CHAPTER 5

A NEW NEURAL NETWORK FOR ADAPTIVE PATTERN RECOGNITION OF MULTICHANNEL INPUT SIGNALS

M. Fernández-Delgado[1,2], **J. Presedo**[1], **M. Lama**[1], and **S. Barro**[1]
[1]Department of Electronics and Computer Science
University of Santiago de Compostela
15706 Santiago de Compostela, Spain
[2]Department of Informatics
University of Extremadura
Cáceres, Spain
eldelga@usc.es

MART (Multichannel ART) is a neural computational architecture which is based on ART architecture and aimed at pattern recognition on multiple simultaneous information input paths. This chapter describes the characteristic aspects of MART architecture, how it operates in pattern recognition and its flexibility in adapting itself to the temporal evolution of input information fed into the network. Finally, a real application is presented demonstrating its potential for solving complex problems, above all in the field of multichannel signal processing.

1 Introduction

The neural computational model ART was introduced by Carpenter and Grossberg in the 1980s, and developed through various versions, such as ART1 [5], ART2 [4], ART-2A [8] and ART3 [6]. These networks have contributed a number of valuable properties with respect to other neural architectures, amongst which could be mentioned its on-line and self-organizing learning. On the other hand, these networks make it possible to resolve the dilemma between plasticity and stability, allowing both the updating of the classes learned and the immediate learning of new classes without distorting the already existing ones. This property enables it to be used in problems in which the number of classes is not limited a priori, or in which there is an evolution in the classes over time. These characteristics are shared by a great number of variants of the ART architecture, which are acquiring more operational and application possibilities. Thus, mentioning only a small number of the most representative examples, many networks have been proposed such as ARTMAP [9], which enables supervised learning, Fuzzy ART [7] and Fuzzy ARTMAP [10], which adapt themselves to the processing of fuzzy patterns, or HART (Hierarchical ART) [3], integrated by a series of ART1 networks which carry out cascade clustering tasks on input patterns. Nevertheless, this wide range of options lacks certain characteristics which, to our way of thinking, are especially interesting in order to tackle certain problems typical to pattern recognition such as:

- In many pattern recognition applications there are various paths or channels of information about the same event or system under analysis. This is the case, for example, with monitoring of systems by means of multiple sensors that supply complementary or alternative views of the system behavior. In these cases, the joint consideration of the information given by all of the sensors enables us to increase the reliability of the final result, making it easier to detect noise and eliminate ambiguities.

- Supervised learning does not permit the reconfiguration of a neural network during its routine application. In the case of ART networks, however, the representations of the classes to which belong the input patterns are constructed on-line, starting from a total absence of a priori information on these patterns and their associated classes. Nevertheless, this type of network does not adapt the capability of discrimination between classes of patterns during its operation, since they operate with a fixed vigilance parameter. Thus they cannot adapt their behavior to the typical characteristics of each class to be recognized nor to their possible temporal evolution.

- ART networks carry out a partitioning or clustering operation on the input space, on the basis of a vector of features that describe the input patterns, a measure of the distance between these patterns and the classes to be discerned, and a (vigilance) threshold to be applied to the distance obtained between the input pattern to be classified and previously recognized classes. Each one of these classes (associated to a cluster in the input space) includes a series of patterns. Thus, it is possible for classes to exist whose patterns are much more similar amongst themselves than with those associated to other classes. ART networks presume that the patterns of all classes have the same variability, as they use a single vigilance for all of them.

- In many pattern recognition problems the different classes have a different representativeness about the input patterns. The relevance of a class usually depends on the specific criteria of the problem tackled (appearance frequency of input patterns associated to this class, level of danger associated to this class, etc.). ART networks do not supply any mechanisms for selectively evaluating the different classes learned.

The relevance, which to our understanding these properties have, leads us to propose a new model of neural computational architecture, which we have called MART [14]-[16]. Maintaining those aspects of ART architecture that are especially relevant, we have incorporated the capability for the adaptation of those parameters that determine the operation of the network, adapting its values to the application and to the set of patterns to be processed. At the same time, MART deals jointly, although in a selective and adaptive manner, with multiple channels of information and internal representations of the classes learnt from the input data submitted to the network during its operation. In the following section we will describe the structure and operation of MART, in order to later deal with its properties for the learning of information. We will then present an illustrative example of the operation on artificially generated patterns, in order to later deal with its application in a real problem such as the recognition of morphological patterns of heart beats on multiple derivations of electrocardiographic signals. Finally we will summarize the main contributions of MART to neural computation, along with the future lines of study on which our work is focused.

2 Architecture and Functionality of MART

As we said before, MART is a neural computation architecture for the pattern recognition, which operates simultaneously over multiple channels or inputs of information. Figure 1 shows its structure, made up of I blocks, each one of which is associated to the processing of the input pattern in a different signal channel. Each one of these blocks determines the local similarity (relative to channel i) between the input pattern and the expected values of the learnt classes, through processing of the patterns showed previously to the network in that channel. The $F4$ block, located on the top of the figure, is governed by a competitive "*winner-takes-all*" mechanism, in which the only active output, $u_k=1$, is associated to the class with maximum global (over the set of channels) similarity (we will call this class "winning class"). This class propagates its expected value downwards through the single channel blocks, with the aim of determining the local differences, relative to each channel, with the input pattern. The Orientation System evaluates these differences, determining whether they are or are not sufficiently low in order to assign the input pattern to the winning class (resonance or reset respectively). Finally, the Class Manager controls the dynamic creation and suppression of classes as it becomes necessary in the classification process.

2.1 Bottom-Up Propagation in a Single-Channel Block

Figure 2 shows the structure of a single-channel block of MART. The input pattern in channel i, $\mathbf{E}_i=(E_{i1},\ldots,E_{iJ})$, $E_{ij}\in [0,1]$, $\forall j$, is presented in the units of layer $F1_i$, where it is propagated towards $F2_i$. The connection weight vector $\mathbf{z}_{ik}=(z_{i1k},\ldots,z_{iJk})$ between the units of $F1_i$ and unit $F2_{ik}$ stores the expected value of class k in channel i. The output of unit $F2_{ik}$ is determined by the expression:

$$L_{ik} = f(\mathbf{E}_i, \mathbf{z}_{ik}),\ \forall k \tag{1}$$

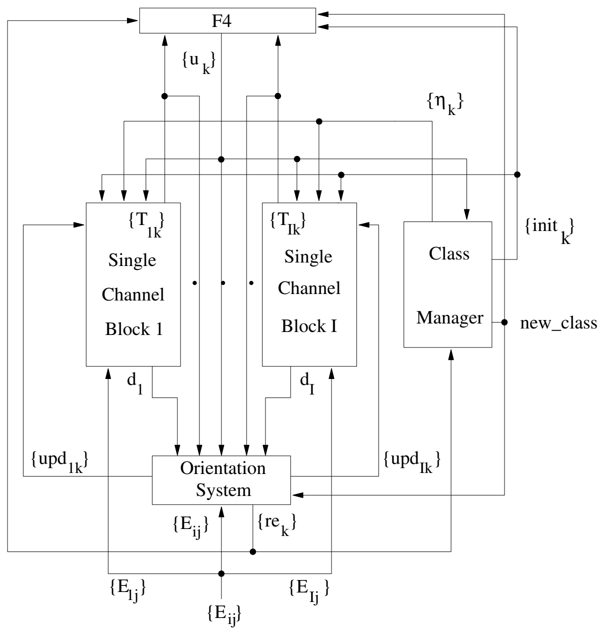

Figure 1. Block diagram of MART.

where the function $f(\mathbf{x},\mathbf{y})$ evaluates the dissimilarity between the vectors $\mathbf{x}$ and $\mathbf{y}$, $0 \leq f(\mathbf{x}, \mathbf{y}) \leq 1$, being 0 the value associated to a total coincidence between them. The vector $\mathbf{L}_i=(L_{i1},\ldots,L_{iK})$ is propagated towards layer $F3_i$, whose output is:

$$T_{ik} = \eta_k (1 - L_{ik}), \ \forall k \tag{2}$$

In this expression, η_k is an output item of the Class Manager (see Figure 4) which verifies $\eta_k=1$ when the unit k in $F4$ is committed (i.e., associated to a learnt class), otherwise $\eta_k=0$.

