Self-organising maps for music style perception*

Pedro J. Ponce de León and José M. Iñesta

Universidad de Alicante
Departamento de Lenguajes y Sistemas Informáticos
pjleon@mail.ono.es,inesta@dlsi.ua.es

Abstract. In this paper the capability of using self-organising neural maps (SOM) as music style classifiers from symbolic specifications of musical fragments is studied. From MIDI file sources, the monophonic melody track is extracted and cut into fragments of equal length. From these sequences, melodic and harmonic numerical descriptors are computed and presented to the SOM. The performance is analysed in terms of separability in different music classes from the activations of the map, obtaining different degrees of succeed. This scheme has a number of applications like indexing and selecting musical databases or the evaluation of style-specific automatic composition systems.

1 INTRODUCTION

There are a number of applications in computer music to the possibility of melodic fragment comparison. Two main representations of music can be found: sounds (recorded from human or computer interpretation of a music score) and symbols (representation codes independent of the sonic outcome of an interpretation). The automatic machine learning and pattern recognition techniques available, successfully employed in other fields, can be also applied in music analysis. Immediate applications are the classification, indexation and content-based search of digital musical libraries, where digitized (MP3), sequenced (MIDI) or structurally represented (XML) music can be found.

One of the tasks that can be posed is the modelisation of the music style. This means to provide the computer with the capability of discrimination between musical styles or sub-styles, or even between different composers. Even more, the computer could be trained in a given user musical taste in order to look for that kind of music over large musical databases. Such a model could be used in cooperation with automatic composition procedures to guide this process according to some stylistic profile provided by the user.

The aim of this work is to develope a system able to distinguish among a set of musical styles from a symbolic representation of a melody. We have chosen random, jazz and classical melodies for our experiments. We will investigate whether those symbols by themselves have enough information to achieve this

* This work has been funded by the Spanish CICYT project TAR: TIC2000–1703–CO3–02. The authors would like to thank F. Moreno-Seco for his valuable help.

goal or, on the contrary, there is also timbric information that has to be included for that purpose.

In this paper the input to the system are melodic lines represented symbolically in MIDI file tracks. This can be extended to feed the system with XML representations of music, in addition to MIDI files.

The key point of this work is to test the ability of self-organizing maps (SOM) [1], to automatically perform this task. SOM are neural methods able to obtain approximate projections of high-dimensional data distributions in low-dimensional spaces, usually bidimensional. With the map, different clusters in the input data can be located. These clusters can be semantically labelled to characterize the training data and also hopefully future new inputs.

1.1 Related works

In a recent paper, Rauber and Frühwirth [2] pose the problem of organising music digital libraries according to the sound features of each musical piece, in such a way that similar themes can be located clustered. This would permit the user to locate sections within the library according to stylistic similarities. The authors utilize a SOM in order to create a map of the digital library, where similar music themes can be found in zones close to one another. After finding a given music in the map, others related can be found with an exploration of the surroundings of that point, permitting an intuitive exploration of the library. This is, therefore, a content-based classification of the data (sounds in that case).

Other related work is that of Whitman and Flake [3] in which they present a system named Minnonwatch, based on neural nets and support vector machines, able to classify a sound musical fragment into a given source or artist. The system achieves a success rate of 91% with 5 different artists or sources, of 70% with 10 artists and of 46% with 21 artists.

In a similar work [4], the authors describe a system to recognize music types using an explicit-time modelling neural net that codes an abstraction of acoustic events in the hidden layer of the net representing temporal structures of the musical parts. This abstractions are then used to discriminate among different types of music. The experiments show that the system improves the recognition rate of other methods like recurrent neural nets or hidden Markov models.

In [5] the authors present a hierarchical SOM able to analyze time series of musical events. The model can recognize instances of a reference sequence (a fugue by J.S. Bach) in presence of noise, and even discriminate those instances in a different musical context. In this work, the SOM are used as sequence recognizers, using a time integration mechanism in the input layer of two SOM, arranged one on top of the other, to represent the reference monophonic melodic sequence in order to provide the SOM with the ability of processing time sequences.

In the work by Thom [6] pitch histograms (measured in semitones relative to the central pitch of the tonality and independent of the octave) are used to describe blues fragments. The pitch frequencies are used to train a SOM.

All these works pose the same problem that we face here, and most of them use digital sounds as input. Only the latter two use symbolic representations for

recognizing musical parts, not styles. The approach we propose here is to use the symbolic representation of music as the input to self-organizing maps for classification of musical fragments into a initially small set of styles. The success of this approach would permit to extend it to other styles and to apply this methodology to the huge amount of symbolic data stored in music databases all over the Internet.

