

Military Technical College  
Kobry El-Kobbah,  
Cairo, Egypt



ICEENG

7<sup>th</sup> International Conference  
on Electrical Engineering  
ICEENG 2010

## Decision Threshold Setting for Radar System Identification Using HMM

By

Hossam Eldin Abou-Bakr Hassan\*

### Abstract:

The principal objective of applying the Hidden Markov Model (HMM) in Radar System Identification (RSI) is to show that the recognition performance of a HMM exceeds that of the conventional methods such as cross-correlation. An important offshoot of this research is to provide a method for choosing the optimal model parameters for an actual radar signal so that a library of HMMs can be created and used for practical EW tasks. Therefore, optimal threshold settings between competing HMMs should be investigated to improve overall recognition performance. In this paper, a new method for predicting the false recognition rates and deriving optimal decision regions between competing HMMs that model the dynamic behavior of the radars stored in the threat library of the Electronic Warfare (EW) is proposed. The proposed method uses only the prior pulse repetition interval (PRI) statistics of a known radar and template matching which can lead to a qualitative understanding into radar correlation. Moreover, the paper investigates the idea of a threshold setting so that the receiver can have a reject option and decide that an observation sequence does not belong to any of the HMMs in its library. Computer simulations are performed through the paper to validate the obtained theoretical analysis.

### Keywords:

Hidden Markov Model (HMM), Radar System Identification (RSI), Pulse Repetition Interval (PRI).

### 1. Introduction:

In [1], [2], it is assumed that any pulsed radar observation sequence necessarily belong to one of the HMMs in the threat library of the Electronic Warfare (EW) system. Unfortunately this is not a practical assumption because there will be invariably unknown foreign signals in the environment. It is, therefore, required that a

---

\* Egyptian Armed Forces

decision on the signal source also includes the possibility of classifying it as unidentifiable [3]. In other words, a reject option must be included if the respective recognition probabilities of all the HMMs fall below a given threshold. This idea is analogous to a HMM that can not recognize a spoken word or phoneme in speech recognition. This paper, investigates first the decision threshold for observation sequences generated by HMMs in the threat library. Afterwards, a threshold will be discussed for foreign sequences.

A search for analytic method for determining an optimum decision threshold of a given HMM is difficult because there is no simple analytic probability density function (PDF) for a HMM. Ways to overcome this problem is proposed in this paper. The paper is organized as follows: Section 2 will introduce the idea of a distance vector for pseudo-random sources and show that it is possible to use this vector as a method for analytically calculating a probability of error (false recognition) for a given HMM. Section 3 will follow the method used in [4] in an effort to show that any prescribed threshold will have to be dependent on the HMM being used. That is there is no optimal universal threshold for all HMMs within a given threat library. Section 4 presents the application of the proposed method to obtain a decision threshold for unknown signals. Finally, Section 5 summarizes the main conclusions of the paper.

## **2. Pseudo-Random Source Distance Vector:**

Because we are only using the PRI of the received signal, it is difficult to extract an extensive information about the source. The parameters of a HMM we use to model the source work well as long as the observed sequence is not too corrupted and not too short. How corrupted and how short an observation depend not only on the correct model's parameters, but also on those competing HMMs in the threat library. Presumably it would be advantageous to know exactly how long a certain observation sequence must be before there is little or no possibility of an incorrect model obtaining the highest recognition probability since this could be a cut-off length. Any received signal shorter than this length would automatically be classified as "unknown" or rejected since it has a high likelihood of being erroneously classified. In addition, knowing how corrupted a sequence can be before false recognition begins to occur could aid in deciding whether to reject the sequence or not. Unfortunately, it is not possible for a receiver to determine what the error level of an observation sequence is. In any event, the false recognition rate of a particular HMM in a threat library is a function of its parameters  $A, B, \pi$  the parameters of every other competing model, the length of the observation sequence, and the number of errors in a sequence. All these variables make it extremely difficult to predict a false recognition rate and thus an optimal decision threshold for a HMM.

The idea of distance vector for a pseudo-random source attempts to simplify the problem by determining a threshold that is independent of the HMM parameters and concentrates solely on the source's prior PRI statistics. The only variables we now need to concern ourselves with are the length of the observation sequence and the characteristics of the competing models. In addition, it is assumed that there are no errors in the observation sequence.

The distance vector of a pseudo-random source is defined as a vector containing the expected values of all possible transitions per word from time  $t$  to  $t+T$ . For example take  $T=1$  so that from time  $t$  to  $t+1$  there are the following four 2-symbol transitions:

Table 1: 2-Symbol Transitions

Received Symbols	
Time= $t$	Time= $t+1$
0	0
0	1
1	0
1	1

It is possible for these signal transitions to occur within a given word (intra-word transition) or between two different words (inter-word transition). By calculating the expected value of each transition per word, it is possible to define a distance vector  $d$ , that can be used to find the mean squared Euclidian distance between two different pseudo-random sources.

Take for example, the following artificial radar (radar 4) containing two equiprobable words. Word 1: [0 1] with probability 0.5 and Word 2: [0 0 1] with probability 0.5  
 So that the conceivable generated sequence could be:01 001 001 01 001

The notation  $E[o_t, o_{t+1}]$  will be used to define the expected value of going from observed symbol  $o_t$  at time  $t$  to the next symbol  $o_{t+1}$  at time  $t+1$  per word. Therefore,  $E[0, 0]$  can be found by summing all possible ways in which '0' can be followed by another '0' multiplied by the probability of such an event happening. Here, the only way this event can occur is within word 2 and so  $E[0,0] = 1 \times \frac{1}{2} \times 1 = \frac{1}{2}$  (number of occurrences of the event  $\times$  probability of the word  $\times$  probability of going from symbol  $o_t$  to  $o_{t+1}$ ). The event '0' followed by a '1' can occur within both words so  $E[0,1] = \left(1 \times \frac{1}{2} \times 1\right) + \left(1 \times \frac{1}{2} \times 1\right) = 1$ . The event '1' followed by a '0' occurs only during inter-word transitions so the probability becomes the joint probability of being in the first word and going to the second word. Since these events are statistically

independent, the joint probability simply becomes the product of the two word probabilities. There are four possible ways of this event occurring: first word back to the first word, the first word back to the second word, the second word back to the second word, the second word back to the first word. Thus,  $E[1,0] = 4 \times \left(\frac{1}{2} \times \frac{1}{2}\right) \times 1 = 1$ . Finally, there is no way for a '1' to be followed by another '1' so  $E[1,1] = 0$ .

It is now possible to define the radar's distance vectors as

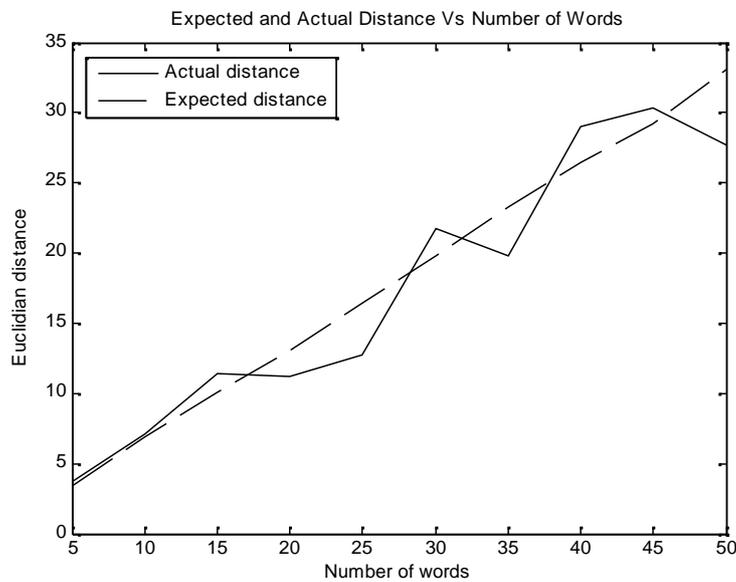
$$d_4 = [E[0,0] \quad E[0,1] \quad E[1,0] \quad E[1,1]] = [0.5 \quad 1 \quad 1 \quad 0] \quad (1)$$

The distance vector for the second source (radar 3) is found to be  $d_3 = [0.375 \quad 1.125 \quad 1.125 \quad 0.625]$  and the total square distance between these two is defined as:

$$d_{43}^2 = \sum_{i=1}^4 (x_{4i} - x_{3i})^2 \quad (2)$$

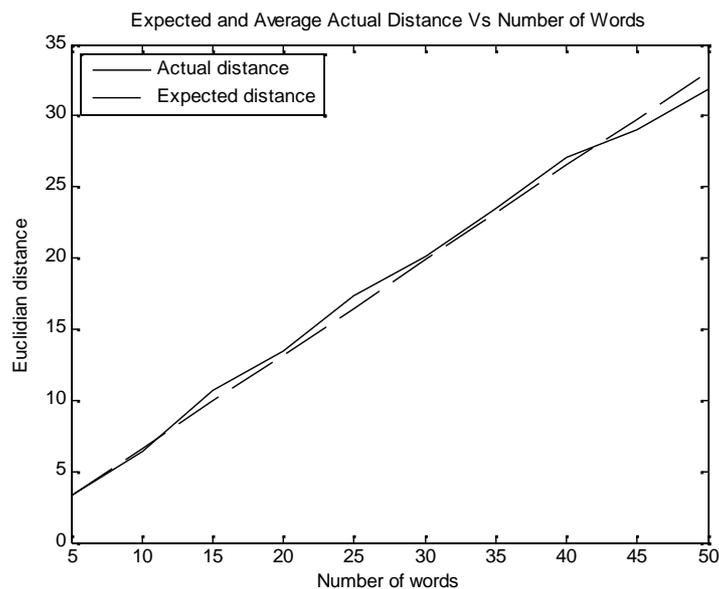
The square root of this distance is found to be 0.6614 (see Appendix A, B).