2.2 Class Selection

The input items for block $F4$ (local similarities) are integrated for the different channels generating the global similarities P_k:

$$P_k = \overline{re}_k \sum_{i=1}^{I} T_{ik} \tag{3}$$

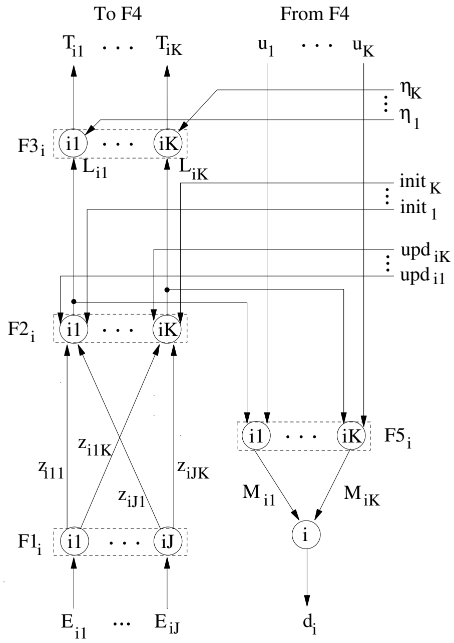

Figure 2. Diagram of the single-channel block *i*.

The input items re_k come from the Orientation System and, as will be seen, their value is *1* for those classes which have been temporarily reset for having shown an excessively high difference with the input pattern, for which $P_k = 0$. During the first comparison between the input pattern and the learnt classes no unit is reset, in such a manner that $P_k = \sum_{i=1}^{I} T_{ik}$, $\forall k$. The binary output items u_k of *F4* are governed by the expression:

$$u = init_k \wedge \left\{ \overline{new\_class} \wedge \left[\bigwedge_{l<k} \tau(P_k - P_l) \right] \wedge \left[\bigwedge_{l>k} \Gamma(P_k - P_l) \right] \right\},\ \forall k \tag{4}$$

where the symbols $\vee$ and $\wedge$ designate the OR and AND binary operators, respectively, and the functions $\tau(x)$ and $\Gamma(x)$ are governed by the following expressions:

$$\tau(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x \end{cases} \qquad\qquad \Gamma(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x \end{cases}$$

The *new_class* and ***init***$=(init_1,...,init_K)$ input items are binary, and come from the Class Manager: *new_class=1* is associated to the creation of a new class, because the input pattern does not resonate with any of the learnt ones. The ***init*** vector is only non-zero when *new_class=1*, in which case its only non-zero component is the one associated to the unit *k'* selected by the Class Manager to be initialized. Thus, if *new_class=1*, then $u_k=init_k$ (from expression (4)) and the unit *k'* is the only active unit in *F4*. Otherwise, *new_class*$=0=init_k$, $\forall\, k$, and the value of u_k will be determined by the functions τ and Γ, from expression (4), being zero except for the minimum index unit with the maximum P_k (a class which shows the greatest global similarity with the input pattern). We will denote this unit by means of the index k_1, due to it being the winner in this first cycle of comparison between the input pattern and the classes learnt: in this way, $u_{k1}=1$, $u_k=0$, $\forall\, k \neq k_1$.

2.3 Top-Down Propagation in a Single-Channel Block

The vector $\boldsymbol{u}=(u_1,...,u_K)$ is propagated towards the $F5_i$ layers of the I single-channel blocks, whose output takes the form of:

$$M_{ik} = u_k L_{ik}, \ \forall k \tag{5}$$

These output items are zero, except for the unit k_1, for which it is $M_{ik1}=L_{ik1}=f(\mathbf{E}_i,\mathbf{z}_{ik1})$ (difference between the input pattern and the expected value of the active class k_1 in channel i). Finally, the output d_i of the single-channel block i is determined according to the expression:

$$d_i = \sum_{k=1}^{K} M_{ik} = L_{ik_1} = f(\mathbf{E}_i, \mathbf{z}_{ik_1}) \tag{6}$$

which represents the difference between the input pattern and the expected value of class k_1 in that channel. This output is propagated towards the Orientation System, whose structure and functioning will be dealt with in the following section.

2.4 The Orientation System

The basic objective of this module, as shown in Figure 3, is to determine the global result of the pattern-class comparison. For this reason, the block "Channel Credits" determines the global difference d between the input pattern and a given class from the local differences d_i and the channel credits x_i, which constitute an indirect measure of the signal quality in the different channels. The global difference is calculated through the following expression:

$$d = \sum_{i=1}^{I} x_i d_i \tag{7}$$

The value obtained for the global difference d is compared with the global vigilance ρ_{k1}^{g} associated to the class k_1 (one of the K outputs of the block "Global Vigilances", in Figure 3). This parameter establishes the discrimination capability of the system in the comparisons with the k_1 class, having an adaptive value which is limited to the range $\rho_{min} \leq \rho_{k1}^{g} \leq 1$. This comparison takes place in the block "Reset Evaluation", which determines the output items re_k of the Orientation System based on the expression:

$$re_k(t+1) = \left[re_k(t) \wedge \overline{new\_pattern(t)}\right] \vee \left[\overline{re_k(t)} \wedge u_k \wedge \Gamma(d - \rho_k^g)\right] \quad , \quad \forall k \tag{8}$$

in such a manner that if $d < \rho_{k1}^{g}$ and $re_{k1}(t)=0$ then $re_{k1}(t+1)=0$ and the state of the system, with the unit k_1 active in $F4$, is maintained stable, reaching resonance with the input pattern. In this case, as shall be seen in the following section, the information associated with the resonant class k_1 is updated and the resonance is maintained until the presentation of a new input pattern. On the other hand, if $d \geq \rho_{k1}^{g}$ and $re_{k1}(t)=0$, then $re_{k1}(t+1)=1$ and, from the expression (3), $P_{k1}=0$, class k_1 being inhibited for the competition in $F4$ (re_{k1} remains active until the presentation of a new input pattern, the instant in which the *new_pattern(t)* output of block "Detection of New Pattern" is activated, taking again re_{k1} to zero). When this occurs, a new winner k_2 is determined in *F4* and the comparison cycle is repeated, re_{k2} being evaluated according to $\rho_{k2}{}^{g}$ and the new value of *d*. If the result is a new reset, the process is repeated until either a resonance with one of the learnt classes is reached, or until all are reset, in which case the Class Manager determines the creation of a new class. This class will be, if in *F4* there exist uncommitted nodes (not associated to learnt classes), the one with the minimum *k'* index. If there is no uncommitted node, it will select the class with minimum relevance (determined through its class credit), as we will see in the next section.

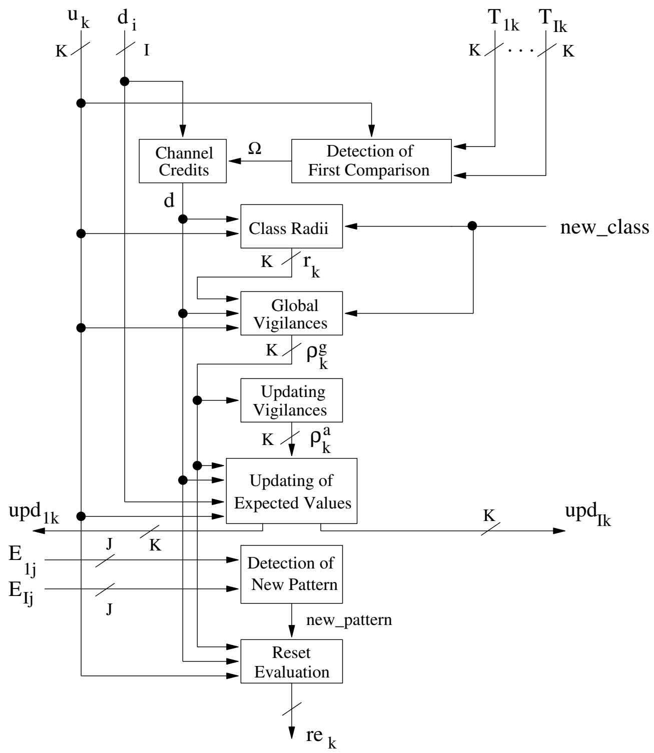

Figure 3. Structure of the Orientation System. For clarity the figure does not show the neural organization of the different blocks.

