Music Rhythm Recognition Through Feature Extraction and Neural Networks

Giovanna Morgavi, Mauro Morando, Daniela Baratta*

Institute for Electronic Circuits, National Research Council, via De Marini 6, 16149 Genova Italy
*PhD at Institute for Electronic Circuits, National Research Council, via De Marini 6, 16149 Genova Italy

Fax: 39-010-6475200; e-mail: **morgavi@ice.ge.cnr.it;** http//ge.cnr.it

Abstract

In this paper a procedure to solve the problem of recognition and classification of sampled musical rythms is presented. The lack of precise rules for doing this analysis makes difficult and often ambiguous the automatic execution of a cognitive process naturally performed by human brain. This procedure can be extended to the classification of any signals showing similar characteristic (i.e. EEG or ECG). Due to the complexity of the time dependence, standard procedures used for chaos characterisation (i.e. correlation dimension, Lyapunov exponents, etc) can fail. Moreover a direct usage of artificial neural network can introduce too many optimization variables. The proposed procedure can be organized in two phases: the extraction of some new type of invariant from the sampled time series and the usage of this extracted features as input for a classifying standard neural network. This system was able to distinguish between binary and ternary signals with a precision of 99%. The single rhythm was classified within an error of 5%. This system seems to be able to deal with the behaviour that characterises a musical rhythmic sequence, and to classify patterns independently of the musical instrument and tempo.

Keyword: chaotic signals, invariant extraction, signal recognition, neural networks

1 Introduction

The chaotic signal recognition and classification is a relevant scientific problem in a wide range of practical applications: from the failure detection in mechanical equipment (Malher, 1994) to illness diagnosis in medicine and biology. The analysis and classification of chaotic signal involves extracting significant features from sampled time series that lend themselves to easier interpretation. Real chaotic signals are usually difficult to be directly processed. In the classification field, the complexity of this type of signal cannot be sufficiently represented by the computation of some parameters like dimensionality, Lyapunov exponents and entropy. The large number of zeros collapse

the usual procedures to compute these parameters. On the other hand, the large number of noisy data doesn't allow the direct usage of artificial neural network. Dimensionality reduction as a pre-processing step for classification of signals before using a neural network could be more effective than trying to classify unprocessed signals because smaller networks can be used and it leads to easier classification.

2. Data

In this paper a procedure to solve the problem (Rabiner et al., 1975) of musical rhythm recognition and classification is presented.

