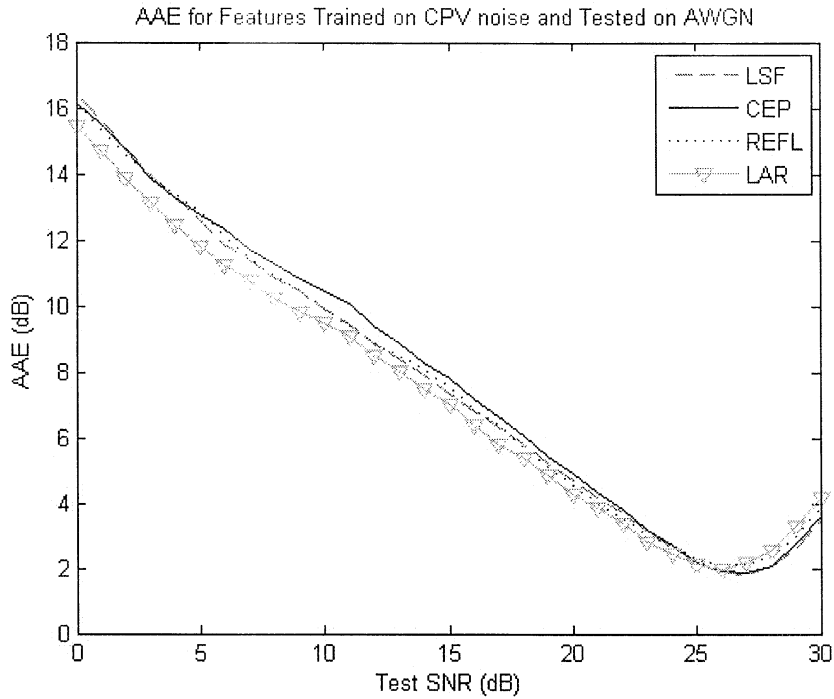


Figure 40. AAE for All Features when Trained on CPV Noise and Tested on AWGN with 1 dB Resolution

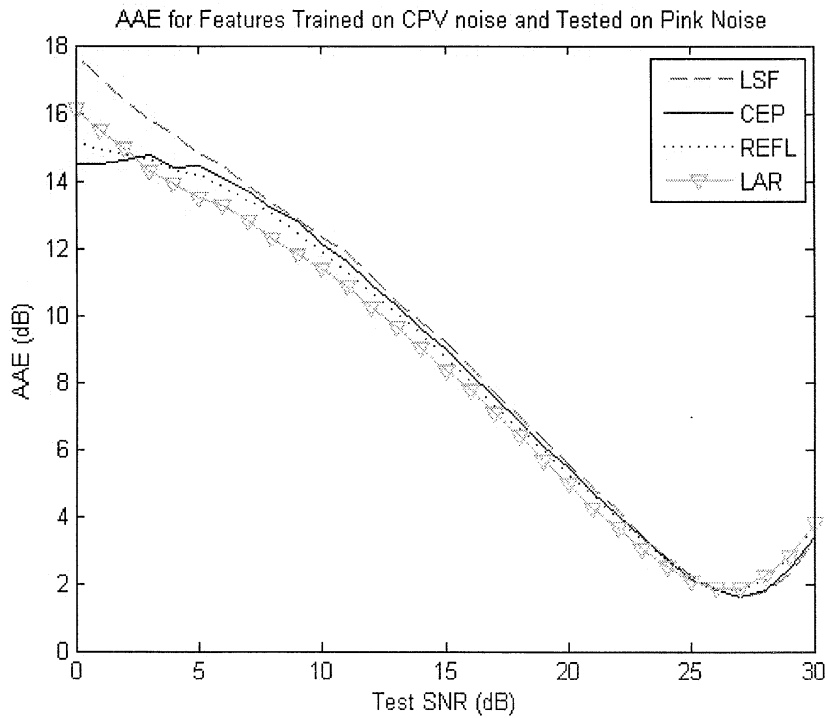


Figure 41. AAE for All Features when Trained on CPV Noise and Tested on Pink Noise with 1 dB Resolution

4.2.4 Multiple Spectrum System

Based on the results of the cross spectrum testing it was observed that the VQ pattern recognition system does not perform well with noise types that are not used to corrupt signals during training. A robust system was designed by training with multiple noise types. During training each sentence is corrupted by AWGN, Pink Noise, and CPV Noise individually for feature extraction, effectively tripling the amount of training data. The robust system was designed using size 256 codebooks and 1 dB codebook resolution for all features. It was expected that this system will perform well for all trained noise types, however a robust system must perform well for untrained noise spectra as well. Speech signals corrupted with CMV noise were used to determine the robust system's performance on untrained noise.

4.2.4.1 Robust Spectrum Results

The robust system was tested by corrupting speech signals with AWGN, Pink Noise, CPV Noise, and CMV noise. Tables 17 through 20 show the OAAE results for AWGN, Pink Noise, CPV Noise, and CMV Noise for the robust system with confidence intervals calculated for 95% confidence. The confidence intervals are small due to the low standard deviation in the total error and the large test population size of 380 speech signals for each of the 31 SNR levels tested creating 11780 total speech signals tested. The OAAE results are shown for both the hard decision classification method and soft decision estimation method. The OAAE results show that the estimation method again performs better than the classification method. The AWGN OAAE results, Pink Noise OAAE results, and CPV OAAE results show that the robust system has decreased

performance and greater error than the systems trained on individual noise spectrum types and tested on just that noise spectrum. For AWGN the best feature is now the Line Spectral Frequencies with an OAAE of 3.08 dB for the soft decision estimation method, for Pink Noise the best feature is the LP Cepstrum with and OAAE of 2.97 dB for the soft decision estimation method, and for CPV noise the best feature is the LP Cepstrum with an OAAE of 3.16 dB for the soft decision estimation method.

Table 17: AWGN OAAE in dB for Robust System

	Classifier	CI	Estimator	CI
LSF	3.64	0.06	3.08	0.05
CEP	3.67	0.06	3.20	0.05
REFL	3.68	0.06	3.23	0.05
LAR	3.84	0.06	3.37	0.05
ACW	4.33	0.07	3.63	0.05
PFL	3.91	0.06	3.25	0.05

Table 18: Pink Noise OAAE in dB for Robust System

	Classifier	CI	Estimator	CI
LSF	3.57	0.06	3.00	0.05
CEP	3.47	0.05	2.97	0.05
REFL	3.47	0.06	2.98	0.05
LAR	3.48	0.06	2.98	0.05
ACW	4.47	0.07	3.80	0.05
PFL	4.31	0.06	3.58	0.05

Table 19: CPV Noise OAAE in dB for Robust System

	Classifier	CI	Estimator	CI
LSF	3.67	0.06	3.23	0.05
CEP	3.57	0.06	3.16	0.05
REFL	3.59	0.06	3.22	0.05
LAR	3.70	0.06	3.29	0.05
ACW	4.37	0.07	3.70	0.06
PFL	3.99	0.06	3.47	0.05

The OAAE results based on the untrained noise spectrum type CMV Noise show a large contrast to feature results of the trained noise spectra, which will need to be taken into consideration when deciding specifications for the final system parameters to ensure system robustness. The typically better performing features, the Line Spectral

Frequencies and LP Cepstrum show that they are the least robust to untrained noise spectra. The ACW Cepstrum, and PFL Cepstrum, which had always had the highest system errors are the most robust to untrained noise spectra. The OAAE of the Reflection Coefficients and Log Area Ratios is greater than when trained noise spectra are tested, however these features are still robust to untrained noise spectra.

Table 20: CMV Noise OAAE in dB for Robust System

	Classifier	CI	Estimator	CI
LSF	6.07	0.10	5.54	0.09
CEP	7.30	0.13	6.49	0.10
REFL	4.16	0.06	3.66	0.05
LAR	4.19	0.06	3.77	0.05
ACW	4.32	0.07	3.51	0.05
PFL	4.09	0.07	3.28	0.05

The soft decision estimation method AAE results for the robust system are shown in figures 42 through 47 for the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum respectively. The AAE for each feature on the trained noise spectrum types behave the same as they did for the systems trained on just those noise spectra. The AAE increases as the SNR increases and the amount of noise in the system decreases with peaks at the extrema of the test range. The main difference in AAE behavior from previous results exists in the AAE for each feature's performance with CMV Noise. The Line Spectral Frequencies and LP Cepstrum behave in the same way as the systems trained on one single noise spectrum type and tested on another. These features perform the worst at lower SNR levels corresponding to an inability to estimate the SNR of a signal which is more heavily corrupted with noise. The AAE results of these features also show that the influence of the noise on the feature becomes more apparent within the 15 to 20 dB range. The

Reflection Coefficients and Log Area Ratios show a decreased feature performance as the amount of noise in the signal increases for CMV Noise. The Reflection Coefficients and Log Area Ratios still perform better at the lower SNR levels with more highly corrupt signals than in the range of higher SNR levels with cleaner signals. This means the confusion caused when comparing clean signals is greater than the effect of classifying a noise type that is not trained. The ACW Cepstrum, and PFL Cepstrum AAE behaves the same for CMV noise as it does for the trained noise spectrum types. This shows the the ACW cepstrum and PFL cepstrum, which were designed to create pattern recognition systems robust to noise, are robust to noise spectrum type.

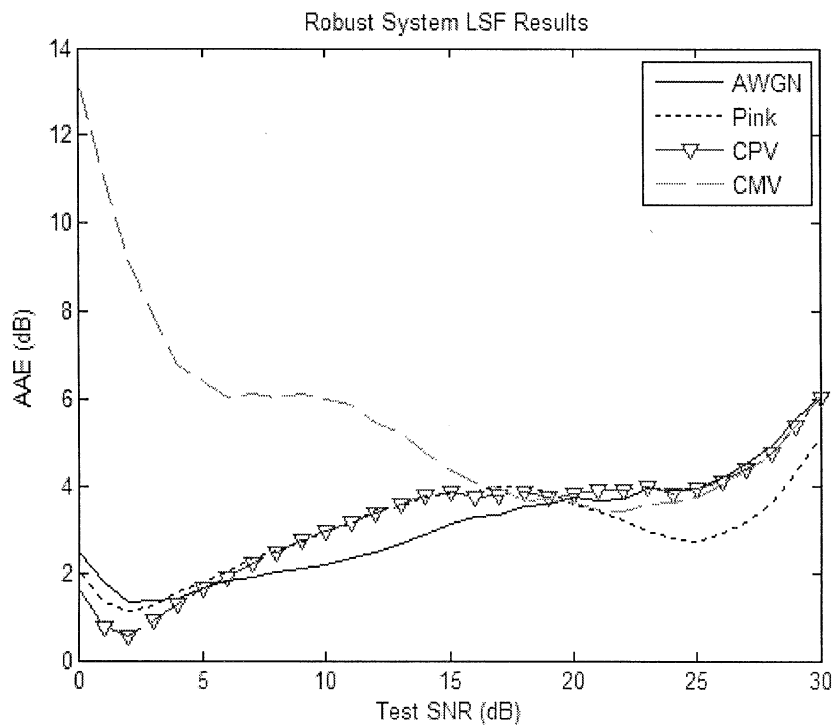


Figure 42. Comparison of Robust System LSF Feature AAE for All Tested Noise Types

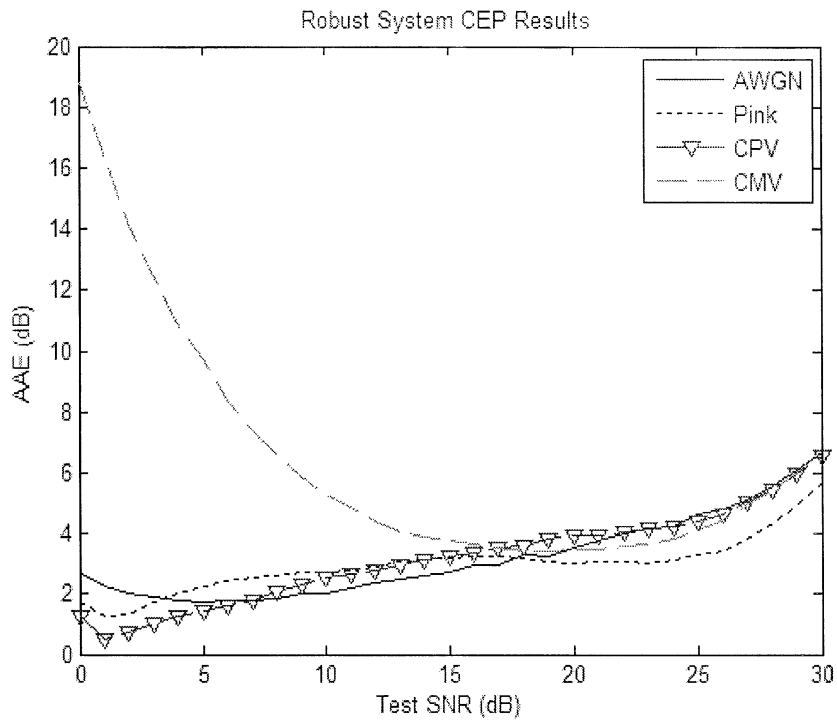


Figure 43. Comparison of Robust System CEP Feature AAE for All Tested Noise Types