This value, 0.6614 is the *average* distance between these two radars for one word including the transition to the next word. The average distance will increase linearly with this value as the number of words increases. Obviously, the greater the distance the more different the two sequences will be and thus, the less likely they will be confused as having originated from the wrong source. This is equivalent to the false alarm rate being low for long test sequences and being high for short ones. Figure 1 shows the actual and expected distance between these two radars as the number of words increases. Note that there are no errors in both sequences. The distance vectors for both radars are computed by counting the number of 2-symbol transitions that occur in each sequence divided by the number of the words in the sequence. The Euclidian distance is then calculated by taking the square root of equation (2).



**Figure (1).** Expected and actual euclidian distance between radar 3 and radar 4

Although the actual distance as a function of words may appear erratic, if it is averaged over 50 independent trials (50 different sequences generated) for every sequence length, it converges to the expected distance as shown in Figure 2



**Figure(2):** Averaged euclidean distance between radar 3 and radar 4

In theory, increasing the value of  $T$  would provide more information about the source. That is, knowing the number of expected transitions from time  $t$  to  $t+10$  would certainly yield a more specific distance vector than the 2-symbol transition vector used previously. Unfortunately, the vector would become increasingly sparse and its entries would become increasingly small making it difficult to accurately calculate the source's theoretical distance vector. For this reason it was decided to use distance

vector with  $T=2$  (3-symbol transitions implying an 8-component distance vector).

**2.1. Distance Vector as Probability Distributions:**

If we normalize the distance vector described in Section 2, then it will be possible to obtain a PDF of the pseudo-random source. More precisely,

$$P(\Omega = w_j) = P_j = \frac{d_j}{\sum_{i=1}^N d_i} \tag{3}$$

where  $\Omega = \{w_1, w_2, \dots, w_n\}$  is the set of all possible 2-symbol transitions that can occur and  $d_j$  is the  $j^{\text{th}}$  index of the distance vector. For example if  $T=1$  then we would have the 4-component distribution vector shown in equation 1 ( $w_1 = E[0,0], w_2 = E[0,1], \dots$ ). The probability that we receive a '0' followed by another '0' in this case would be  $\frac{0.5}{0.5+1+1+0} = 0.2$ .

**2.2 Mean Squared Error (MSE) Distribution:**

The mean squared error (MSE) between the received and theoretical pseudo-random source's PDF can be shown to follow one of two possible distributions depending on whether the received probability vector belongs to a given theoretical probability vector or not. For example, if the received probability vector is of the form  $p_{received} = [p_1, p_2, \dots, p_n]$  and we know that it originated from radar A then the probability distribution of the MSE between the received vector and the theoretical vector for radar A will follow different distribution from other competing theoretical vectors. This will become clearer in the following pages. Most of the theory presented in equations 4 through 15 as well as 18 and 19 has either been taken from [5] or [6] define MSE as:

$$\begin{aligned} MSE = d_{MSE}^2 &= E \left[ \sum_{i=1}^n \left( p_i - \hat{p}_i \right)^2 \right] \\ &= \sum_{i=1}^n \left( E \left[ \hat{p}_i^2 \right] - 2 p_i E \left[ \hat{p}_i \right] + p_i^2 \right) \\ &= \sum_{i=1}^n \left( \text{var} \left[ \hat{p}_i \right] + \left( p_i - E \left( \hat{p}_i \right) \right)^2 \right) \end{aligned} \tag{4}$$

Where  $\hat{p}_i$  is the received probability of event i occurring and  $p_i$  is the theoretical (predetermined) probability. If all of our received probability are unbiased

$E\left[\hat{p}_i\right] = p_i \quad \forall i$ , and  $\sigma_i^2 = \text{var}\left[\hat{p}_i\right]$  then, the MSE simply becomes

$$d_{MSE}^2 = \sum_{i=1}^n \sigma_i^2 = n\sigma^2 \tag{5}$$

In the special case where the  $n$  estimates all have the same variance  $\sigma^2$ .

Relating this back to the probability vector, each, of the  $\hat{p}_i$  corresponds to the  $i^{\text{th}}$  component of the received probability vector (obtained from the observation sequence). It will be shown that these components will tend to a normal distribution with mean  $\mu_i$  and variance  $\sigma^2$ . Note that  $n$  refers to the number of components in the probability vector while  $m$  is the number of words in the observation sequence.

Define the normalized sum (called the *sample mean*) as

$$Y = \frac{1}{m} \sum_{i=1}^m X_i \tag{6}$$

Where  $X_i, i=1,2,3,\dots,m$  are statistically independent and identically distributed random variables. The mean of  $Y$  is [7 pp. 56-57]

$$E[Y] = \frac{1}{m} \sum_{i=1}^m E[X_i] = \frac{mE[X]}{m} = E[X] \tag{7}$$

and the variance of  $Y$  is

$$\text{var}[Y] = \frac{1}{m^2} \sum_{i=1}^m \text{var}[X_i] = \frac{m \text{var}[X_i]}{m^2} = \frac{\sigma_x^2}{m} \tag{8}$$

which implies that the variance associated with our estimate,  $\hat{p}$  (shown as  $X$  in equations 6 to 8), decreases as  $\frac{1}{m}$  where  $m$  is the number of received words. This, in turn, means that the MSE will also decrease as the number of samples gets larger or, equivalently as the observation sequence gets longer.

With the above theory, we can say the following related to the 3-symbol transition probability vector ( $n = 8$ ).

- 1- For the proper model (the theoretical and the received probability vectors originate from the same source), the MSE between them will approach zero as the number of words (samples),  $m$ , tends to infinity.
- 2- For all other models the MSE will approach the bias of the model as compared to the proper model as the number of words (samples),  $m$ , tends to infinity.

The second statement is more readily understood if we write it mathematically.

$$\begin{aligned} d_{MSE}^2 &= n\sigma^2 + n\left(p - E\left(\hat{p}\right)\right)^2 \\ &= n\sigma^2 + n\beta^2 \end{aligned} \tag{9}$$

where we have assumed that the  $n$  observations are independent and identically distributed (i. i. d) and we have denoted the bias term  $\left(p_i - E\left(\hat{p}_i\right)\right)^2$  as  $\beta^2$ .

Now, as shown in equation (8), as the number of words increases, the variance inversely decreases as  $\frac{1}{m}$ . Therefore, equation (9) becomes

$$\lim_{m \rightarrow \infty} d_{MSE}^2 = n\beta^2 \tag{10}$$

It is now possible to determine the probability distribution of the MSE for the proper and competing models. Assuming the eight components

$\left(P(w_i) = \hat{p}_i, i = 1, 2, 3, \dots, 8\right)$  of the distribution vector are independent and normally distributed with mean  $\mu_i, i = 1, 2, \dots, 8$  and variance  $\sigma_x^2$  (from the central limit theorem) then we define the first transformation

$$Y_i = p_i - \hat{p}_i \quad i = 1, 2, \dots, 8 \quad p_i \in R \tag{11}$$

where  $p_i$  and  $\hat{p}_i$  are the theoretical and received values for the  $i^{th}$  component of the given probability vector respectively. Note that the expected value of this transformation is equal to zero for the proper model and  $\beta_i$  for the  $l^{th}$  model. The second transformation is defined as

$$Z = \sum_{i=1}^n Y_i^2 \tag{12}$$

The combination of these two transformations is equivalent to the MSE calculation shown in equation (4). That is,  $Z$  is equivalent to  $d_{MSE}^2$ .