2 METHODOLOGY

The monophonic melodies are extracted from the rest of the musical content in the MIDI files and preprocessed to extract melodic and harmonic descriptors. This way we have a sequence of note events. Other kind of MIDI events are filtered out. Each note can take a value from 0 to 127 (the pitch) and the duration is the distance from the event that onsets the sound of a note to the event that finishes it (there is no limit to this in theory). Note that this symbolic representation implies the lack of timbre information. It is just like a music score, containing information about music events but not about the instrument that is playing. This way, the situation is like an expert trying to classify scores.

Even dealing with monophonic melodies the search space is very vast. Nevertheless, the hypothesis is that melodies from a same musical genre may share some common features that make possible that a experienced listener is able to assign a musical style to them.

In order to have more restricted data, only with melodies written in 4/4 have been considered. A window 8 bars wide has been defined for analysis (enough to get a good sense of the melodic line), and for each window position a vector of musical descriptors is computed from the notes in the window. These vectors contain melodic and harmonic information from the melody in each window and will be the inputs for training and testing the SOM.

For the experiments we have considered, along with real melodies, other randomly-generated melodies in order to test the ability to separate well structured melodies from other non-sense musical constructions. For generating this kind of melodies each bar was divided into Q pulses (quantization) and the melody was considered as formed by three kinds of events that can appear at each pulse: note onsets, silences and continuation of the previous event.

We have restricted the note pitches to a range of [45,82], heuristically determined after an analysis of a large number of real melodies. In 8 bars we will have $8 \times Q$ events. Each melody was generated with a proportion of notes / silences / continuations among this possibilities:

1-1-1	1-1-2	1-2-1	1-2-3	2-1-1	2-1-3	2-3-1	3-1-2	3-2-1

where N-S-C indicates the probability of generating a note onset (N), a silence (S) or a continuation event (C), according to the expression $X/(N+S+C)$ where X can be N, S or C. Therefore a melody generated according to the pattern 2-3-1 will have nearly a 33% of note onset events, a 50% of silence events and a 17% of continuation events.

In the next experiments we have initially considered this set of descriptors:

- Overall descriptors:
  - Number of notes and number of silences in the melody.
- Pitch descriptors:
  - Lowest, highest (these values provide information about the pitch range of the melody), average, standard deviation (provide information about how the notes are distributed in the score).
- Note duration descriptors (these descriptors are measured in pulses):
  - Minimum, maximum, average, and standard deviation.
- Silence duration descriptors (in pulses):
  - Minimum, maximum, average, and standard deviation.
- Interval descriptors (distance in pitch between two consecutive notes, measured in semitones):
  - Minimum, maximum, average, and standard deviation.

These 18 features are melodic descriptors. This number will be increased in further experiments in order to evaluate harmonic aspects of the melodies.

3 EXPERIMENTS AND RESULTS

The experiments are divided into two phases: first, a set of random melodies with different proportions of notes, silences and continuation events are generated, and a set of real melodies, extracted from jazz standards, is built. We put the capability of SOM for this task to a test with an, a priori, easy task: to separate random musical sequences from melodies with real musical feeling. The second phase consists of substituting music of other type different from jazz for the random melodies in order to test the ability for style discrimination. Classical music was chosen and melodic samples were taken from works by Mozart, Bach, Schubert, Chopin, Grieg, Vivaldi, Schumann, Brahms, Beethoven, Dvorak, Haendel, Pagannini and Mendhelson.

For SOM implementation and graphic representations the SOMPAK software [7] has been used. For the experiments a hexagonal geometry for unit connections and a bubble neighbourhood for training have been selected. In this paper, two main kinds of map representations are shown: the Sammon projection, as a way to display in 2D the organization of the weight vectors in the weight space, and the U-map representation, where the units are represented by hexagons with a dot or label in their center. The grey level of unlabelled hexagons represents the distance between neighbour units (the clearer the closer they are). The grey level of labelled units is an average of those distances. This way, clear zones are clusters of units in the SOM, sharing similar weight vectors. The labels are a result of calibrating the map with a series of test samples and indicate the class of the sample that activates that unit more times.

3.1 Random versus jazz melodies

400 random samples have been generated and 430 jazz samples have been extracted from 54 MIDI sequences of jazz standards, all of them made up of 8 bars with a quantization of $Q = 8$ pulses per bar (64 events per melody). From them, the 18 descriptors listed above were computed. Using these sets a SOM of 16 neurons for the OX axis and 8 for the OY axis was trained. The training consisted of two stages: a coarse one of 1,000 iterations with wide neighbourhoods (12 units) and a high learning rate (0.1) and then a fine one of 10,000 iterations with smaller neighbourhood ratio (4 units) and learning rate (0.05).

