NUWC-NPT Technical Report 11,681
7 June 2005

Development of a Class-Specific Module for Hyperbolic, Frequency-Modulated Signals

Brian F. Harrison
Sensors and Sonar Systems Department

NEWPORT

Naval Undersea Warfare Center Division
Newport, Rhode Island

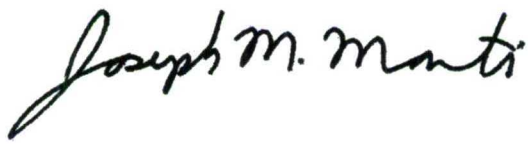
PREFACE

This work described in this report was sponsored by the Office of Naval Research under the Sonar Automation Technology Project, principal investigator Stephen G. Greineder (Code 159B).

The technical reviewer for this report was Stephen G. Greineder.

The author acknowledges the guidance of Paul M. Baggenstoss (Code 1522) in applying the class-specific method to this problem.

Reviewed and Approved: 7 June 2005



Joseph M. Monti
Head, Sensors and Sonar Systems Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	7 June 2005	

4. TITLE AND SUBTITLE

Development of a Class-Specific Module for Hyperbolic, Frequency-Modulated Signals

5. FUNDING NUMBERS

PR X101545

6. AUTHOR(S)

Brian F. Harrison

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Undersea Warfare Center Division
1176 Howell Street
Newport, RI 02841-1708

8. PERFORMING ORGANIZATION REPORT NUMBER

TR 11,681

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Ballston Centre Tower One
800 North Quincy Street
Arlington VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The class-specific method (CSM) is a novel technique for signal classification. CSM operates by computing low-dimensional feature sets, each tailored to a given signal class. In this manner, CSM avoids the difficulties inherent with classical Bayesian signal classification methods that attempt to estimate probability distributions in high-dimensional feature space. The fundamental building block of a CSM classifier is the feature extraction module. Each module is designed for a specific signal class. This report presents the development of a module for hyperbolic, frequency-modulated signals. This development also serves to acquaint the CSM novice with the general considerations in module design.

14. SUBJECT TERMS

Class-Specific Classifiers
Signal Classification
Signal Processing

15. NUMBER OF PAGES

30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

TABLE OF CONTENTS

Section **Page**

LIST OF ILLUSTRATIONS

Figure **Page**

DEVELOPMENT OF A CLASS-SPECIFIC MODULE FOR HYPERBOLIC, FREQUENCY-MODULATED SIGNALS

1. INTRODUCTION

The class-specific method (CSM) (reference 1) is a novel approach to signal classification. Classical Bayesian signal classification approaches use a common feature set for all signal classes from which an estimate of the probability distribution of the features is computed and decision boundaries are constructed. If the feature dimension is too high, severe errors in the probability-distribution estimate will occur, which will lead to classification errors. It has been shown (reference 2) that if the probability distribution meets certain smoothness assumptions, the amount of training data required for nonparametric estimators rises exponentially with feature dimension. If the feature dimension is too low, the insufficient information will cause the signal classes to become overlapped in feature space and cause classification errors. CSM, on the other hand, uses individual low-dimensional feature sets tailored to each signal class to overcome these difficulties. The key components in a class-specific classifier are the feature extraction modules designed for each signal class of interest. Understanding the process required in designing a module is fundamental to building a class-specific classifier.

In this report, the development of a module for hyperbolic, frequency-modulated (HFM) signals is presented. The objective is to describe the development of the HFM module and to acquaint the CSM novice with the general considerations in module design.

Section 2 provides a brief summary of CSM fundamentals, and section 3 provides a detailed description of the development of the HFM module.

2. FUNDAMENTAL CSM CONCEPTS

This section summarizes some of the important fundamental concepts in developing a class-specific feature module. For a detailed description of CSM, see reference 1, which also contains examples of other types of feature modules.

The CSM operates by extracting feature sets $\mathbf{z}_i$ from the raw data $\mathbf{x}$, where the features computed by module i are specific to the i^{th} data class of M class hypotheses H_i. In this manner, the feature spaces are of low dimension, thus avoiding the "curse of dimensionality" inherent when applying the classical Bayesian approach to classification. For low-dimensional feature spaces, the probability density functions (PDFs) $p(\mathbf{z}_i \mid H_i)$, $i = 1,\ldots,M$, for each of the M data classes can be accurately estimated using training data.

In classifying a raw data event $\mathbf{x}$, CSM computes features corresponding to each class $\mathbf{z}_i = T_i(\mathbf{x})$, evaluates the likelihoods $\hat{p}(\mathbf{z}_i \mid H_i)$, and then converts the likelihoods back to the raw data domain using the PDF projection given as

$$p_p(\mathbf{x} \mid H_i) = \left[\frac{p(\mathbf{x} \mid H_{0,i})}{p(\mathbf{z}_i \mid H_{0,i})} \right] \hat{p}(\mathbf{z}_i \mid H_i), \tag{1}$$

which is an approximation of $p(\mathbf{x} \mid H_i)$. Substituting equation (1) into the expression for the optimal Bayesian classifier given by

$$i^* = \arg\max p(\mathbf{x} \mid H_i) p(H_i) \quad \text{for} \quad i = 1,\ldots,M, \tag{2}$$

where $p(H_i)$ is the prior probability for class H_i, results in the CSM classifier,

$$i^* = \arg\max \left[\frac{p(\mathbf{x} \mid H_{0,i})}{p(\mathbf{z}_i \mid H_{0,i})} \right] \hat{p}(\mathbf{z}_i \mid H_i) p(H_i) \quad \text{for} \quad i = 1,\ldots,M. \tag{3}$$

The ratio

$$\mathbf{J}(\mathbf{x}, T_i, H_{0,i}) = \frac{p(\mathbf{x} \mid H_{0,i})}{p(\mathbf{z}_i \mid H_{0,i})} \tag{4}$$

in equation (3) is called the "J-function," which is the correction factor necessary to convert the feature PDFs to data PDFs. The J-function is based on a class-specific reference hypothesis $H_{0,i}$ and allows for the fair comparison of likelihoods computed

from different feature sets. Guidelines for choosing an appropriate $H_{0,i}$ are given in reference 1. In general, $H_{0,i}$ should be selected so that both the numerator and the denominator of the J-function can be determined in closed form, or to a good approximation, even in the (far) tails of the distribution. It is extremely important that the J-function be accurate to ensure that $p_p(\mathbf{x} \mid H_i)$ results in a valid PDF. As a goal, one should strive to identify a feature set $\mathbf{z}$ of low dimension and $H_{0,i}$ combination that results in $\mathbf{z}$ being an approximately sufficient statistic for differentiating $H_{0,i}$ from H_i. The better this sufficiency condition can be approximated, the more accurately the projected PDF will approximate $p(\mathbf{x} \mid H_i)$. The general form of a class-specific feature module contains both feature computation $\mathbf{z} = T(\mathbf{x})$ and J-function calculation $\mathbf{J}(\mathbf{x}, T, H_0)$.