2.5 Class Manager

Figure 4 shows the structure of the Class Manager block, which controls the creation and dynamic suppression of classes throughout the processing, being activated when the input pattern belongs to an as of yet unlearned class. In this case, all the classes learned beforehand (for which $\eta_k=1$) have been reset ($re_k=1$) and the "Creation of a New Class" block activates its *new_class* output, based on the following expression:

$$new\_class = \bigwedge_{k=1}^{K} \overline{\eta_k} \vee re_k \tag{9}$$

When $new\_class=1$, the "Class Selection" block determines the class k' which is going to be created, establishing $init_{k'}=1$, $inic_k=0$, $\forall\ k\neq k'$. Other input data to this block are the credits of class μ_k ($k=1,...,K$), which evaluate the associated relevance of the different classes. As previously mentioned, these parameters allow us to perform a selective evaluation of the classes learned according to their relevance. In a later section we will describe the rule which governs the evolution of the

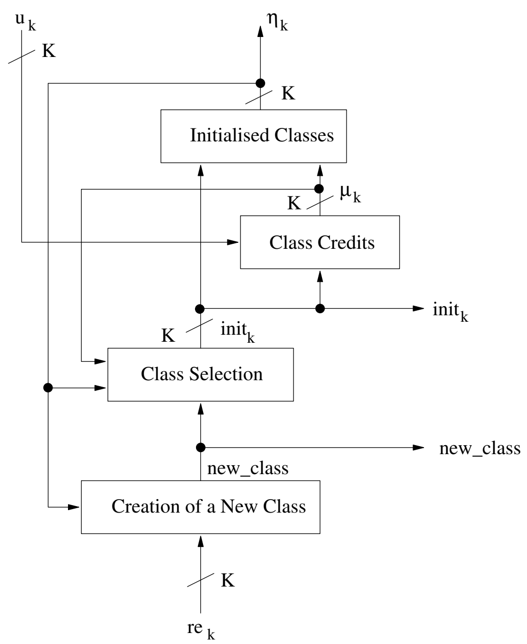

Figure 4. Structure of the Class Manager.

class credits; for the present suffice it to know that $\mu_k \in [0,1]$, $\mu_k=1$ being the credit associated with the maximum relevance. Finally, the input data η_k also take part in the creation of a new class (we should remember that $\eta_k=1$ for those units k of $F4$ associated to learned classes). The output of the "Class Selection" block is determined by the following expression:

$$inic_k = new\_class \wedge \left[\bigwedge_{l<k} \tau(\varsigma_k - \varsigma_l) \right] \wedge \left[\bigwedge_{l>k} \Gamma(\varsigma_k - \varsigma_l) \right], \ \forall k \qquad (10)$$

where ς_k is defined in the following manner:

$$\varsigma_k = (1-\mu_k)\xi + \overline{\eta_k}, \ \forall k \qquad\qquad \xi = \bigwedge_{l=1}^{K} \eta_l$$

and the functions $\tau(x)$ and $\Gamma(x)$ are the ones previously defined. In this manner if $new\_class=0$, $init_k=0$, $\forall\ k$; i.e., a new class is created only when $new\_class=1$. In this case, if there exists some unit in *F4* that is not associated to learned classes (for which $\eta_k=0$), then $\varsigma_k= \overline{\eta_k}$ and $init_k=1$ for this unit (if there are various, the unit selected is the one with the minimum index k). On the contrary, if all the units in *F4* are associated to learned classes ($\eta_k=1$, $\forall\ k$), $\varsigma_k=(1-\mu_k)$ and the class selected is the one with the minimum class credit μ_k, i.e., the one that has the least relevance at that time.

At this point we end the description of the structure and working of MART in order to tackle aspects associated with its plasticity and adaptation according to the information extracted for the input patterns over time. This description is included in the following section.

3 Learning in MART

The MART network uses an unsupervised, "*on-line*" learning, typical of ART networks, for the determination and updating of the expected values of the different classes which are being identified during the processing. Nevertheless, it also provides other learning mechanisms, including channel credits and credits, radii and global vigilances associated to the classes learnt, as will be seen in this section.

3.1 Expected Values

The expected value $\mathbf{z}_{ik}$ of class k in channel i is updated with each input pattern which matches this class. Nevertheless, in order to avoid possible classification errors provoking strong distortions in this expected value, it is desirable to use a threshold ρ_k^u (updating vigilance), more restrictive than global vigilance ρ_k^g, and to update only in those channels i in which the local difference d_i is lower than the global vigilance. The following rule governs class expected value learning:

$$z_{ijk}(new) = inic_k E_{ij} + \overline{inic_k}\left[(1-\alpha_z upd_{ik})z_{ijk}(old) + \alpha_z upd_{ik} E_{ij}\right],\ \forall i, j, k \tag{11}$$

When a new unit k is created, $init_k=1$ and $z_{ijk}(new)=E_{ij}$. In the case of resonance with class k, $\mathbf{z}_{ik}$ only evolves if $upd_{ik}=1$, which occurs when $d<\rho_k^u$ in those channels i where $d_i<\rho_k^g$. We can see that, if $upd_{ik}=1$, then we have, from the expression (11):

$$z_{ijk}(new) = (1-\alpha_z)z_{ijk}(old) + \alpha_z E_{ij} \quad \forall i, j, k \tag{12}$$

where the parameter $0\leq\alpha_z\leq 1$ determines the speed of change in the expected value of the resonant class.

3.2 Channel Credits

Channel credits, associated to the weights inside the block "Channel Credits", in the Orientation System (see Figure 3), have an initial value of $x_i=1/I$, $\forall i$. The credit of a channel represents its weight in the global classification developed by MART over every input pattern. Channel credits update according to the result of the comparison between the pattern presented to the network and the class which has results the most similar to it. This measure of similarity does not take into account the own channel credits, and the most similar class, k_1, will be that with greatest $\sum_{i=1}^{I} T_{ik1}$, during the first comparison (in this stage the output Ω of the block "Detection of First Comparison", in the Orientation System, is activated). In effect, except for specific cases associated with the appearance of patterns which belong to classes not learnt by the network yet, the local difference in each channel with the most similar class should be reduced. In order to do this, a channel with repeatedly high local differences d_i can be considered to be associated with a higher noise content and lower signal quality, and as such, its credit should be lowered. On the contrary, if the local differences in this first comparison are reduced, this can be considered as a reliable channel with regard to the classification process, and as such, its credit should be increased. In this way, an indirect estimation of the noise-to-signal ratio is used to determine the credit or weight factor of each channel in the integration and evaluation of multi-channel information.

Obviously, the full functionality of the channel credits is reached when there is a temporal continuity in the input signals, which allows to use the information associated to previous times in order to make a prediction, in this case, about the signal quality in every signal channel. Learning is carried out on the basis of the following expression:

$$x_i(new) = \Theta\left[x_i(old) + \Delta x_i(d_i)\right], \ \forall i \tag{13}$$

The function $\Delta x_i(x)$ determines the value of the increase in x_i on the basis of the following expression:

$$\Delta x_i(x) = \begin{cases} \Delta x\left[1 - \dfrac{x}{\delta_{i1}}\right] & 0 \le x < \delta_{i1} \\ -\Delta x\left[\dfrac{x - \delta_{i1}}{\delta_{i2} - \delta_{i1}}\right] & \delta_{i1} \le x < \delta_{i2} \\ -\Delta x & \delta_{i2} \le x \le 1 \end{cases} \tag{14}$$

where Δx is a fixed amount, independent of the signal channel. In this way, the increase $\Delta x_i(x)$ is positive for reduced values of x ($0 \le x \le \delta_{i1}$) and negative for high values of x ($\delta_{i1} \le x \le 1$), reproducing in this manner the behavior outlined previously. The parameters Δx, δ_{i1} and δ_{i2}, $\forall\, i$, must be determined for each application, if we have a set of input patterns which are representative of those which will be presented to the network during its normal operation. Lastly, the function $\Theta_i(x)$ limits the values of x in the range $0 \le x \le 2/I$.

