SENSITIVITY OF NEURAL NETWORKS WHICH APPROXIMATE THE NEYMAN-PEARSON DETECTOR TO THRESHOLD VARIATIONS

P. Jarabo-Amores , R. Gil-Pita, M. Rosa-Zurera, F. López-Ferreras

Departamento de Teoría de la Señal y Comunicaciones,
Escuela Politécnica Superior, Universidad de Alcalá
Ctra. Madrid-Barcelona, km. 33.600, 28805, Alcalá de Henares - Madrid (SPAIN)
phone: + (34) 91 885 67 43, fax: + (34) 91 885 66 90, email: mpilar.jarabo@uah.es
web: www2.uah.es/teose/

ABSTRACT

The application of adaptive systems trained in a supervised manner to approximate the Neyman-Pearson detector is considered. The general expression of the function approximated when using the LMSE criterion is calculated. To evaluate the sensitivity of the decision rule based on this function to threshold variations, a novel strategy is proposed based on the calculus of the partial derivative of the probabilities of detection and false alarm with respect to detection threshold. Results allow us to explain the dependence of the decision rule performance on design parameters such as the prior probabilities, the desired outputs and the signal to noise ratio selected for training (TSNR). In previous works based on a trial and error strategy, TSNR has appeared as a critical parameter, but until now, no effort had been made to explain it. As an example, the detection of gaussian signals in gaussian interference is considered.

1. INTRODUCTION

In this paper the application of adaptive systems trained in a supervised manner to approximate the Neyman-Pearson detector is considered. This detector maximizes the probability of detection (P_D), while maintaining the probability of false alarm (P_{FA}) lower than or equal to a given value. The characteristics of such a detector are reflected in its ROC (Receiver Operating Characteristic) curve, that relates P_D to P_{FA} [1].

Although Ruck et al. [2], and Wan [3], demonstrated that a neural network can be used to approximate the optimum bayessian classifier when trained using the least mean squared-error (LMSE) criterion, and extended this result to any adaptive system trained using this error criterion, they did not consider the problem of approximating the Neyman-Pearson detector.

In previous works, neural networks were proposed as a solution for detecting radar echoes in different environments [4, 5, 6]. Results highlighted a strong dependence of the neural network-based detector performance on the signal to noise ratio selected for training (TSNR). They also observed that the detection capabilities and the influence of the TSNR depended on the desired P_{FA}. Recently, some attempts to reduce the dependence of the neural detector performance on TSNR have been carried out [7], based on the use of a complex pre-processing stage that reduces this dependence at the expense of a high computational cost. Nevertheless, no effort has been made to explain the reasons of such a dependence.

This work has been supported by the "Consejería de Educación de la Comunidad de Madrid" (SPAIN), under Project 07T/0036/2003 1

In relation to the capability of an adaptive system trained using the LMSE criterion to approximate the Neyman-Pearson detector, a novel method is proposed in this paper, based on the calculus of the expression approximated by the system during training. On the other hand, we use the function approximated by the system to study the sensitivity of the detection rule based on it to approximation errors. This novel approach not only allows us to explain the influence of TSNR on detector performance, but also the influence of other design parameters, such us the prior probabilities and the desired outputs.

2. THE APPROXIMATED DISCRIMINANT FUNCTION AND THE DECISION RULE

D. W. Ruck et al. [2] demonstrated that a multilayer perceptron (MLP) converges to a mean squared-error approximation of the Bayes optimal discriminant function, when trained using the LMSE criterion. They study two-class and multiclass problems. For binary detection and desired outputs $\{-1,1\}$ (1 for input vectors from class H_1 and -1 for input vectors from class H_0), they proved that the neural network output approximates the Bayes optimal discriminant function $g_0(\mathbf{z}) = P(H_1|\mathbf{z}) - P(H_0|\mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^n$ is the feature vector, and $P(H_1|\mathbf{z})$ and $P(H_0|\mathbf{z})$ are the posterior probabilities of the classes.