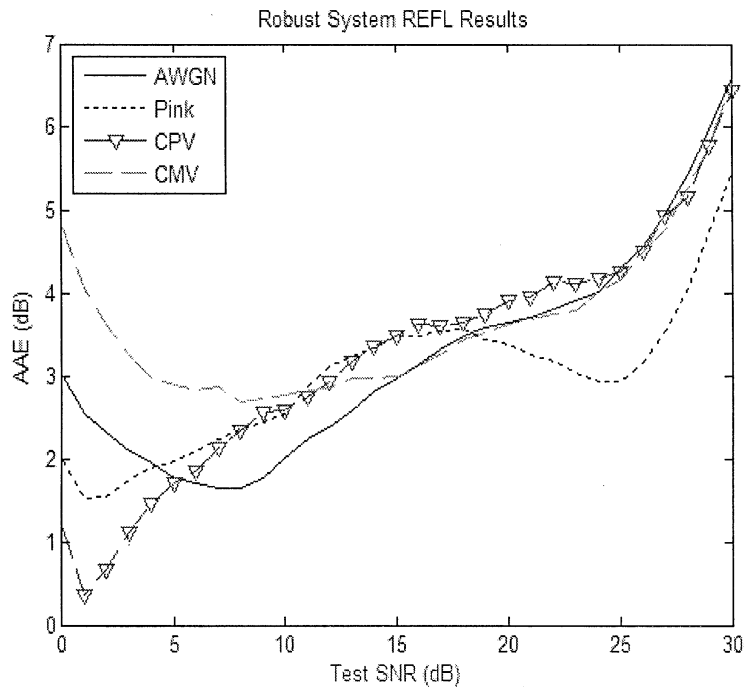


Figure 44. Comparison of Robust System REFL Feature AAE for All Tested Noise Types

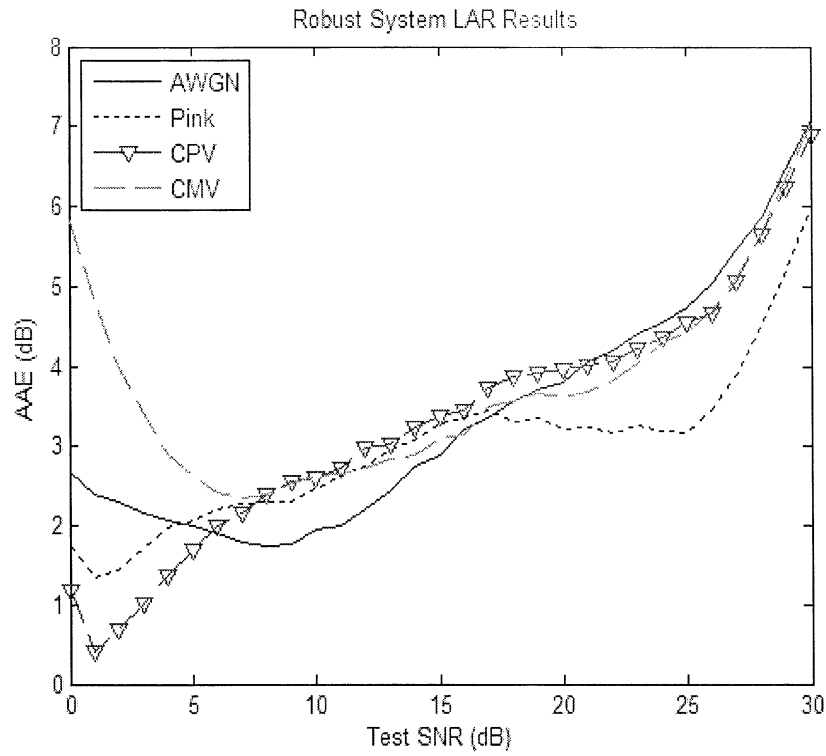


Figure 45. Comparison of Robust System LAR Feature AAE for All Tested Noise Types

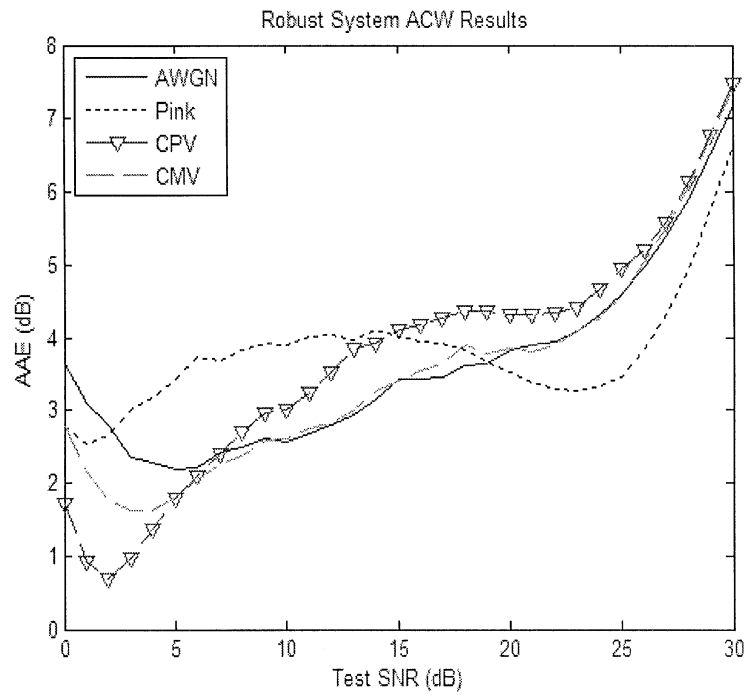


Figure 46. Comparison of Robust System ACW Feature AAE for All Tested Noise Types

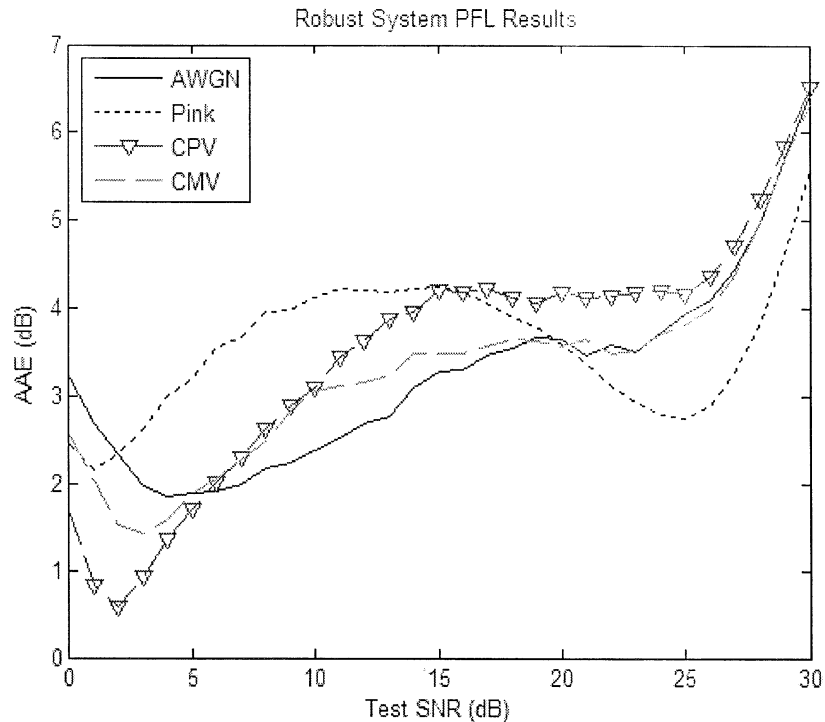


Figure 47. Comparison of Robust System PFL Feature AAE for All Tested Noise Types

4.2.4.2 Beyond Range Codebook System

The last change for systems designed with single features was the addition of codebooks beyond the test range of 0 to 30 dB SNR. A codebook was added at -1 dB and 31 dB to improve the soft decision estimation method in the robust systems. It was expected that adding the extra codebooks would remove the peaks at the extrema of the test range. These peaks are caused by the inability to estimate a signal at the extrema of the codebook range. The soft decision estimation method averages three weighted codebook values to obtain the SNR estimation. If the weights are approximately equal, the estimation is close to the average of the three codebook SNR values. This averaging technique will therefore never estimate a signal at the extrema of the test range without

codebooks placed outside the test range. Tables 21 through 24 show the OAAE results for the 33 codebook system for AWGN, Pink Noise, CPV Noise, and CMV Noise respectively. The system was designed to improve the soft decision estimation method, by smoothing the AAE error curve and removing the peaks at the extrema so improvement is expected in the Estimator systems and not the Classifier systems. There is little improvement in the Classifier systems, but the addition of the extra codebooks does reduce error more for the Estimator system. For AWGN the best feature, the Reflection Coefficients, the error is reduced from 3.08 to 3.03. The LP Cepstrum is still the best feature for Pink Noise and CPV Noise, but only shows improvement from 3.16 dB to 3.13 dB for CPV Noise. The PFL Cepstrum only improves by a factor of .01 dB in OAAE for SNR estimation, from 3.28 dB to 3.27 dB. The best performing features for each noise spectrum type do not show great improvement in OAAE, however improvement is shown in all features.

The soft decision estimation method AAE results for the 33 codebook robust system are shown in figures 48 through 53, for the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum respectively. The main difference in behavior is observed at the lowest SNR levels in the test range. The peak caused by the estimation method at 0 dB SNR is reduced or removed in many cases. In other cases the error curve is smoothed. In the special case of the Line Spectral Frequencies and LP Cepstrum for CMV noise there is no discernible difference in AAE behavior and no useful improvement OAAE by adding the extra codebooks. There is no observable difference in behavior for any feature at the higher SNR levels of

the test range from adding an extra codebook beyond the range.

Table 21: AWGN OAAE in dB for Robust 33 Codebook System

	Classifier	CI	Estimator	CI
LSF	3.66	0.06	3.03	0.05
CEP	3.67	0.06	3.16	0.05
REFL	3.68	0.06	3.15	0.05
LAR	3.84	0.06	3.31	0.05
ACW	4.33	0.07	3.55	0.05
PFL	3.92	0.06	3.17	0.05

Table 22: Pink Noise OAAE in dB for Robust 33 Codebook System

	Classifier	CI	Estimator	CI
LSF	3.66	0.06	2.99	0.05
CEP	3.52	0.05	2.97	0.05
REFL	3.53	0.06	2.97	0.05
LAR	3.53	0.06	2.97	0.05
ACW	4.51	0.07	3.75	0.05
PFL	4.37	0.07	3.57	0.05

Table 23:CPV Noise OAAE in dB for Robust 33 Codebook System

	Classifier	CI	Estimator	CI
LSF	3.74	0.06	3.21	0.05
CEP	3.60	0.06	3.13	0.05
REFL	3.65	0.06	3.21	0.05
LAR	3.74	0.06	3.27	0.05
ACW	4.42	0.07	3.65	0.06
PFL	4.02	0.06	3.42	0.05

Table 24: CMV Noise OAAE in dB for Robust 33 Codebook System

	Classifier	CI	Estimator	CI
LSF	6.12	0.10	5.59	0.09
CEP	7.38	0.13	6.55	0.11
REFL	4.19	0.06	3.65	0.05
LAR	4.21	0.06	3.76	0.05
ACW	4.33	0.07	3.45	0.05
PFL	4.12	0.07	3.27	0.05