3.3 Class Radii

For each class learnt, MART establishes an adaptive average of the global differences which have been obtained with those patterns assigned to that class over the time. This average is an approximate measurement of the radius of the cluster associated to the class in the input space and, as will be seen, constitutes the basis for the updating of the global vigilance for this class. Class radii correspond to the weights inside the block "Class Radii", in the Orientation System (see Figure 3). The learning of the radius associated to a class takes place with each resonance between a new input pattern and that class or, for the first time, when this class is created, and it is governed by the following expression:

$$r_k(new) = \overline{u_k} r_k(old) + u_k \left\{ \overline{new\_class} \left[(1-\alpha_r)(r_k(old) + \delta(r_k(old))d) + \alpha_r d \right] \right\} \quad \forall k \tag{15}$$

where the function $\delta(x)$ is 1 if $x=0$ and 0 in the opposite case. The radius remains constant for non-winning classes ($u_k=0$). If $new\_class=1$, $r_k(new)=0$ for the class k' created, since in this case $u_{k'}=1$. On the other hand, in the case of resonance, radius is equal to the global difference, if $r_k(old)=0$, otherwise being updated as a weighted sum of its previous value and the global difference d (expression (16)), weighted by a variation factor α_r. In this manner, the radius of a class is adapted to the variability existent between the patterns belonging to it.

$$r_k(new) = (1-\alpha_r) r_k(old) + \alpha_r d, \ \forall k \tag{16}$$

3.4 Global Vigilances

The learning of global vigilances for each class allows the adaptation of the system's discrimination capacity to the level of variability of the input patterns. For this, for each resonance MART compares the global difference d with the radius r_k of the resonant class. The increments in the variability are translated into increases in d with respect to r_k, and it is then advisable to increase ρ_k^g in order to avoid false negatives when the input pattern is not assigned to class k and creates a redundant class. On the contrary, reductions in variability reduce d against r_k, which allows the reduction of the global vigilance in order to adapt the discrimination capability to the new situation. Figure 5 shows an example of the time evolution in the variability of the input patterns. Initially (*t=0*) the variability is small, and the vigilance is low for that class, from the radius associated to it. Later (*t=N*), this variability grows up, and it is advisable to increase the vigilance in order to avoid that input patterns belonging to that class lead to the creation of redundant classes. Finally (*t=M*), the variability of the input patterns reduces again, which decreases the radius associated to that class and its vigilance.

Global vigilances are associated to the weights of the block "Global Vigilances", in the Orientation System (see Figure 3). The expression that controls the learning of global vigilances is the following:

$$\rho_k^g(new) = \overline{u_k}\,\rho_k^g(old) + u_k\left\{\rho_{ref}\ new\_class + \right.$$
$$\left. + \theta\left[\rho_k^g(old) + \mathrm{sgn}(d - r_k)\Delta\rho\right]\overline{new\_class}\right\}\ \forall k \qquad (17)$$

where the functions *sgn(x)* and *θ(x)* are defined by the following equations:

$$\mathrm{sgn}(n) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases} \qquad \theta(x) = \begin{cases} \rho_{min} & x < \rho_{min} \\ x & \rho_{min} \le x < 1 \\ 1 & 1 \le x \end{cases}$$

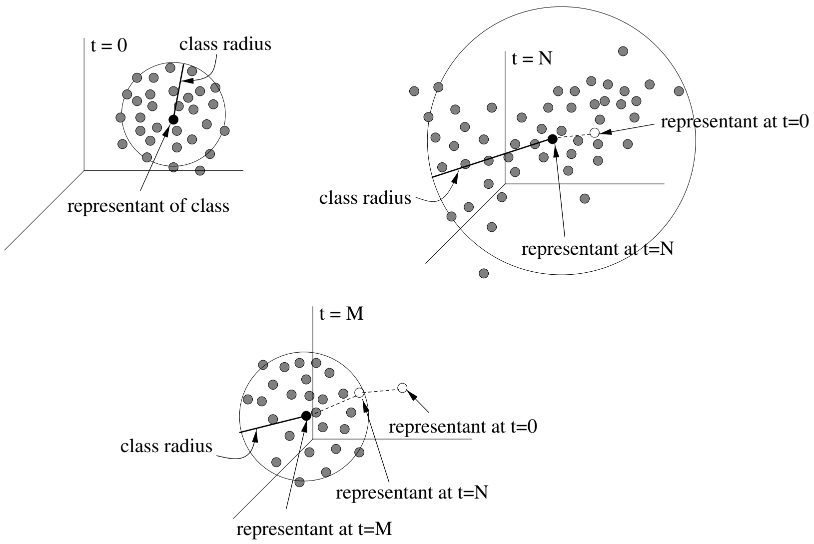

Figure 5. Illustration of the temporal evolution of the variability of the input patterns associated to a class and updating of its representative.

Parameter ρ_k^g remains constant, except in the creation of a new class *k'*, in which case ρ_k^g has the initial value ρ_{ref}, and when resonance is achieved with class *k''*. In this case, $\rho_{k''}^g$ increases its value in $\Delta\rho$ if $d>r_{k''}$ and decreases in the same amount if $d<r_{k''}$, maintaining a lower limit ρ_{min} which is associated to the maximum discrimination capability that a class may possess. In this way, MART is able to selectively evaluate the different classes learnt by means of an individualized and adaptive consideration in its discrimination capability.

3.5 Other Characteristics

As commented in Section 1, another path for the selective evaluation of the classes learnt is associated with "class credits", which corresponds with the connection weights inside the block "Class Credits", in the Class Manager (see Figure 4). These credits allow the evaluation of the relevance of each class throughout the operation of the network. The credit μ_k associated to class *k* has an initial value $\mu_k=1$ at the moment of its creation, increasing in a constant factor Δ_p with each input pattern assigned to it, and decreasing in a constant factor Δ_n in the opposite case, always within the range $0\leq\mu_k\leq 1$.

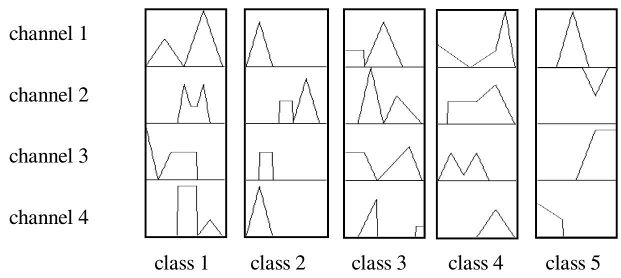

Figure 6. Classes used in the generation of patterns.

On the other hand, MART dynamically manages the classes learnt, creating a class when faced with the appearance of a pattern not belonging to any of them, or deleting a class when its credit is set to zero. For this reason, the output items η_k from the block "Initialized Classes", in the Class Manager (see Figure 4), allow discrimination between those committed units in *F4* ($\eta_k=1$) and those which are not assigned to a class yet. For each unit k, $\eta_k=1$ when it is committed, maintaining this value while its credit μ_k is not zero. When $\mu_k=0$, this class is "forgotten" and unit k remains uncommitted until it is re-committed to a new class, thus enabling the dynamic creation/elimination of classes as and when necessary.