In a more general problem, if the network is trained to produce t_{H_1} when the feature vector is from class H_1 and t_{H_0} when the feature vector is from class H_0, the mean squared-error between the network output, $F(\mathbf{z})$, and the desired outputs is given by $E_m = \lim_{N\to\infty}(E_s/N)$, where:

$$E_s = \Big[\sum_{\mathbf{z}\in Z_1}(F(\mathbf{z})-t_{H_1})^2 + \sum_{\mathbf{z}\in Z_0}(F(\mathbf{z})-t_{H_0})^2\Big] \qquad (1)$$

E_s/N is the sample mean squared-error calculated for a set of N pre-classified feature vectors; $Z_1 \subset Z \subseteq \mathbb{R}^n$ and $Z_0 \subset Z \subseteq \mathbb{R}^n$ are the regions where H_1 and H_0 are chosen, respectively ($Z_0 \cup Z_1 = Z$, $Z_0 \cap Z_1 = \varnothing$, Z being the input space). If the training set represents a reasonable approximation to the input space, E_m will be minimized when the network is trained to minimize E_s.

Using the Strong Law of Large Numbers, E_m can be expressed as:

$$E_m = P(H_1)\int_Z (F(\mathbf{z})-t_{H_1})^2 f(\mathbf{z}|H_1)d\mathbf{z}$$
$$+P(H_0)\int_Z (F(\mathbf{z})-t_{H_0})^2 f(\mathbf{z}|H_0)d\mathbf{z} \qquad (2)$$

The function $F(\mathbf{z})$ that minimizes E_m, which will be denoted as $F_0(\mathbf{z})$, is given in (3). It is obtained by simple derivation of E_m with respect to $F(\mathbf{z})$.

$$F_0(\mathbf{z}) = \frac{P(H_1)f(\mathbf{z}|H_1)t_{H_1} + P(H_0)f(\mathbf{z}|H_0)t_{H_0}}{P(H_1)f(\mathbf{z}|H_1) + P(H_0)f(\mathbf{z}|H_0)} \tag{3}$$

Taking into consideration that $f(\mathbf{z}) = P(H_1)f(\mathbf{z}|H_1) + P(H_0)f(\mathbf{z}|H_0)$, the following expression fulfills:

$$F_0(\mathbf{z})f(\mathbf{z}) = P(H_1)f(\mathbf{z}|H_1)t_{H_1} + P(H_0)f(\mathbf{z}|H_0)t_{H_0} \tag{4}$$

and expression (2) can be expressed as:

$$E_m = \int_Z (F(\mathbf{z}) - F_0(\mathbf{z}))^2 f(\mathbf{z})d\mathbf{z}$$
$$+ \int_Z (P(H_1)f(\mathbf{z}|H_1)t_{H_1}^2 + P(H_0)f(\mathbf{z}|H_0)t_{H_0}^2 - F_0^2(\mathbf{z}))d\mathbf{z} \tag{5}$$

Since the second integral is independent of $F(\mathbf{z})$, minimizing E_m is equivalent to minimize (6). So the network output is an approximation of $F_0(\mathbf{z})$ in the mean squared-error sense, for any pair of desired outputs.

$$\int_Z (F(\mathbf{z}) - F_0(\mathbf{z}))^2 f(\mathbf{z})d\mathbf{z} \tag{6}$$

For $t_{H_1} = 1$ and $t_{H_0} = -1$, expression (3) is equal to that obtained by Ruck et al. [2]. If a hard threshold detector is placed at the output of the system, the decision rule based on $F_0(\mathbf{z})$ is expressed in (7).

$$\frac{P(H_1)f(\mathbf{z}|H_1)t_{H_1} + P(H_0)f(\mathbf{z}|H_0)t_{H_0}}{P(H_1)f(\mathbf{z}|H_1) + P(H_0)f(\mathbf{z}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \tag{7}$$

The question that has not been answered yet is: does the decision rule (7) implement the Neyman-Pearson detector, when η_0 is fixed attending to P_{FA} conditions? By means of simple transformations, expression (7) comes into (8), where $\Lambda(\mathbf{z}) = f(\mathbf{z}|H_1)/f(\mathbf{z}|H_0)$ is the likelihood ratio. Finally, $\Lambda(\mathbf{z})$ can be cleared up in (8), to obtain the equivalent expression (9), proving that the rule (7) is an implementation of the Neyman-Pearson detector.

$$\frac{P(H_1)\Lambda(\mathbf{z})t_{H_1} + P(H_0)t_{H_0}}{P(H_1)\Lambda(\mathbf{z}) + P(H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \tag{8}$$

$$\Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{cv} = \frac{P(H_0)(\eta_0 - t_{H_0})}{P(H_1)(t_{H_1} - \eta_0)} \tag{9}$$

Expression (9), where η_{cv} is the detection threshold for the decision rule based on $\Lambda(\mathbf{z})$, shows that η_0 for a given P_{FA} is not only a function of the likelihood functions, but also depends on the prior probabilities and on the desired outputs. These parameters can be selected by the designer when generating the training set, and when determining the activation function of the output neuron, respectively.