Once a module has been developed, it must be validated to ensure that the J-function is accurate. A test, referred to as the "acid test," has been devised that provides an end-to-end validation of the module. This is done by designing a hypothesis H_v for which the PDF $p(\mathbf{x} \mid H_v)$ is exactly known and for which a large number of synthetic raw data samples can be generated. The data are converted to features and used to estimate the PDF $\hat{p}(\mathbf{z} \mid H_v)$. By applying the PDF projection in equation (1) to this estimate, the projected PDF $p_p(\mathbf{x} \mid H_v)$ is obtained. The accuracy of the J-function can then be validated by comparing the exact PDF values of the data $p(\mathbf{x} \mid H_v)$ with the projected PDF values $p_p(\mathbf{x} \mid H_v)$. This is done by plotting the $p_p(\mathbf{x} \mid H_v)$ values on one axis and the $p(\mathbf{x} \mid H_v)$ values on the other axis for each synthetic data event. The J-function is deemed accurate if the points lie near the $y = x$ line.

Feature modules can be linked together in series to form more sophisticated processing chains. In this case, the PDF projection is applied recursively, resulting in the overall J-function being the product of the individual J-functions of all the modules in the chain. Typically, computations on PDFs are done in the log domain, so the overall J-function would be the sum of the log J-function values. For example, assuming a chain comprising of three feature modules, equation (1) would become

$$\log p_p(\mathbf{x} \mid H_i) = j_1 + j_2 + j_3 + \log \hat{p}(\mathbf{z} \mid H_i), \tag{5}$$

where j_k is the log of the J-function of module k, and $\mathbf{z}$ is the feature vector at the output of module 3.

3. CLASS-SPECIFIC HFM MODULE DEVELOPMENT

The design of a class-specific module is an intricate process requiring significant attention to detail. This section presents the steps taken in the development of a module for HFM signals. The module uses a matched filter (MF) as one of its components, as well as the progression of modules developed.

3.1 HFM SIGNAL MODEL AND MATCHED FILTER

One of the components of the HFM module is an MF. The coefficients of the MF consist of samples of an HFM replica signal modeled after the HFM signal of interest. The HFM replica was developed from analysis of an experimentally obtained set of training signals. Matched filtering an HFM signal with its replica produces an impulse-like signal. This signal compression is a desirable property in that it facilitates the statistical modeling of feature sequences using a hidden Markov model (reference 3). For this case, the signal of interest was an HFM signal that swept from frequencies F1 to F2 in T milliseconds. Figure 1 is a spectrogram of a representative HFM signal, with time on the x-axis and frequency on the y-axis. Figure 2 is a spectrogram of the HFM replica signal. The HFM signal model can readily generate signals of any bandwidth and duration by simply changing the input parameters.

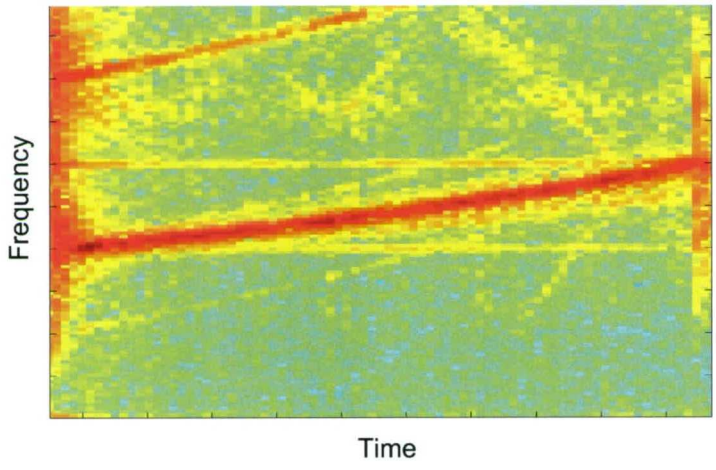


Figure 1. Spectrogram of HFM Signal

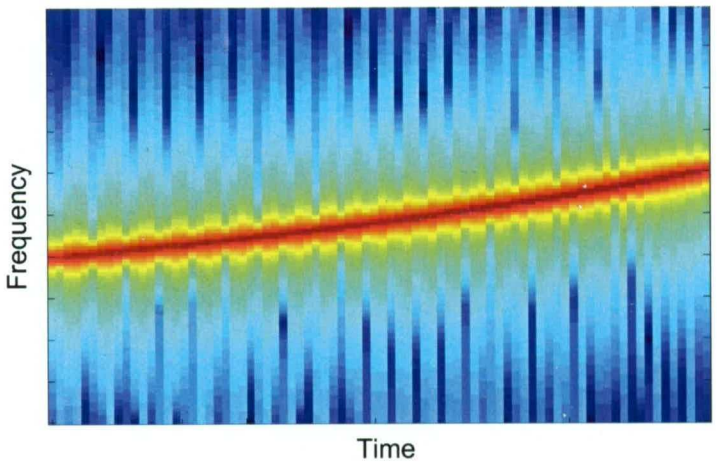


Figure 2. Spectrogram of HFM Replica

The power spectral density (PSD) of this HFM signal has a negative slope across its signal band as seen in figure 3. This is because the instantaneous frequency changes more rapidly as it increases. Therefore, since the signal dwells for less time at the higher frequencies, the signal power decreases with increasing frequency. One of the considerations when developing a class-specific module is the requirement to have filters with flat spectral responses, so that when a white process is input, a band-limited white process will be produced at the output. This is a necessary requirement for deriving the J-function for the module. Given this, the PSD of the HFM signal was flattened by multiplying the signal by $\sqrt{\partial f_i / \partial t}$, the square-root of the derivative of its instantaneous frequency f_i. The PSD of the HFM signal after flattening is shown in figure 4.