If we are dealing with the proper model then it can be shown that [6, pp41-44]  $Z$  is a central Chi-Square distribution with  $n$  degrees of freedom and has the following first two moments:

$$\begin{aligned} E[Z] &= n\sigma_y^2 \\ \text{var}[Z] &= 2n\sigma_y^4 \end{aligned} \tag{13}$$

Otherwise, if it is a competing model,  $Z$  is a non-central Chi-Square distribution with  $n$  degrees of freedom and the first two moments

$$\begin{aligned} E[Z] &= n\sigma_y^2 + S^2 \\ \text{var}[Z] &= 2n\sigma_y^4 + 4\sigma_y^2 S^2 \end{aligned} \tag{14}$$

where  $S^2$ , the non-centrality parameter, is defined by

$$S^2 = \sum_{i=1}^n \beta_i^2 \tag{15}$$

The probability that the MSE of a competing model is smaller than that of proper model, a false recognition can be defined in terms of the log likelihood ratio,  $\Lambda(Z)$  for equal prior probabilities and uniform costs as:

$$P(\log P_{\lambda_{proper}}(Z) < \log P_{\lambda_{competing}}(Z)) = P(\log \Lambda(Z) < 0) \tag{16}$$

where  $\Lambda(Z)$  is defined as

$$\frac{P_{H_0}(Z)}{P_{H_1}(Z)} \tag{17}$$

with  $P_{H_0}(Z)$  representing the MSE probability conditioned on the proper model and  $P_{H_1}(Z)$  is the MSE distribution conditioned on the competing model. This is actually a binary hypothesis test with  $H_0$  and  $H_1$  being the following two guesses:

$H_0$ : Received vector belong to radar A

$H_1$ : Received vector does not belong to radar A

Assuming the received signal originates from radar A, the probability distribution of MSE conditioned on the event  $H_0$  is [6]

$$P_{H_0}(Z) = \frac{Z^{\left(\frac{n}{2}\right)-1} e^{\frac{-Z}{2\sigma_z^2}}}{\sigma_z^n 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \tag{18}$$

where,  $n$  is the number of i.i.d Gaussian random variables ( the components of the normalized distance vector) and  $\Gamma(x)$  is the gamma function. This is the Chi-Square distribution with  $n$  degrees of freedom.

The distribution of the MSE conditioned on  $H_1$  is given as

$$P_{H_1}(Z) = \frac{1}{2\sigma_z^2} \left(\frac{Z}{S^2}\right)^{\frac{(n-2)}{4}} e^{\frac{-(S^2+Z)}{2\sigma_z^2}} I_{\frac{n}{2}-1}\left(\sqrt{Z} \frac{S}{\sigma_z^2}\right) \tag{19}$$

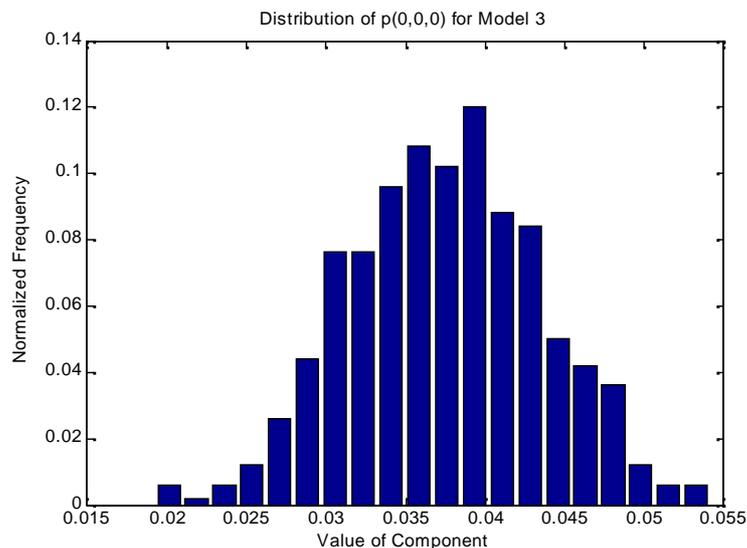
where  $I_\alpha(x)$  is the  $\alpha^{\text{th}}$ -order modified Bessel function of the first kind.

After some algebraic manipulations, we get the following log-likelihood ratio test

$$\frac{n-2}{4} \text{Log } Z - \text{Log } I_{\frac{n-1}{2}} \left( \sqrt{Z} \frac{S}{\sigma_z^2} \right) \stackrel{>}{<} \text{Log} \left[ 2^{\frac{n-2}{2}} \times \sigma_z^{n-2} \times S^{(2-n)/2} \times \Gamma\left(\frac{n}{2}\right) \right] - \frac{S^2}{2\sigma_z^2} = \tau_1 \quad (20)$$

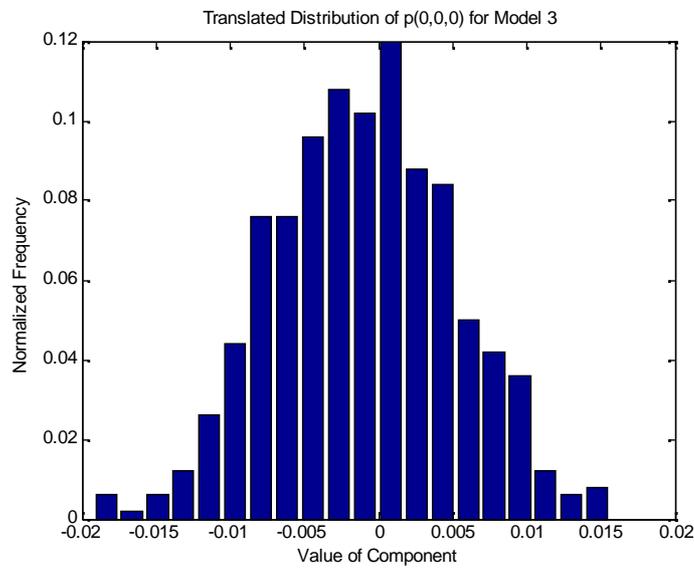
where we would choose  $H_0$  if the left hand side were greater than the threshold,  $\tau_1$ ,  $H_1$  if it were less than  $\tau_1$ , and either hypothesis could be chosen if there were equality between the two sides.

Several simulations were run to verify the many assumptions that have been made and the probability properties that have been used in the preceding pages regarding the distribution and independence of the eight components of the probability distribution vector. Figure 3 shows the distribution of  $p_1^{\text{model } 3}$ , that is the probability distribution of the first component  $p(0,0,0)$  for model 3. Note that the theoretical value for this model and component is 0.0385. Complete tables of the theoretical distribution vectors are provided in Appendix C.

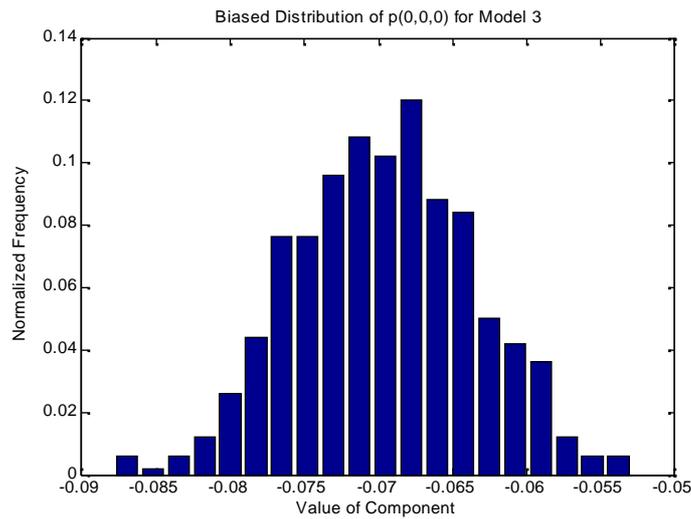


**Figure (3):** Normalized distribution of  $p(0,0,0)$

Figure 3 is the result of 500 independent runs using radar 3 as a test sequence with a 200-word length. In this case it has a mean of 0.0383 and a variance of  $3.7103 \times 10^{-5}$ . Figure 4 shows the average value of the same component after being translated by  $Y_1 = p_1 - \hat{p}_1$ . The translated random variable now has a mean of 0.0002 with the same variance as before ( $3.7103 \times 10^{-5}$ ). If we look at the translated distribution using a competing model (the theoretical vector for radar 5) we see the mean of the distribution as the difference between the theoretical components of the two models. Note that in Figure 5, the mean is -0.0688 which corresponds to the difference in the expected values of the first components for radar 3 and radar 5. More specifically,  $E[p_3(0,0,0)] = 0.0385$  and  $E[p_5(0,0,0)] = 0.1071$ . The difference is  $0.0385 - 0.1071 = -0.0686$ .