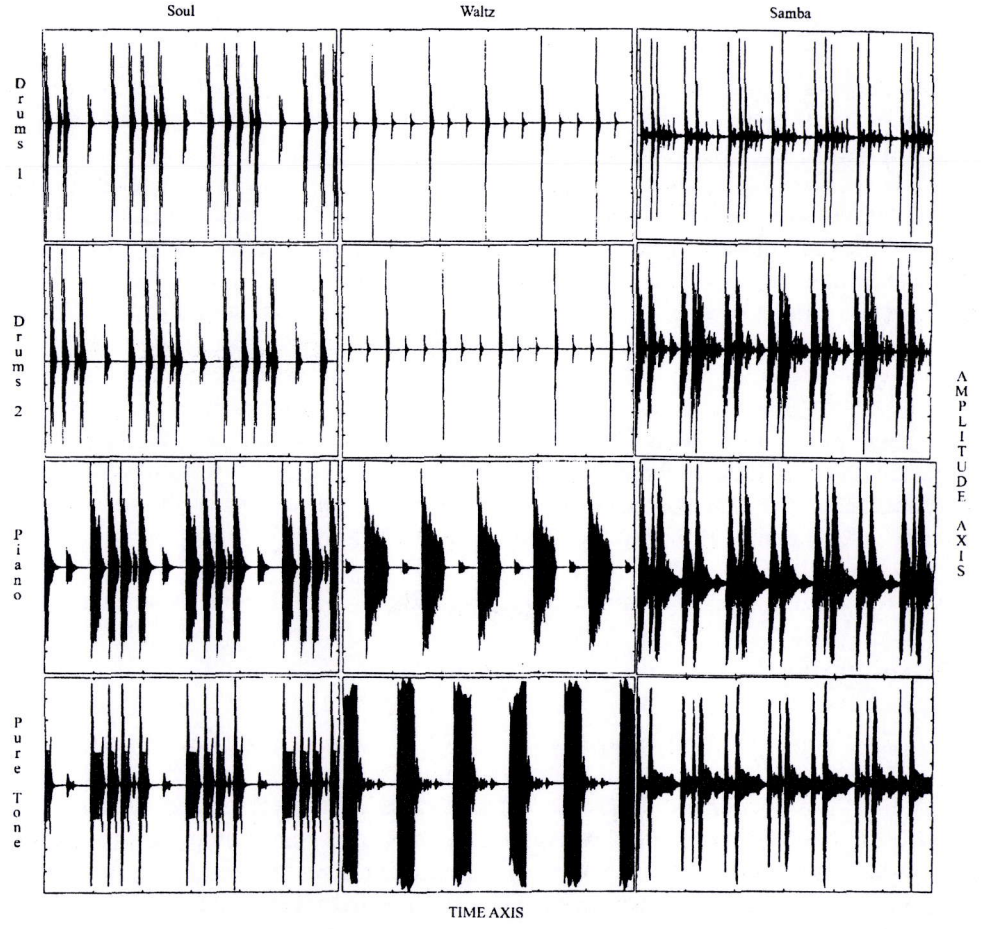

Figure 1 Time sampled signal of soul, waltz and samba rhythms played with different instruments

This procedure can be extended to the classification of any signals showing similar characteristic (i.e. E.E.G. or E.C.G.). Input signals are sequences of discrete values obtained by sampling at frequency of 44100 Hz musical rhythm chosen among six different types: tango, soul, samba (duple rhythms), waltz, joropo and march (triple rhythms) (Martini et al.,1995). These rhythms were generated by a MIDI system, playing different instruments (two kind of drum, one piano, one pure tone (sinusoidal generator)). The music was played with four different tempos: 108, 120,138 and 160 beats per minute and with 3 different levels of rhythm complexity. Each sequence was played for 60 seconds. In figure 1 and in figure 2, as an example, the plot of some rhythms with minimum complexity are shown. The values of sampled signals were normalised within the range [-1,1] and each plot shows the behaviour of the first 300000 samples.

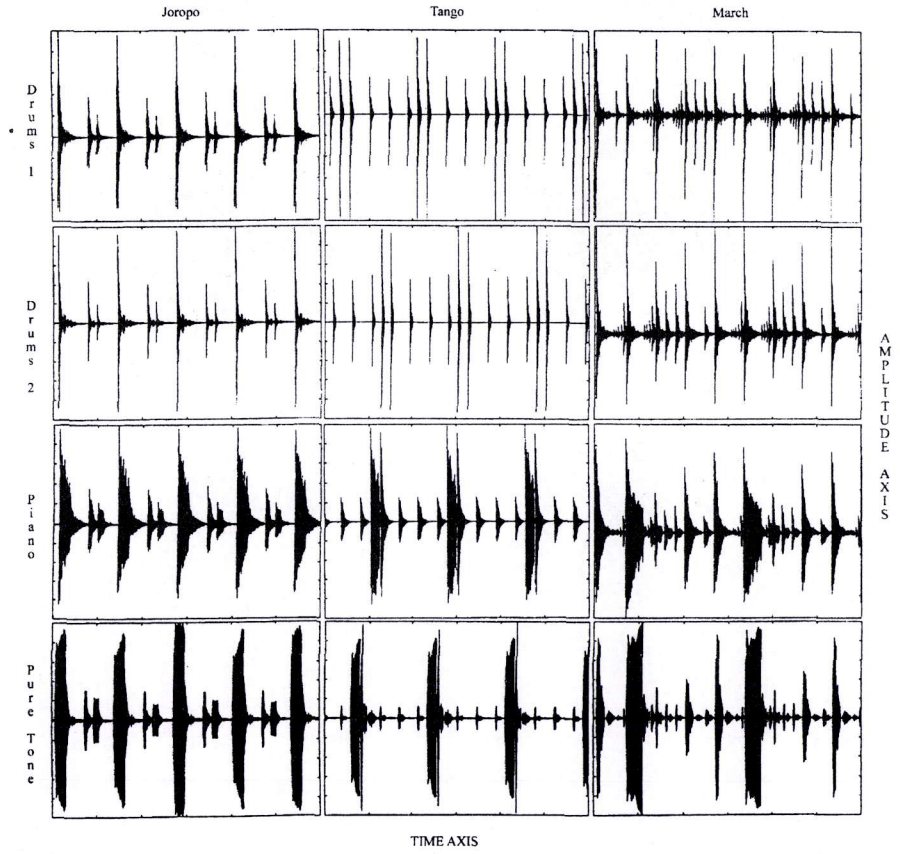

Figure 2 Time sampled signal of joropo,tango and march rhythms played with different instruments

Figure 1 and Figure 2 show that the signals are not periodical. The lack of precise rules for doing the analysis of such a signal makes difficult and often ambiguous the automatic execution of a cognitive process naturally performed by human brain. The emergence of the essential role of the temporal dimension in the dynamics of the sensory cortex for invariant extraction and dynamic reconstruction of a complex input, suggest an alternative to the standard classification systems.