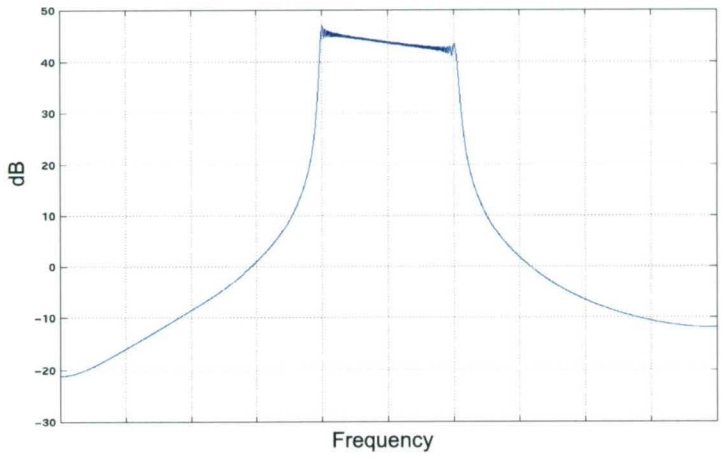


Figure 3. PSD of HFM Replica

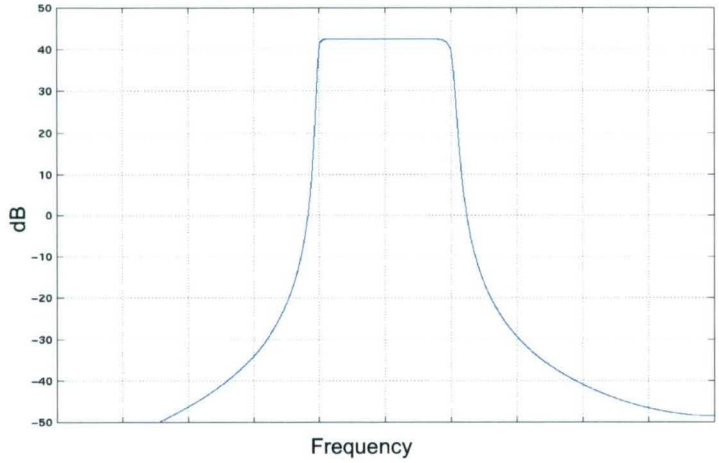


Figure 4. PSD of HFM Replica After Flattening

Using the replica, the matched-filtering operation can be viewed as the matrix-vector product $\mathbf{y} = \mathbf{W}\mathbf{x}$, where $\mathbf{y}$ is the vector of samples at the MF output, $\mathbf{x}$ is a vector containing the input signal samples, and $\mathbf{W}$ is a circulant matrix:

$$\mathbf{W} = \begin{bmatrix} b_N & \cdots & b_3 & b_2 & b_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_N & \cdots & b_3 & b_2 & b_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_N & \cdots & b_3 & b_2 & b_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_N & \cdots & b_3 & b_2 & b_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_N & \cdots & b_3 & b_2 & b_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_N & \cdots & b_3 & b_2 & b_1 \\ b_1 & 0 & 0 & 0 & 0 & 0 & b_N & \cdots & b_3 & b_2 \\ b_2 & b_1 & 0 & 0 & 0 & 0 & 0 & b_N & \cdots & b_3 \\ b_3 & b_2 & b_1 & 0 & 0 & 0 & 0 & 0 & b_N & \cdots \\ \cdots & b_3 & b_2 & b_1 & 0 & 0 & 0 & 0 & 0 & b_N \end{bmatrix}. \tag{6}$$

Each row of the matrix consists of a circularly shifted set of the MF coefficients b_i zero padded out to the length of $\mathbf{x}$. This expression can be simplified by replacing $\mathbf{W}$ with its eigen-decomposition to produce $\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{U}^H\mathbf{x}$, where the columns of $\mathbf{U}$ are the eigenvectors and $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{W}$. The eigenvectors of a circulant matrix are discrete Fourier transform (DFT) basis vectors, and the eigenvalues are equal to the DFT of its first row. Premultiplying both sides of the equation by $\mathbf{U}^H$ results in $\mathbf{U}^H\mathbf{y} = \mathbf{D}\mathbf{U}^H\mathbf{x}$. Observe on the left side of the equation that $\mathbf{U}^H\mathbf{y} = \tilde{\mathbf{y}}$ is the DFT of $\mathbf{y}$ and on the right $\mathbf{U}^H\mathbf{x} = \tilde{\mathbf{x}}$ is the DFT of $\mathbf{x}$. The matched-filtering operation in the frequency domain is then $\tilde{\mathbf{y}} = \mathbf{D}\tilde{\mathbf{x}}$. If $\mathbf{h}$ is a vector consisting of the diagonal elements of $\mathbf{D}$, then the MF operation is simply the element-by-element product of $\tilde{\mathbf{x}}$ with $\mathbf{h}$.

3.2 HFM MODULE DEVELOPMENT

As is often the case in the development of CSM modules, the development of the HFM module went through a number of iterations before the final design was reached. It is instructive to observe the process that led to the final design. Therefore, the initial module designs and the motivation for changing them, which led to the final HFM module design, are described.

Design 1

The initial idea for the HFM module feature computation was to separate the input signal into in-band and out-band components, compute features from each separately, and combine them to compose the feature set. The in-band component is that portion of the input signal whose frequency content is within the HFM replica band, i.e., F1 to F2, with the remainder of the signal considered to be the out-band component.

As a first step in verifying the operation of the module and deriving the J-function, the features were chosen to simply be the total in-band power P_{IB} and the total out-band power P_{OB}. A block diagram of the process is shown in figure 5. The upper portion of the diagram shows the pre-processing performed on the replica prior to it being input to the MF block. The replica signal is first transformed to the frequency domain using a fast Fourier transform (FFT). Its frequency spectrum is then conditioned so that the magnitudes of the FFT bins in the signal band are all normalized to one and the FFT bins outside of this band are set to zero, thus approximating a "brick wall" filter. The phase of the FFT bins remains unchanged. This conditioned spectrum is shown in figure 6.

One reason for conditioning the spectrum in this manner is to maintain orthogonality (independence) between the in-band and out-band bins. In other words, it is desired to have no sharing of energy between the in-band and out-band regions at the band edges. With CSM, accounting for all of the energy in the event is very important. Independence between the in-band and out-band feature spaces is also a desirable property since it allows a much easier derivation of the denominator of the J-function. Under the assumption of independence,

$$p(\mathbf{z} \mid H_0) = p([P_{IB}, P_{OB}] \mid H_0) = p(P_{IB} \mid H_0)\, p(P_{OB} \mid H_0). \tag{7}$$

Thus, the joint PDF of the features under H_0 is simply equal to the product of their individual PDFs under H_0.