**Figure (4):** Translated distribution for proper model



**Figure (5):** Translated distribution for the competing model

Now, if we look at the transformation  $Z = \sum_{i=1}^n Y_i^2$  for the proper model (radar 3), we get the following plot. Recall that we are expecting a Chi-Square distribution with mean equal to  $n\sigma_y^2$  and a variance  $2n\sigma_y^4$

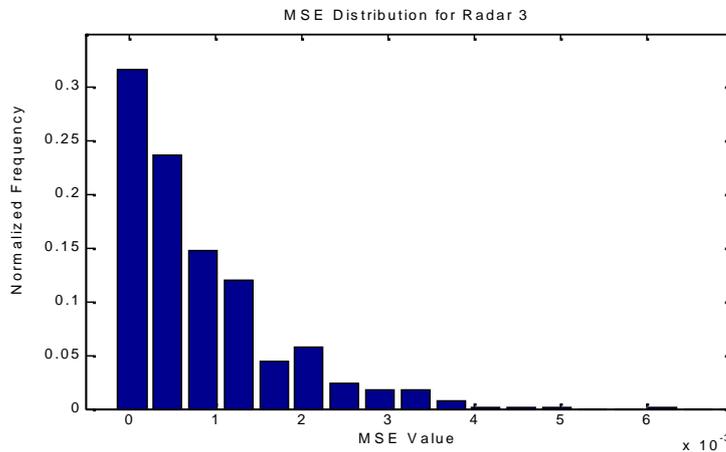


Figure (6): MSE Distribution for proper model

The calculated (output) values from this simulation for the mean and variance of Figure 6 are:

$$E[Z] = 1 \times 10^{-3}$$

$$Var[Z] = 9.4953 \times 10^{-7}$$

Assuming that all eight components are identically distributed (same variance) we get the following output mean and variance for Z.

$\sigma_y^2 = Var[p_3(0,0,0)] = 3.7103 \times 10^{-5}$  as was shown before which implies that we get for Z

$$E[Z] = n\sigma_y^2 = 8(3.7103 \times 10^{-5}) = 2.9682 \times 10^{-4}$$

$$\sigma_z^2 = 2n\sigma_y^4 = 16(3.7103 \times 10^{-5})^2 = 2.2026 \times 10^{-8}$$

If, instead, we use the output variances for each of the eight components we get the following:

$$\sigma_z^2 = \sum_{i=1}^8 \sigma_i^2 = \left( 0.0371 + 0.0930 + 0.2866 + 0.0995 + 0.0908 + 0.2736 \right. \\ \left. + 0.1004 + 0.0644 \right) \times 10^{-3} = 0.001 \text{ So}$$

that now we get for the mean and the variance of Z

$$E[Z] = \sum_{i=1}^8 \sigma_i^2 = \sigma^2 = 1 \times 10^{-3}$$

$$Var[Z] = 2 \sum_{i=1}^8 (\sigma_i^2)^2 = 2\sigma^4 = 2 \times 1.9939 \times 10^{-7} = 3.9879 \times 10^{-7}$$

which are much closer to the actual output values of 0.001 and  $9.4953 \times 10^{-7}$  for the mean and variance respectively. This demonstrates that the eight components are not identically distributed, but do have, however, very similar variances. Although a second simulation will not produce the exact same results (due to the random nature of signals), it would still yield similar conclusions.

Figure 7 shows the distribution of the MSE for a competing model which, in this case, is radar1. The theoretical bias between models 1 and 3 is calculated to be

0.0286.

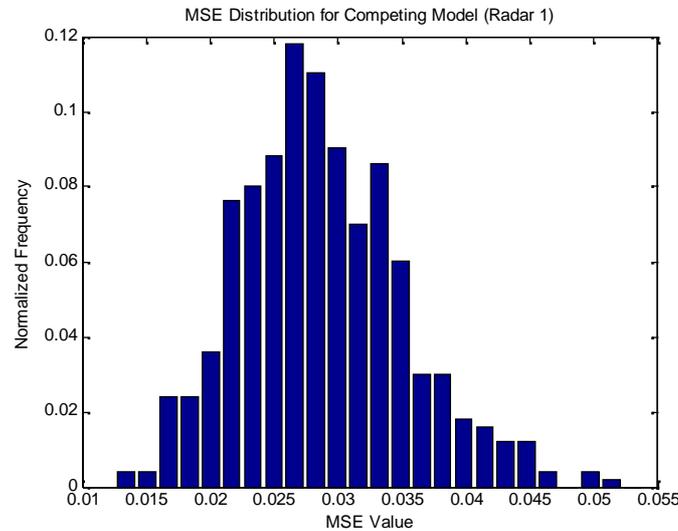


Figure (7): MSE Distribution for competing model

The general conclusion is that it is possible to derive an analytical expression for the probability of error (false recognition) between two pseudo-random sources one being the correct model and one being a competitor-based only on a *priori* PRI statistics of each source.

**2.3 False Recognition Probability Example:**

Figure 8 shows the theoretical MSE distribution of both central and non-central Chi-Square distribution with  $n=8$  degrees of freedom where the non-centrality parameter,  $S^2$ , is equal to 2 and the variance,  $\sigma^2$  is equal to 0.25.

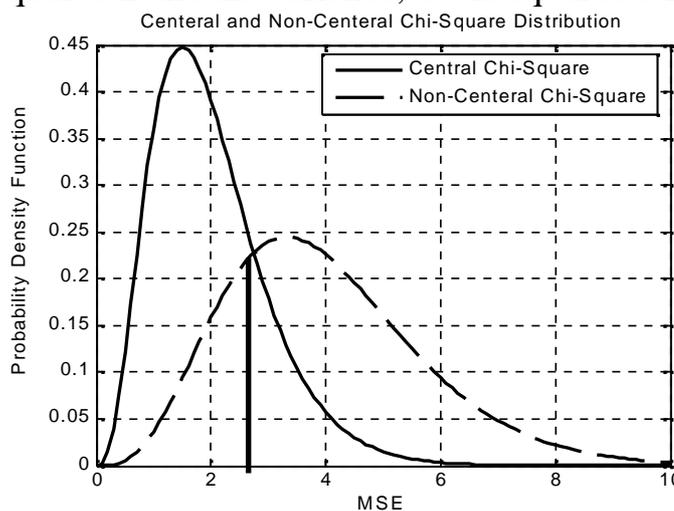


Figure (8): Central and Non-Central Chi-Square probability distribution

Figure 9 shows the log likelihood function,  $\log \Lambda(Z)$ . Note that the zero-crossing in Figure 9 occurs at the same MSE value when the two curves intersect in Figure 8.

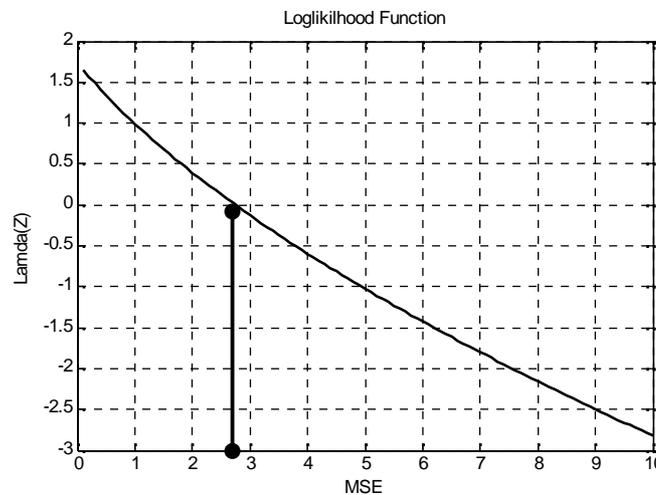


Figure (9) : Log-Likelihood function

Figure 10 shows how the probability of the false recognition (choosing  $H_1$  when  $H_0$  is true) increases as the variance (and, accordingly the MSE) increases. The probability of false recognition can be described mathematically as

$$P(\text{choose } H_1 | H_0 \text{ is true}) = \int_{\tau}^{\infty} P_{H_0}(Z) dz \tag{21}$$

where  $\tau$  is chosen as the MSE value when the log likelihood function crosses zero. The false recognition probability was calculated using MATLAB where equations 18 and 19 were set to equal each other in order to determine the threshold and equation 21 was used for the actual probability calculations.