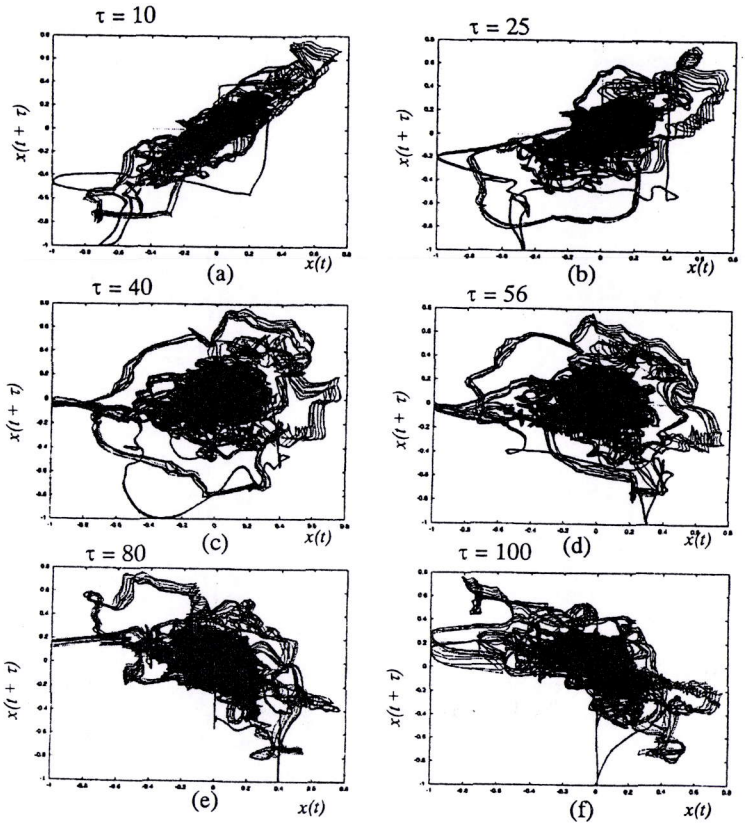

Figure 3: Phase portraits for rhythm t120t1m (tango, tempo120, Drum number 1, minimum complexity) changing the τ value

The alternative consists in considering the input as a dynamic system (at least chaotic) from which the cognitive system extract interesting invariants to construct an inner dynamic representation of the input. Since in music sampled signals the time dependence is very complex, standard procedures used for chaotic signal characterisation (i.e. correlation dimension, Lyapunov exponent, entropy, etc) can be

insufficient. Moreover the large number of zeros can induce false high dimensions in usual algorithms. The proposed procedure can be subdivided in two consecutive phases: the first of which is the extraction of invariant from the sampled time series and the second one is their usage for classification through a standard neural network.