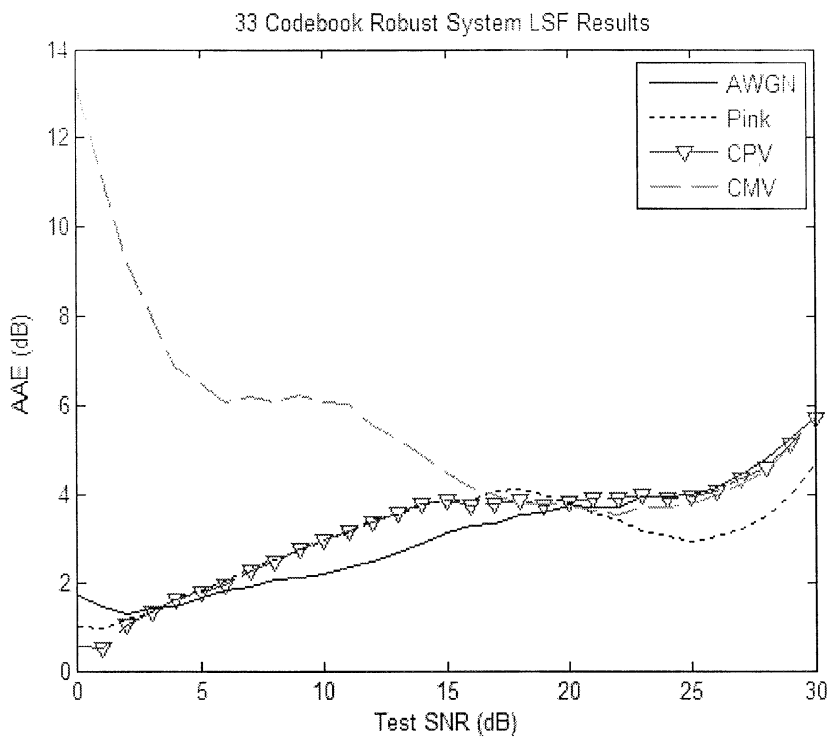


Figure 48. Comparison of 33 Codebook Robust System LSF Feature AAE for All Tested Noise Types

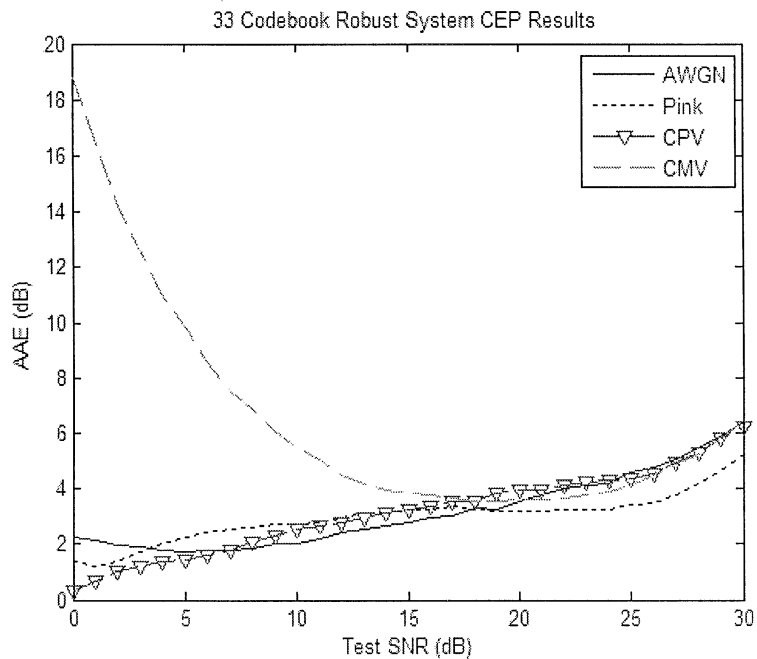


Figure 49. Comparison of 33 Codebook Robust System CEP Feature AAE for All Tested Noise Types

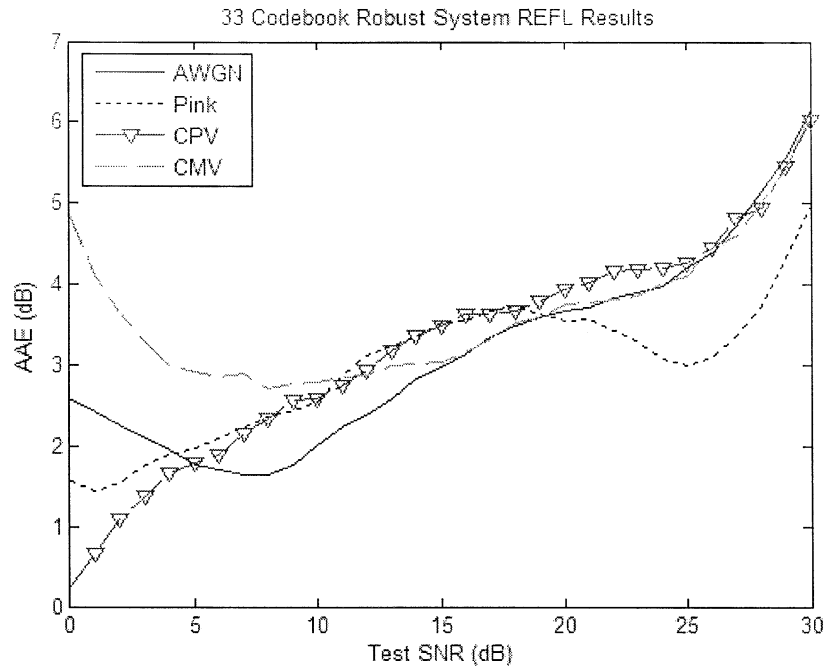


Figure 50. Comparison of 33 Codebook Robust System REFL Feature AAE for All Tested Noise Types

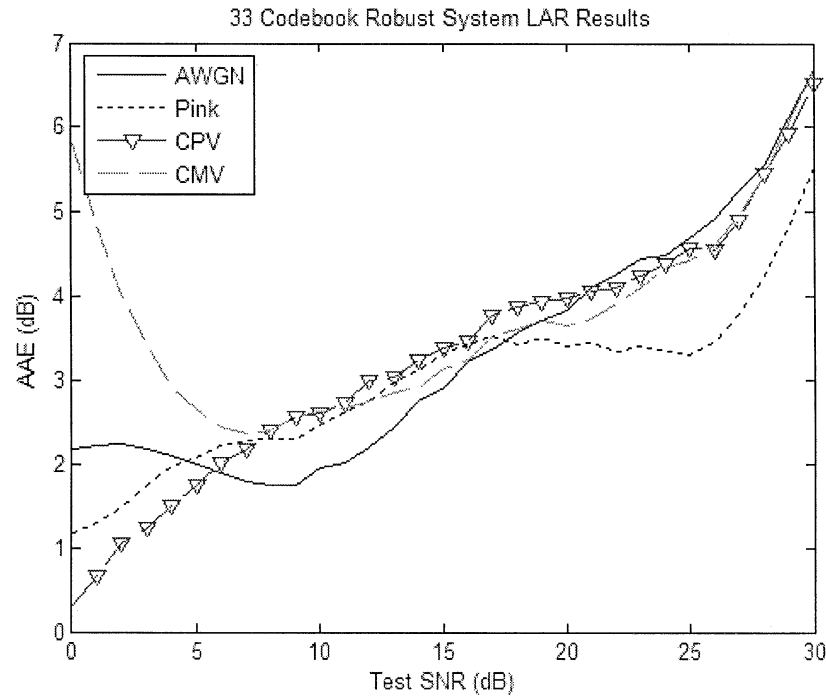


Figure 51. Comparison of 33 Codebook Robust System LAR Feature AAE for All Tested Noise Types

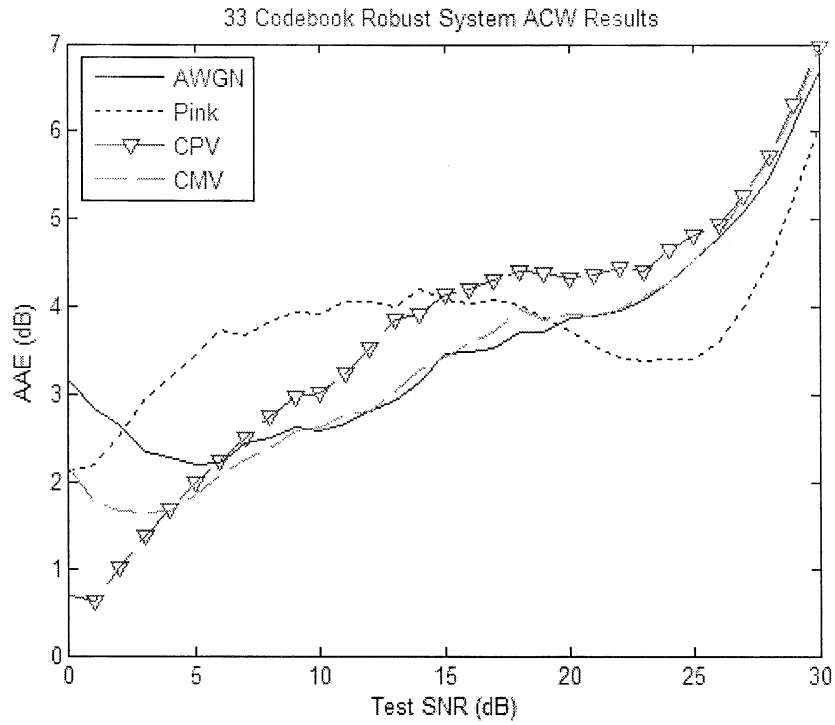


Figure 52. Comparison of 33 Codebook Robust System ACW Feature AAE for All Tested Noise Types

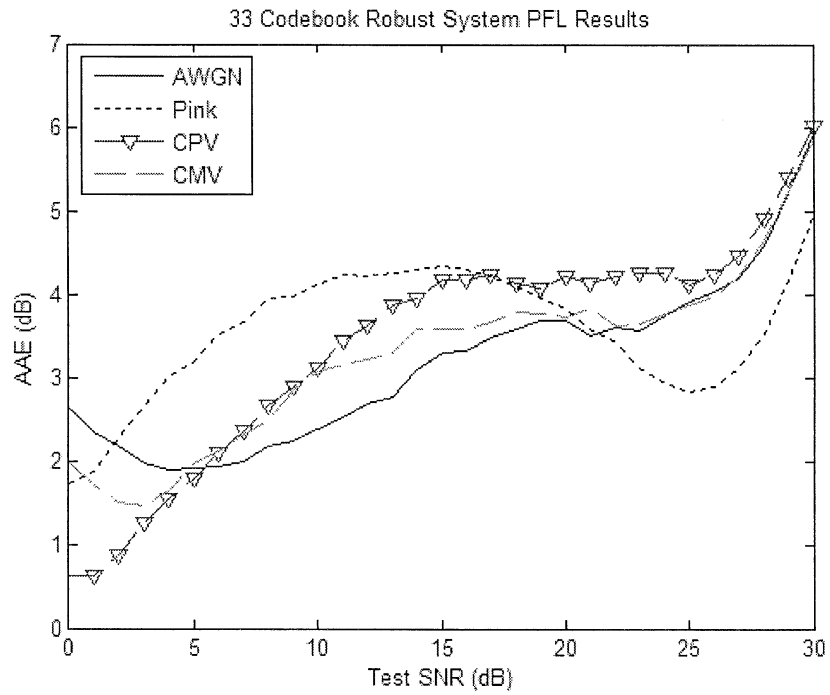


Figure 53. Comparison of 33 Codebook Robust System PFL Feature AAE for All Tested Noise Types

4.3 Feature Fusion

The last method investigated to improve the VQ based pattern recognition system for estimating the SNR of a speech signal was feature fusion. Currently the VQ system offers a separate estimate for the SNR of a sentence from each of the six studied features, determined by the sum of the Euclidean distances of each feature vector of the sentence from each VQ codebook. Decision and distance level fusion were both tested. Decision level fusion involves determining the best way to combine the final estimate of each feature for a sentence to reduce the error of a final estimate. A genetic algorithm was used as an optimization technique to further improve decision level fusion by applying weights to each feature's SNR estimate. Distance level fusion involves combining the weights of each feature from each codebook to make a single final decision based on the distance of all involved features from each codebook.