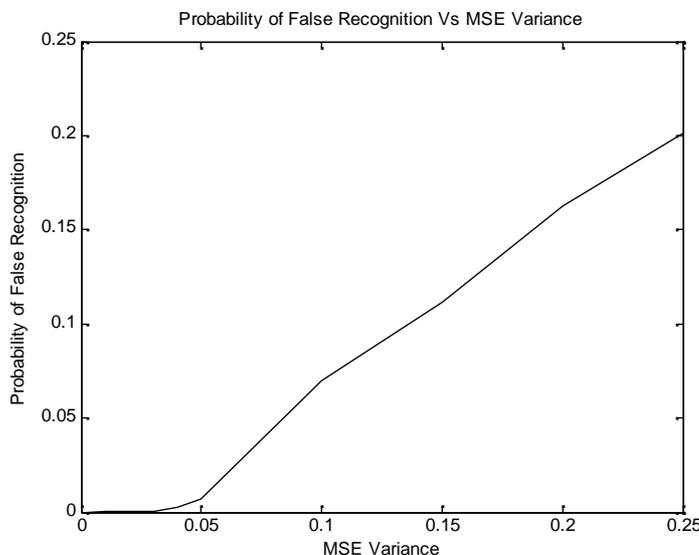
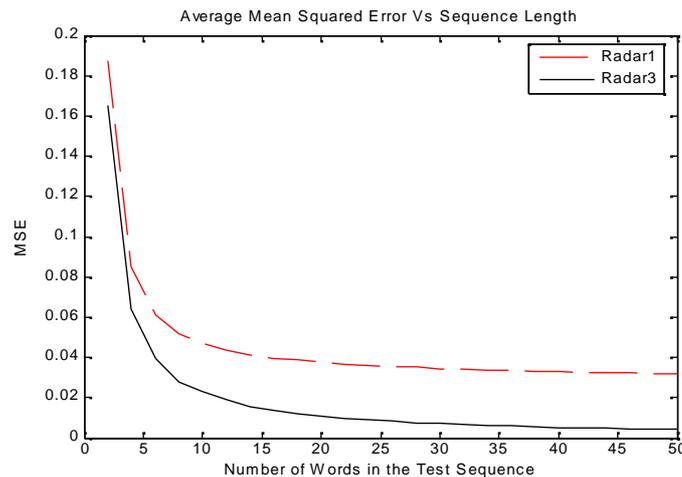


Figure (10): Probability of false recognition

As would be expected, the false recognition rate goes up as the variance in our MSE estimate increases. The variance increases as fewer samples (words) are received which explains why short sequences produce high false recognition rates than do long ones.

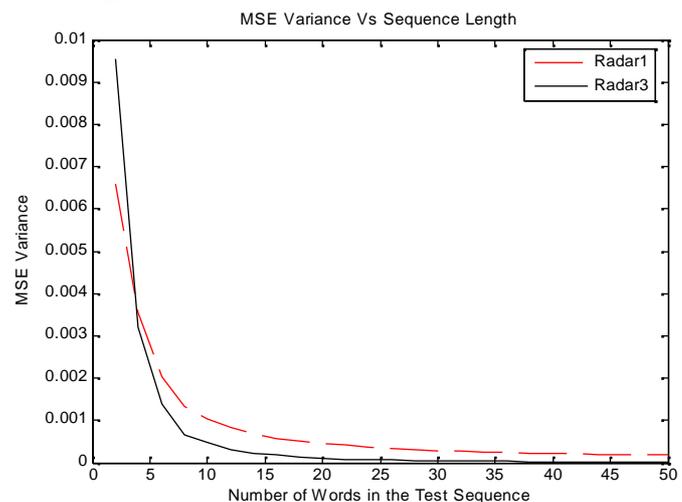
**2.4 Simulation Results:**

This subsection compares the theoretical probability of false recognition to actual simulation results. The simulation uses radar 3 as the correct model and radar 1 as a competitor. The theoretical non-centrality parameter,  $S^2 = \sum_{k=1}^{\infty} \beta_k^2$  between these radars is 0.0286 as indicated in Appendix C. Figure 11 illustrates how the MSE converges to 0 for the proper model (radar 3) and towards the non-centrality parameter for radar 1 as the observation sequence increases in length.



**Figure (11): MSE VS Test Sequence Length (Radar 3 Being the Correct Model)**

Figure 12 shows how the variance of the MSE estimate decreases as a function of the observation sequence length.



**Figure (12): MSE variance Vs test sequence length.(radar 3 being the correct model)**

Figure 13 shows the relationship between MSE estimate of the proper model and the variance of this estimate. This plot is generated from 500 different test sequences

ranging in length from 1 to 50 words. The MSE values shown are those between the received and theoretical vectors from radar 3. They are the average values for the MSE calculated from each of the 500 test sequences of a given length while the MSE variance is the variance of all the values. This plot, in essence combines Figure 11 and Figure 12. Recall that, theoretically, the MSE should equal 8 times the MSE variance (equation 5)

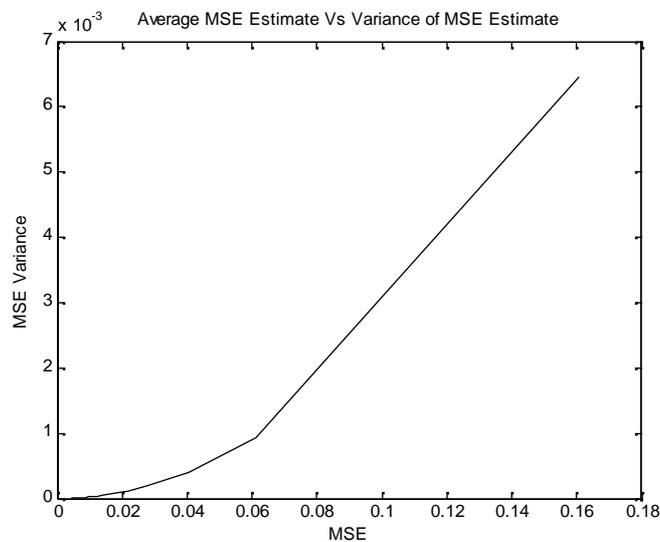


Figure (13): MSE Vs variance of estimate

Figure 14 compares the theoretical calculation for the false recognition probability to the simulation results. The actual false recognition probability is calculated by counting the number of times the MSE between the received and correct probability vectors is not the smallest (a competing model has a smaller MSE) as a function of the observation sequence length, and, correspondingly, the MSE variance (equation 8). The theoretical curve is computed using MATLAB in a similar manner to Figure 10

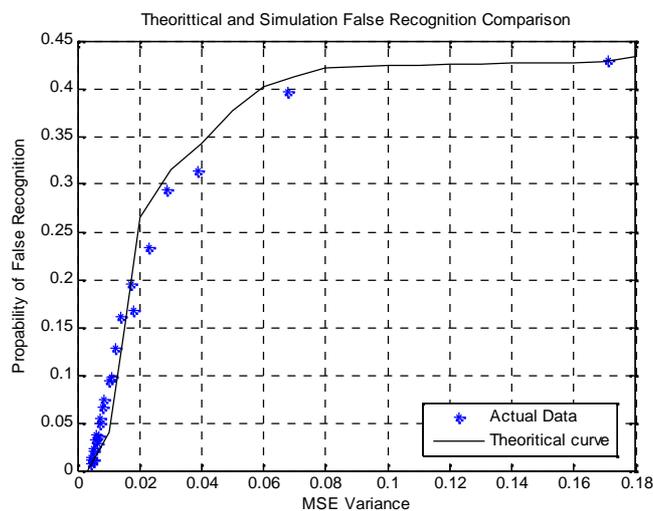


Figure (14): Theoretical and Simulation Comparison of the Probability of False Recognition.

Although the theoretical curve is only a coarse approximation for the actual false recognition probabilities, it does still offer an insight into the relationship between the MSE and false recognition rate. The curve could probably seem to more closely resemble the actual data by lowering the value of  $S^2$ , but 0.0286 is already very close to the lower bound ability of the MATLAB program to solve the equation  $\int_{\tau}^{\infty} P_{H_0}(Z) dz$  since such small variances are involved this could, in turn, lead to round-off error in the calculations.

**3. Gaussian Distribution of HMM Recognition Probability:**

Table 2 shows the actual average recognition probability of 8 HMMs for 1500 different test sequences each containing 20 words. The rows correspond to the radar that was used to generate the observation sequence while the columns indicate the recognition probability of each HMM for that sequence. Note that HMM<sub>1</sub> is the trained model for radar 1, HMM<sub>2</sub> is the trained model for radar 2 and so on.

**Table 2: Log Recognition Probability for 8 Radars**

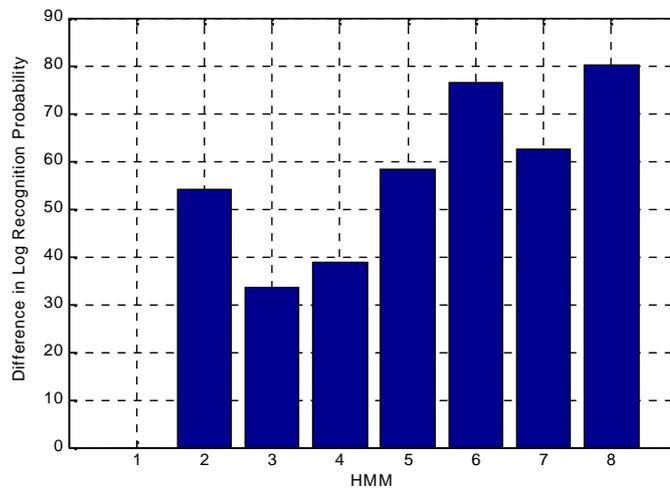
Threat Library								
Sequence source	HMM <sub>1</sub>	HMM <sub>2</sub>	HMM <sub>3</sub>	HMM <sub>4</sub>	HMM <sub>5</sub>	HMM <sub>6</sub>	HMM <sub>7</sub>	HMM <sub>8</sub>
Radar 1	-11.2406	-65.2379	-44.8571	-50.0822	-69.4772	-87.7990	-73.6274	-91.4689
Radar 2	-51.8197	-12.0434	-56.5937	-72.6261	-71.1339	-86.6473	-59.8420	-123.510
Radar 3	-39.7264	-77.3997	-10.0861	-56.8642	-61.8544	-79.1460	-74.4219	-106.045
Radar 4	-11.8935	-56.2459	-35.0319	-6.0227	-49.6079	-56.3802	-50.9857	-60.5805
Radar 5	-87.4821	-160.722	-87.2430	-114.021	-12.0473	-161.194	-147.275	-192.637
Radar 6	-128.781	-154.973	-105.834	-122.378	-75.4119	-5.5229	-33.9215	-68.5426
Radar 7	-74.6023	-99.0511	-77.1047	-69.0025	-61.3744	-55.1038	-10.0957	-71.6117
Radar 8	-184.468	-186.212	-178.309	-175.970	-171.123	-86.8578	-84.7649	-13.5782