3. The feature extraction

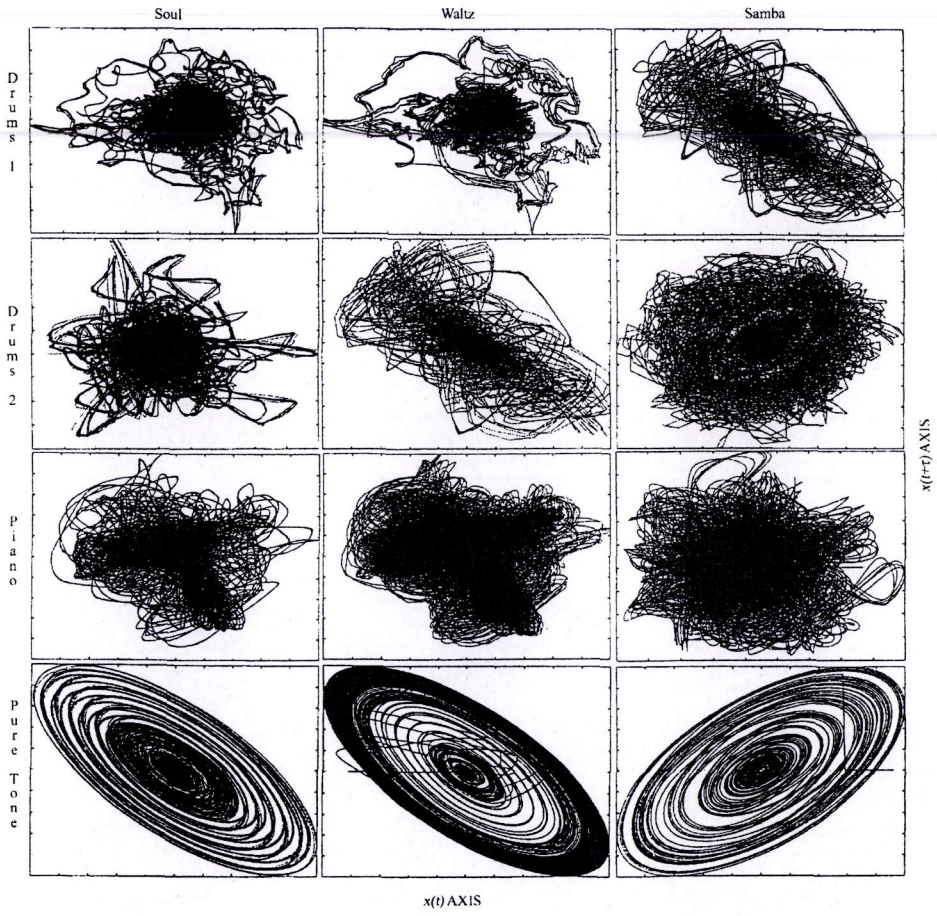

Figure 4 Phase portrait for soul, waltz and samba rhythms.

A real dynamic system may be considered as governed by a large number of freedom degrees. It can be described as: $\dot{x} = F(x)$,were $\dot{x}$ is a vector in $\Re^m$ such that each component represents a dynamic system characteristic variable and F is a vector function. If a component $x_1(t)$ of x can be measured, it is possible to extract information on the model of the physical phenomenon if a system $\dot{y} = G(y)$ can be written, were y is a vector in $\Re^k$ such that $y(t)$ is coincident with the measured $x_1(t)$ (m, is usually not known). A good reconstruction process can be mathematically described as an embedding.

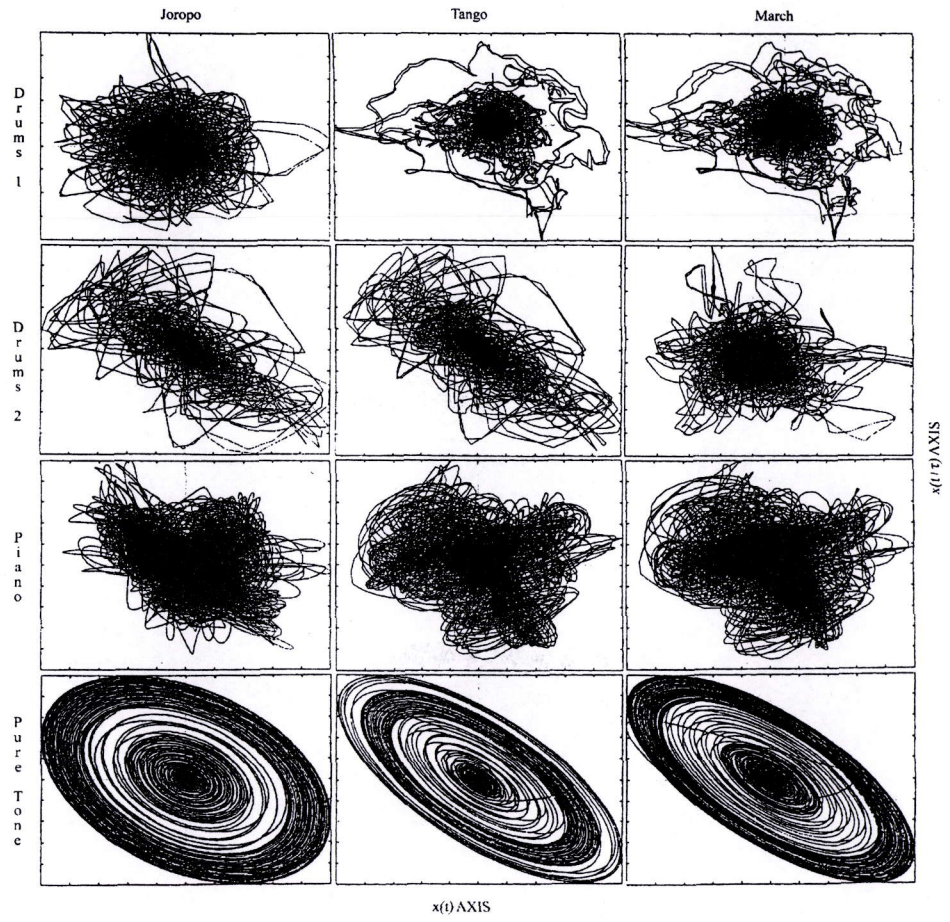

Figure 5 Phase portrait for joropo, tango and march rhythms.