4.3.1 Unweighted Fusion

The Combination Rules studied for unweighted decision level fusion include mean fusion, median fusion, and trimmed mean fusion, and the combination rules for distance level fusion include minimum distance fusion, mean distance fusion, median distance fusion, and trimmed mean distance fusion. There were six features totaling 57 possible combinations for feature fusion. Tables 25 through 31 show the OAAE obtained for AWGN, Pink Noise, CPV Noise, an average of all three trained noise types, and CMV Noise for each feature combination using mean decision combination, median decision combination, trimmed mean decision combination, minimum distance combination, mean distance combination, median distance combination, and trimmed mean distance

combination respectively. These results were obtained by combining soft decision estimations or distances in the robust 33 codebook system for each test sentence. Based on these results mean decision combination provides the best method of feature combination for reducing error in the SNR estimation, giving the greatest reduction in error for each of the tested noise spectrum types. Decision level fusion is observed to work better than distance level fusion. Median decision combination and trimmed mean decision combination also reduce the error, however not as well as mean decision combination. The lowest OAAE obtained for AWGN individually is 2.7 dB provided by mean decision combination of the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, and PFL Cepstrum. The lowest OAAE obtained for Pink Noise individually is 2.61 dB provided by median decision combination of the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, and Log Area Ratios. The lowest OAAE obtained for CPV Noise individually is 2.92 dB provided by mean decision combination of the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, and PFL Cepstrum. The lowest OAAE obtained for an average of all three trained noise spectra is 2.76 dB provided by mean decision combination of the Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, and PFL Cepstrum. The lowest OAAE obtained for CMV individually is 2.93 dB provided by mean decision combination of the Reflection Coefficients, ACW Cepstrum, and PFL Cepstrum. As expected, while the LSF and LP Cepstrum generally aid in improving the results for the three trained noise types, their existence in a feature combination typically results in a high OAAE for CMV noise. Table 32 shows the three overall best feature combinations. Figure 54 and figure 55 show

the AAE for the two best unweighted fusion combinations. These combinations were obtained by the improvement each feature combination had as compared to the best performing feature for each noise spectrum type. The best combination is the mean decision level fusion of the Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum with an average improvement of .19 dB over all noise spectrum types. Intuitively this combination works best because it combines all features that perform well for CMV Noise.

Table 25: Decision Level Mean Fusion Combination OAAE Results in dB

Features	Mean				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	2.87	2.74	3.00	2.87	5.83
LSF, REFL	2.84	2.71	3.03	2.86	4.32
LSF, LAR	2.92	2.71	3.07	2.90	4.35
LSF, ACW	2.87	2.97	3.14	2.99	3.96
LSF, PFL	2.77	2.93	3.07	2.92	3.97
CEP, REFL	2.92	2.70	2.98	2.87	4.82
CEP, LAR	3.02	2.73	3.04	2.93	4.85
CEP, ACW	2.98	2.99	3.09	3.02	4.45
CEP, PFL	2.84	2.93	3.03	2.93	4.41
REFL, LAR	3.03	2.75	3.10	2.96	3.49
REFL, ACW	2.99	3.00	3.17	3.05	3.14
REFL, PFL	2.86	2.95	3.09	2.96	3.07
LAR, ACW	3.08	2.99	3.18	3.08	3.20
LAR, PFL	2.92	2.93	3.12	2.99	3.14
ACW, PFL	3.05	3.36	3.27	3.23	3.01
LSF, CEP, REFL	2.79	2.62	2.94	2.79	4.90
LSF, CEP, LAR	2.85	2.64	2.98	2.82	4.92
LSF, CEP, ACW	2.78	2.78	2.99	2.85	4.61
LSF, CEP, PFL	2.72	2.76	2.95	2.81	4.61
LSF, REFL, LAR	2.85	2.64	3.01	2.83	3.95
LSF, REFL, ACW	2.77	2.77	3.02	2.85	3.66
LSF, REFL, PFL	2.72	2.75	2.99	2.82	3.65
LSF, LAR, ACW	2.83	2.76	3.04	2.88	3.68
LSF, LAR, PFL	2.76	2.74	3.02	2.84	3.67
LSF, ACW, PFL	2.77	2.96	3.07	2.93	3.46
CEP, REFL, LAR	2.92	2.64	2.98	2.85	4.29
CEP, REFL, ACW	2.84	2.77	2.99	2.87	3.99
CEP, REFL, PFL	2.77	2.76	2.96	2.83	3.96
CEP, LAR, ACW	2.91	2.78	3.01	2.90	4.02
CEP, LAR, PFL	2.83	2.75	2.99	2.85	3.99
CEP, ACW, PFL	2.84	2.98	3.03	2.95	3.75
REFL, LAR, ACW	2.93	2.80	3.07	2.93	3.16
REFL, LAR, PFL	2.84	2.78	3.04	2.88	3.12
REFL, ACW, PFL	2.85	2.98	3.08	2.97	2.93
LAR, ACW, PFL	2.90	2.97	3.10	2.99	2.98
LSF, CEP, REFL, LAR	2.81	2.59	2.95	2.78	4.46
LSF, CEP, REFL, ACW	2.74	2.68	2.94	2.78	4.22
LSF, CEP, REFL, PFL	2.70	2.67	2.92	2.76	4.22
LSF, CEP, LAR, ACW	2.79	2.67	2.96	2.81	4.24
LSF, CEP, LAR, PFL	2.74	2.66	2.95	2.78	4.24
LSF, CEP, ACW, PFL	2.71	2.81	2.96	2.83	4.03
LSF, REFL, LAR, ACW	2.79	2.68	3.00	2.82	3.54
LSF, REFL, LAR, PFL	2.74	2.67	2.98	2.80	3.53
LSF, REFL, ACW, PFL	2.71	2.80	3.00	2.84	3.35
LSF, LAR, ACW, PFL	2.75	2.79	3.02	2.85	3.36
CEP, REFL, LAR, ACW	2.85	2.69	2.97	2.84	3.79
CEP, REFL, LAR, PFL	2.79	2.68	2.96	2.81	3.77
CEP, REFL, ACW, PFL	2.76	2.82	2.97	2.85	3.57
CEP, LAR, ACW, PFL	2.81	2.81	2.99	2.87	3.60
REFL, LAR, ACW, PFL	2.82	2.82	3.03	2.89	2.98
LSF, CEP, REFL, LAR, ACW	2.76	2.63	2.94	2.78	4.01
LSF, CEP, REFL, LAR, PFL	2.73	2.62	2.93	2.76	4.01
LSF, CEP, REFL, ACW, PFL	2.69	2.72	2.93	2.78	3.84
LSF, CEP, LAR, ACW, PFL	2.72	2.71	2.95	2.80	3.85
LSF, REFL, LAR, ACW, PFL	2.73	2.72	2.98	2.81	3.31
CEP, REFL, LAR, ACW, PFL	2.77	2.73	2.96	2.82	3.50
LSF, CEP, REFL, LAR, ACW, PFL	2.71	2.66	2.93	2.77	3.72

Table 26: Decision Level Median Fusion Combination OAAE Results in dB

Features	Median				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	2.87	2.74	3.00	2.87	5.83
LSF, REFL	2.84	2.71	3.03	2.86	4.32
LSF, LAR	2.92	2.71	3.07	2.90	4.35
LSF, ACW	2.87	2.97	3.14	2.99	3.96
LSF, PFL	2.77	2.93	3.07	2.92	3.97
CEP, REFL	2.92	2.70	2.98	2.87	4.82
CEP, LAR	3.02	2.73	3.04	2.93	4.85
CEP, ACW	2.98	2.99	3.09	3.02	4.45
CEP, PFL	2.84	2.93	3.03	2.93	4.41
REFL, LAR	3.03	2.75	3.10	2.96	3.49
REFL, ACW	2.99	3.00	3.17	3.05	3.14
REFL, PFL	2.86	2.95	3.09	2.96	3.07
LAR, ACW	3.08	2.99	3.18	3.08	3.20
LAR, PFL	2.92	2.93	3.12	2.99	3.14
ACW, PFL	3.05	3.36	3.27	3.23	3.01
LSF, CEP, REFL	2.85	2.67	2.98	2.83	4.93
LSF, CEP, LAR	2.90	2.70	3.03	2.88	4.96
LSF, CEP, ACW	2.86	2.80	3.04	2.90	4.88
LSF, CEP, PFL	2.81	2.80	3.00	2.87	4.88
LSF, REFL, LAR	2.92	2.69	3.06	2.89	3.68
LSF, REFL, ACW	2.86	2.78	3.07	2.90	3.46
LSF, REFL, PFL	2.79	2.78	3.03	2.87	3.44
LSF, LAR, ACW	2.91	2.78	3.11	2.93	3.49
LSF, LAR, PFL	2.83	2.75	3.06	2.88	3.46
LSF, ACW, PFL	2.88	3.05	3.14	3.02	3.10
CEP, REFL, LAR	2.98	2.73	3.06	2.92	3.77
CEP, REFL, ACW	2.91	2.82	3.05	2.93	3.50
CEP, REFL, PFL	2.85	2.80	3.02	2.89	3.48
CEP, LAR, ACW	2.97	2.79	3.09	2.95	3.56
CEP, LAR, PFL	2.90	2.78	3.04	2.91	3.54
CEP, ACW, PFL	2.93	3.06	3.12	3.04	3.10
REFL, LAR, ACW	3.00	2.81	3.14	2.98	3.26
REFL, LAR, PFL	2.93	2.79	3.10	2.94	3.23
REFL, ACW, PFL	2.95	3.08	3.16	3.06	3.00
LAR, ACW, PFL	2.99	3.06	3.19	3.08	3.02
LSF, CEP, REFL, LAR	2.83	2.61	2.98	2.81	4.23
LSF, CEP, REFL, ACW	2.77	2.67	2.97	2.80	4.07
LSF, CEP, REFL, PFL	2.74	2.67	2.95	2.79	4.06
LSF, CEP, LAR, ACW	2.82	2.67	3.00	2.83	4.09
LSF, CEP, LAR, PFL	2.78	2.66	2.97	2.80	4.08
LSF, CEP, ACW, PFL	2.75	2.81	2.99	2.85	3.80
LSF, REFL, LAR, ACW	2.83	2.67	3.04	2.85	3.36
LSF, REFL, LAR, PFL	2.78	2.66	3.01	2.82	3.34
LSF, REFL, ACW, PFL	2.76	2.81	3.03	2.87	3.10
LSF, LAR, ACW, PFL	2.80	2.79	3.05	2.88	3.11
CEP, REFL, LAR, ACW	2.89	2.70	3.02	2.87	3.42
CEP, REFL, LAR, PFL	2.84	2.69	3.00	2.84	3.40
CEP, REFL, ACW, PFL	2.80	2.83	3.01	2.88	3.12
CEP, LAR, ACW, PFL	2.84	2.81	3.03	2.89	3.16
REFL, LAR, ACW, PFL	2.86	2.82	3.08	2.92	3.00
LSF, CEP, REFL, LAR, ACW	2.82	2.64	3.00	2.82	3.61
LSF, CEP, REFL, LAR, PFL	2.78	2.64	2.98	2.80	3.60
LSF, CEP, REFL, ACW, PFL	2.75	2.73	2.99	2.82	3.40
LSF, CEP, LAR, ACW, PFL	2.79	2.71	3.01	2.84	3.41
LSF, REFL, LAR, ACW, PFL	2.80	2.73	3.03	2.85	3.17
CEP, REFL, LAR, ACW, PFL	2.85	2.74	3.03	2.87	3.19
LSF, CEP, REFL, LAR, ACW, PFL	2.75	2.65	2.98	2.79	3.34