Notice how highest recognition probability per row occurs when the HMM used is the same as the radar source generating the sequence. The table, however, is not symmetric. As an example, the recognition probability (-44.8571) is computed by HMM<sub>3</sub> with radar 1 generating the sequence is not the same as the recognition probability (-39.7264) is computed by HMM<sub>1</sub> for radar 3. In other words, HMM<sub>1</sub> is more likely to falsely classify a signal from radar 3 than HMM<sub>3</sub> is likely to incorrectly classify a signal from radar 1. Table 3 shows the relative difference in recognition probabilities between all 8 competing models.

**Table 3: Relative log recognition probabilities**

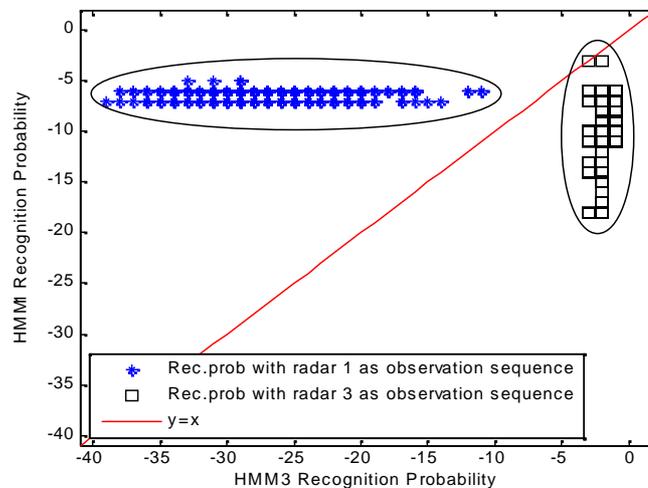
Threat Library								
Sequence source	HMM <sub>1</sub>	HMM <sub>2</sub>	HMM <sub>3</sub>	HMM <sub>4</sub>	HMM <sub>5</sub>	HMM <sub>6</sub>	HMM <sub>7</sub>	HMM <sub>8</sub>
Radar 1	0	53.9973	33.6165	38.8416	58.2366	76.5584	62.3868	80.2283
Radar 2	39.7763	0	44.5503	60.5827	59.0905	74.6039	47.7986	111.4667
Radar 3	29.6403	67.3136	0	46.7781	51.7683	69.0599	64.3358	95.9589
Radar 4	5.8708	50.2232	29.0092	0	43.5852	50.3575	44.9636	54.5578
Radar 5	75.4348	148.675	75.1957	101.974	0	149.147	135.228	180.590
Radar 6	123.2581	149.450	100.311	116.855	69.8890	0	28.3986	63.0197
Radar 7	64.5066	88.9554	67.0090	58.9068	51.2787	45.0081	0	61.5160
Radar 8	170.8898	172.634	164.731	162.392	157.545	73.2796	71.1867	0

Figure 15 provides a graphical representation of the first row in Table 5.3, that is, the relative difference in the recognition probabilities between the correct and competing models with radar 1 as the test sequences generator. Obviously, the smaller the relative difference the greater the likelihood that a false recognition could occur.



**Figure (15): Relative Log Recognition Scores**

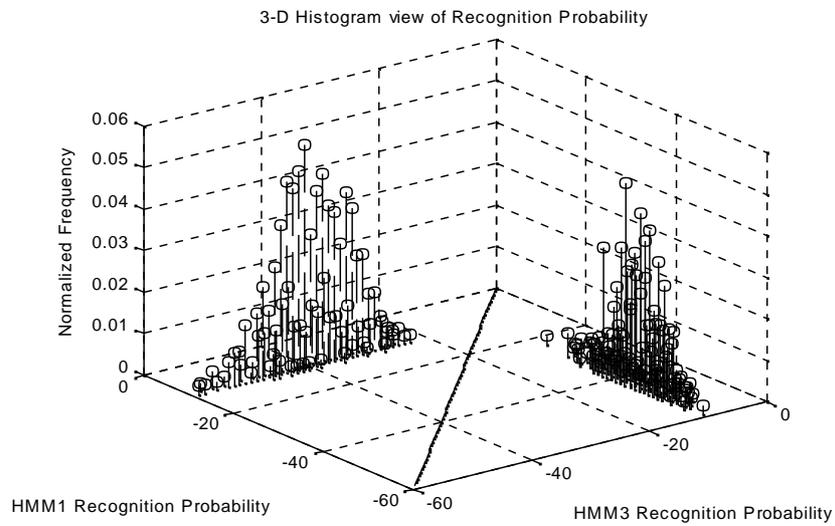
Figure 16 uses radar 1 and radar 3 as the generators of 500, 10-word length test sequences and displays the actual recognition probabilities for HMM<sub>1</sub> (shown as stars) and HMM<sub>3</sub> (shown as squares) as coordinates in the x-y plane. The x-axis represents the recognition probability of HMM<sub>1</sub> and the y-axis is the recognition probability of HMM<sub>3</sub>. Notice how, on the whole, both HMMs correctly classify the observation sequences. Those sequences that are incorrectly classified lie on the wrong side of the decision boundary line  $y = x$  (the maximum-likelihood (ML) threshold) which is drawn as a reference to more easily see when false recognition occurs.



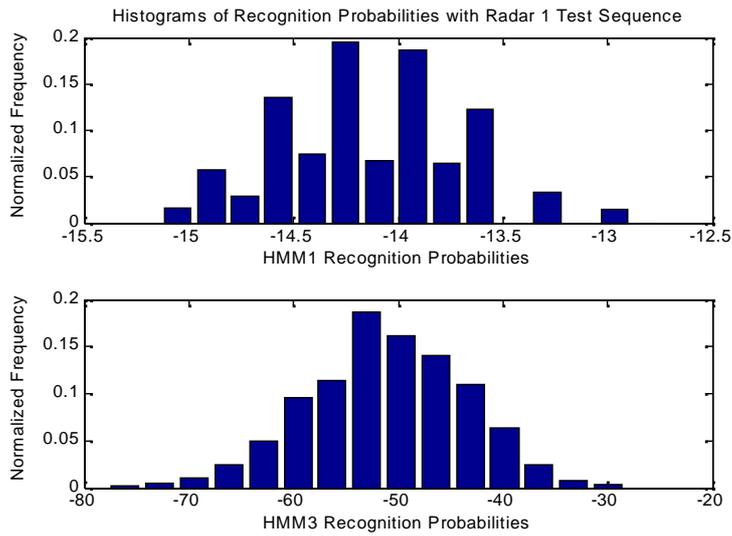
**Figure (16):** Recognition probabilities for HMM 1 and 3

Recall from Table 2 that  $HMM_1$  is more likely to incorrectly classify a signal from radar 3 than  $HMM_3$  could classify a signal from radar 1. It is seen in Figure 16 that some squares are lying above the decision line, but no stars lying below it. The scatter patterns, for each HMM, have also been elliptically drawn to show the general distribution of the recognition probabilities. The area between the two ellipses when they intersect each other has a closed form expression and represents a 'reject' region. A recognition point  $(P(O|\lambda^{(3)}), P(O|\lambda^{(1)}))$  lying in the intersection of the two ellipses if exists would be unclassifiable due to a lack of sufficient information in the observation sequence.