The embedding is a smooth mapping that is a diffeomorphism from the manifold to a sub manifold of $\Re^k$ space, where k is the embedding dimension. In the literature it is known that multidimensional phase-portraits could be constructed from measurements of a single scalar time series (Takens, 1980). Practically, the values of $y(t)$ are k different values of x_1 sampled at different time steps before t.

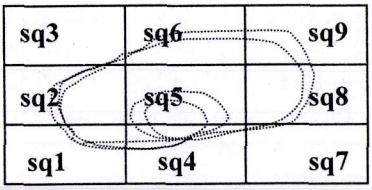

Figure 6 Grid on the phase protrait diagram

Time delays τ such that: $y_i(t) = x_1(t + (i-1) * \tau), \quad (i = 1, k)$ can be used. The choice

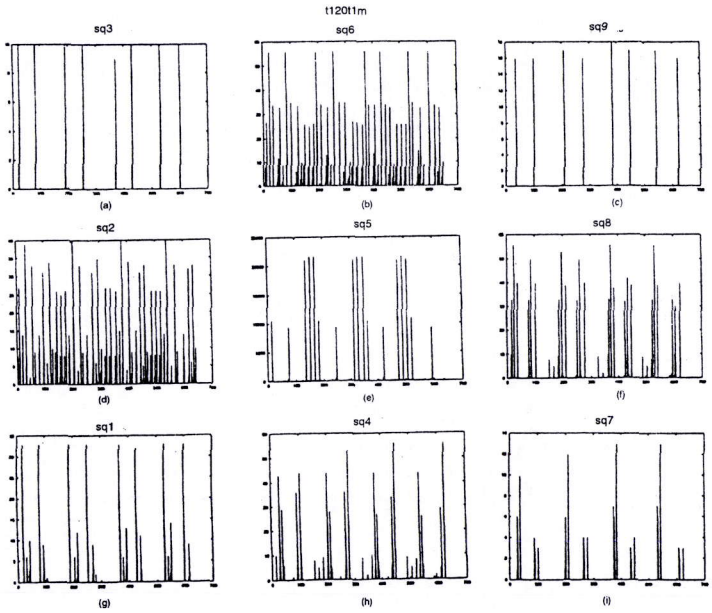

Figure 7 New time series generated for a tango rhythm played with drum1

of a 'good' τ is, in principle, arbitrary as long as it is not related to the samples, but, in practice, if it is too small the coordinates become singular, so that $x_1(t)$ is coincident with $x_1(t+(k-1)*\tau)$. If τ is too big, chaos makes $x_1(t)$ and $x_1(t+(k-1)*\tau)$ casually disconnected. In practice τ is often chosen by trial and error, starting with a low value and increasing it, searching optimal results. In literature both zero of autocorrelation function and minimum of mutual information (ott et al., 1994) are suggested. Both these suggestion were not usable due to the characteristics of this type of signal (i.e. number of zero values). From the other hand, our goal is to find a good τ to underline differences between rhythms. The best τ will not be the best respect the single signal, but a good one to classify the differences between them. The classification based on differences is a typical human behaviour: no one is able to recognise white noise, but everybody can identify even a small change. In Figure 3 phase portraits for a tango played by drum with different time delay are shown: they show that if τ is too small or too large the plot unrolls on a diagonal. The empiric research of a good τ required many trial: finally we chosen τ=56.