3. SENSITIVITY OF THE DETECTOR

The performance of a system that approximates the Neyman-Pearson detector must be evaluated from the difference between its ROC curve and the Neyman-Pearson detector one. For a given P_{FA}, the difference between the probabilities of detection must be as lower as possible.

In practical situations, the detection threshold is adjusted to achieve the desired P_{FA}, so P_{FA} requirements can be fulfilled. But, due to approximation errors, the associated P_D will be lower than the optimum detector one for the same P_{FA}. As we know the discriminant function approximated by the system, we can calculate the difference between the thresholds required by both, the system and the decision rule based on the corresponding $F_0(\mathbf{z})$, and use it as a measure of the approximation error.

The quantity of interest is not the approximation error, but how the detector performance is affected by it. Taking into consideration the previous reasoning, the objective is to evaluate how the differences between the detection thresholds required by the trained system and the decision rule based on the corresponding $F_0(\mathbf{z})$, affect the global performance of the detector based on the adaptive system. As this performance is evaluated in terms of P_{FA} and P_D, a method based on the calculus of the partial derivatives of both probabilities with respect to the detection threshold is proposed.

3.1 Proposed method

If P_i is used for denoting P_{FA} or P_D, the partial derivative of P_i with respect to η_0 can be calculated using the chain rule (10).

$$\frac{\partial P_i}{\partial \eta_0} = \frac{\partial P_i}{\partial \eta_{cv}} \frac{\partial \eta_{cv}}{\partial \eta_0} \tag{10}$$

The second factor of the right side of (10) can be calculated from (9) to obtain (11). It depends on the prior probabilities of the classes and the desired outputs selected for training, factors that can be controlled by the designer.

$$\frac{\partial \eta_{cv}}{\partial \eta_0} = -\frac{P(H_0)(t_{H_0} - t_{H_1})}{P(H_1)(t_{H_1} - \eta_0)^2} \tag{11}$$

t_{H_1} is greater than t_{H_0}, and $\eta_0 \in [t_{H_0}, t_{H_1}]$. Because of that, the partial derivative of η_{cv} with respect to η_0 is always positive. Also, for η_0 values close to t_{H_1}, that is, for very low P_{FA} values, $\partial \eta_{CV}/\partial \eta_0$ is very high. We can try to compensate it in some degree, by increasing the difference between the desired outputs, or using training sets where feature vectors from hypothesis H_1 are more likely than those from hypothesis H_0.

The first factor of the right side of (10) depends on the likelihood functions of the problem to be solved. Its value is calculated in (12) ($H_i = H_1$ for P_D, and $H_i = H_0$ for P_{FA})

$$\frac{\partial P_i}{\partial \eta_{cv}} = \frac{\partial}{\partial \eta_{cv}}\Big[1 - \int_{-\infty}^{\eta_{cv}} f(\Lambda(\mathbf{z})|H_i)d(\Lambda(\mathbf{z}))\Big]$$
$$= -f(\Lambda(\mathbf{z})|H_i)|_{\Lambda(\mathbf{z}|H_i)=\eta_{cv}} \tag{12}$$

To gain an insight into $\partial P_i/\partial \eta_{cv}$, we transform the rule (9) into an equivalent one, based on a simpler statistic, $\digamma(\mathbf{z})$, and the corresponding detection threshold, η_s. The relation between η_s and η_{cv} is determined by the relation that exists between the likelihood ratio and the selected statistic. Expression (10) can be re-written as:

$$\frac{\partial P_i}{\partial \eta_0} = \frac{\partial P_i}{\partial \eta_s}\frac{\partial \eta_s}{\partial \eta_{cv}}\frac{\partial \eta_{cv}}{\partial \eta_0} \tag{13}$$

Expression (13) shows that $\partial P_i/\partial \eta_0$ can be expressed as the product of three factors:

- $\partial P_i/\partial \eta_s$ and $\partial \eta_s/\partial \eta_{cv}$ depend on the problem to be solved.
- $\partial \eta_{cv}/\partial \eta_0$, not only depends on the a prior probabilities of the classes and the desired outputs selected from training, because the value of η_0 depends on the problem to be solved.