Table 27: Decision Level Trimmed Mean Fusion Combination OAAE Results in dB

Features	Trimmed Mean				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	2.87	2.74	3.00	2.87	5.83
LSF, REFL	2.84	2.71	3.03	2.86	4.32
LSF, LAR	2.92	2.71	3.07	2.90	4.35
LSF, ACW	2.87	2.97	3.14	2.99	3.96
LSF, PFL	2.77	2.93	3.07	2.92	3.97
CEP, REFL	2.92	2.70	2.98	2.87	4.82
CEP, LAR	3.02	2.73	3.04	2.93	4.85
CEP, ACW	2.98	2.99	3.09	3.02	4.45
CEP, PFL	2.84	2.93	3.03	2.93	4.41
REFL, LAR	3.03	2.75	3.10	2.96	3.49
REFL, ACW	2.99	3.00	3.17	3.05	3.14
REFL, PFL	2.86	2.95	3.09	2.96	3.07
LAR, ACW	3.08	2.99	3.18	3.08	3.20
LAR, PFL	2.92	2.93	3.12	2.99	3.14
ACW, PFL	3.05	3.36	3.27	3.23	3.01
LSF, CEP, REFL	3.16	2.97	3.13	3.09	6.55
LSF, CEP, LAR	3.16	2.97	3.13	3.09	6.55
LSF, CEP, ACW	3.16	2.97	3.13	3.09	6.55
LSF, CEP, PFL	3.16	2.97	3.13	3.09	6.55
LSF, REFL, LAR	3.15	2.97	3.21	3.11	3.65
LSF, REFL, ACW	3.15	2.97	3.21	3.11	3.65
LSF, REFL, PFL	3.15	2.97	3.21	3.11	3.65
LSF, LAR, ACW	3.31	2.97	3.27	3.18	3.76
LSF, LAR, PFL	3.31	2.97	3.27	3.18	3.76
LSF, ACW, PFL	3.55	3.75	3.65	3.65	3.45
CEP, REFL, LAR	3.15	2.97	3.21	3.11	3.65
CEP, REFL, ACW	3.15	2.97	3.21	3.11	3.65
CEP, REFL, PFL	3.15	2.97	3.21	3.11	3.65
CEP, LAR, ACW	3.31	2.97	3.27	3.18	3.76
CEP, LAR, PFL	3.31	2.97	3.27	3.18	3.76
CEP, ACW, PFL	3.55	3.75	3.65	3.65	3.45
REFL, LAR, ACW	3.31	2.97	3.27	3.18	3.76
REFL, LAR, PFL	3.31	2.97	3.27	3.18	3.76
REFL, ACW, PFL	3.55	3.75	3.65	3.65	3.45
LAR, ACW, PFL	3.55	3.75	3.65	3.65	3.45
LSF, CEP, REFL, LAR	2.92	2.70	2.98	2.87	4.82
LSF, CEP, REFL, ACW	2.92	2.70	2.98	2.87	4.82
LSF, CEP, REFL, PFL	2.92	2.70	2.98	2.87	4.82
LSF, CEP, LAR, ACW	3.02	2.73	3.04	2.93	4.85
LSF, CEP, LAR, PFL	3.02	2.73	3.04	2.93	4.85
LSF, CEP, ACW, PFL	2.98	2.99	3.09	3.02	4.45
LSF, REFL, LAR, ACW	3.03	2.75	3.10	2.96	3.49
LSF, REFL, LAR, PFL	3.03	2.75	3.10	2.96	3.49
LSF, REFL, ACW, PFL	2.99	3.00	3.17	3.05	3.14
LSF, LAR, ACW, PFL	3.08	2.99	3.18	3.08	3.20
CEP, REFL, LAR, ACW	3.03	2.75	3.10	2.96	3.49
CEP, REFL, LAR, PFL	3.03	2.75	3.10	2.96	3.49
CEP, REFL, ACW, PFL	2.99	3.00	3.17	3.05	3.14
CEP, LAR, ACW, PFL	3.08	2.99	3.18	3.08	3.20
REFL, LAR, ACW, PFL	3.08	2.99	3.18	3.08	3.20
LSF, CEP, REFL, LAR, ACW	2.92	2.64	2.98	2.85	4.29
LSF, CEP, REFL, LAR, PFL	2.92	2.64	2.98	2.85	4.29
LSF, CEP, REFL, ACW, PFL	2.84	2.77	2.99	2.87	3.99
LSF, CEP, LAR, ACW, PFL	2.91	2.78	3.01	2.90	4.02
LSF, REFL, LAR, ACW, PFL	2.93	2.80	3.07	2.93	3.16
CEP, REFL, LAR, ACW, PFL	2.93	2.80	3.07	2.93	3.16
LSF, CEP, REFL, LAR, ACW, PFL	2.85	2.69	2.97	2.84	3.79

Table 28: Distance Level Minimum Fusion Combination OAAE Results in dB

Features	Minimum Distance				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	3.03	2.99	3.21	3.08	5.59
LSF, REFL	3.03	2.99	3.21	3.08	5.59
LSF, LAR	3.03	2.99	3.21	3.08	5.59
LSF, ACW	3.03	2.99	3.21	3.08	5.59
LSF, PFL	3.03	2.99	3.21	3.08	5.59
CEP, REFL	3.16	2.97	3.13	3.09	6.55
CEP, LAR	3.16	2.97	3.13	3.09	6.55
CEP, ACW	3.16	2.97	3.13	3.09	6.55
CEP, PFL	3.16	2.97	3.13	3.09	6.55
REFL, LAR	3.15	2.97	3.21	3.11	3.65
REFL, ACW	3.15	2.97	3.21	3.11	3.65
REFL, PFL	3.15	2.97	3.21	3.11	3.65
LAR, ACW	3.31	2.97	3.27	3.18	3.76
LAR, PFL	3.31	2.97	3.27	3.18	3.76
ACW, PFL	3.55	3.75	3.65	3.65	3.45
LSF, CEP, REFL	3.03	2.99	3.21	3.08	5.59
LSF, CEP, LAR	3.03	2.99	3.21	3.08	5.59
LSF, CEP, ACW	3.03	2.99	3.21	3.08	5.59
LSF, CEP, PFL	3.03	2.99	3.21	3.08	5.59
LSF, REFL, LAR	3.03	2.99	3.21	3.08	5.59
LSF, REFL, ACW	3.03	2.99	3.21	3.08	5.59
LSF, REFL, PFL	3.03	2.99	3.21	3.08	5.59
LSF, LAR, ACW	3.03	2.99	3.21	3.08	5.59
LSF, LAR, PFL	3.03	2.99	3.21	3.08	5.59
LSF, ACW, PFL	3.03	2.99	3.21	3.08	5.59
CEP, REFL, LAR	3.16	2.97	3.13	3.09	6.55
CEP, REFL, ACW	3.16	2.97	3.13	3.09	6.55
CEP, REFL, PFL	3.16	2.97	3.13	3.09	6.55
CEP, LAR, ACW	3.16	2.97	3.13	3.09	6.55
CEP, LAR, PFL	3.16	2.97	3.13	3.09	6.55
CEP, ACW, PFL	3.16	2.97	3.13	3.09	6.55
REFL, LAR, ACW	3.15	2.97	3.21	3.11	3.65
REFL, LAR, PFL	3.15	2.97	3.21	3.11	3.65
REFL, ACW, PFL	3.15	2.97	3.21	3.11	3.65
LAR, ACW, PFL	3.31	2.97	3.27	3.18	3.76
LSF, CEP, REFL, LAR	3.03	2.99	3.21	3.08	5.59
LSF, CEP, REFL, ACW	3.03	2.99	3.21	3.08	5.59
LSF, CEP, REFL, PFL	3.03	2.99	3.21	3.08	5.59
LSF, CEP, LAR, ACW	3.03	2.99	3.21	3.08	5.59
LSF, CEP, LAR, PFL	3.03	2.99	3.21	3.08	5.59
LSF, CEP, ACW, PFL	3.03	2.99	3.21	3.08	5.59
LSF, REFL, LAR, ACW	3.03	2.99	3.21	3.08	5.59
LSF, REFL, LAR, PFL	3.03	2.99	3.21	3.08	5.59
LSF, REFL, ACW, PFL	3.03	2.99	3.21	3.08	5.59
LSF, LAR, ACW, PFL	3.03	2.99	3.21	3.08	5.59
CEP, REFL, LAR, ACW	3.16	2.97	3.13	3.09	6.55
CEP, REFL, LAR, PFL	3.16	2.97	3.13	3.09	6.55
CEP, REFL, ACW, PFL	3.16	2.97	3.13	3.09	6.55
CEP, LAR, ACW, PFL	3.16	2.97	3.13	3.09	6.55
REFL, LAR, ACW, PFL	3.15	2.97	3.21	3.11	3.65
LSF, CEP, REFL, LAR, ACW	3.03	2.99	3.21	3.08	5.59
LSF, CEP, REFL, LAR, PFL	3.03	2.99	3.21	3.08	5.59
LSF, CEP, REFL, ACW, PFL	3.03	2.99	3.21	3.08	5.59
LSF, CEP, LAR, ACW, PFL	3.03	2.99	3.21	3.08	5.59
LSF, REFL, LAR, ACW, PFL	3.03	2.99	3.21	3.08	5.59
CEP, REFL, LAR, ACW, PFL	3.16	2.97	3.13	3.09	6.55
LSF, CEP, REFL, LAR, ACW, PFL	3.03	2.99	3.21	3.08	5.59

Table 29: Distance Level Mean Fusion Combination OAAE Results in dB

Features	Mean Distance				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	2.97	2.83	3.12	2.97	5.90
LSF, REFL	2.98	2.78	3.10	2.95	4.06
LSF, LAR	3.01	2.83	3.17	3.00	4.12
LSF, ACW	2.96	3.02	3.28	3.09	3.70
LSF, PFL	2.90	2.97	3.22	3.03	3.80
CEP, REFL	3.06	2.77	3.14	2.99	4.54
CEP, LAR	3.17	2.79	3.20	3.05	4.61
CEP, ACW	3.02	2.95	3.16	3.04	4.08
CEP, PFL	2.95	2.92	3.10	2.99	4.18
REFL, LAR	3.16	2.81	3.26	3.08	3.53
REFL, ACW	3.06	3.00	3.28	3.11	3.20
REFL, PFL	2.96	2.96	3.21	3.04	3.16
LAR, ACW	3.12	2.95	3.27	3.11	3.31
LAR, PFL	3.02	2.96	3.24	3.07	3.26
ACW, PFL	3.19	3.42	3.50	3.37	3.16
LSF, CEP, REFL	2.97	2.74	3.12	2.94	4.66
LSF, CEP, LAR	3.02	2.73	3.16	2.97	4.73
LSF, CEP, ACW	2.90	2.82	3.13	2.95	4.30
LSF, CEP, PFL	2.88	2.80	3.10	2.92	4.43
LSF, REFL, LAR	3.02	2.74	3.19	2.99	3.73
LSF, REFL, ACW	2.92	2.81	3.17	2.97	3.43
LSF, REFL, PFL	2.87	2.78	3.12	2.92	3.42
LSF, LAR, ACW	2.93	2.78	3.17	2.96	3.49
LSF, LAR, PFL	2.90	2.78	3.16	2.95	3.47
LSF, ACW, PFL	2.92	2.95	3.28	3.05	3.34
CEP, REFL, LAR	3.10	2.74	3.22	3.02	3.97
CEP, REFL, ACW	2.99	2.79	3.19	2.99	3.67
CEP, REFL, PFL	2.95	2.78	3.16	2.96	3.66
CEP, LAR, ACW	3.05	2.76	3.21	3.00	3.72
CEP, LAR, PFL	2.98	2.79	3.17	2.98	3.72
CEP, ACW, PFL	2.94	2.94	3.15	3.01	3.41
REFL, LAR, ACW	3.04	2.81	3.30	3.05	3.27
REFL, LAR, PFL	3.02	2.85	3.27	3.04	3.27
REFL, ACW, PFL	2.98	3.00	3.28	3.09	3.06
LAR, ACW, PFL	3.00	2.96	3.28	3.08	3.11
LSF, CEP, REFL, LAR	3.02	2.70	3.16	2.96	4.11
LSF, CEP, REFL, ACW	2.92	2.73	3.17	2.94	3.84
LSF, CEP, REFL, PFL	2.91	2.72	3.13	2.92	3.85
LSF, CEP, LAR, ACW	2.94	2.70	3.16	2.93	3.89
LSF, CEP, LAR, PFL	2.93	2.74	3.16	2.95	3.93
LSF, CEP, ACW, PFL	2.86	2.81	3.12	2.93	3.67
LSF, REFL, LAR, ACW	2.95	2.71	3.20	2.95	3.39
LSF, REFL, LAR, PFL	2.93	2.74	3.19	2.95	3.36
LSF, REFL, ACW, PFL	2.86	2.81	3.17	2.94	3.22
LSF, LAR, ACW, PFL	2.89	2.78	3.18	2.95	3.22
CEP, REFL, LAR, ACW	3.01	2.73	3.25	2.99	3.55
CEP, REFL, LAR, PFL	3.01	2.73	3.21	2.98	3.52
CEP, REFL, ACW, PFL	2.92	2.80	3.19	2.97	3.32
CEP, LAR, ACW, PFL	2.95	2.78	3.20	2.97	3.36
REFL, LAR, ACW, PFL	2.97	2.85	3.29	3.04	3.11
LSF, CEP, REFL, LAR, ACW	2.94	2.69	3.19	2.94	3.66
LSF, CEP, REFL, LAR, PFL	2.96	2.68	3.18	2.94	3.66
LSF, CEP, REFL, ACW, PFL	2.87	2.73	3.17	2.92	3.49
LSF, CEP, LAR, ACW, PFL	2.91	2.72	3.16	2.93	3.53
LSF, REFL, LAR, ACW, PFL	2.91	2.74	3.21	2.95	3.22
CEP, REFL, LAR, ACW, PFL	2.95	2.74	3.24	2.98	3.30
LSF, CEP, REFL, LAR, ACW, PFL	2.92	2.69	3.19	2.93	3.43