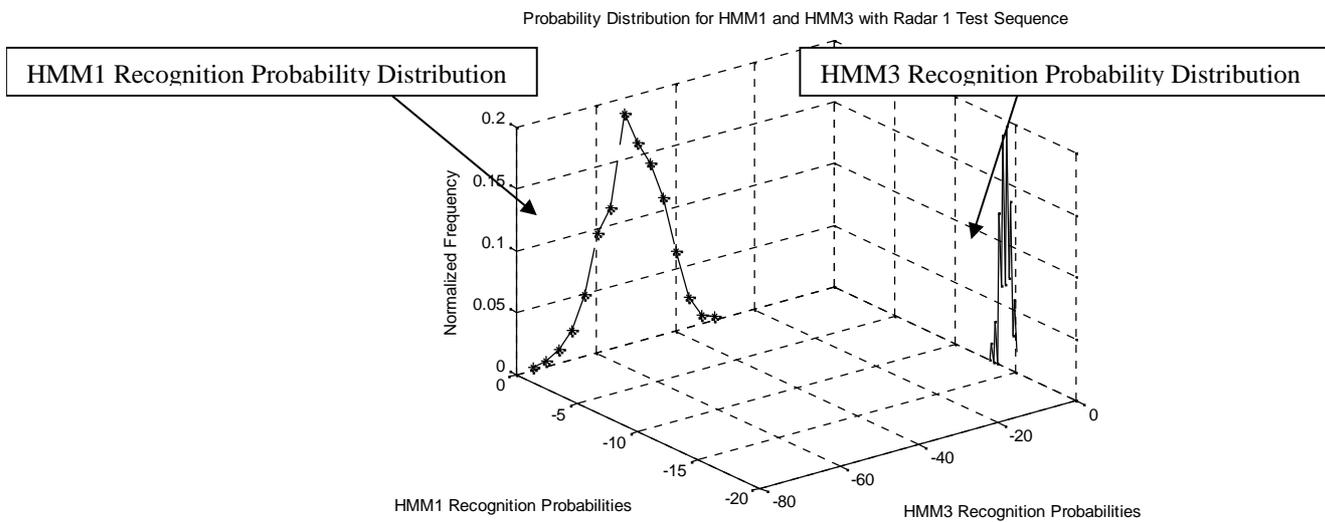
Figure 17 gives a graphical 3-dimensional histogram view of Figure 16, The observation sequences used in this simulation are 25-words long, which explains why there are no false recognition points as compared to Figure 16 where the sequences were only 10-words long. It is interesting to note that the histograms provide a jointly Gaussian distribution for the two scatter patterns. This is even more evident in Figure 18 where the distributions of the radar 1 and 3 test sequence patterns are shown for each axis respectively. Figure 19 gives a spatial view of these distributions.



**Figure (17): 3-D Histogram of recognition probabilities**



**Figure (18): Radar 1 test sequence recognition probabilities**



**Figure (19): 3-D View of radar 1 test sequence recognition probabilities**

It is evident from the previous figures that the distribution of the recognition probabilities for both scatter patterns in Figure 16 closely resembles a jointly log-normal distribution in the HMM<sub>1</sub> and HMM<sub>3</sub> axes. In general, an L-dimensional recognition probability space (probability space with L competing HMMs) could be described by the multivariate Gaussian density function. Denote each of the Gaussian random variables (recognition probabilities)  $P(O|\lambda^{(l)}), l=1,2,\dots,L$  as  $\theta = [\theta_1, \theta_2, \dots, \theta_L]^T$  with mean vector  $\mu = [\mu_1, \mu_2, \dots, \mu_L]^T$  and the covariance matrix  $\Sigma_\theta = E[\theta^T \theta]$  where T represents the matrix transpose function. The Gaussian PDF is described by [8] as

$$P_\theta(\theta) = \frac{1}{(2\pi)^{L/2} |\Sigma_\theta|^{1/2}} \exp\left[-\frac{1}{2}(\theta - \mu)\Sigma_\theta^{-1}(\theta - \mu)^T\right] \quad (22)$$

In the special 2-dimensional (L=2) case described in the preceding sections, the distributions of the scatter patterns in Figure 16 would follow a bivariate Gaussian density function with covariance matrix

$$\Sigma_\theta = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad ; |\rho| < 1 \quad (23)$$

where  $\rho$  is the correlation coefficient defined as

$$\rho_{12} = \rho_{21} = \rho = \frac{E[\theta_1\theta_2]}{\sqrt{\sigma_1^2\sigma_2^2}} \quad (24)$$

where  $\sigma_1^2$  and  $\sigma_2^2$  are the variance of  $\theta_1$  and  $\theta_2$  respectively. The determinant of the covariance matrix is

$$|\Sigma_\theta| = \sigma_1^2\sigma_2^2(1 - \rho^2) \quad (25)$$

and the matrix inverse

$$\Sigma_\theta^{-1} = \frac{1}{|\Sigma_\theta|} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \quad (26)$$

In order to see if equation 22 does, in fact, accurately model the data points distribution, another simulation was run using radars 7 and 8 as the test sequence generators. 1500 different sequences were used and the recognition probabilities are plotted in Figure 20 (analogous to Figure 16).

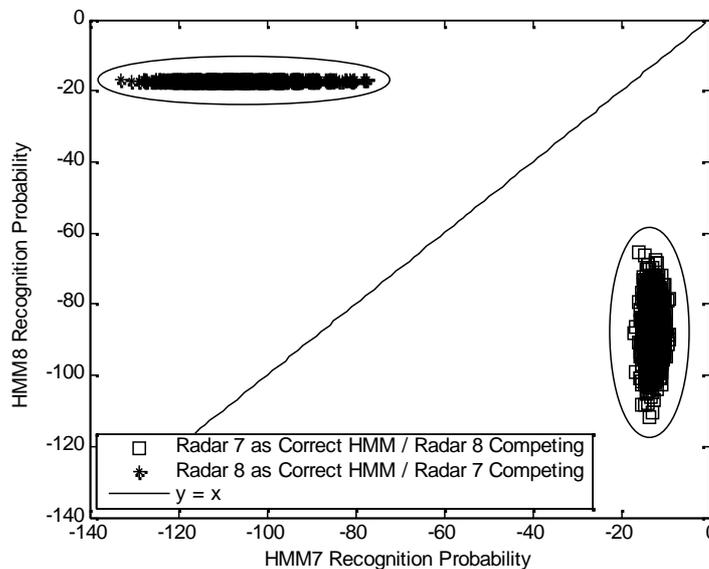


Figure (20) : Recognition Probability for HMMs 7, 8

**4. Decision Threshold with Unknown Signals:**

The decision threshold shown in the preceding sections assumes that the only two possible signals that can be received originate from either radar 1 or radar 3. If a recognition point lies above the threshold then the classification output will be for radar 1; if it lies below the threshold then the output will be in favour of radar 3. What would happen, however, if the received signals were from an unknown source—neither radar 1 nor radar 3? Both HMMs would still compute recognition probabilities (one being larger than the other) which would imply that the signal would be erroneously classified since it do not belong to either. Instead, it would be highly desirable for the receiver to say that it can not recognize the signal because it has no representative HMM in the threat library. That is, how small must the recognition probabilities computed by HMMs 1 and 3 be before the receiver rejects the observation sequence? Recall from Section 4 that the intersection of the two ellipses in Figure 16 could be a region in which the receiver would say that it does not have enough information (too short test sequence) to decide from which radar the received signal is originated. In a similar manner, if a recognition point fall outside the ellipse, it would make sense to decide that the observation sequence was generated from unknown source. Figure 21 shows the scatter patterns for HMMs 1 and 3 with three different radar sources: radar 1, radar 3 and unknown radar.

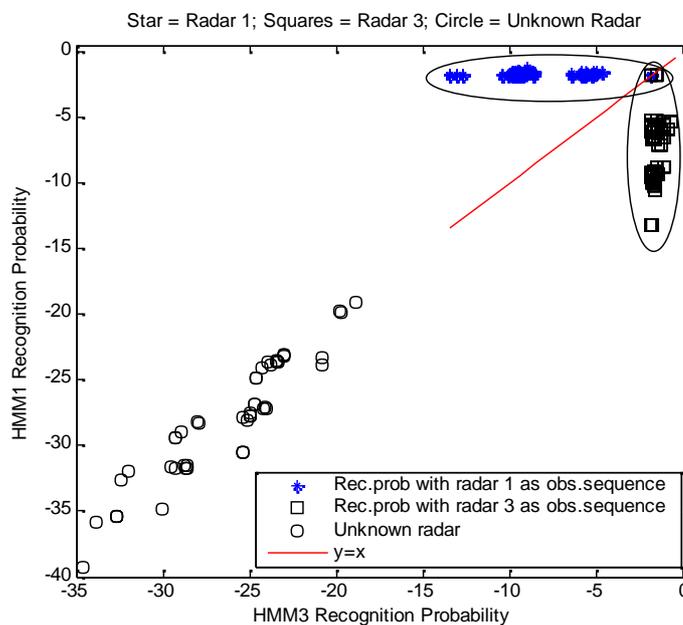


Figure (21): Scatter patterns with unknown radars

Since the unknown radar's scatter pattern lies below the line,  $y = x$ , decision threshold, it would be incorrectly classified as radar 3 if there were no reject option for HMM<sub>3</sub>. Because in a practical situation it would not be possible to construct such a scatter plot for a foreign signal since the PRI statistics are unknown, a threshold must be constructed based on our confidence of the positions of the recognition points for the known signals. The probabilities of false recognition of unknown signal and missed detection of known signal would both depend on our confidence level. Have to rely solely on the known statistics of the scatter patterns for the HMMs in the threat library. A preliminary approach would be to use the ellipse in Figure 21 as a foreign signal reject threshold; any recognition point lying outside the ellipse would be rejected. A typical scatter plot such as that one shown in Figure 21 would therefore consist of two reject regions. The first would be the region where we decide that the observation sequence is not long enough to classify and can be described by the intersection of the two ellipses if exist. The second region would be that area lying outside each ellipse in order to avoid incorrect classification of unknown signals.