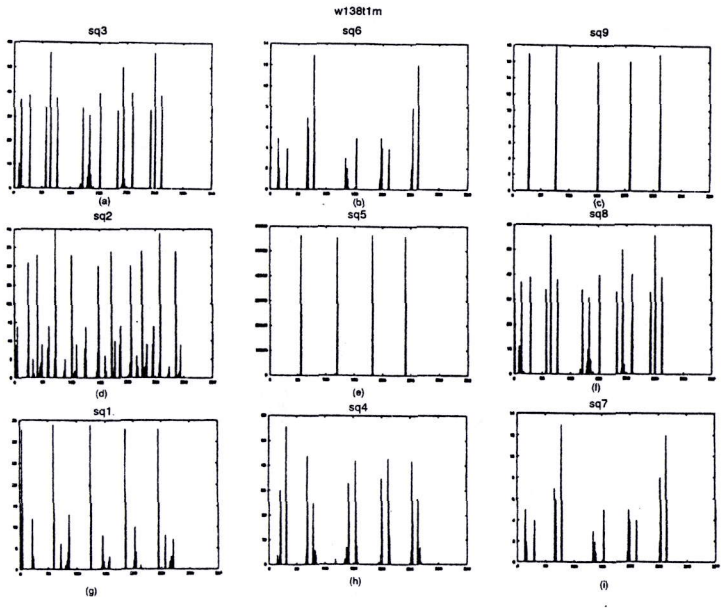

Figure 8 New time series generated for a waltz rhythm played with drum1

Figure 4 and in Figure 5 phase portraits for rhythms of Figure 1 and Figure 2 with τ=56 are shown. In the rhythm the concept of time is foundamental: the state space has been

divided by 9 squares as shown in Figure6. A new time series containing the information of the time step occupancy in each square have been generated: for each square, and each sampled signal, 9 new time series containing the occupancy time in a square and the number of visiting in the others have been created. The most significant reduced time series (the one around the axis origin) has been chosen as input to the artificial neural network.

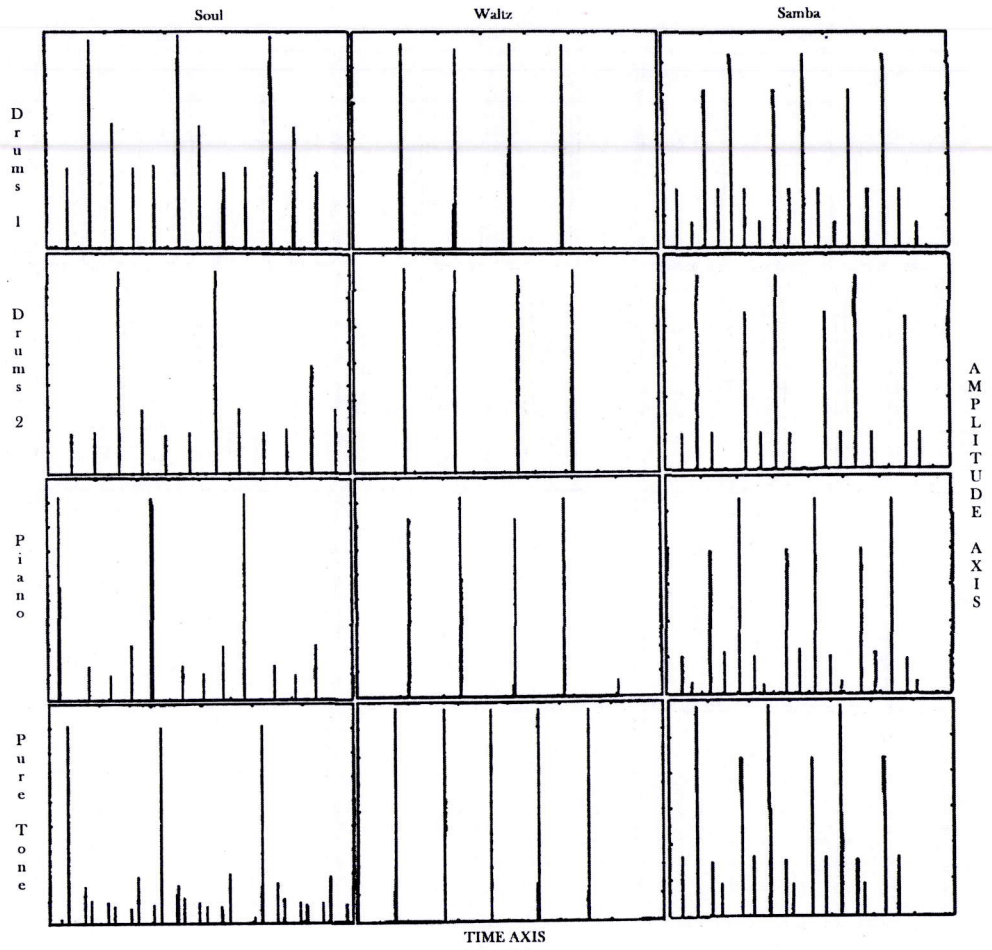

Figure 9 Central square time series for soul, waltz and samba rhythms played with different instruments

4. The neural network

The classification of different rhythms has been carried on by an artificial neural network implemented with the Back Propagation algorithm modified to reduce the convergence time. Then a Multi-Layer Perceptron was trained by epoch: the weight values were updated after the presentation of the whole training set.