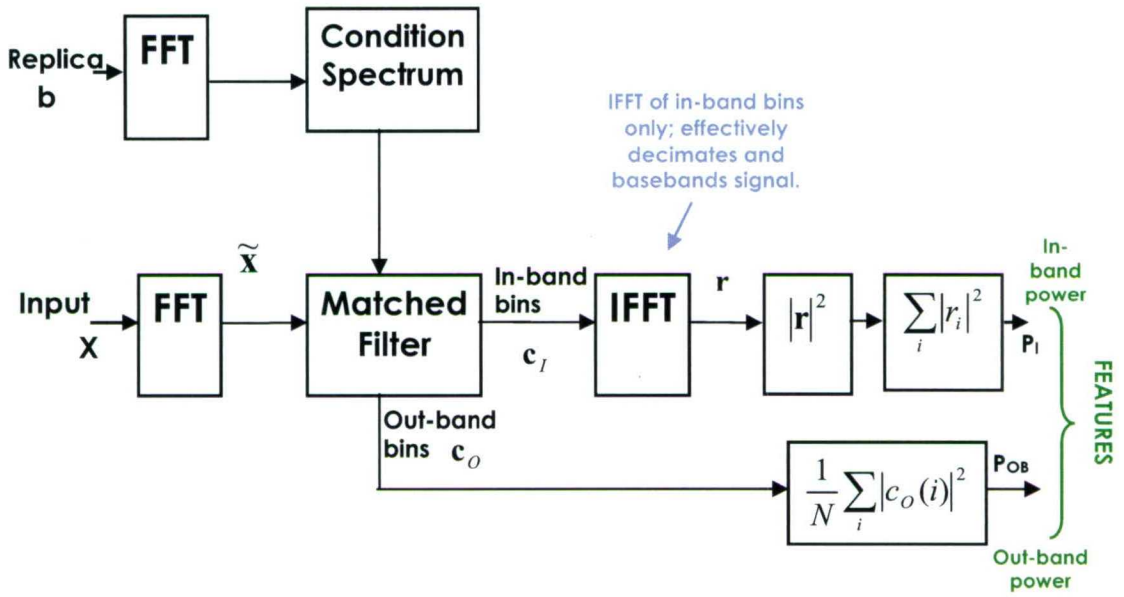


Figure 5. HFM Module Computing In-Band Power and Out-Band Power Features

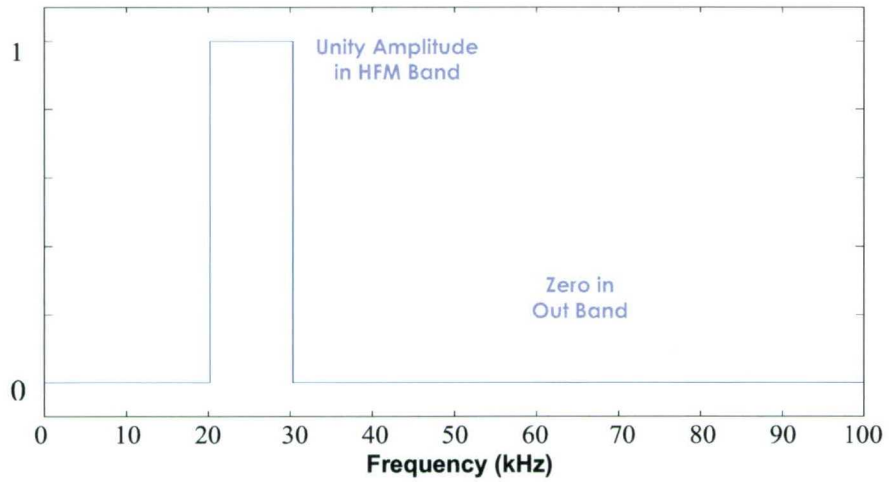


Figure 6. Magnitude Spectrum of HFM Replica After Conditioning

Processing of the input signal **x** begins by transformation to the frequency domain using an N-point FFT denoted $\tilde{\mathbf{x}}$, followed by matched-filtering with the conditioned replica. (The matched-filtering operation is performed as described in section 3.1.) At the output of the MF, the FFT bins have been separated into a set of M in-band bins and a set of $N/2 - M$ out-band bins. Since the MF has a single-sided spectrum, which results in an output signal with a single-sided spectrum, it was decided to compute the relative powers using only the positive frequencies indexed 0 to $N/2$. The in-band bins $\mathbf{c}_I$ are those encompassing the HFM signal band, and they contain the matched-filtering result. The out-band bins $\mathbf{c}_O$ are the remaining bins, and they contain their coefficients *prior* to the

matched-filtering operation, i.e., they are not all zeros. The out-band power is then computed as

$$P_{OB} = \frac{1}{N} \sum_{i=1}^{N/2-M} \left| c_O(i) \right|^2 . \tag{8}$$

The M in-band bins are input to an inverse FFT (IFFT) and transformed back to a time series **r**. Note that since the M bins are used in the IFFT, the result is equivalent to base-banding and decimating the MF output by N/M. To preserve the proper signal power during this operation, the $\mathbf{c}_I$ are scaled by $\sqrt{M/N}$ prior to the IFFT operation. To illustrate, if the input signal had a white spectrum with magnitude A, the magnitudes of $\mathbf{c}_I$ would all equal A at the output of the MF. Total power prior to decimation would then be

$$P = \frac{1}{N} \sum_{M} A^2 = \frac{MA^2}{N}. \tag{9}$$

Without scaling, the total power after base-banding and decimation would be

$$P' = \frac{1}{M} \sum_{M} A^2 = \frac{MA^2}{M} = A^2; \tag{10}$$

therefore, scaling the $\mathbf{c}_I$ by $\sqrt{M/N}$ preserves signal power.

Total in-band power is then computed from **r**, the time-domain MF output, as

$$P_{IB} = \sum_{i=1}^{M} \left| r(i) \right|^2 . \tag{11}$$

Now, consider the derivation of the J-function for this module. Since the choice of a reference hypothesis is left to the module developer, let the reference hypothesis H_0 be independent, zero-mean, Gaussian noise of variance 1. The numerator of the log J-function is then

$$\log p(\mathbf{x} \mid H_0) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \left(\sum_{i=1}^{N} x(i)^2 \right). \tag{12}$$

A library of MATLAB functions called the Class-Specific Toolkit exists at the Naval Undersea Warfare Center Division, Newport, RI; this library contains a collection of pre-tested class-specific modules, including a function to evaluate equation (12) for a given data event **x**.

To derive the denominator of the J-function, one needs to determine the statistics at the output of each of the processing blocks in the module. Under H_0, at the output of the FFT the bins are independent, zero-mean, Gaussian random variables (RVs) with variance N. Bins 0 and $N/2$ are real-valued, while the rest are complex valued. The statistics remain the same for the in-band and out-band bins at the output of the MF.

Focusing first on the out-band power computation, the magnitude-squaring operation on the $\mathbf{c}_O$ results in independent chi-squared RVs $\tilde{\mathbf{c}}_O$. The bins 0 and $N/2$ are chi-squared distributed with 1 degree of freedom scaled by N and denoted by $p_r(\tilde{c})$:

$$p_r(\tilde{c}) = \frac{1}{N\sqrt{2\pi}} \left(\frac{\tilde{c}_i}{N} \right)^{-1/2} \exp\left\{ -\frac{\tilde{c}_i}{2N} \right\}. \tag{13}$$

Bins 1 through $N/2-1$ are chi-squared with 2 degrees of freedom scaled by $N/2$ and denoted by $p_c(\tilde{c})$:

$$p_c(\tilde{c}) = \frac{1}{N} \exp\left\{ -\frac{\tilde{c}_i}{N} \right\}. \tag{14}$$

The complete log-PDF of $\tilde{\mathbf{c}}_O$ under H_0 is then

$$\log p(\tilde{\mathbf{c}}_O \mid H_0) = \log p_r(\tilde{c}_0) + \sum_{i \in O} \log p_c(\tilde{c}_i) + \log p_r(\tilde{c}_{N/2}), \tag{15}$$

where the summation in the second term is over all of the out-band bins, excluding bins 0 and $N/2$.

The result of summing the magnitude-squared bins P_{OB} is also a chi-squared distributed RV. However, there is a problem with determining the statistics of P_{OB} because it is the sum of chi-squared RVs of different scale factors. In order for the PDF of P_{OB} to be a proper chi-squared distribution, the distributions of all of the magnitude-squared bins must have the same scale factor. One can accomplish this by scaling the bins $\mathbf{c}_O$ prior to the magnitude-squaring operation. If the real-valued bins are multiplied by $1/\sqrt{N}$ and the complex-valued bins by $\sqrt{2/N}$, all of the bins after magnitude-squaring will be chi-squared distributed scaled by 1. With that, the PDF $p(P_{OB} \mid H_0)$ is determined to be chi-squared with $2(N/2 - M - 2) + 2$ degrees of freedom scaled by $1/N$. The Class-Specific Toolkit also contains a function to evaluate the log-PDF of a chi-squared RV given its degrees of freedom and scale.