4 Analysis of the Behavior of Certain Adaptive Parameters

In order to illustrate the operation of MART we now give an application example on a set of 2,000 artificially generated patterns with $I=4$ signal channels, a pattern length of $J=125$ values in each channel, and input data in the range $0 \le E_{ij} \le 1$. These patterns were generated from 5 basic morphologies, each one of these being labeled as belonging to a class identified with the morphology from which each pattern derived. The distance function $f(\mathbf{x}, \mathbf{y})$ used is the city-block distance between the input pattern $\mathbf{x}$ and the representative of class $\mathbf{y}$. Figures 6 and 7 show the original classes and some of the patterns generated from them, respectively.

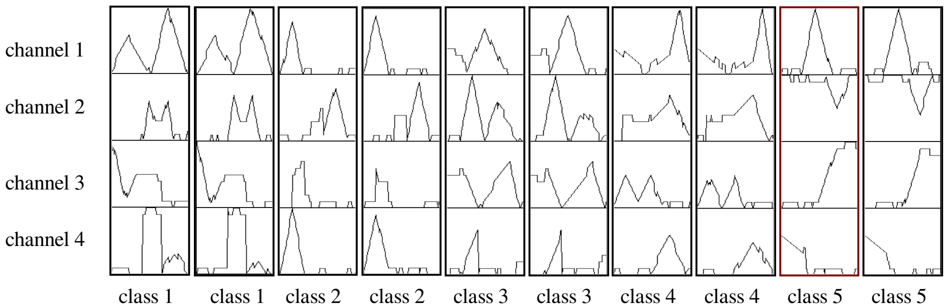

Figure 7. Examples of patterns.

Figure 8 shows an example operation, although in this case, for reasons of clarity, it is represented on 2 signal channels, the credit x_2 being notably lower than x_1. The lower left-hand section shows the input pattern in $F1_1$ and $F1_2$. This pattern propagates to $F2_1$ and $F2_2$: the competitive process in $F4$ selects class 1 as the one that is most similar to the input pattern, its representative (expected value) being shown in both channels with the upper left-hand section. At the same time, the right-hand section shows the area resulting from subtracting the input pattern and the expected value of the winning class in both channels, together with the channel credits and global difference and vigilance. The area of difference (local difference) is lower in channel 1 and somewhat higher in channel 2. Nevertheless, the reduced credit value x_2 attenuates the contribution of channel 2 to the global difference *d*, the value of which is lower than the global vigilance associated to class 1, and brings about a situation of resonance with the input pattern.

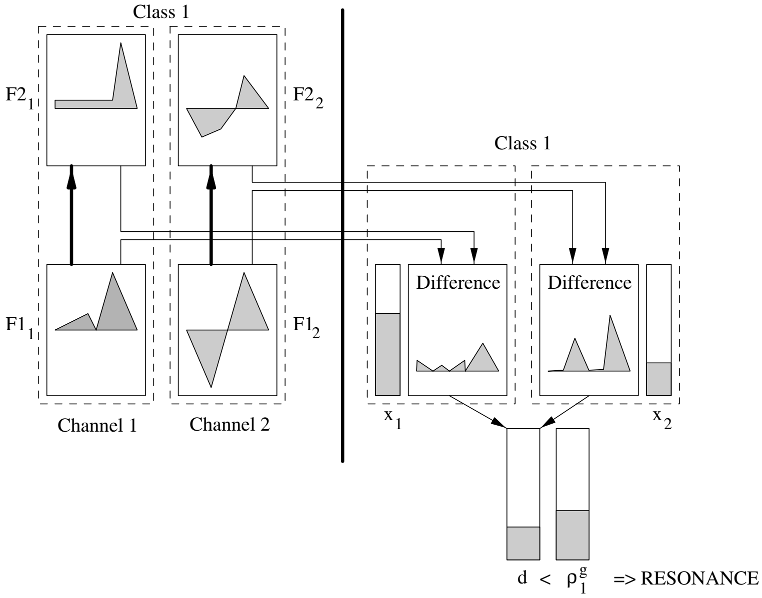

Figure 8. Example of resonance between an input pattern and its most similar class.

Figure 9 shows a second example in which the input pattern undergoes a reset with its most similar class (class 2). The latter wins the competitive process in *F4*, but the descending propagation generates a greater difference in channel 2, the credit of which is higher than the one associated with channel 1, leading to a situation of reset (first comparison). In the second comparison class 3 is the winning class in *F4*. The descending propagation provokes a higher difference in channel 1, but lower in channel 2, reaching resonance on the basis of the weighting relative to the respective channel credits.

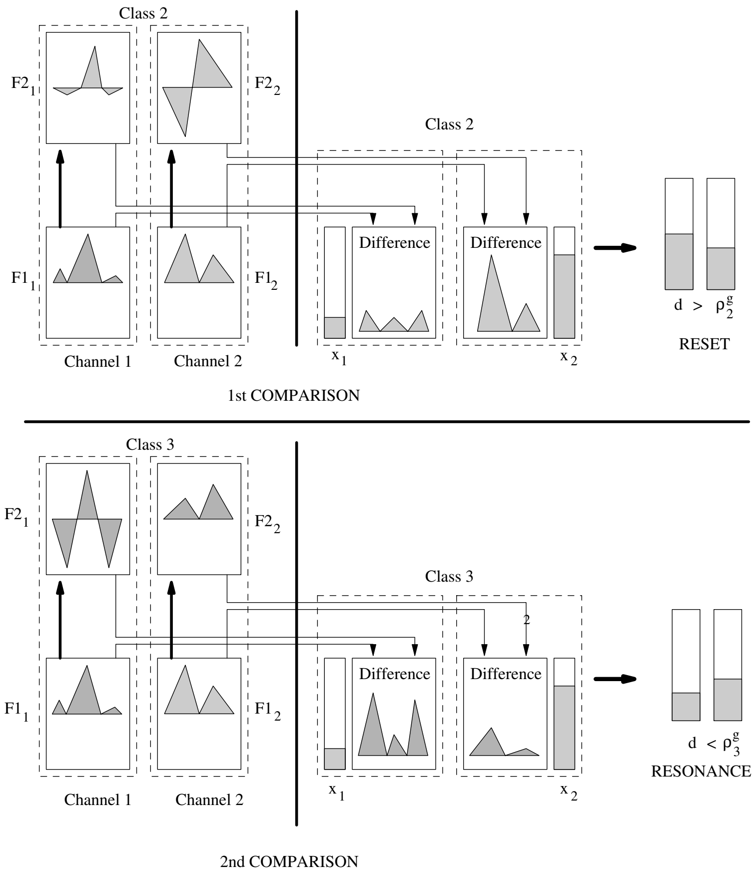

Figure 9. Example of reset between an input pattern and its most similar class, and the subsequent resonance after the comparison of the pattern with another class.

With the aim of illustrating the working of the channel credits x_i in the MART operation we distorted the input pattern, although only in channel 1, by means of the substitution of a part of the original input pattern E_{ij} for randomly generated values. Figure 10 demonstrates how the appearance of noise leads to significant rises in the local differences in channel 1 in the first comparison, differences which are thus useful in order to estimate the high noise/signal ratio associated to this channel. The figure also demonstrates the reduction in the channel credit x_1, which is produced in those intervals where noise is added to the signal (elevations in d_1) suitably reflecting the drop in reliability in channel 1 with regard to the final classification process of the input patterns. We should also emphasize the high level of stability in those credits associated with the remaining channels, in which the signal quality does not undergo any noticeable difference throughout the process.

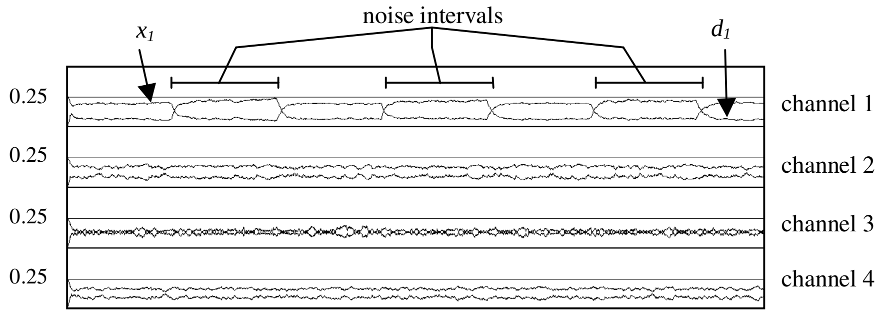

Figure 10. Evolution of the channel credits x_i in response to the addition of noise in channel 1.