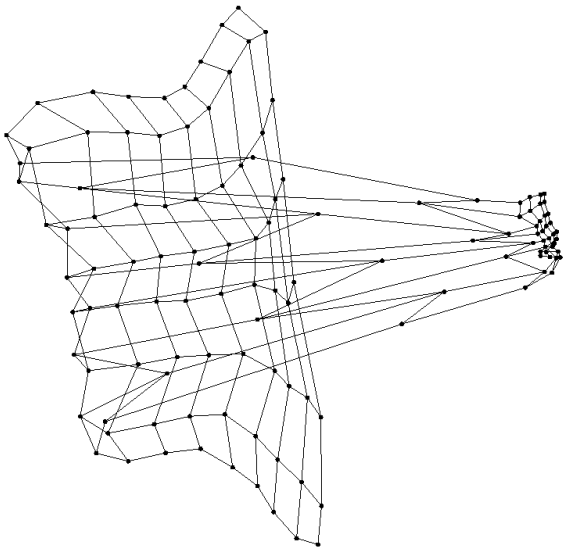

Fig. 1. Sammon projection of the 16×8 map of figure 2: random versus real melodies.

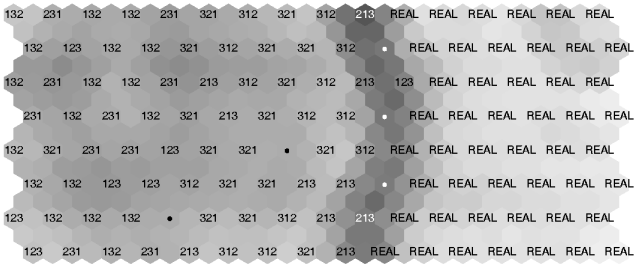

Fig. 2. SOM map for the same weights as in figure 1.

In figures 1 and 2 the Sammon projection and the SOM map are displayed after training. Note that there exists a clear gap between two zones in the map.

The small cluster on the right in Fig. 1 corresponds to the real melodies and that of the left to the random melodies. In the map, the same can be observed for the two areas clearly separated: random samples on the right and real samples on the left. The dark strip represents the separation between both zones. The SOM has been labelled using the training samples. The "REAL" cluster has less extension than that of random samples (labelled according to the event proportions), because the latter have more variability. There was an almost total lack of overlapping (units labelled with both styles) between the zones.

It is clear that the distinction between both zones in the map corresponds to real differences between the random and jazz melodies, and the SOM has been able to capture those differences.

3.2 Jazz versus classical music

So far we have shown that the SOM can easily discriminate between melodies with musical feeling and random ones, using quantitative melodic descriptors. Now we will step forward and subtitute real melodies for the random samples. The new set is composed of monophonic fragments of classical music and a number of changes in the experimental setup are going to be made.

522 classical music melody fragments of eight bars of length were extracted from MIDI files for the training set along with the previous 430 jazz samples, now quantized to $Q = 48$ pulses by bar in order to have more resolution. Initial experiments showed that the overlapping degree for the SOM unit labels of both musical styles was rather high (39.0% of the units were activated by samples from both music styles). This fact suggests that differences were detected between both styles but maybe there is a lack of information to take decisions. For this, harmonic features were added to the set of 18 descriptors already considered.

Addition of harmonic descriptors Most of western music is based on a number of scales (sets of notes ordered by pitch), and melodies can be formed taking notes from those sets. A *diatonic* melody is made up of the natural notes, without sharp or flat notes (named *accidentals*). In western music most of the melodies belong to one of two main scale types: major or minor. The first note of a scale determines its 'tonality' or 'key' and in any melody diatonic and accidental notes can appear.

If the overall key and kind of scale (major or minor) of a melody are known, the set of diatonic pitches is also known and any note event can be classified into diatonic or accidental, and some harmonic information can be evaluated, like the proportion of diatonic notes with respect to the total. If the proportion is high then it is an indication of small key changes or modulations, if any. On the other hand, a low proportion indicates that there are a lot of key changes.

The detection of the key and the definition of the diatonic scale utilized is based on musicological criteria outside the scope of this paper.

We number the accidental notes of a given scale from 1 to 5 according to their distance in pitch from the key note of the scale. We will call this the *accidental degree*. According to this criterion, three harmonic descriptors are defined:

- *Number of accidental notes.* An indication of frequent excursions outside tonality or modulations.
- *Average degree of accidental notes.* Describes the kind of excursions.
- *Standard deviation of degrees of accidental notes.* Indicates a higher variety in the modulations.

From each MIDI file the key is extracted and then, the harmonic descriptors are computed. A new experiment is designed using the new set of 21 descriptors. The same training set of melodies has been used.