The usefulness of adding a new factor in (10) only can be proved if a particular case is considered. Next section deals with the problem of detecting gaussian signals in gaussian interference.

4. EXAMPLE: THE NEYMAN-PEARSON DETECTOR FOR GAUSSIAN SIGNALS IN GAUSSIAN INTERFERENCE

Lets assume that the feature vector is composed of n independent gaussian samples of zero mean and unity variance under hypothesis H_0, and zero mean and a variance σ_s^2+1 under hypothesis H_1. The signal-to-noise ratio is defined as $SNR = 10log_{10}(snr) = 10log_{10}(\sigma_s^2)$ and the value selected for generating the training set is denoted as $tsnr$.

For a given $tsnr$, the likelihood functions are expressed in (14) and (15); the decision rule based on the likelihood ratio is given by (16).

$$f(\mathbf{z}/H_0) = \frac{1}{\sqrt{(2\pi)^n}}\exp\Big(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\Big) \tag{14}$$

$$f(\mathbf{z}/H_1) = \frac{1}{\sqrt{(2\pi)^n(tsnr+1)^n}}\exp\Big[-\frac{1}{2(tsnr+1)}\mathbf{z}^T\mathbf{z}\Big] \tag{15}$$

$$\Lambda(\mathbf{z}) = \frac{1}{(tsnr+1)^{\frac{n}{2}}}\exp\Big[\frac{tsnr}{2(tsnr+1)}\mathbf{z}^T\mathbf{z}\Big] \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{cv} \tag{16}$$

A simpler sufficient statistic can be obtained applying logarithms and re-arranging terms:

$$\mathcal{F}(\mathbf{z}) = \mathbf{z}^T\mathbf{z} \underset{H_0}{\overset{H_1}{\gtrless}} 2\frac{tsnr+1}{tsnr}ln[\eta_{cv}(tsnr+1)^{\frac{n}{2}}] = \eta_s \tag{17}$$

As the variance of the samples generated under hypothesis H_0 is fixed to unity, the probability density function of $\mathcal{F}(\mathbf{z}|H_0)$ does not depend on $tsnr$, and for a given P_{FA}, η_s is independent of it. Because of that, the performance of the Neyman-Pearson detector is independent of $tsnr$.

The partial derivative of η_s with respect to η_{cv} is given in (18).

$$\frac{\partial \eta_s}{\partial \eta_{cv}} = \frac{2(1+tsnr)}{\eta_{cv}tsnr} \tag{18}$$

Taking into consideration that $\mathcal{F}(\mathbf{z}|H_0)$ is a chi-square random variable with n degrees of freedom and $\mathcal{F}(\mathbf{z}|H_1)$ is a gamma random variable [1] of parameters $a = n/2$ y $b = 2(snr+1)$, the partial derivatives of P_{FA} and P_D with respect to η_s are calculated in (19), where $\Gamma()$ is the gamma function.

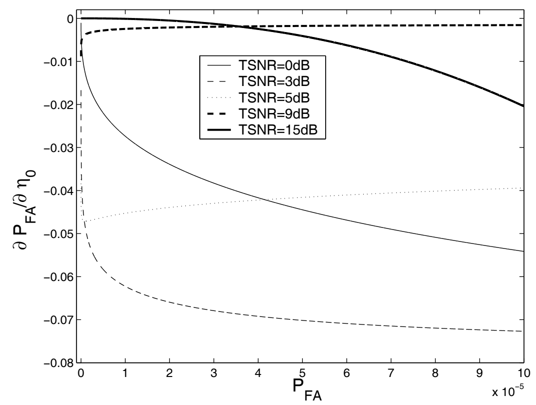

Figure 1: $\partial P_{FA}/\partial \eta_0$ *for* $n = 16$, $P(H_1) = P(H_0) = 0.5$, $t_{H_1} = 1$, $t_{H_0} = 0$, *different* $TSNR$ *values and* $P_{FA} \leq 10^{-4}$

$$\frac{\partial P_{FA}}{\partial \eta_s} = -\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\eta_s^{(\frac{n}{2}-1)}\exp(\frac{-\eta_s}{2})$$
$$\frac{\partial P_D}{\partial \eta_s} = -\frac{1}{(2(snr+1))^{\frac{n}{2}}\Gamma(\frac{n}{2})}\eta_s^{(\frac{n}{2}-1)}\exp(\frac{-\eta_s}{2(snr+1)}) \tag{19}$$