Turning now to the statistics for the in-band bins, applying the scale factor $\sqrt{M/N}$ to the bins at the MF output results in the IFFT output $\mathbf{r}$ consisting of samples of independent zero-mean Gaussian RVs with variance 1. The RVs are independent due to

the conditioning of the replica to have a flat spectrum coupled with using only the M in-band bins in the IFFT. Under H_0, this results in a white spectrum being input to the IFFT. Note that the time series **r** is complex-valued because its spectrum is not symmetric about 0 Hz (even spectrum). The spectrum of the real-valued input signal **x** is even, but the MF output retains only the positive frequencies that are then used in the decimation/base-banding approach to computing **r**. Therefore, the elements of $|\mathbf{r}|^2$ are independent chi-squared RVs with 2 degrees of freedom scaled by 1/2. Computing the in-band power using equation (11) results in the PDF $p(P_{IB} \mid H_0)$ being chi-squared distributed with $2M$ degrees of freedom scaled by 1/2. The complete log J-function is then

$$j = \log p(\mathbf{x} \mid H_0) - \log p(P_{IB} \mid H_0) - \log p(P_{OB} \mid H_0). \tag{16}$$

These theoretically derived statistics were verified by applying a large number of samples of independent, zero-mean, unit-variance, Gaussian RVs to the input of the module, and then computing estimates of the statistics at the outputs of each of the processing blocks. However, this module did not pass the acid test, which indicates that this feature set is not a sufficient statistic for differentiating H_i from $H_{0,i}$. A possible reason for this is that the processing did not result in exact orthogonal sets of in-band and out-band bins used to compute P_{IB} and P_{OB}, thus violating the independence assumption. Nevertheless, this development provided a good illustration in the steps required in deriving the J-function for a simple case of two features.

Design 2

The next design is a modification to the HFM module that computes a larger dimension feature set that is an approximate sufficient statistic. This module design uses P_{OB} and parameters of $|\mathbf{r}|^2$ as features instead of P_{IB}. For this case, in addition to P_{OB}, the selected features are the N largest peaks of $|\mathbf{r}|^2$, denoted $|\mathbf{r}|^2_N$, the associated indices λ of $|\mathbf{r}|^2_N$, and the residual power $P_r = |\mathbf{r}|^2 - |\mathbf{r}|^2_N$. The joint PDF under H_0 of the order statistics for $|\mathbf{r}|^2_N$ and P_r has been derived (reference 4) assuming that the process they are computed from is chi-squared with 2 degrees of freedom. A module exists in the Class-Specific Toolkit to compute this joint PDF under H_0. The joint PDF of the indices is uniformly distributed between 1 and L, where L is the number of samples in $|\mathbf{r}|^2$ and is given by

$$p(\boldsymbol{\lambda} \mid H_0) = \frac{1}{L(L-1)(L-2)\cdots(L-N+1)}. \tag{17}$$

The complete log J-function for this module is then

$$j = \log p(\mathbf{x} \mid H_0) - \log p(\boldsymbol{\lambda} \mid H_0) - \log p(|\mathbf{r}|^2_N, P_r \mid H_0) - \log p(P_{OB} \mid H_0), \tag{18}$$

where $\log p(|\mathbf{r}|_N^2, P_r \mid H_0)$ is computed by the order-statistics module.

However, the order-statistics module suggests the use of a floating-reference hypothesis—one that depends on the data. (This concept is explained in more detail in reference 1.) In general, a floating-reference hypothesis is used to prevent numerical problems when computing the J-function. It is used to position H_0 to simultaneously maximize the numerator and denominator of the J-function. This prevents $\mathbf{x}$ and $\mathbf{z}$ from being evaluated in the tails of the respective PDFs. The floating-reference hypothesis assumed at the input to the order-statistics module, denoted by $H_0(\mu)$, is chi-squared with 2 degrees of freedom and a mean equal to the mean of $|\mathbf{r}|^2$, denoted as μ. Therefore, one needs to determine a new reference hypothesis H_0' so that the assumed statistics for $H_0(\mu)$ are met. Consider the processing string producing $|\mathbf{r}|^2$ to be a module whose output is the input to the order-statistics module. Working backwards through the processing flow diagram in figure 5 (beginning with the computation of $|\mathbf{r}|^2$), one can determine the statistics for $\mathbf{x}$ assuming $p(|\mathbf{r}|^2 \mid H_0(\mu))$ at the input to the order-statistics module. The new H_0' is found to be independent, zero-mean, Gaussian noise with variance equal to μ. From this, one can now re-derive the statistics for P_{OB} given H_0'. Proceeding in the same manner used in design 1 to derive $p(P_{OB} \mid H_0)$, it is determined that the PDF is chi-squared with $2(N/2 - M - 2) + 2$ degrees of freedom scaled by μ/N. The log J-function can then be computed using these reference hypotheses; thus,

$$j = \log p(\mathbf{x} \mid H_0') - \log p(\boldsymbol{\lambda} \mid H_0') - \log p(|\mathbf{r}|_N^2, P_r \mid H_0(\mu)) - \log p(P_{OB} \mid H_0'). \qquad (19)$$

With the J-function derived, the module must be validated using the acid test. For the acid test, the feature set is assumed to consist of P_{OB}, P_r, and the two largest amplitudes of $|\mathbf{r}|^2$ and their associated indices. Therefore, $\mathbf{z}$ is a six-dimensional feature vector. Estimation of the feature PDF $\hat{p}(\mathbf{z} \mid H_v)$ was done using a Gaussian mixture (GM) model whose general form, given an N-dimensional feature vector $\mathbf{z}$, is

$$p(\mathbf{z}) = \sum_{i=1}^{L} \alpha_i \mathrm{N}(\mathbf{z}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \qquad (20)$$

where

$$\mathrm{N}(\mathbf{z}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = (2\pi)^{-N/2} |\boldsymbol{\Sigma}_i|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{z} - \boldsymbol{\mu}_i)\right\}, \qquad (21)$$

is the multivariate Gaussian function with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$. The GM PDF $p(\mathbf{z})$ is the weighted sum of L Gaussian functions, each parameterized by $\Lambda_i = \{\alpha_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$, where α_i is the mixing weights. The iterative expectation-maximization (EM) algorithm (reference 5) can be used to estimate the Λ_i for a given L-component mixture with respect

to a set of feature observations $\mathbf{z}_k$ (training data). The acid test hypothesis H_v was independent, zero-mean, Gaussian noise of variance 100. The results of the acid test are shown in figure 7. In the upper plot of figure 7 for $p(\mathbf{x} \mid H_v)$ versus $p_p(\mathbf{x} \mid H_v)$, it can be seen that the points line up closely to the $y = x$ line, indicating that the J-function is accurate. The lower plot shows the projected PDF error.