Another interesting aspect of the operation of MART is associated with the evolution of the global vigilances associated to the different classes. In order to illustrate this point we added noise to the patterns belonging to class 1 at certain instances of the processing. Figure 11 shows the evolution of the vigilances ρ_1^g and ρ_2^g; it can be seen how addition of noise leads to a noticeable increase in the global difference associated with those patterns that resonate with class 1; the one associated with class 2 remains approximately constant. These increases lead to a rise in the radius associated with class 1 (r_1) and, thus, in the vigilance ρ_1^g of this class, whilst ρ_2^g undergoes no significant alterations. In turn, the disappearance of the added noise brings about an immediate decrease in r_1 and ρ_1^g. Consequently, this figure shows the capability of MART's vigilance to adapt itself to the properties of the input data, more specifically to the variability that these patterns demonstrate over time.

5 A Real Application Example

In the previous section we analyzed the operation of MART on a set of artificial patterns, with the aim of highlighting some of the most interesting characteristics of learning in MART. We now show, albeit briefly, the application of MART in a real problem.

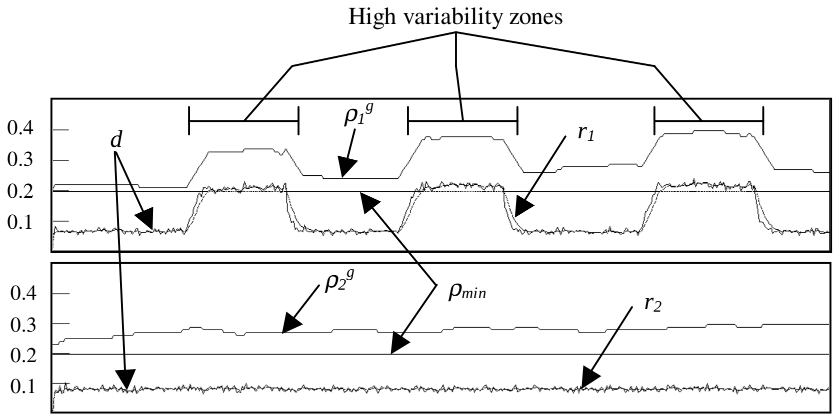

Figure 11. Evolution of the global differences, radii and global vigilances associated to classes 1 and 2.

ART networks have indeed been employed in a wide variety of problems. Amongst these we could mention the extraction of rules from massive data for weather forecasting [21], recognition of aerial images [20] or written characters [23], identification of patterns in turbulent fluids [17], creation of semantic associations between terms on a text database [24], detection of patterns in satellite images [28], recognition and retrieval of aircraft parts in databases [12], etc. There also exist applications aimed at the automatic monitoring of signals in chemical plants [29] and nuclear power stations [22].

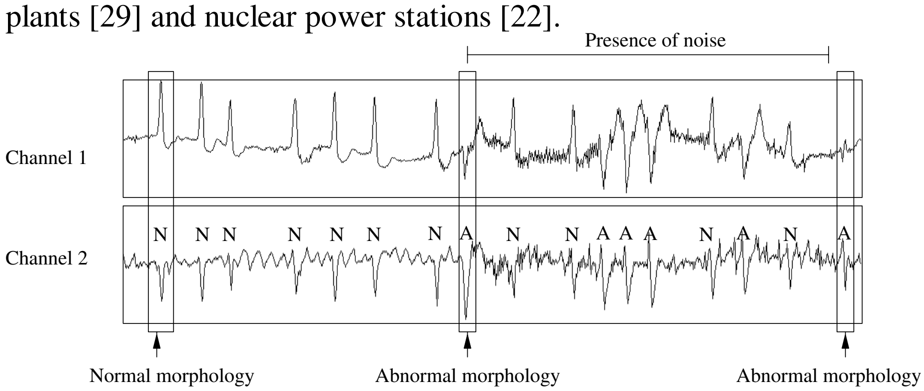

Figure 12. ECG interval on 2 derivations in which beats with normal (N) and aberrant (A) morphologies are indicated. A signal interval with noise present is also indicated.

Finally, one of the principal application environments of these networks is biomedicine, especially cardiovascular medicine, where we find examples aimed at the extraction of information from ECG signals ([26],[27],[18]) or the prediction of the risk of myocardial infarct on the basis of clinical and electrocardiographic data [13]. It is in this setting that the problem we deal with in this section is taken, the recognition of morphological patterns in heart beats, more specifically from ventricular activation complexes or QRS complexes, on multiple ECG derivations ([1],[15]). Figure 12 shows a signal interval over 2 derivations, in which different beats can be seen. Those beats associated to normal cardiac activity (N) have similar morphologies (normal morphology). On the other hand, cardiac complication generally become apparent at the electrocardiographic level, associated to beats with aberrant morphologies, such as the ones labeled "A" in the figure. The presence of different points of origin of the heart beats generally means an unfavorable prognosis for the patient, due to which it is of great importance to recognize this situation in real time. Different origins normally give rise to different morphologies in the electrical manifestation of the beat associated to them, due to which it is essential to detect all the beat morphologies that are produced over time and classify each new beat detected in accordance with them. In Figure 12, a signal interval with significant noise can also be seen; this distorts the morphology of beats and makes its correct morphological characterization difficult.

Although this problem belongs to the already classic field of pattern recognition, it has a series of typical properties which make it especially complex and, on the other hand, suitable for the use of MART in its resolution:

- The morphological characterization of beats on various ECG derivations is very important, given that it significantly increases the information available in the single-channel case. This, then, is a problem for which the use of a multichannel pattern recognition system such as MART is suitable.

- Heartbeat morphologies vary substantially between different patients and according to the derivation of the ECG under consideration, and even for the same patient over time. Thus it is impossible to construct a sufficiently representative training set, which renders supervised approaches inadequate for this problem.

- Any new morphology is important and should be learned immediately, since it may reflect the appearance of complications in the mechanism of generation and propagation of beats. Furthermore, any morphology may change substantially throughout processing. All this requires the learning of new classes as well as the adaptation of the ones that have already been learned, i.e., to resolve the dilemma between stability and plasticity. This need precludes solutions involving off-line learning, making ART architecture-based networks prime candidates for this problem.

- The determination of a reduced set of representative characteristics of the heart-beat morphology may turn out to be extremely complex, which would make operation directly onto the ECG signal itself prudent.

- The signal may be contaminated by noise from different origins, which may significantly alter the morphology of the beats. Thus the pattern recognition system must be able to adapt its discrimination capability according to the noise level or, alternatively, to the morphological variability shown by the input patterns at each instant. As they use a fixed vigilance during the processing, ART networks do not allow such an adaptation, as opposed to MART, which does so by means of its capability to adapt its vigilance parameters.

- Classes formed by morphological patterns that are very similar amongst themselves may co-exist with others made up of patterns with a high level of variability in their morphology. As a consequence, it is advisable to adapt the discrimination capability of the system to the typical characteristics of each class. MART demonstrates this property, as it uses a different vigilance for each class, with an evolution that depends on the patterns assigned to it.

- The appearance of artifacts that imitate true beats leads to the creation of spurious morphological classes, which should be rejected. With this aim, MART's ability to dynamically suppress classes is interesting, because it favors the elimination of those classes with the lowest degree of representativeness (class credit) amongst the input patterns.