Combining expressions (11), (18) and (19), the partial derivatives of P_{FA} and P_D with respect to η_0 can be calculated. Although the ROC curves do not depend on $tsnr$, $\partial P_{FA}/\partial \eta_0$ and $\partial P_D/\partial \eta_0$ depend on it. Because of that, the sensitivity of the detector based on $F_0(\mathbf{z})$ to detection threshold variations depends on $tsnr$.

For $n = 16$, $P(H_1) = P(H_0) = 0.5$, $t_{H_1} = 1, t_{H_0} = 0$ and different $tsnr$ values, $\partial P_{FA}/\partial \eta_0$ versus P_{FA} curves are represented in figures 1 and 2. In figure 1, P_{FA} values lower than 10^{-4} are considered, because higher values of P_{FA} have no interest in practical situations. The variation of $\partial P_{FA}/\partial \eta_0$ is very complex, especially for very low P_{FA} values, but, in general, we can conclude that high values of $TSNR$ are preferable. A study for higher P_{FA} values reveals that in this region the sensibility is higher and increases dramatically with the TSNR (note that only the curves for $TSNR = 0, 3$ and $5dB$ have been represented in order to show the variation with P_{FA} and $TSNR$).

The corresponding $\partial P_D/\partial \eta_0$ versus P_{FA} curves for $SNR = 3$ and $9dB$, and $P_{FA} \leq 10^{-4}$, are represented in figures 3 and 4. Again, the influence of $TSNR$ is very important. Also, the SNR must be taken into consideration. Comparing figures 3 and 4, we can conclude that the sensitivity of P_D to detection threshold variations reduces significantly when SNR increases. For a given value of SNR, in general, high $TSNR$ values are recommended in order to reduce the sensitivity.

$\partial P_D/\partial \eta_0$ curves for higher values of P_{FA} are not included because they present a variation similar to that observed for the $\partial P_{FA}/\partial \eta_0$, although the sensitivity is lower.

[1]The Gamma probability density function is given by: $f(x|a,b) = \frac{1}{b^a\Gamma(a)}x^{a-1}\exp(-\frac{x}{b})$

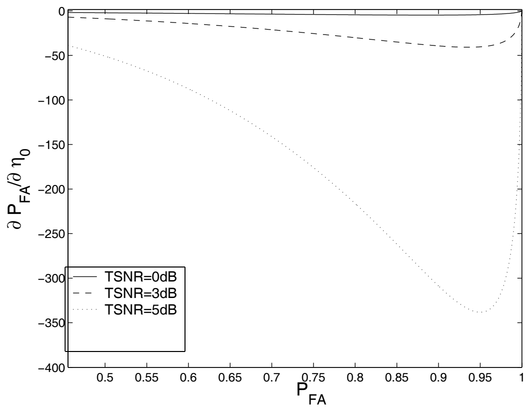

Figure 2: $\partial P_{FA}/\partial \eta_0$ *for* $n = 16$*,* $P(H_1) = P(H_0) = 0.5$*,* $t_{H_1} = 1$*,* $t_{H_0} = 0$*, different* $TSNR$ *values and* $0.5 \leq P_{FA} \leq 1$

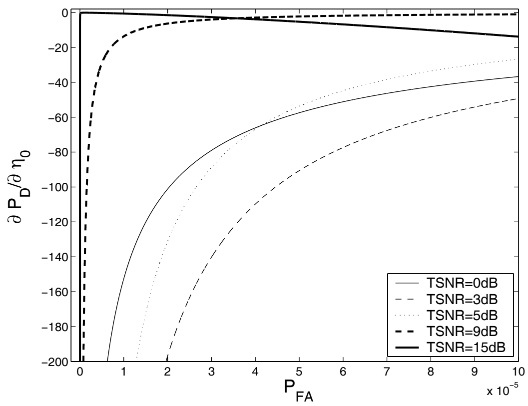

Figure 3: $\partial P_D/\partial \eta_0$ *for* $n = 16$*,* $P(H_1) = P(H_0) = 0.5$*,* $t_{H_1} = 1$*,* $t_{H_0} = 0$*,* $SNR = 3dB$*, different* $TSNR$ *values and* $P_{FA} \leq 10^{-4}$

