MIXED ACOUSTIC EVENTS CLASSIFICATION USING ICA AND SUBSPACE CLASSIFIER

Georges Linares, Pascal Nocera, Henri Meloni

C.E.R.I. 339 Chemin des Meinajariès
BP 1228-84911 Avignon Cedex 9
France
e-mail: linares,nocera,meloni@univ-avignon.fr

ABSTRACT

This paper describes a new neural architecture for unsupervised learning of a classification of mixed transient signals. This method is based on neural techniques for blind separation of sources and subspace methods. The feed-forward neural network dynamically builds and refreshes an acoustic events classification by detecting novelties, creating and deleting classes. A self-organization process achieves a class prototype rotation in order to minimise the statistical dependence of class activities. Simulated multidimensional signals and mixed acoustic signals in real noisy environment have been used to test our model. The results on classification and detection model properties are encouraging, in spite of structured sound bad modeling.

1. INTRODUCTION

Contrary to what happens in laboratory conditions, acoustic signal acquisition in real situations may be achieved in noisy and varying environments. Noise modeling by supervised learning assumes noise predictability. Unfortunately, it is difficult to obtain a suficient description of all sounds able to disrupt an acoustic information processing system, because of large noise and context variability. Therefore, low a priori knowledge based techniques provide interesting solutions to this noise unpredictability in complex environments. That is one of the reason why there has been a growing interest in unsupervised learning rules over the past years, especially noticeable in linear and nonlinear hebbian learning rules for blind separation of sources [1] [2] [3] [4] [5].
In this paper, we propose a new neural architecture for mixed acoustic events classification inspired by neural methods for source separation and Subspace Classifier [6] [7].

2. PRINCIPLE

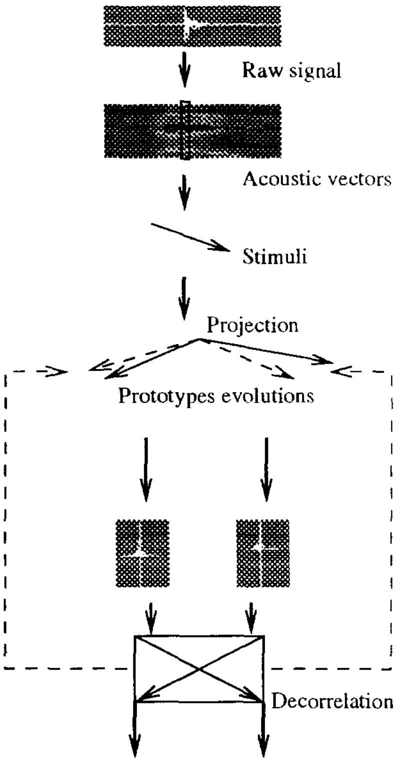

Figure 1: Principle of neural processing and adaptation to input signals

We use a feed-forward neural network for acoustic events detection and classification. Each output cell is associated with an event class, and several output cell activities are considered as simultaneous presence of events of different classes. The unsupervised neural classifier self-organizes in order to adapt itself to environment evolutions. This self organising process is made on line, by detecting novelties, creating and deleting classes. This first data space modeling reduces input space dimension to a smaller one. A second process computes a decorrelation matrix which achieves prototype rotations in order to minimise the second order moments of the network output. Decorrelation operator application on network outputs is equivalent to a class prototype rotation. So, the computed operator is applied directly to prototype matrix, and the system stabilizes itself in a state of uncorrelated output cell activities. The figure 1 shows successives process applied.

3. ARCHITECTURE

The neural net has two fully inter-connected layers. The input layer receives the coefficients of stimuli vectors. The output layer has one cell per class, and another for novelty.

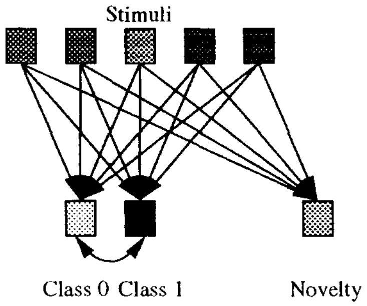

Figure 2: Architecture of neural network

Each class is represented by a prototype $P_i(t)$, an instantaneous activity and an inertia $I_i(t)$. Inertia is computed from the temporal signal of the cell activity by a classical alpha-beta filter:

$$\mathrm{I_i(t+dt)} = \beta||\mathrm{act_i(t)}|| + (1-\beta)\mathrm{I_i(t)}$$

The choise of the parameter β determines the persistency of network memory.

Cell activities are computed by the projection of stimuli vectors on prototype space. It is assumed that class prototypes are linearly independent, that is implicitly respected in the new class acquisition stage. The output vector A_t is computed at time t by

$$\mathrm{A_t} = \mathrm{P^+(t)V_t}$$

where $V(t)$ is the stimulus vector, P the prototype matrix (each column of P is a prototype vector), and P^+ the pseudoinverse of P.