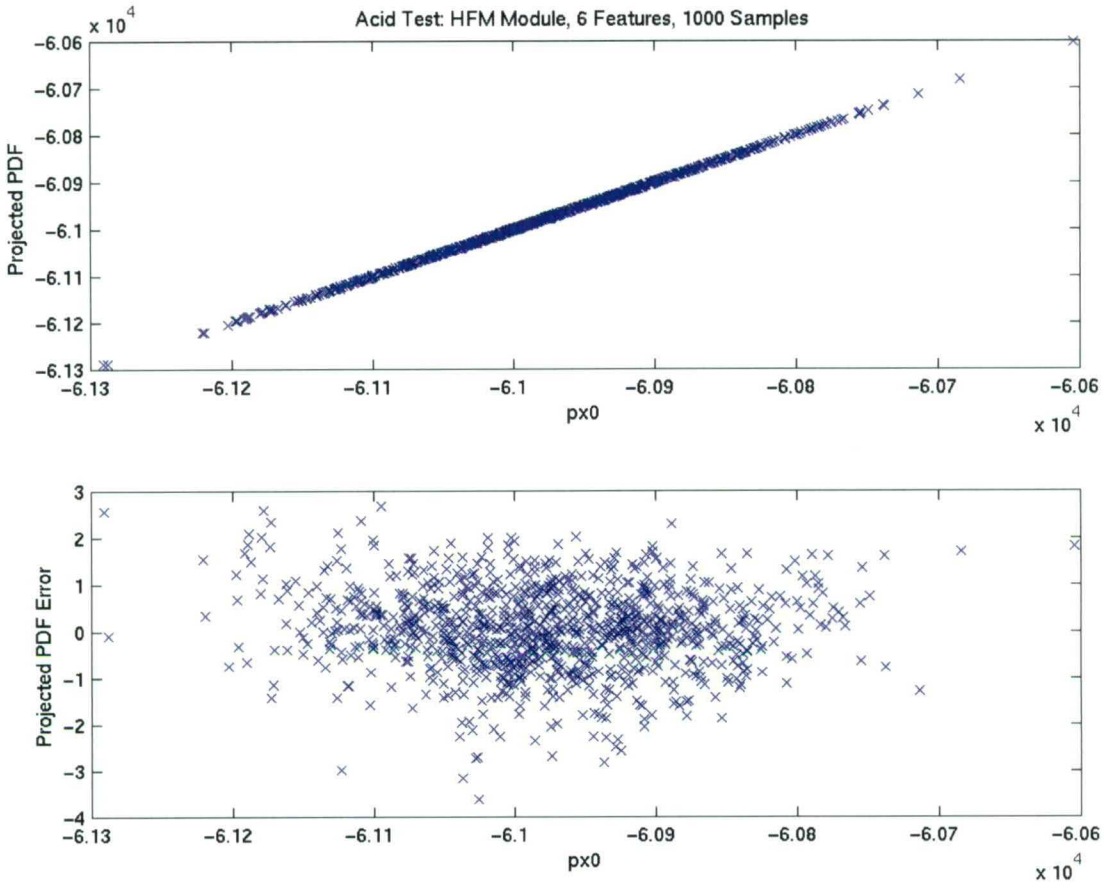


Figure 7. Acid Test Result for Module Computing Six Features

The problem observed when using this module to process experimental event data is that there are a relatively large number of significant peaks in the time-domain MF output $|\mathbf{r}|^2$ (see figure 8). This is likely due to multiple signal arrivals as a result of multipath propagation. Using all of the significant peaks would lead to a high dimensional feature space, which is contrary to the objective of CSM. The next modification (design 3) to the module is an attempt to decrease the feature space dimension.

Design 3

The third design approach was motivated by observing that the $|\mathbf{r}|^2$ obtained from the experimental event data consisting of HFM signals appeared to have a somewhat Gaussian shape. Given that, the next design idea was to model $|\mathbf{r}|^2$ as a GM. Using this approach, the Λ_i would replace the peaks of $|\mathbf{r}|^2$ and their associated indices as features. The implementation of this approach begins by passing $|\mathbf{r}|^2$ through a bank of varying duration Hanning-weighted integrators to detect the peak indices of Gaussian-like pulses and to estimate their widths. Using the indices as the GM means and the widths as GM standard deviations for initial estimates, a maximum-likelihood estimator was used to estimate the parameters of the GM model for the detected pulses. A representative $|\mathbf{r}|^2$ signal computed from experimental HFM signal data, along with its resulting GM model representation, are shown in figure 8. For this case, two Gaussian pulses were detected, resulting in a two-mode GM model.

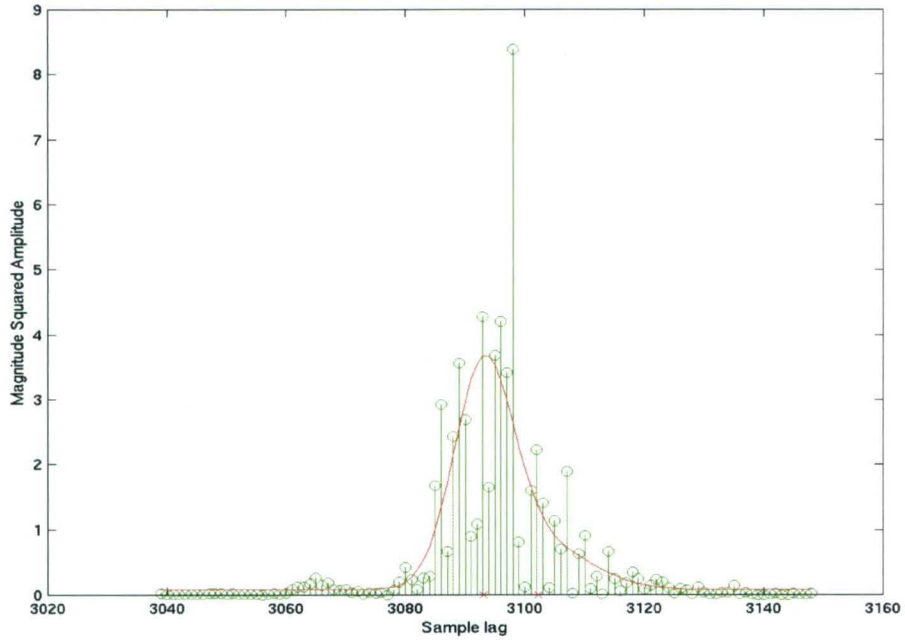


Figure 8. Time-Domain MF Output $|\mathbf{r}|^2$ (Green) and GM Model Representation (Red)

Validation of this module using the acid test was never performed. The standard acid test hypothesis H_v is independent, zero-mean, Gaussian noise of a given variance. Under the standard H_v, $|\mathbf{r}|^2$ would not have the multi-modal character, as seen in figure 8, but would instead appear more uniformly distributed as a function of sample lag. It would, therefore, be inappropriate to model it using a GM and maintain a low-dimensional feature space. An alternate H_v would need to be constructed that would meet the multi-modal requirement and also result in a $p(\mathbf{x} \mid H_v)$, which was known exactly. Rather than pursuing the development of this alternate H_v, it was decided to explore a further modification to the module.