Table 30: Distance Level Median Fusion Combination OAAE Results in dB

Features	Median Distance				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	2.97	2.83	3.12	2.97	5.90
LSF, REFL	2.98	2.78	3.10	2.95	4.06
LSF, LAR	3.01	2.83	3.17	3.00	4.12
LSF, ACW	2.96	3.02	3.28	3.09	3.70
LSF, PFL	2.90	2.97	3.22	3.03	3.80
CEP, REFL	3.06	2.77	3.14	2.99	4.54
CEP, LAR	3.17	2.79	3.20	3.05	4.61
CEP, ACW	3.02	2.95	3.16	3.04	4.08
CEP, PFL	2.95	2.92	3.10	2.99	4.18
REFL, LAR	3.16	2.81	3.26	3.08	3.53
REFL, ACW	3.06	3.00	3.28	3.11	3.20
REFL, PFL	2.96	2.96	3.21	3.04	3.16
LAR, ACW	3.12	2.95	3.27	3.11	3.31
LAR, PFL	3.02	2.96	3.24	3.07	3.26
ACW, PFL	3.19	3.42	3.50	3.37	3.16
LSF, CEP, REFL	2.95	2.79	3.12	2.95	4.79
LSF, CEP, LAR	3.04	2.77	3.16	2.99	4.80
LSF, CEP, ACW	2.95	2.83	3.14	2.97	4.64
LSF, CEP, PFL	2.96	2.84	3.13	2.97	4.70
LSF, REFL, LAR	3.06	2.79	3.17	3.01	3.73
LSF, REFL, ACW	2.97	2.85	3.19	3.00	3.46
LSF, REFL, PFL	2.96	2.82	3.17	2.98	3.45
LSF, LAR, ACW	3.00	2.84	3.19	3.01	3.55
LSF, LAR, PFL	2.99	2.84	3.20	3.01	3.55
LSF, ACW, PFL	2.99	3.09	3.35	3.14	3.23
CEP, REFL, LAR	3.13	2.80	3.23	3.05	3.89
CEP, REFL, ACW	2.99	2.84	3.19	3.01	3.49
CEP, REFL, PFL	2.97	2.83	3.17	2.99	3.50
CEP, LAR, ACW	3.07	2.83	3.22	3.04	3.65
CEP, LAR, PFL	3.05	2.83	3.21	3.03	3.67
CEP, ACW, PFL	2.99	3.07	3.24	3.10	3.18
REFL, LAR, ACW	3.11	2.89	3.29	3.09	3.38
REFL, LAR, PFL	3.07	2.89	3.28	3.08	3.39
REFL, ACW, PFL	2.98	3.10	3.32	3.13	3.10
LAR, ACW, PFL	3.04	3.09	3.31	3.15	3.09
LSF, CEP, REFL, LAR	3.03	2.71	3.16	2.97	4.07
LSF, CEP, REFL, ACW	2.94	2.74	3.15	2.94	3.81
LSF, CEP, REFL, PFL	2.94	2.74	3.14	2.94	3.85
LSF, CEP, LAR, ACW	2.95	2.72	3.16	2.94	3.91
LSF, CEP, LAR, PFL	2.97	2.75	3.17	2.96	3.96
LSF, CEP, ACW, PFL	2.87	2.84	3.15	2.95	3.56
LSF, REFL, LAR, ACW	3.00	2.73	3.19	2.97	3.43
LSF, REFL, LAR, PFL	3.00	2.77	3.22	3.00	3.41
LSF, REFL, ACW, PFL	2.91	2.84	3.21	2.99	3.16
LSF, LAR, ACW, PFL	2.90	2.84	3.20	2.98	3.20
CEP, REFL, LAR, ACW	3.04	2.73	3.24	3.00	3.47
CEP, REFL, LAR, PFL	3.02	2.74	3.21	2.99	3.45
CEP, REFL, ACW, PFL	2.92	2.83	3.20	2.99	3.18
CEP, LAR, ACW, PFL	2.94	2.79	3.19	2.97	3.23
REFL, LAR, ACW, PFL	2.98	2.85	3.28	3.04	3.15
LSF, CEP, REFL, LAR, ACW	2.99	2.72	3.19	2.97	3.67
LSF, CEP, REFL, LAR, PFL	2.99	2.74	3.21	2.98	3.66
LSF, CEP, REFL, ACW, PFL	2.91	2.79	3.18	2.96	3.40
LSF, CEP, LAR, ACW, PFL	2.93	2.77	3.18	2.96	3.50
LSF, REFL, LAR, ACW, PFL	2.97	2.82	3.24	3.01	3.30
CEP, REFL, LAR, ACW, PFL	3.00	2.79	3.26	3.02	3.30
LSF, CEP, REFL, LAR, ACW, PFL	2.95	2.72	3.21	2.96	3.37

Table 31: Distance Level Trimmed Mean Fusion Combination OAAE Results in dB

Features	Trimmed Mean Distance				
	Soft Decision				
	AWGN	Pink	CPV	Trained	CMV
LSF, CEP	2.97	2.83	3.12	2.97	5.90
LSF, REFL	2.98	2.78	3.10	2.95	4.06
LSF, LAR	3.01	2.83	3.17	3.00	4.12
LSF, ACW	2.96	3.02	3.28	3.09	3.70
LSF, PFL	2.90	2.97	3.22	3.03	3.80
CEP, REFL	3.06	2.77	3.14	2.99	4.54
CEP, LAR	3.17	2.79	3.20	3.05	4.61
CEP, ACW	3.02	2.95	3.16	3.04	4.08
CEP, PFL	2.95	2.92	3.10	2.99	4.18
REFL, LAR	3.16	2.81	3.26	3.08	3.53
REFL, ACW	3.06	3.00	3.28	3.11	3.20
REFL, PFL	2.96	2.96	3.21	3.04	3.16
LAR, ACW	3.12	2.95	3.27	3.11	3.31
LAR, PFL	3.02	2.96	3.24	3.07	3.26
ACW, PFL	3.19	3.42	3.50	3.37	3.16
LSF, CEP, REFL	2.95	2.79	3.12	2.95	4.79
LSF, CEP, LAR	3.04	2.77	3.16	2.99	4.80
LSF, CEP, ACW	2.95	2.83	3.14	2.97	4.64
LSF, CEP, PFL	2.96	2.84	3.13	2.97	4.70
LSF, REFL, LAR	3.06	2.79	3.17	3.01	3.73
LSF, REFL, ACW	2.97	2.85	3.19	3.00	3.46
LSF, REFL, PFL	2.96	2.82	3.17	2.98	3.45
LSF, LAR, ACW	3.00	2.84	3.19	3.01	3.55
LSF, LAR, PFL	2.99	2.84	3.20	3.01	3.55
LSF, ACW, PFL	2.99	3.09	3.35	3.14	3.23
CEP, REFL, LAR	3.13	2.80	3.23	3.05	3.89
CEP, REFL, ACW	2.99	2.84	3.19	3.01	3.49
CEP, REFL, PFL	2.97	2.83	3.17	2.99	3.50
CEP, LAR, ACW	3.07	2.83	3.22	3.04	3.65
CEP, LAR, PFL	3.05	2.83	3.21	3.03	3.67
CEP, ACW, PFL	2.99	3.07	3.24	3.10	3.18
REFL, LAR, ACW	3.11	2.89	3.29	3.09	3.38
REFL, LAR, PFL	3.07	2.89	3.28	3.08	3.39
REFL, ACW, PFL	2.98	3.10	3.32	3.13	3.10
LAR, ACW, PFL	3.04	3.09	3.31	3.15	3.09
LSF, CEP, REFL, LAR	3.03	2.71	3.16	2.97	4.07
LSF, CEP, REFL, ACW	2.94	2.74	3.15	2.94	3.81
LSF, CEP, REFL, PFL	2.94	2.74	3.14	2.94	3.85
LSF, CEP, LAR, ACW	2.95	2.72	3.16	2.94	3.91
LSF, CEP, LAR, PFL	2.97	2.75	3.17	2.96	3.96
LSF, CEP, ACW, PFL	2.87	2.84	3.15	2.95	3.56
LSF, REFL, LAR, ACW	3.00	2.73	3.19	2.97	3.43
LSF, REFL, LAR, PFL	3.00	2.77	3.22	3.00	3.41
LSF, REFL, ACW, PFL	2.91	2.84	3.21	2.99	3.16
LSF, LAR, ACW, PFL	2.90	2.84	3.20	2.98	3.20
CEP, REFL, LAR, ACW	3.04	2.73	3.24	3.00	3.47
CEP, REFL, LAR, PFL	3.02	2.74	3.21	2.99	3.45
CEP, REFL, ACW, PFL	2.92	2.83	3.20	2.99	3.18
CEP, LAR, ACW, PFL	2.94	2.79	3.19	2.97	3.23
REFL, LAR, ACW, PFL	2.98	2.85	3.28	3.04	3.15
LSF, CEP, REFL, LAR, ACW	2.97	2.68	3.18	2.94	3.61
LSF, CEP, REFL, LAR, PFL	2.98	2.70	3.18	2.96	3.63
LSF, CEP, REFL, ACW, PFL	2.89	2.74	3.15	2.93	3.40
LSF, CEP, LAR, ACW, PFL	2.88	2.74	3.16	2.93	3.46
LSF, REFL, LAR, ACW, PFL	2.92	2.76	3.22	2.97	3.21
CEP, REFL, LAR, ACW, PFL	2.94	2.73	3.24	2.97	3.22
LSF, CEP, REFL, LAR, ACW, PFL	2.93	2.70	3.18	2.94	3.37

Table 32: Four Best Fusion Combinations OAAE in dB

Combination	AWGN	Pink	CPV	CMV	Avg. Improvement
Best Single Feature	3.03, LSF	2.97, CEP, REFL, LAR	3.13, CEP	3.27, PFL	0
Mean REFL, LAR, ACW, PFL	2.82	2.82	3.03	2.98	0.19
Median LSF, REFL, ACW, PFL	2.76	2.81	3.03	3.1	0.18
Mean REFL, ACW, PFL	2.85	2.98	3.08	2.93	0.14
Mean LAR, ACW, PFL	2.9	2.97	3.1	2.98	0.11

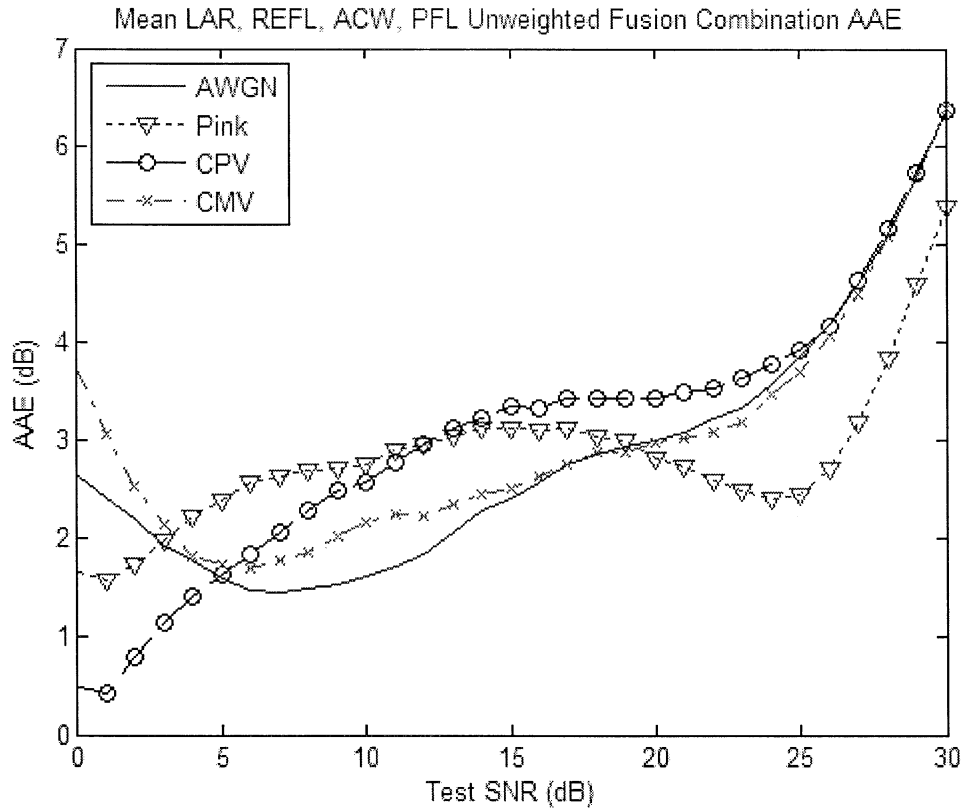


Figure 54. AAE For all Tested Noise Types Under Best Unweighted Fusion Combination