## 5. Conclusions:

This paper proposes a method for determining pre-assigned threshold settings and reject regions for the competing HMMs. The proposed method used the concept of distance vectors which relied purely on a prior knowledge of a radar's PRI statistics. This was a template-matching classification system where the observation vector was compared to theoretical vectors and a correlation made according to the mean squared error. It was shown that the MSE between a theoretical probability vector and a

received vector is distributed according to a central Chi-square distribution if the two vectors are from the same source. Otherwise, the MSE of a competing theoretical vector will follow a non-central Chi-square distribution. Simulations showed that the actual false recognition rate as a function of the MSE variance only coarsely followed the theoretical curve which is most likely due to the fact that even the correct model's distance vector contains a small bias term. This would imply that it is not a perfect central Chi-square random variable and, accordingly, yield distributions of what more closely resemble each other thus causing the false recognition rate to rise. Additionally, since the variances involved are so small, there may be some round-off error in the calculations. Nevertheless, the distance vector concept can still at least provide a qualitative indication about how correlated two different radars are without the use of HMMs and how likely a radar will be mistaken for another.

### **References:**

- [1] Hossam E. Abou-Bakr Hassan, Khairy Elbarbay, Mohamed Asaad Abd-Elrazek "Identification of Pulsed Radar using Hidden Markov Model" 12<sup>th</sup> International Conference Military Technical College May, 2007.
- [2] Hossam E. Abou-Bakr Hassan, Ahmad El-Mahdy, Mohamed Asaad Abd-Elrazek "Performance Evaluation of Pulsed Radar Identification using Hidden Markov Model" 12<sup>th</sup> International Conference Military Technical College May, 2007.
- [3] Rabiner, L.R., "A tutorial on hidden Markov models and selected applications in speech recognition", Proc. IEEE, Vol. 77, No. 2, pp. 257-285, February, 1989.
- [4] Junkowitsch, J., Neubauer, L., Kuske, G., "A new keyword spotting algorithm with pre-calculated optimal thresholds", Proc. 4<sup>th</sup> Int. Con. on Spoken Language Processing, Philadelphia, PA., 3-6, October, 1996.
- [5] John M. Wozencraft, Irwin Mark Jacobs, "Principles of Communication Engineering", John Wiley and Sons, Inc., 1965, pp. 170—171.
- [6] John G. Proakis, "Digital Communications", McGraw-Hill, Inc., New York, 2001, pp. 41-44.
- [7] Merrill Skolnik, "Radar Handbook Second Edition", McGraw-Hill, Inc., 1990, p.17.20.
- [8] Juang, B.-H. and Rabiner, L., "A probabilistic measure for hidden Markov models", AT&T Technical Journal., Vol. 64, No. 2, pp. 391-409, February 1985.

APPENDIX A

Table 1 : PSEUDO-RANDOM RADARS

P-R Radar 1		P-R Radar 2		P-R Radar 3		P-R Radar 4	
Word	P(occur)	Word	P(occur)	Word	P(occur)	Word	P(occur)
0 1	0.25	0 0 0 1	0.25	1 0 1 0	0.25	0 1	0.5
0 0 1	0.25	0 1 0 1	0.25	1 1 0 1	0.25	0 0 1	0.5
0 1 1	0.25	0 1 1 0	0.25	0 0 1	0.25		
1 0 0 1	0.25	0 1 1 1	0.25	1 0	0.25		
P-R Radar 5		P-R Radar 6		P-R Radar 7		P-R Radar 8	
Word	P(occur)	Word	P(occur)	Word	P(occur)	Word	P(occur)
0 0 0 0 1 0 1	0.25	1 0 0 0 0 0 0		0 0 0 1 0 1	0.25	1 0 0 0 0 0	0.2
0 1 0 1	0.25	1 0 1 0 1 0		0 1 0 0 0 0 0	0.25	1 0 0 0 0 0 0	0.2
1 1 0 1 1 0 1 0 1	0.25			0 0 0 0 0 0 0 0 0 1	0.25	1 0 0 0 0 0 0 0 0	0.2
0 0 0 1 1 0 0 1	0.25			0 1	0.25	1 0 0 0 0 0 0 0 0 0	0.2
						1 0 0 0 0 0 0 0 0 0 0	0.2

APPENDIX B : DISTANCE VECTOR INFORMATION

Table 2. The 4 And 8 Components Distance Vector for Each Of The 8 Pseudo-Random Radars

Sequence generator	4-component Distance vector				8-component Distance vector							
	P-R Radar1	0.5	1	1	0.5	0	0.5	0.5625	0.4375	0.5	0.5	0.4375
P-R Radar2	0.75	1.25	1.25	0.75	0.3125	0.4375	0.75	0.5	0.4375	0.8125	0.5	0.25
P-R Radar3	0.375	1.125	1.125	0.625	0.125	0.25	0.625	0.5	0.25	0.875	0.5	0.125
P-R Radar4	0.5	1	1	0	0	0.5	1	0	0.5	0.5	0	0
P-R Radar5	1.5	2.25	2.25	1	0.75	0.75	1.5	0.75	0.75	1.5	0.75	0.25
P-R Radar6	3.333	1.667	1.667	0	2.6667	0.667	1.667	0	0.6667	1	0	0
P-R Radar7	2	1.25	1.25	0	1.5468	0.4531	1.25	0	0.4531	0.7969	0	0
P-R Radar8	6	1	1	0	5	1	1	0	1	0	0	0

The 4-component distance vector for radar I represents the following expected 2-symbol transitions:

$$d_i^4 = [E[0,0], E[0,1]E[1,0], E[1,1]] \tag{1}$$

The 8-component distance vector represents the following expected 3-symbols transitions:

$$d_i^8 = [E[0,0,0], E[0,0,1]E[0,1,0], E[0,1,1], E[1,0,0], E[1,0,1], E[1,1,0], E[1,1,1]] \tag{2}$$

Table 3. Euclidean Distance Chart Calculated from 2-Symbol Transition Distance Vector

	P-R Radar1	P-R Radar2	P-R Radar3	P-R Radar4	P-R Radar5	P-R Radar6	P-R Radar7	P-R Radar8
P-R Radar1	0	0.5000	0.2500	0.5000	2.0917	3.0275	1.6202	5.5227
P-R Radar2	0.5000	0	0.4330	0.8660	1.6202	2.7536	1.6202	5.5227
P-R Radar3	0.2500	0.4330	0	0.6614	1.9843	3.1190	1.7500	5.6624
P-R Radar4	0.5000	0.8660	0.6614	0	2.2638	2.9859	1.5411	5.5000
P-R Radar5	2.0917	1.6202	1.9843	2.2638	0	2.2449	1.8028	4.9371
P-R Radar6	3.0275	2.7536	3.1190	2.9859	2.2449	0	1.4576	2.8289
P-R Radar7	1.6202	1.6202	1.7500	1.5411	1.8028	1.4576	0	4.0156
P-R Radar8	5.5227	5.5227	5.6624	5.5000	4.9371	2.8289	4.0156	0

**Appendix C : Probability vector information**

**Table 4. The 8-Component Theoretical Probability Vectors For Each Of The Eight Radars.**

	P(0,0,0)	P(0,0,1)	P(0,1,0)	P(0,1,1)	P(1,0,0)	P(1,0,1)	P(1,1,0)	P(1,1,1)
P-R Radar1	0	0.1667	0.1875	0.1458	0.1667	0.1667	0.1458	0.0208
P-R Radar2	0.0781	0.1094	0.1875	0.1250	0.1094	0.2031	0.1250	0.0625
P-R Radar3	0.0385	0.0769	0.1923	0.1538	0.0769	0.2692	0.1538	0.0385
P-R Radar4	0	0.2000	0.4000	0	0.2000	0.2000	0	0
P-R Radar5	0.1071	0.1071	0.2143	0.1071	0.1071	0.2143	0.1071	0.0357
P-R Radar6	0.4000	0.1000	0.2500	0	0.1000	0.1500	0	0
P-R Radar7	0.3438	0.1007	0.2778	0	0.1007	0.1771	0	0
P-R Radar8	0.6250	0.1250	0.1250	0	0.1250	0	0	0

For example,  $P(0,0,0) = \frac{d(1)}{\sum_{i=1}^8 d(i)}$ , where  $d$  is the distance vector