Consequently, the weight matrix is the pseudoinverse of the prototype matrix.

$$\mathrm{W_t} = \mathrm{P^+(t)}$$

The novelty cell activity is computed by:

$$\mathrm{act_0(t)} = \frac{||\mathrm{V(t)} - \mathrm{V_c(t)}||}{||\mathrm{V_t}||}$$

where $V_c(t)$ is the component of $V(t)$ inside the prototype space. We do not consider several simultaneous class activations as a competition between classes, with a winner. We rather consider several significative output cell responses as the simultaneous presence of different event classes. This is possible only if the linear mixture of "physical" events corresponds to the same linear combination of input vectors. For acoustic applications, only the additive noise will be correctly modeled and separated from the useful signal. Therefore, in such an acoustic context, linear transformation will be used for converting one-dimensional signal into input acoustic vectors sequence, such as FFT in sliding temporal windows.

4. DYNAMIC CLASSIFICATION LEARNING

The system is initially empty, therefore there are no known classes. A new class is created when novelty cell activation exceeds a fixed vigilance threshold, with the input vector as the new class prototype. The latter is necessarily linearly independant from the current prototype base, because the novelty cell represents a distance betwen prototype subspace and input vector. This distance is computed from the input vector component which is orthogonal to the prototype space. Low inertia means low class representativeness. So the system permanently scans class inertias. If one of them is lower than a deletion threshold, then the class will be killed.

This class integration and deletion on-line process induces a stabilisation of subspace dimension, which depends on the thresholds and the input variability. This stabilisation is necessary, because the system sensitiveness depends from number of class known.

5. PROTOTYPE EVOLUTIONS

In this class acquisition method, the first input vector is always in the prototype base. In complex environments, the presence of several events at system initialisation is likely. In this case, the new prototype is a mixture of different potential prototype classes. In such a situation, using only the above described learning process may lead to bad classification. Here is an example of a catastrophic scenario:

A,B,C,D are acoustic events, V(t) stimuli vector at

time t.
t=0; V(0) = A+B; V0=A+B is integrated
t=1; V(1)=A+C=V0-B; V1=A+C is integrated; V0 and V1 are actives
t=2; V(2)=A+B+C+D; V2=A+B+C+D is integrated; V0, V1, V2 are actives
etc..

This shows that the integration and deletion process is not sufficient for good classification: sufficient inertias remain for the classes not to be killed (in our example, prototype V0 is always active). The bad position of the first class prototype will induce bad global classification, as well as bad clustering. One consequence of such a situation will be high class activities correlation. Therefore, we can assume that different event classes must be statistically independent.

The synaptic weight evolution rule is based on the minimisation of the class statistical dependence. We use a source separation neural method developed by [4]. This method is able to recover original nonstationary and statistically independent signals from their linear mixtures. This neural network has two fully connected layers. Its transfer function is:

$$\mathrm{S(t) = (I + C)^{-1} Y(t)}$$

where S is the network output vector, Y the input vector, and C the network weights matrix. The cost function is computed from second order moments of output signals. The originally learning rule is :

$$\mathrm{T}\frac{\mathrm{dC}}{\mathrm{dt}} = (\mathrm{C}^{\mathrm{T}})^{-1} * (\mathrm{diag} < \mathrm{S(t)}, \mathrm{S(t)}^{\mathrm{T}} >)^{-1} < \mathrm{S(t)S(t)}^{\mathrm{T}} > -\mathrm{I}$$

The proof of this algorithm convergence, and more information about it can be found in [4]. This algorithm can only be used for non-stationary signals separation, and input signals must have zero means and unity variances. Therefore, cell activities are normalised in order to respect these constraints. The non-stationary input signals constraint may not to be respected in the case of stationary class activities. Therefore, we have added a sigmoidal term to the original learning rule, which allows to regularise learning dynamics in locally silent areas. The learning rule used is:

$$\frac{dC}{dt} = F\ (I + C^T)^{-1}\ (diag\ <\ s(t), s(t)^T\ >)^{-1}\ < s(t)s(t)^T > -I)$$

$$with\ F_{i,i} = \frac{(1 - e^{-2\|y_i(t)\|})}{(1 + e^{-2\|y_i(t)\|})},\ F_{i,j} = 0\ if\ i \neq j$$

The application of the uncorrelation operator C^{-1} to classifier outputs is equivalent to a prototype rotation:

$$S(t) = C_t^{-1}.W_t^+ V(t) = (W_t C_t)^+ V(t)$$

Thus, the operator is applied directly to the prototype matrix.

6. EXPERIMENTS

In the first test, the first seven alphabet pattern letters have been randomly mixed. 2000 samples of pattern mixtures have been used as input, without any information concerning original shapes and pattern number. Seven classes have been effectively found, and original patterns were recovered with low noise level (figure 3).