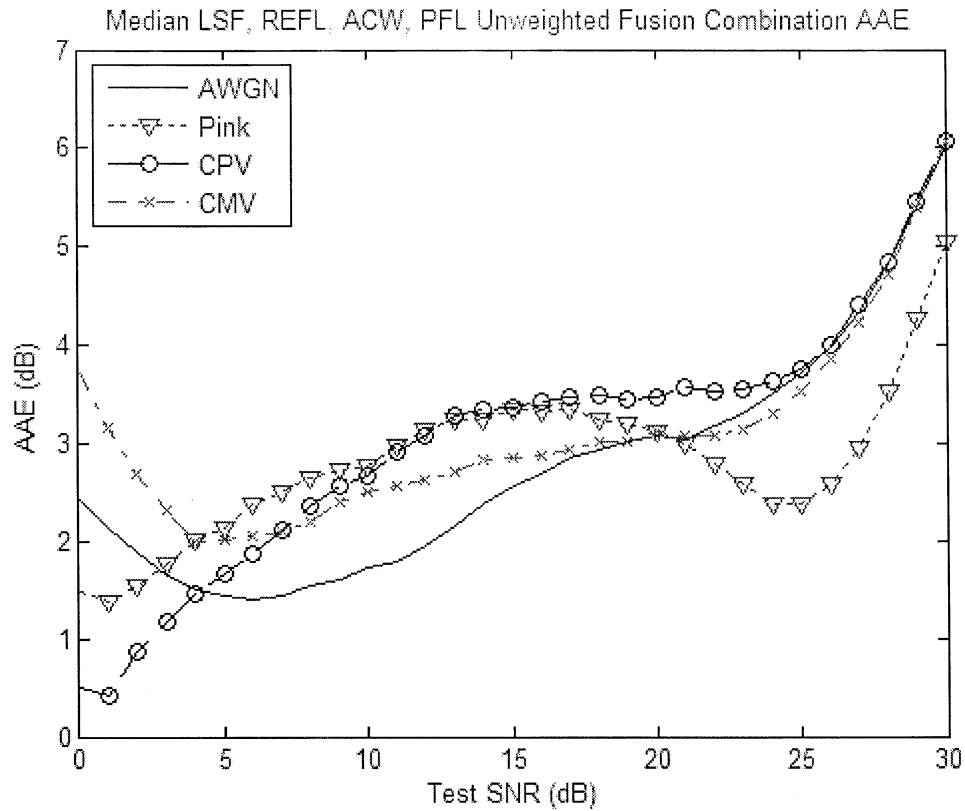


Figure 55. AAE For all Tested Noise Types Under Second Best Unweighted Fusion Combination

4.3.2 Weighted Fusion

Weighted fusion involves mean combination of features after adjusting the SNR estimation provided by each involved feature based on a predetermined weight. One method of obtaining the weights would be the use of an optimization technique. A genetic algorithm was designed to determine the optimal weights for a feature set to improve feature fusion. The weights are provided by the genetic algorithm by optimizing the soft decision estimation of the training data against the system it was used to train and then those weights are used to provide an estimate for the test data. The genetic algorithm was run a total of 100 times per feature combination to find the best set of weights for each

combination. Multiple trials were required due to the random nature of the algorithm. Table 33 provides the OAAE results achieved through weighted combination. The best weights are obtained by first averaging the OAAE provided for each noise type for combination and weight set and then taking the minimum averaged OAAE for each combination. Table 34 shows the best weighted feature combinations and their improvement to each noise spectrum type individually and as an average for total system improvement. Figure 56 and figure 57 show the AAE for the two best performing weighted fusion combinations. The OAAE results show that using a genetic algorithm to generate weights for mean fusion did reduce error especially for the three trained noise spectra, however in many cases CMV noise error increased. The increase in CMV noise estimation OAAE is likely due to the weights being generated on strictly training data which did not include CMV noise. The two best fusion combinations both provide an average improvement of .19 dB on average for all tested noise spectrum types. The combination of the Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum again provides the best combination with the least amount of features. The Line Spectral Frequencies, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum also provides an average improvement in OAAE of .19 dB, and the Line Spectral Frequencies, Reflection Coefficients, ACW Cepstrum, and PFL Cepstrum provide an improvement in OAAE of .16 dB. These two systems however provide little improvement for CMV Noise. These combinations perform best because they utilize mostly features that perform well for the untrained noise spectrum type, CMV Noise. Using this optimization technique to generate weights to improve feature combination

has proven unsuccessful, as the best feature combination provides the same improvement as the unweighted system with the same number of features and same improvement in CMV noise. More improvement in CMV noise regardless of slight penalties against the three trained noise types would motivate use of the weights to improve system performance on noise spectra not used in training.

Table 33: Decision Level Weighted Fusion OAAE Results in dB

GA Based Fusion Features	OAAE			CMV
	W	P	T	
LSF, CEP	2.87	2.77	3.03	5.70
LSF, REFL	2.84	2.71	3.03	4.23
LSF, LAR	2.91	2.72	3.07	4.41
LSF, ACW	2.90	3.02	3.17	3.86
LSF, PFL	2.82	3.04	3.13	3.65
CEP, REFL	3.02	2.81	3.08	3.99
CEP, LAR	3.02	2.73	3.03	5.04
CEP, ACW	3.09	3.16	3.21	3.97
CEP, PFL	2.92	3.16	3.16	3.66
REFL, LAR	3.03	2.76	3.10	3.49
REFL, ACW	2.97	2.89	3.11	3.26
REFL, PFL	2.86	2.96	3.09	3.07
LAR, ACW	3.08	2.88	3.14	3.32
LAR, PFL	2.95	2.87	3.12	3.20
ACW, PFL	3.02	3.35	3.25	3.01
LSF, CEP, REFL	2.83	2.70	3.02	4.28
LSF, CEP, LAR	2.84	2.64	2.97	5.10
LSF, CEP, ACW	2.84	2.92	3.09	4.05
LSF, CEP, PFL	2.79	3.00	3.09	3.72
LSF, REFL, LAR	2.85	2.64	3.01	3.89
LSF, REFL, ACW	2.88	2.90	3.10	3.32
LSF, REFL, PFL	2.77	2.87	3.04	3.25
LSF, LAR, ACW	2.89	2.84	3.09	3.48
LSF, LAR, PFL	2.78	2.80	3.04	3.45
LSF, ACW, PFL	2.82	3.07	3.11	3.24
CEP, REFL, LAR	2.97	2.73	3.05	3.79
CEP, REFL, ACW	2.93	2.91	3.10	3.29
CEP, REFL, PFL	2.85	2.95	3.08	3.09
CEP, LAR, ACW	3.01	2.95	3.13	3.48
CEP, LAR, PFL	2.86	2.92	3.08	3.35
CEP, ACW, PFL	2.89	3.13	3.12	3.27
REFL, LAR, ACW	2.92	2.74	3.05	3.24
REFL, LAR, PFL	2.82	2.86	3.06	3.05
REFL, ACW, PFL	2.84	2.86	3.05	3.07
LAR, ACW, PFL	2.89	2.88	3.08	3.04
LSF, CEP, REFL, LAR	2.82	2.59	2.95	4.40
LSF, CEP, REFL, ACW	2.82	2.81	3.03	3.65
LSF, CEP, REFL, PFL	2.75	2.87	3.03	3.38
LSF, CEP, LAR, ACW	2.79	2.71	2.98	4.07
LSF, CEP, LAR, PFL	2.77	2.83	3.02	3.63
LSF, CEP, ACW, PFL	2.78	2.97	3.05	3.46
LSF, REFL, LAR, ACW	2.80	2.73	3.02	3.42
LSF, REFL, LAR, PFL	2.75	2.77	3.01	3.27
LSF, REFL, ACW, PFL	2.75	2.87	3.02	3.12
LSF, LAR, ACW, PFL	2.78	2.86	3.04	3.18
CEP, REFL, LAR, ACW	2.89	2.77	3.03	3.37
CEP, REFL, LAR, PFL	2.85	2.96	3.10	3.04
CEP, REFL, ACW, PFL	2.82	2.96	3.05	3.02
CEP, LAR, ACW, PFL	2.83	2.92	3.03	3.33
REFL, LAR, ACW, PFL	2.80	2.83	3.03	2.98
LSF, CEP, REFL, LAR, ACW	2.82	2.74	3.01	3.48
LSF, CEP, REFL, LAR, PFL	2.72	2.73	2.96	3.65
LSF, CEP, REFL, ACW, PFL	2.77	2.87	3.02	3.30
LSF, CEP, LAR, ACW, PFL	2.74	2.77	2.98	3.64
LSF, REFL, LAR, ACW, PFL	2.74	2.75	2.99	3.18
CEP, REFL, LAR, ACW, PFL	2.79	2.82	3.00	3.15
LSF, CEP, REFL, LAR, ACW, PFL	2.71	2.64	2.93	3.79

Table 34: Three Best Weighted Fusion Combinations OAAE in dB

Combination	AWGN	Pink	CPV	CMV	Avg. Improvement
Best Single Feature	3.03, LSF	2.97, CEP, REFL, LAR	3.13, CEP	3.27, PFL	0
REFL, LAR, ACW, PFL	2.8	2.83	3.03	2.98	0.19
LSF, REFL, LAR, ACW, PFL	2.74	2.75	2.99	3.18	0.19
LSF, REFL, ACW, PFL	2.75	2.87	3.02	3.12	0.16

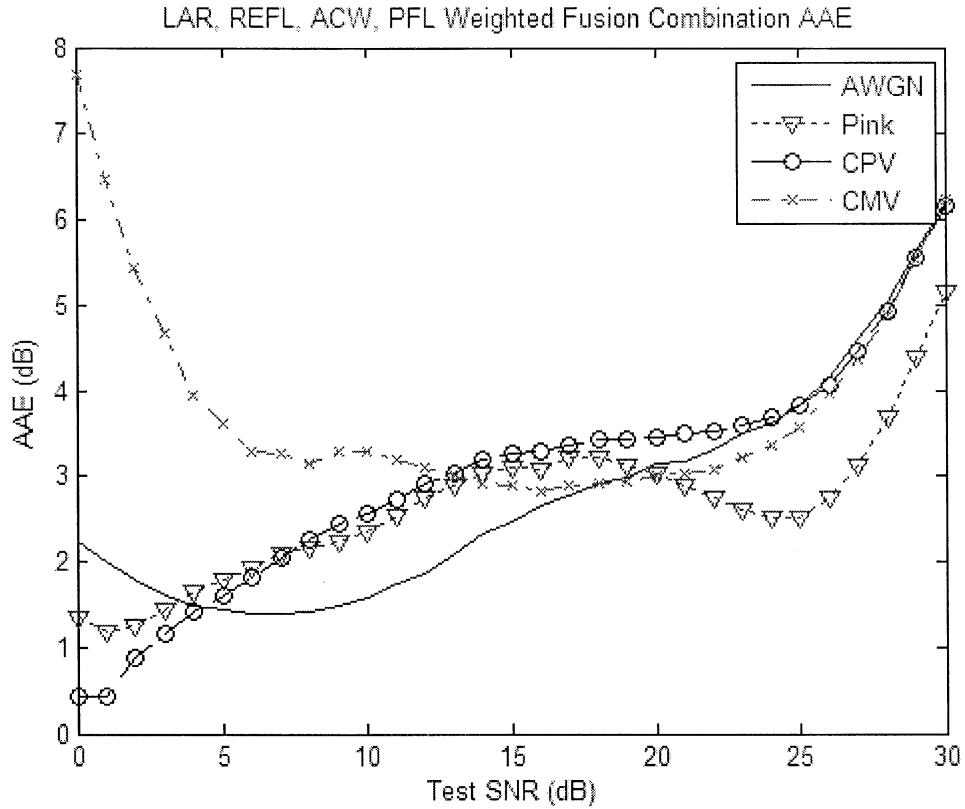


Figure 56. AAE For all Tested Noise Types Under Best Weighted Fusion Combination

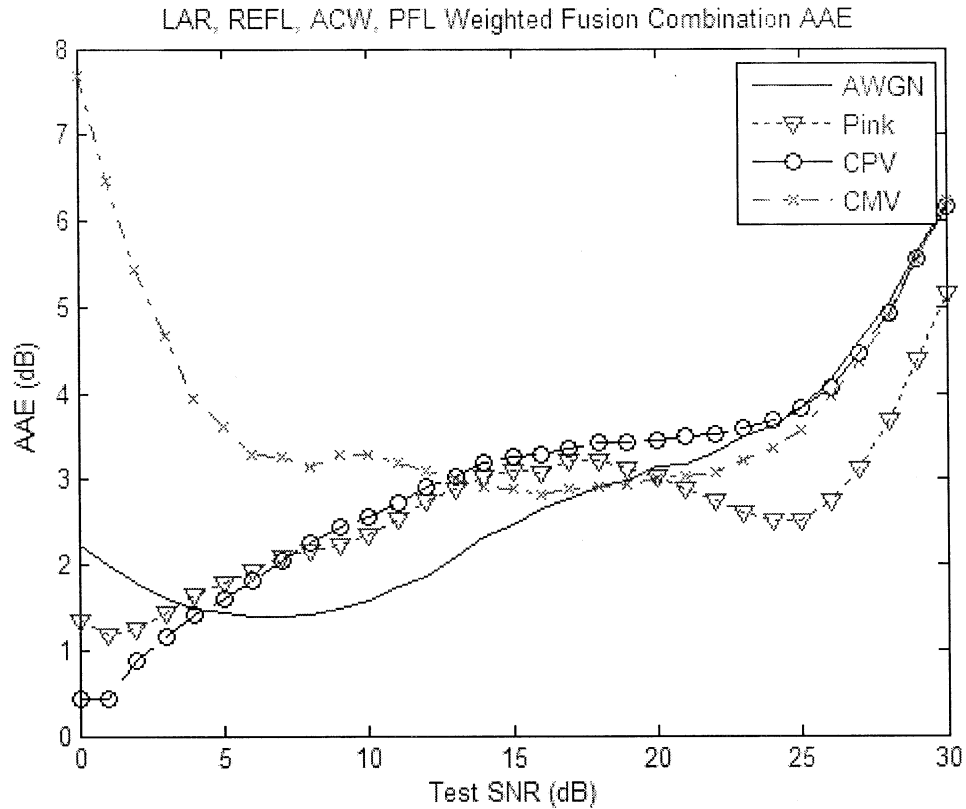


Figure 57. AAE For all Tested Noise Types Under Second Best Weighted Fusion Combination