We now go on to describe how this problem has been resolved using MART. Our data set consists of a group of 20 electrocardiographic registers from the MIT-BIH Arrhythmia Database [19]. This database was chosen due to it being widely known, and due to the high number of beat morphologies that it contains. Furthermore, a morphological labeling of its beats is supplied, which allows a rigorous validation of the solution applied to the problem. The input to MART in each channel (the MIT-BIH database has $I=2$ signal channels) is $J=128$ ECG samples corresponding to an interval of 256 msec., which included the whole of the QRS complex associated with each of the beats analyzed. The maximum number of classes to be learned was set at $K=15$. Amplitude scaling was performed on the input patterns so that its maximum value was 1 and its minimum value 0. The distance measure used $f(\mathbf{x},\mathbf{y})$ is the city-block distance between the input pattern in each channel and the expected value of each class, as is shown in the following expression:

$$f(\mathbf{E}_i, \mathbf{z}_{ik}) = \frac{1}{J}\sum_{j=1}^{J}\left| E_{ij} - \frac{z_{ijk} - \min\limits_{l=1,\ldots,J}\{z_{ilk}\}}{\max\limits_{l=1,\ldots,J}\{z_{ilk}\} - \min\limits_{l=1,\ldots,J}\{z_{ilk}\}} \right| \tag{18}$$

This distance, traditionally used in electrocardiographic pattern recognition [11], was chosen due to its intuitive character (area between the vectors to be compared), its relative ease of calculation and the existence of works that prove its superiority with regard to other distance measurements on the ECG signal [25]. It should be borne in mind that the expected value of the class k in the channel i, $\mathbf{z}_{ik}=(z_{i1k},\ldots,z_{iJk})$, is not scaled to $[0,1]$, as occurs with the input, given the learning rule associated to it (expression (12)), which is the reason why it is necessary to include this scaling into the distance calculation. Figure 13 shows how the morphologies of the expected values of the different morphological classes are codified in the weights z_{ijk} associated to each one of them, being used in order to determine the distance with the input data.

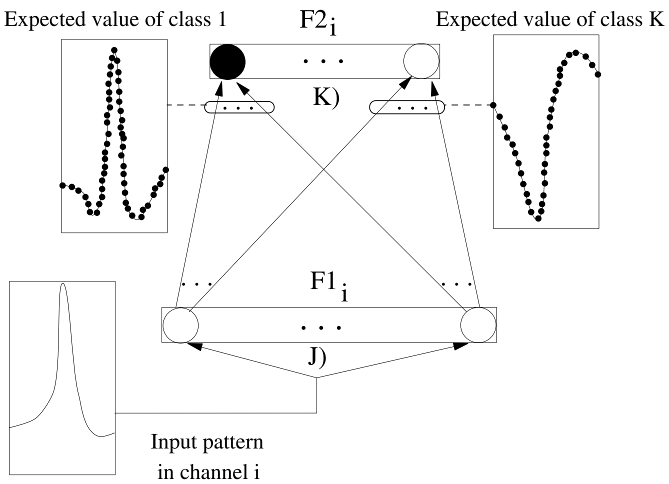

Figure 13. Values of the connection weights z_{ijk}, where the expected value of class k in channel i is stored. The shaded unit in $F2_i$ is the one that is associated to the class whose expected value has the greatest morphological similarity with the input pattern in channel i.

It is important to point out that, given that MART dynamically creates and suppresses classes, the classes created by MART do not generally coincide with those determined by the database. In this sense, the appearance of noise usually leads to the creation of redundant classes (which occurs when beats belonging to the same morphological class in the database divide into two or more classes in the MART network) although the use of adaptive channel vigilances and credits help to reduce this phenomenon.

Figure 14 shows the evolution of the channel credits and local differences, together with the number of classes created since the beginning of the processing of the register 105 of the MIT-BIH database. It can be seen that the principal increment in the number of classes is produced in instants (1) and (2), where noise appears in both channels. Nevertheless, the first of the increments, which occurs when x_1 is high, is notably greater than the second one, where x_1 has already been reduced in response to the reduction in quality in channel 1.

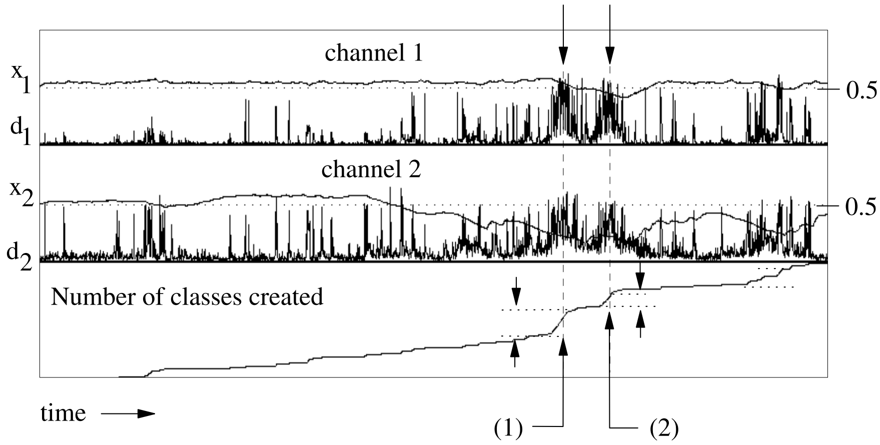

Figure 14. Temporal evolution of channel credits, local differences and the number of classes created by MART on register 105 of MIT-BIH Arrhythmia Database.

Figure 15 shows the number of classes created by MART, with and without channel credits, throughout the processing of register 200 of the MIT-BIH database. Both figures make it evident that the use of channel credits to a great degree prevents the proliferation of classes by lowering the contribution to the final result of less reliable or noisy channels. In this case, the low and high ranges for the local differences are determined by the parameters $\delta_{11}=\delta_{21}=0.05$ and $\delta_{12}=\delta_{22}=0.20$, where $\Delta x=0.0025$.

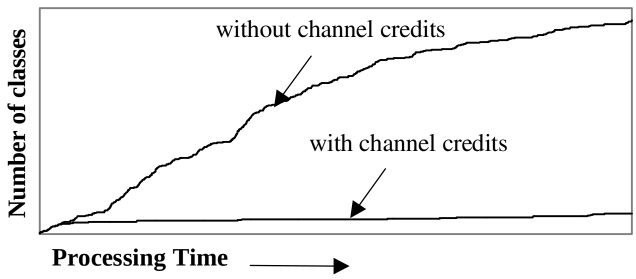

Figure 15. Temporal evolution of the number of classes created by MART on register 200, with and without channel credits.

With regard to the evolution of the class vigilances, the values used were $\rho_{min}=\rho_{ref}=0.15$, with $\Delta\rho=0.002$. More specifically, the adaptation capability of the vigilances was another reducing factor in the number of classes created by MART, as can be seen in Figure 16, by relaxing the discrimination capability in the presence of noise, thus preventing the creation of redundant classes.

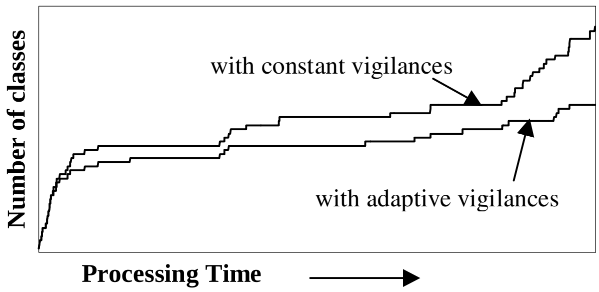

Figure 16. Temporal evolution of the number of classes created by MART during the processing of register 200, using constant and adaptive vigilances.