Design 4

The next approach was to segment the IFFT output **r** into *M*-sample segments, compute the log power of each segment, and then compute the differences of the log powers as features. A hidden Markov model (HMM) was used to statistically model the sequence of features. This processing string is depicted in figure 9, where $\mathbf{r}_n^s$ is the n^{th} *M*-sample sequence derived from the segmentation of **r**, $\mathbf{P}^s$ is a vector consisting of the sequence of log powers computed from the segments $\mathbf{r}_n^s$, and the output $\mathbf{P}_d$ is a vector of log-power differences computed from $\mathbf{P}^s$ as $P_d(n) = P_n^s - P_{n-1}^s$. Determining the differences of the log-power sequence was performed for conditioning to make it appear more like a sequence of discrete states appropriate for modeling using an HMM.

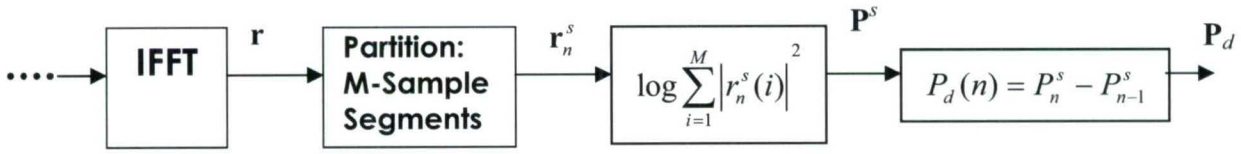


Figure 9. Computation of Log-Power Differences from Segments of **r**

The complete set of features using this approach is then $\mathbf{z} = [P_{OB}, \mathbf{P}_d]$. With the PDF of P_{OB} modeled using a GM, and the PDF of $\mathbf{P}_d$ modeled using an HMM, one needs to use a mixed modeling approach in computing $\hat{p}(\mathbf{z} \mid H_1)$. Since P_{OB} is independent of $\mathbf{P}_d$, their PDFs can be estimated independently from training data using the different models to form the complete feature PDF as

$$\hat{p}(\mathbf{z} \mid H_1) = \hat{p}(P_{OB} \mid H_1)\hat{p}(\mathbf{P}_d \mid H_1), \tag{22}$$

where hypothesis H_1 is the HFM signal.

With regard to the J-function, it was shown that under an H_0 of independent, zero-mean, Gaussian noise of variance 1, each element of $|\mathbf{r}|^2$ is independent chi-squared with 2 degrees of freedom scaled by 1/2. Therefore, prior to taking the log, the sum $\rho_n = \sum_{i=1}^{M} \left| r_n^s(i) \right|^2$ over each segment results in $p(\rho_n \mid H_0)$ being chi-squared distributed with $2M$ degrees of freedom scaled by 1/2. The complete PDF over all the segments is then the product of the $p(\rho_n \mid H_0)$ over *n*. The log function was computed using the log module in the Class-Specific Toolkit by passing it the ρ_n. The complete log J-function is then

$$j = \log p(\mathbf{x} \mid H_0) - \sum_n \log p(\rho_n \mid H_0) - p(P_{OB} \mid H_0) + j_{\log}, \tag{23}$$

where $p(P_{OB} \mid H_0)$ is the same PDF as that used in equation (16), and $j_{\log}$ is the log J-function for the log module. Note that the differencing operation has no impact on the J-function.

Once the accuracy of the J-function for this module was verified using the acid test, the next step was to use the event data to compute the likelihood in equation (1), in log space, under H_1. This was done by separating the events into two subsets, denoted E1 and E2. First, the feature PDF $\hat{p}(\mathbf{z} \mid H_1)$ must be trained using features extracted from E1. Next, equation (1) must be evaluated for each event in E2 using features extracted from the given events and combining the likelihood values. The next step is to reverse the roles of E1 and E2—train using E2 and evaluate using E1.

All of the above obtained likelihood values are then combined to form a total likelihood value for $p(\mathbf{x} \mid H_1)$ over all the events $\mathbf{x}_k$. This procedure is useful in comparing the performance of different modules in processing a given class of events. The module producing the highest likelihood is deemed the best for use in classifying the given class.

To benchmark the total likelihood value obtained using this module, it was compared to that obtained from processing the events using the autoregressive (AR) module in the Class-Specific Toolkit. The AR module is one of the more robust in that it performs well for a wide variety of signal classes. The basic operation of the AR module proceeds by dividing the input signal into M-sample segments and computing a P^{th}-order AR model (reference 6) for each segment as features. This sequence of AR features is statistically modeled using an HMM. The AR module is optimized for a given signal class by finding the values of M and P that jointly maximize the total likelihood value.

In comparing the total likelihood values obtained from both the HFM and AR modules, the AR module produced the higher value. The reason for this appears to be that the background noise is non-stationary over the duration of each event. Correspondingly, the spectrum of the event is changing over its duration. Out-band power P_{OB} is the sum of FFT bins computed using the entire event, which implicitly assumes stationarity—an assumption that is violated in this case. Alternately, the AR module computes all of its features from segmented data. It can, therefore, adapt to changes in the spectrum during the event, and its feature set better represents this signal class.

To test this conjecture, the procedure for computing the total likelihoods for both the HFM and AR modules was repeated, but this time synthetic white Gaussian noise was added to each of the events. It should be pointed out that the signal-to-noise ratios (SNRs) of each of the events are extremely large, so the addition of a relatively small level of noise will not obscure the signal. One needs to add just enough noise so that the additive noise becomes dominant over the background noise present in the events. At that point, the resulting background signal will be stationary. As the level of the additive

noise was increased, the difference between the total likelihoods for the modules decreased, and the HFM module likelihood eventually was greater than that of the AR module's. The SNR at this point was still extremely large. With the stationarity assumption met, the HFM module is better. However, the preceding observation regarding computing all features from segmented data affording better adaptability led to the final HFM module design (design 5).

Design 5

The final HFM module design is the most simple of all of the preceding versions. In fact, it simply comprises a matched-filtering operation cascaded with the AR module. It uses a hybrid linear-hyperbolic FM (LHFM) signal as the replica in the MF. A block diagram of this module is shown in figure 10. The operations in the dashed rectangle can be viewed as a pre-processor to the AR module since it has a J-function equal to 1. To illustrate, under H_0 of independent, zero-mean, Gaussian noise of variance 1, the FFT output bins will be zero-mean, Gaussian-distributed with variance N. The magnitude of the spectrum of the replica is 1 across all bins; therefore, the MF operation does not change the statistics. At the output of the IFFT, the samples are independent, zero-mean, Gaussian RVs with variance 1; thus, $p(\mathbf{x} \mid H_0) = p(\mathbf{z} \mid H_0)$ for the pre-processor. Therefore, the complete J-function for this HFM module is that of the AR module.