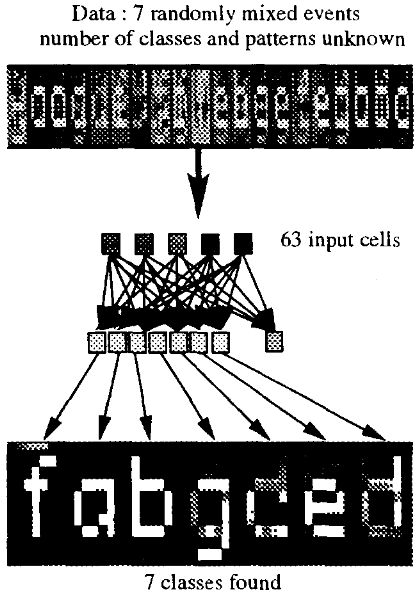
Data : 7 randomly mixed events
number of classes and patterns unknown
63 input cells
7 classes found

Figure 3: Detection and clasification of multi-dimensional simulated events

The second test (figure 4) shows a signal in a real subaquatic environment, and system responses. The first line is the signal spectrogram obtained by a FFT computed in a sliding temporal window. There is a vector of 256 coefficients for each 10 ms. The second line shows the novelty cell activity. New classes are detected from the local maximums of that curve. The third line represents the prototype space dimension relatively to the input space dimension. The other lines shows most meaningful class activities.

Recurent events have been well detected. Their first occurence set off high novelty cell response, and other occurences of class 3 and class 4 events have been well classified. At time 350, a class 3 event and a class 4 event have been occured simultaneously, and the system have detected correctly the two events. The fifth class has been created from the first (and only) isolated event occurence. The last isolated event has occured a time 400. It is a noise which have large temporal structure. These temporal evolutions in large frequency domain have produced several classes creation. It has been badly modeled. That illustrates bad time representation in our model.

7. CONCLUSION AND FUTURE PROSPECTS

The first results concerning model properties for classification and novelty detection seem to be encouraging, in spite of the structured sound bad modeling. Using other source separation techniques based on high order statistics could improve the system's performances. We are currently working in grouping low level class in meta-classes encoding temporally correlated acoustic events.

8. REFERENCES

[1] G. Burel. Blind separation of sources: a nonlinear neural algorithm. *Neural Networks*, 5:937–947, 1994.

[2] P. Comon. Independent component analysis - a new concept ? *Signal Processing*, 36:287–314, 1994.

[3] C. Jutten & J. Herault. Blind separation of sources, part i: an adaptative algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[4] K. Matsuoka & M. Ohya & M. Kawamoto. Neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):441–419, 1995.

[5] E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935, 1992.

[6] T. Kohonen. *Self organisation and Associative Memory*. Springer Series in Information Sciences, third edition, 1989.

[7] E. Oja & T. Kohonen. The subspace learning algorithm as a formalism for pattern recognition and neural networks. In *International Conference on Neural Networks*, pages 227–284, San Diego, 1988. IEEE.

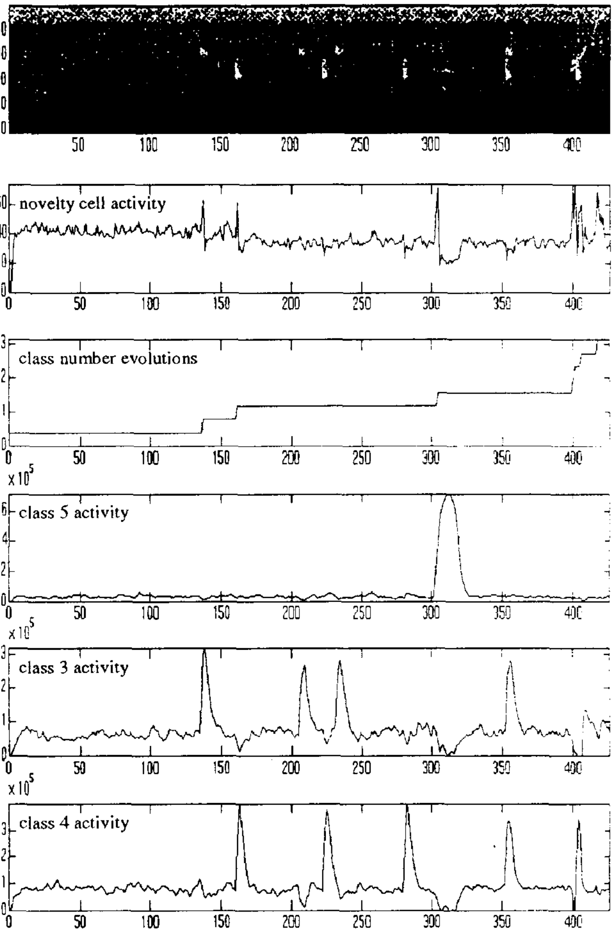

Figure 4: Detection and clasification of acoustic events