Finally we will briefly comment on the results obtained in the application of MART to this problem ([1],[15]). The most revealing piece of data is the lower percentage of characterization errors (1.1%), which illustrates MART's high capacity for morphological discrimination. This parameter only measures errors due to the assignation of beats to classes to which in reality they do not belong, considering any beat assigned to a redundant, but morphologically similar class to be correct. Another illustrative parameter is the redundant class creation ratio: the number of classes created by MART is on average 5.7 times higher than the real number. Although this ratio may appear to be high, it is in fact reasonable, bearing in mind, on one hand, the discrimination capability that is necessary in order to distinguish between classes, and, on the other hand, the noise level, and the high level of variability associated to some of them. However, the percentage of error due to the assignation of beats to redundant classes is 23%, in spite of the number of classes created by MART being almost six times higher than the original one. This means that the majority of the beats (over 76%) are assigned to the class to which they belong, it becoming evident that there is a noticeable capacity for limiting the population of redundant classes. Lastly, we would point out that the use of MART for the morphological pattern recognition on multichannel ECG is operative at the present inside a monitoring system of patients in a ICCU developed by our research group ([2],[15]).

6 Discussion

We do not wish to finish off this chapter without having emphasized some of the principal characteristics that MART contributes in context of neural networks and, more specifically, in that of networks based on the ART model. Amongst these, there are its express orientation towards the recognition of multichannel patterns and its ability for the continuous adaptation to the characteristics of the input pattern. In this sense, besides the obvious capacity for learning and updating of the classes to be discerned, MART offers greater operational flexibility than other neural networks aimed at pattern recognition, allowing the dynamic suppression of classes, selective evaluation of the different signal channels according to the signal quality at each moment in time and adaptation to the variability of the patterns associated to the different classes.

MART's adaptation possibilities reach the height of their functionality in the operation on patterns associated to signals that evolve continually over time. It is here where, for example, the channel credits are associated with an indirect measurement of the signal/noise ratio. These characteristics of MART, which make it unique in the field of neural computation, endow this architecture with an important practical scope. This has become evident when tackling a real and complex problem, such as that of the real time morphological characterization of beats on multichannel ECG.

Lastly, our current objectives include developing new MART-based applications, principally in the field of processing signals of a physiological origin, an area in which our group has wide experience, and to continue improving the performance of MART with regard to its capabilities for the recognition of multichannel patterns and on-line learning guided by network input patterns.

Acknowledgments

Authors wish to acknowledge the support from Xunta de Galicia, under project XUGA20608B97, and from the Spanish Ministry of Education and Culture (CICYT), under project 1FD97-0183.

References

[1] Barro, S., Fernández-Delgado, M., Vila, J.A., Regueiro, C.V. and Sánchez, E., (1998), "An adaptive neural network for the classification of electrocardiographic patterns on multichannel ECG," *IEEE Eng. in Medicine and Biology Magazine*, vol. 17, pp. 139-150.

[2] Barro, S., Presedo, J., Castro, D., Fernández-Delgado, M., Fraga, S., Lama, M. and Vila, J., (1999), "Intelligent tele-monitoring of critical patients," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 4, pp. 80-88.

[3] Bartfai, G., (1996), "An ART-based modular architecture for learning hierarchical clusterings," *Neurocomputing*, vol. 13, pp. 31-45.

[4] Carpenter, G.A. and Grossberg, S., (1987), "ART2: Self-organizing of stable category recognition codes for analog input patterns," *Applied Optics*, vol. 16, no. 23, pp. 4919-4930.

[5] Carpenter, G.A. and Grossberg, S., (1988), "The ART of adaptive pattern recognition by a self-organizing neural network," *Computer*, vol. 21, pp. 77-87.

[6] Carpenter, G.A. and Grossberg, S., (1990), "ART3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures," *Neural Networks*, vol. 3, pp. 129-152.

[7] Carpenter, G.A. and Grossberg, S., (1991), "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759-771.

[8] Carpenter, G.A., Grossberg, S. and Rosen, D.B., (1991), "ART-2A: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition," *Neural Networks*, vol. 4, pp. 493-504.

[9] Carpenter, G.A., Grossberg, S. and Reynolds, J.H., (1992), "ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, vol. 4, pp. 565-588.

[10] Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H. and Rosen, D.B., (1992), "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698-712.

[11] Cashman, P., (1978), "A pattern-recognition program for the continuous ECG processing in accelerated time," *Computers and Biomedical Research*, no. 11, pp. 311-323.

[12] Caudell, T.P., Smith, S.D.G., Escobedo, R. and Anderson, M., (1994), "NIRS: large scale ART-1 neural architectures for engineering design retrieval," *Neural Networks*, vol. 7, no. 9, pp. 1339-1350.

[13] Downs, J., Harrison, R.F., Kennedy, R.L. and Cross, S.S., (1996), "Application of the fuzzy ARTMAP neural network model to medical pattern classification tasks," *Artificial Intelligence in Medicine*, vol. 8, pp. 428-493.

[14] Fernández-Delgado, M. and Barro, S., (1998), "MART: A multichannel ART-based neural network," *IEEE Transactions on Neural Networks*, vol. 9, pp. 139-150.

[15] Fernández-Delgado, M., (1999), *MART: una nueva arquitectura de computación neuronal y su aplicación al diseño de un sistema inteligente de monitorización de pacientes*, Ph.D. Thesis, Dept. of Electronics and Computer Science, University of Santiago de Compostela. In Spanish.

[16] Fernández-Delgado, M., Presedo, J. and Barro, S., (1999), "Multichannel pattern recognition neural network," *Proceedings of the International Work-Conference in Artificial and Natural Neural Networks (IWANN-99)*, Alicante (Spain), pp. 719-729.

[17] Ferre-Gine, J., Rallo, R., Arenas, A. and Giralt, F., (1996), "Identification of coherent structures in turbulent shear flows with a fuzzy ARTMAP neural network," *International Journal of Neural Systems*, vol. 7, no. 5, pp. 559-568.

[18] Ham, F.M. and Han, S., (1996), "Classification of cardiac arrhythmias using ARTMAP," *IEEE Transactions on Neural Networks*, vol. 43, no. 4, pp. 425-430.

[19] Harvard-MIT division of Health Sciences and Technology, "The MIT-BIH arrhythmia database CD-ROM," (1992), 2nd ed.

[20] Healy, M.J., Caudell, T.P. and Smith, D.G.S., (1993), "A neural architecture for pattern sequence verification through inferencing," *IEEE Transactions on Neural Networks*, vol. 4, no. 1, pp. 9-20.

[21] Healy, M.J. and Caudell, T.P., (1997), "Acquiring rule sets as a product of learning in a logical neural architecture," *IEEE Transactions on Neural Networks*, vol. 8, no. 3, pp. 461-473.

[22] Keyvan, S., Durg, A. and Nagaraj, J., (1997), "Application of artificial neural networks for the development of a signal monitoring system," *Expert Systems*, vol. 14, no. 2, pp. 69-79.

[23] Kim, H.J., Jung, J.W. and Kim, S.K., (1996), "On-line Chinese character recognition using ART-based stroke classification," *Pattern Recognition Letters*, vol. 17, pp. 1311-1322.

[24] Muñoz, A., (1996), "Creating term associations using a hierarchical ART architecture," *Lecture Notes on Computer Science*, Springer-Verlag, vol. 1112, pp. 171-177.

[25] Rappaport, S.H., Gillick, L., Moody, G.B. and Mark, R.G., (1982), "QRS morphology classification: quantitative evaluation of different strategies," *Computers in Cardiology*, pp. 33-38.

[26] Suzuki, Y. and Ono, K., (1992), “Personal computer system for ECG ST-segment recognition based on neural networks,” *Medical & Biological Engineering & Computing*, vol. 30, pp. 2-8.

[27] Suzuki, Y., (1995), “Self-organizing QRS-wave recognition in ECG using neural networks,” *IEEE Transactions on Neural Networks*, vol. 6, no. 6, pp. 1469-1477.

[28] Waldemark, J., (1997), “An automated procedure for cluster analysis of multivariate satellite data,” *International Journal of Neural Systems*, vol. 8, no. 1, pp. 3-15.

[29] Whiteley, J.R., Davis, J.F., Mehrotra, A., and Ahalt, S.C., (1996), “Observations and problems applying ART2 for dynamic sensor pattern interpretation,” *IEEE Transactions on Systems, Man and Cybernetics*, Part A: Systems and humans, vol. 26, no. 4, pp. 237-423.