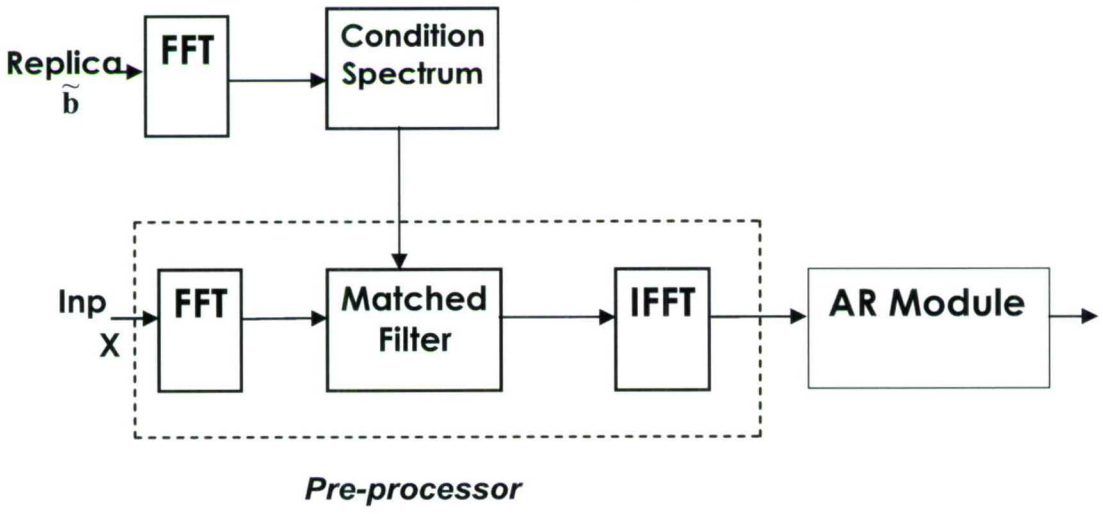


Figure 10. Final HFM Module Design

The LHFM replica was designed to maintain the HFM replica in the F1 to F2 band and to fill the bands above and below with LFM signals. The PSD of the replica is shown in the upper portion of figure 11 and its spectrogram is shown in the lower portion. In designing the replica, care was taken to ensure that the signal phase was continuous

between the LFM-HFM transitions. The use of the LHFM replica was driven by the need to have a signal with a white spectrum at the output of the IFFT under H_0. This was accomplished by conditioning the spectrum of the LHFM replica prior to the MF (see figure 10). The conditioning consists of normalizing the magnitudes of all the FFT bins to one, resulting in an all-pass filter. The entire spectrum of the input signal will be passed through the MF, thereby eliminating the separation into in-band and out-band components. All of the features can then be computed on a per-segment basis using the AR module. As before, an HMM is used to statistically model the features. Because the pre-processor produces the desired compression for the HMM, the likelihood using this module is greater than that using the AR module alone. As discussed before, the module can be optimized for a given signal class by finding the values of M and P that jointly maximize the total likelihood. The pre-processor can also be used in conjunction with other class-specific feature modules to compress any type of frequency-modulated input signal, provided that a reasonably accurate replica can be designed. In fact, this pre-processor is used as part of a recently developed module that computes a joint set of broadband and narrowband features using AR for broadband and a spectral projection technique for narrowband components.

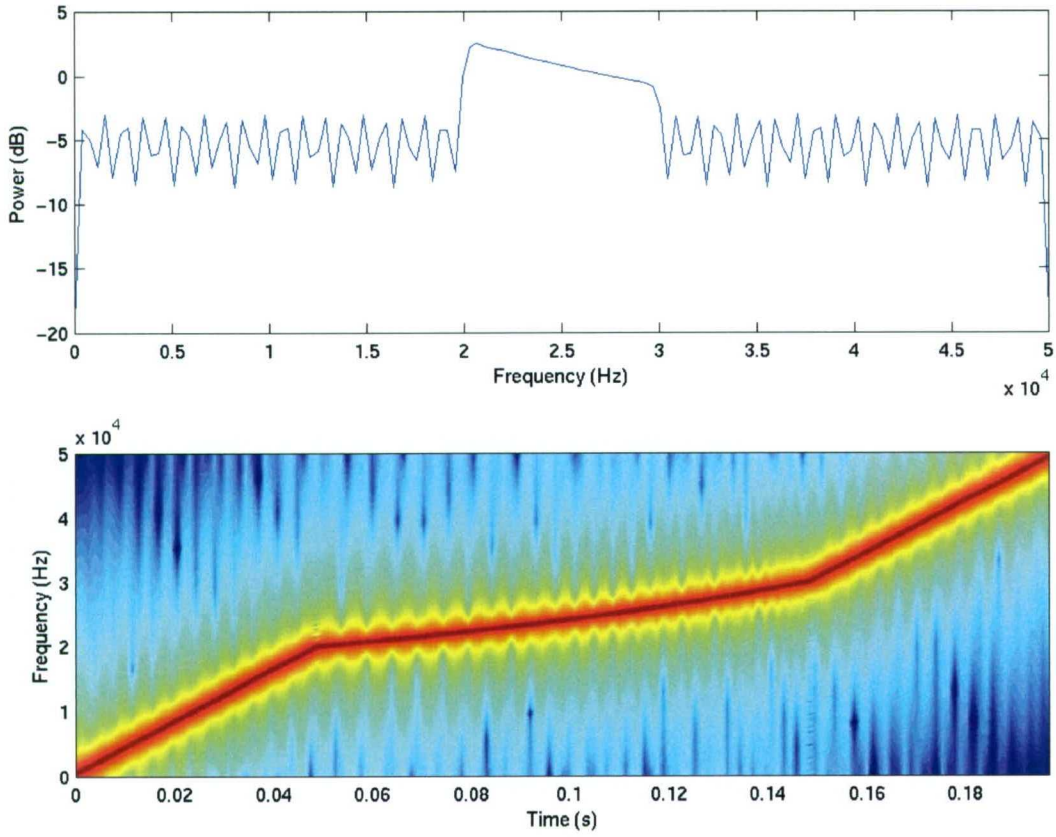


Figure 11. PSD (Top) and Spectrogram (Bottom) of LHFM Replica

4. CONCLUSIONS

Class-specific method (CSM) module design requires careful attention to detail. Fundamental CSM concepts were presented as an introduction to the signal classification technique. As an illustrative example in module design, the detailed steps in the development of a class-specific HFM module were described. Four designs were pursued and rejected until the final design—using a matched-filter pre-processor combined with the autoregressive (AR) module—was determined to achieve a higher likelihood than using the AR module alone. This approach proved to be the simplest, and it outperformed the robust general AR approach that is often used as a benchmark module. Additionally, the final design provides a generic module that can be used for any signal for which a replica can be developed.

5. REFERENCES

1. P. M. Baggenstoss, "Class-Specific Classifier: Avoiding the Curse of Dimensionality," *IEEE A&E Systems Magazine*, vol. 19, no. 1, January 2004, pp. 37-52.

2. C. J. Stone, "Optimal Rates of Convergence for Nonparametric Estimators," *Annals of Statistics*, vol. 8, pp. 1348-1360, 1980.

3. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, pp. 257-286, February 1989.

4. A. H. Nuttall, "Joint Probability Density Function of Selected Order Statistics and the Sum of the Remaining Random Variables," NUWC Technical Report 11,345, Naval Undersea Warfare Center Division, Newport, RI, October 2001.

5. D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York 1985.

6. S. Kay, *Modern Spectral Estimation: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Office of Naval Research (ONR 321: H. South, J. Tague, D. Abraham)	3
Naval Sea Systems Command (PEO-IWS-5A)	1
Defense Technical Information Center	2
Center for Naval Analyses	1