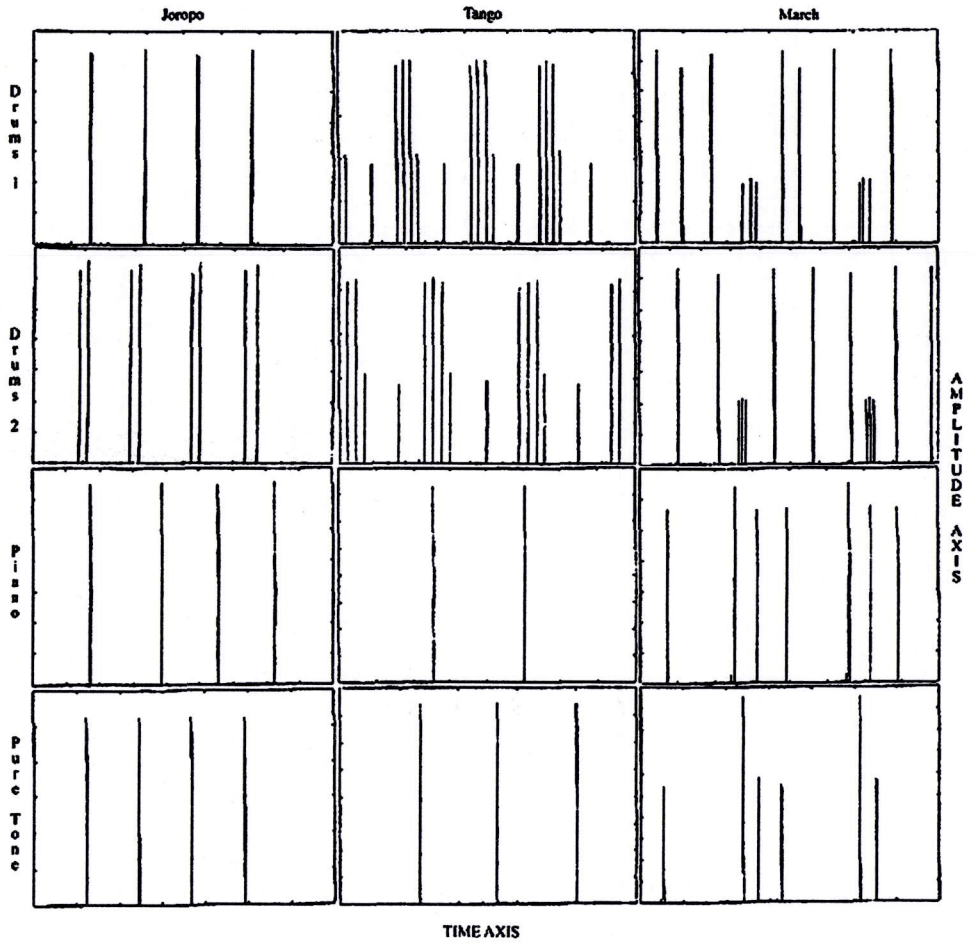

Figure 10 Central square time series for joropo,tango and march rhythms played with different instruments

The Vogl (Vogl et al., 1988) acceleration algorithm has been introduced: this procedure update also the learning rate η and the momentum α based on the network performances on the whole training set. Let define C_{new} the cost of the actual epoch and C_{old} the previous cost:

- if $C_{new} < C_{old}$ then the weight update is accepted and $\eta_{new} = K_1 \eta_{old}$; $\alpha = \alpha_0$
- if $C_{old} \leq C_{new} \leq K_2 C_{old}$ then the weight update is accepted and $\eta_{new} = K_3 \eta_{old}$; $\alpha = 0$
- if $C_{new} > K_2 C_{old}$ then the weight update is not accepted and $\eta_{new} = K_3 \eta_{old}$; $\alpha = 0$

where K_1, K_2 (>1) and K_3 (<1) are constant

The problem of the definition of the architecture of the MLP is hardly discussed in literature. It is well known that a feedforward network with two layers is sufficient to store any number of patterns if a sufficient number of hidden neuron is used (Hertz et al. 1991) .The number of neurons in the hidden layer is a concern in the application of neural networks to signal classification. A rule of thumb (Baum et al., 1989), known as the Baum-Haussler rule, is used to determine the number of hidden neurons to be used:

$$N_{hidden} \leq \frac{N_{train} E_{tolerance}}{N_{pts} + N_{output}}$$

where N_{hidden} is the number of hidden neurons, N_{train} is the number of training examples, $E_{tolerance}$ is the error tolerance, N_{pts} is the number of data points per training example, and N_{output} is the number of output neurons. This rule generally ensures that neural networks generalize, rather than memorize.