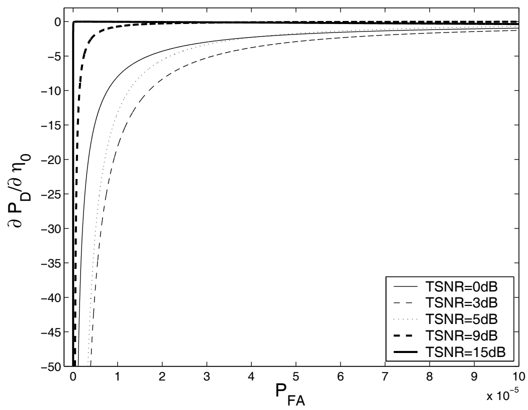

Figure 4: $\partial P_D/\partial \eta_0$ *for* $n = 16$*,* $P(H_1) = P(H_0) = 0.5$*,* $t_{H_1} = 1$*,* $t_{H_0} = 0$*,* $SNR = 9dB$*, different* $TSNR$ *values and* $P_{FA} \leq 10^{-4}$

5. CONCLUSIONS

This paper deals with the application of neural networks to approximate the Neyman-Pearson detector. A general expression of the discriminant function approximated by an adaptive system trained using the LMSE criterion is calculated and analyzed, proving that the decision rule based on it is an implementation of the Neyman-Pearson detector. This general expression reveals the influence of the prior probabilities and the desired outputs selected for training.

Also, a novel method is proposed to evaluate the sensitivity of the Neyman-Pearson detection rule based on this discriminant function, to detection threshold variations. The calculus of the partial derivatives of P_{FA} and P_D with respect to the detection threshold are proposed, as tools to analyze the robustness of the detector and to identify the values of the training parameters that can reduce the effect of approximation errors on the performance of the neural network based detector. Among these parameters, in previous works, the *tsnr* appeared as a critical one, but no effort was made to explain the dependence of the detector performance on it. The proposed method has been applied to a case of study in order to evaluate the influence of tsnr. Results demonstrate that, for the case of study, tsnr infuence is a function of P_{FA}, and, for low values of P_{FA}, high values of *tsnr* are more suitable.

REFERENCES

[1] Van Trees, H.L.: Detection, estimation, and modulation theory, Vol. **1**. Wiley, (1968)

[2] Ruck, D.W., Rogers, S.K., Kabrisky, M., Oxley, M.E., Suter, B.W.: The multilayer perceptron as an aproximation to a Bayes optimal discriminant function. IEEE Trans. on Neural Networks, vol. 1, no. 4, p.p. 296–298, Dec. 1990.

[3] Wan, E.A.: Neural Network Classification: A Bayesian Interpretation. IEEE Trans. on Neural Networks, vol.1, no. 4, pp. 303-305, Dec. 1990.

[4] Gandhi, P.P., Ramamurti, V.: Neural networks for signal detection in non-gaussian noise. IEEE Trans. on Signal Processing, vol. 45, no. 11, pp. 2846-2851, Nov. 1997.

[5] Andina, D., Sanz-González, J.L.: On the problem of binary detection with neural networks. Proc. of the 38th Midwest Symposium on Circuits and Systems, vol. 1, pp. 13-16, 1995.

[6] Andina, D., Sanz-González, J.L.: Comparison of a neural network detector vs. Neyman-Pearson optimal detector. Proc. of the 1996 IEEE ICASSP, vol. 6, pp. 3573-3576, 1995.

[7] Jarabo-Amores, P., Rosa-Zurera, M., López Ferreras, F.: Design of a Pre-processing Stage for Avoiding the Dependence on TSNR of a Neural Radar Detector. Lecture Notes in Computer Sciences, Vol. 2085. Springer-Verlag, (2001) 652–659,