4.4 MCRA Benchmark

The MCRA benchmark test was designed to be comparable to the VQ based pattern recognition approach. No training was required as this is not a pattern recognition based algorithm. All 380 sentences from the New England Dialect portion of the TIMIT database were used for testing at all tested SNR levels. The same range was used for the benchmark test as well and signals were tested at integer SNR levels within the range of 0 to 30 dB SNR inclusive. The base decision method involved obtaining the signal's SNR level estimate from an average of the individual frame estimates. The first modified method involved setting a floor at 0 dB and averaging all frame estimates above the floor

not inclusive. The second modified method involved setting a floor at 65% of the maximum frame SNR estimate and finding the average of all frame estimates above the floor not inclusive. Table 35 shows the OAAE provided for each noise spectrum type used to test the MCRA system. AWGN, Pink Noise, CPV Noise, and CMV Noise were tested to adequately compare this method of SNR estimation to the robust VQ based system. Figure 58 shows the AAE for just the 65% floor method for all four noise types. It is observed that setting the floor at 0 dB actually increased the MCRA system OAAE. Setting the floor at 65% of the maximum frame estimate reduced the OAAE and gave the best SNR estimated for the MCRA system. The MCRA system does not require training and provides SNR estimations with similar accuracy for all tested noise spectra. Referring to table 32, the best combination for the VQ system performs better than the MCRA system in OAAE between 1.28 dB for CPV Noise and 1.57 dB for Pink Noise.

Table 35: MCRA Benchmark OAAE in dB

Noise Type	Base Method	0 dB Floor Method	65% Floor Method
AWGN	6.07	7.01	4.31
Pink Noise	6.15	6.97	4.39
CPV Noise	6.44	7.07	4.31
CMV Noise	6.37	7.08	4.31

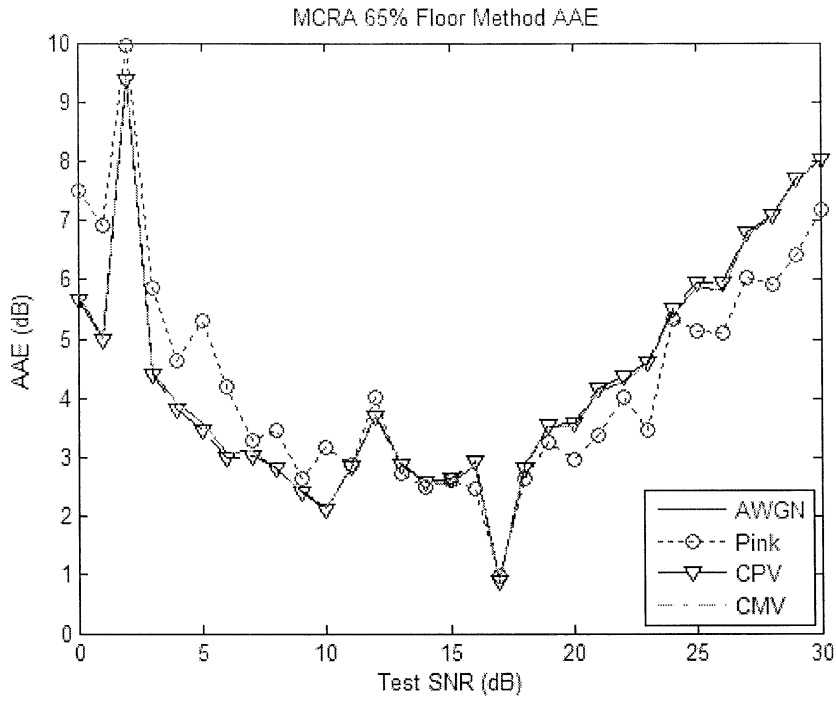


Figure 58. AAE For all Tested Noise Types Under Best MCRA Decision Method

CHAPTER 5 – CONCLUSIONS

5.1 Synopsis of Thesis

Chapter 1 provides an introduction to the problem, to estimate the SNR of a speech signal with no prior knowledge of the original speech signal or the noise corrupting it. Chapter 2 provides background information for the SNR estimation method proposed by this thesis. This background includes other noise estimation methods and an explanation of algorithms used in this specific approach. Chapter 3 provides an synopsis of the approach to developing a robust SNR estimation system for speech signals. Chapter 4 presents and discusses the results obtained by performing the experiments detailed in the approach. The results justify the motivation for the steps defined in the approach to create a robust SNR estimation system.

5.2 Summary of Accomplishments

The goal of this thesis was to research and implement a system for estimating the SNR of a speech signal using suitable features and a VQ pattern recognition approach. From the results the system will estimate the SNR of a speech signal within the range of 0 to 30 dB with an approximate error between 2.8 dB and 3 dB on average. A comparison of the objectives of this thesis as listed in the Introduction chapter with the accomplishments of the approach is listed:

- 7. To investigate six linear predictive based speech signal features identified from use in speaker recognition systems and their contribution to a signal to noise*

ratio estimating system.

The Line Spectral Frequencies, LP Cepstrum, Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum were studied at every level of the approach. The findings conclude, only after thorough experimentation, that the Line Spectral Frequencies and LP Cepstrum are not suitable for a robust SNR estimation system and the PFL Cepstrum is the most suitable for a robust SNR estimation system.

8. *To implement a pattern recognition approach to signal to noise ratio estimation using vector quantization classifiers.*

Vector Quantization classification proved to have the ability to classify SNR of speech signals when using adequate features.

9. *To investigate classifier parameters when designing the VQ system and their effect on signal to noise ratio estimation.*

The parameters of this VQ classifier system that were studied include the size of the VQ codebooks and the spacing between the codebooks. It was shown that increasing the size of the codebooks improved classification performance and reducing the spacing between the created codebooks smoothed the error curve reducing overall error.

10. *To study the ability of a VQ based pattern recognition system to estimate signal to noise ratio of multiple noise types including additive white Gaussian noise, pink noise, Continental Poor Voice (CPV) noise, and Continental Mid Voice (CMV) noise.*

Systems trained on individual noise spectrum types are capable of estimating the SNR of the input speech signal corrupted by the trained noise type but are unable to estimate the SNR of a speech signal corrupted by an untrained noise type.

11. To train a system robust to all trained and untrained noise spectrum types.

A robust system was created with training data from signals corrupted with multiple noise spectra. This system was able to estimate the SNR of a speech signal corrupted by any trained noise spectrum using any of the six studied features. The system showed that the PFL Cepstrum and ACW Cepstrum were the only features robust enough to estimate the SNR of a speech signal corrupted with an untrained noise spectrum.

12. To study feature fusion techniques to identify the best feature combination and fusion scheme for a robust system.

Feature fusion proved to reduce the error of the robust system by finding the best combination of features and the best technique. Due mostly to their ability to estimate the untrained CMV noise spectrum type, the decision level mean combination of the Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum was the best feature combination and method. Use of a genetic algorithm to implement a weighted combination method proved to be not useful and provided results only as good as unweighted combination.

5.3 Final Recommendations

The VQ pattern recognition approach has been thoroughly investigated for its ability to estimate the SNR of a speech signal. Based on the results yielded from the approach,

conclusions can be drawn for a system implementing this method. Codebook size should be further investigated for the demands of a specific system, but codebooks designed with 256 code vectors performed the best for all features studied in the initial phase of experiments. A resolution of 1 dB was the best codebook spacing found in our experimentation. Use of a soft decision estimation method helps to reduce error found in the hard decision classification approach. A system trained on one type of noise will not be robust to input signals corrupted by other types of noise, however if a system is trained on multiple noise spectra the PFL Cepstrum is the best single feature candidate. Use of decision level mean fusion of the estimates provided by the Reflection Coefficients, Log Area Ratios, ACW Cepstrum, and PFL Cepstrum reduces overall system error, showing that these four features provide complementary data. Use of an optimization technique to perform weighted fusion did not reduce system error on a scale greater than unweighted fusion, though other optimization techniques could be studied. The VQ classifier pattern recognition method proved to be a feasible approach for robust SNR estimation in speech signals if provided with training on a adequate number of noise spectra.

REFERENCES

- [1] M. C. Huggins and J. J. Grieco, "Confidence Metrics for Speaker Identification", Int. Conf. on Spoken Language Processing, Denver, Colorado, 2002
- [2] M. C. Huggins and J. J. Grieco, "Speaker Identification Confidence Metrics For Heterogeneous Model Spaces", Proc. of the 8th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, Florida, pp. 440--443, July, 2004
- [3] J. Tchorz, B. Kollmeier, "SNR Estimation Based on Amplitude Modulation Analysis With Applications to Noise Suppression", IEEE Transactions on Speech and Audio Processing, vol. , pp. 184-192, May, 2003
- [4] Ethem Alpaydin, Introduction to Machine Learning, MIT Press, 2004
- [5] K. Yamashita, T. Shimamura, "Nonstationary Noise Estimation Using Low-Frequency Regions for Spectral Subtraction", IEEE Signal Processing Letters, vol. 12, pp. 465-468, June, 2005
- [6] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-27 no.2, pp. 113-120, April, 1979
- [7] E. Nemer, R. Goubran, S. Mahmoud, "SNR Estimation of Speech Signals Using Subbands and Fourth-Order Statistics", IEEE Transactions on Signal Processing, vol. 6 No. 7, pp. 171-174, July, 1999
- [8] I. Cohen, B. Berdugo, "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement, IEEE Signal Processing Letters, vol. 9 No. 1, pp. 12-15, January, 2002
- [9] T.F. Quatieri, "Discrete Time Speech Signal Processing Principles and Practice", Prentice Hall, 2002
- [10] Mammone, R.J., Zhang, X., Ramachandran, R.P., Robust Speaker Recognition, IEEE Signal Processing Magazine, September, 1996
- [11] K. T. Assaleh and R. J. Mammone, "New LP-derived Features for Speaker Identification", IEEE Trans. on Speech and Audio Proc., vol. 2, pp. 630-638, October, 1994

- [12] Reynolds, D.L., Head, L.M., Ramachandran, R.P., ASIC Implementation of Efficient Line Spectral Frequency Computation for Speech Coding Applications
- [13] Y. Linde, A. Buzo, and R.M. Gray, "An Algorithm for Vector Quantizer Design", IEEE Trans. Commun., vol. COMM-28, pp. 84-95, January, 1980
- [14] M. N. Howell, T. J. Gordon, F. V. Brandao, "Genetic Learning Automata for Function Optimization", IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, vol. 32, pp. 804-815, December, 2002
- [15] Chang Wook Ahn, R. S. Ramakrishna,, "Elitism-Based Compact Genetic Algorithms", IEEE Transactions on Evolutionary Computing, vol. 7, pp. 367-385, August, 2003
- [16] Max Planck Institute for Psycholinguistics, "Computer Corpora and Databases:Timit Speech Database", June 1997, <http://www.mpi.nl/world/tg/corpora/timit/timit.html>
- [17] R. Polikar, "Ensemble Based Systems in Decision Making", IEEE Circuits and Systems Mag., 6, No. 3, 21-45, 2006
- [18] X. CUI, A. Alwan, Noise Robust Speech Recognition Using Feature Compensation Based Polynomial Regression of Utterance SNR, IEEE Trans. on Speech and Audio Processing, vol. 13, NO. 6, pp. 1161-1172, November, 2005