5. Results

Since the resulting classification system should be able to recognize a musical rhythm independently from the starting point, we extracted several input patterns from each input signal by shifting a window along the time axis.

Table 1 Classification percentage errors for tango, samba, march and joropo rhythms.

dimension		**Number of iteration**	α	η	**% errors on**	
input	**hidden**				**Training set**	**Test set**
20	10	7819	0.8	0.09	0	1.6
20	18	14317	0.9	0.5	0	2.1

In such a way we build 1500 training and 1500 test time series for each sampled signal.

First we analysed the classification of tango, samba, march and joropo with 120 steps per minutes. Many starting points for wait values have been tried: best results are shown in table 1.

Table 2 Classification percentage error in the training set with the whole data base for single rhythm types

	soul	**waltz**	**samba**	**joropo**	**tango**	**march**
Drum 1	1.2	1.5	1.1	1.2	1.5	0.8
Drum 2	1.3	0.5	1.2	1.7	0.7	1.1
Piano	0.7	1.2	1.7	1.7	0.9	1.6
Pure tone	1.1	1.3	1.5	0.6	1.1	1.5

With the whole data set in input the classifying system was able to distinguish between duple (tango, soul and samba) and triple (Waltz, joropo and march) rhythms with a precision of 99%. As an example in table 2 the training percentage error of classification for the single signal (played tempo120 steps per minute) is shown. This MLP was composed by 20 inputs, 800 hidden layers: the test percentage error was 8.9%. With the whole data set, single rhythm was classified within an error of 5%, the percentage generalization error was 10.1%.

6. Conclusion

In this paper, a scheme for time series classification with a neural network has been proposed: it consists of two phases: feature extraction from input patterns, and construction of the neural network classifier. A new procedure to extract features has been proved to be effective in characterizing input time series. A neural network classifier has been constructed that takes into account input features and achieves accurate results. The Baum-Haussler rule has been used to determine the number of neurons in the hidden layer. The proposed procedure seems to be able to deal with the behaviour that characterizes a difficult signal such as musical rhythmic sequence, and to classify patterns independently from the musical instrument and tempo within an acceptable error.

References

R. Mahler, (1994) Fundamental frequency Estimation of Musical Signals Using a Two-way Mismatch Procedure-, J. Acoust. Soc. of America, vol. 4, pp. 2254-2263.

L. R. Rabiner, B. Gold, (1975) Theory and Application of Digital Signal Processing, Prentice-Hall Inc., Englewood Cliffs.

C. Martini, M. Morando, M. Muselli, (1995) "Period Extraction in a Sampled Rhythmic Sound Sequence", Proc. of the 14th IASTED Conf. *on* Mod., Ident. and Control, Igis, Austria, ed. M. H. Hamza, IASTED - Acta Press Pub., pp. 89-92.

Takens (1980) "Detecting strange attractors in turbulence" Warwick Lecture Notes in Math, vol.898, Berlin, W Germany: Springer pp.366-381

Ott E., Sauer T.,Yorke J.A. (1994) "Coping with Chaos: Analysis of Chaotic Data and Exploitation of Chaotic Systems" Wiley-Interscience Pub. N.Y.

D. E. Rumelhart, G. E. Hinton and R. J. Williams, (1986) "Learning Representations by Backpropagating Errors", Nature, vol. 32), pp. 533-536.

Vogl, T.P., J.K. Mangis, A.K. Rigler, W.K. Zink, and D.L. Alkon, (1988) Accelerating the convergence of back-propagation method, Biol. Cybern., 59, 257-263.

Hertz, J., A. Krogh, R. G. Palmer, (1991) "Introduction to the theory of neural computation", Lecture notes Vol. 1, Santa Fe Institute Studies in the sciences of complexity, Addison-Wesley, Redwood City, CA 94065.

Baum, E. Haussler, D. (1989). "What size net gives valid generalization?" Neural Computation, 1, 151-160