The size of the map has been also increased according to the higher dimensionality of the input vectors. The number of units is now 30×12. The neighbourhood radius has also been adapted to the new dimensions. After training and labelling, the maps in figure 3 have been obtained. The labelling process has located the "JAZZ" labels mainly in the right and upper zone, and those corresponding to classical composers mainly in the lower left zone. The percentage of overlapping was in this case very low: 11.1%. Now a clear distinction of styles has been achieved. In the Sammon projection of figure 4 a knot separates both zones in the map.

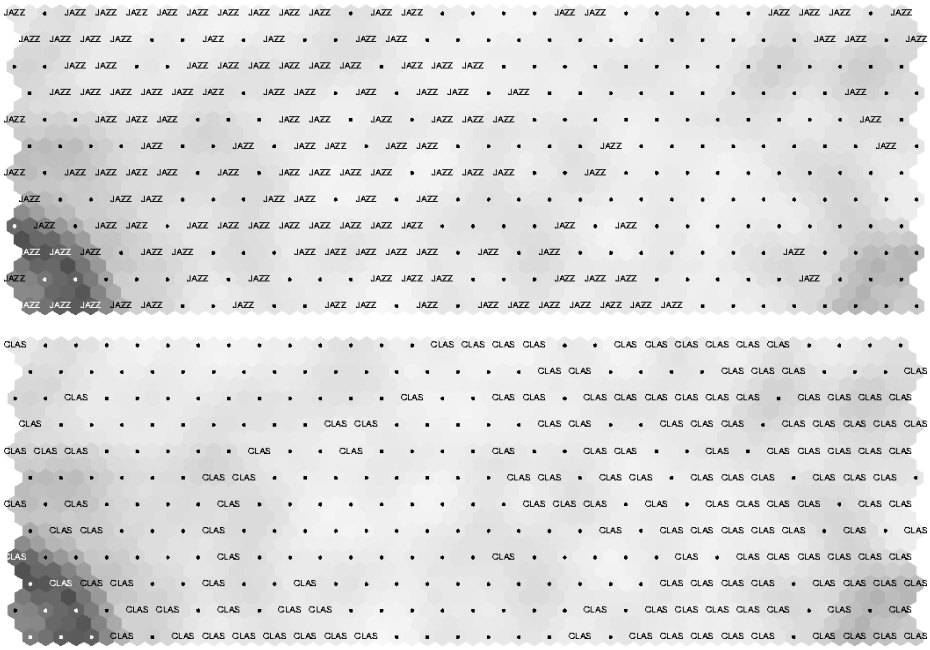

Fig. 3. SOM map after being labelled with jazz (top) and classical (down) melodies.

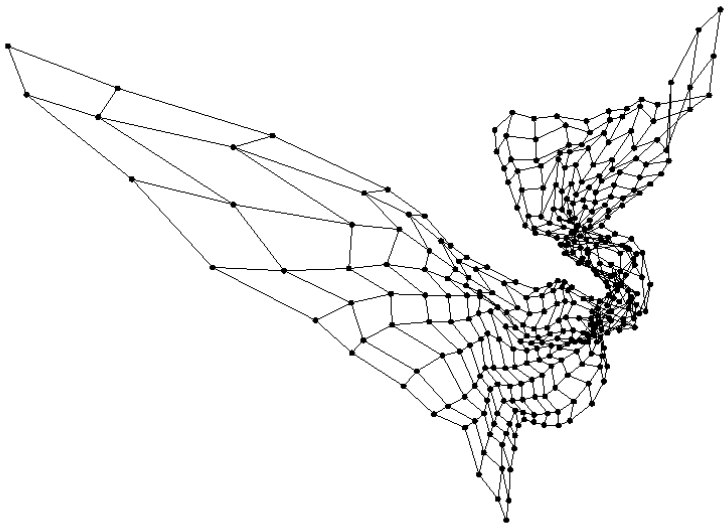

Fig. 4. Sommon projection of the SOM map in figure 3.

3.3 Classification results

The results of classifying new melodic fragments, not contained in the training sets, using the different SOM described above are presented in table 1. These data are obtained with two different maps but both trained with melodic and harmonic descriptors. *One label success* indicates the proportion of melodies to which the map has assigned a unit containing just one label and it was the right one. *First label success* percentage is related to melodies assigned to units with two labels but being the first the right one. Therefore, we are considering as favourable decisions these two criteria: just one correct label or two labels but the first is the correct one (*class success* row). Second label success is if the second was the right one for the assigned unit. Error means that the map has assigned a unit not containing the right label. These two answers define *class error*. Unclassified melodies were those assigned to a unit not containing any label.

The best performance was obtained with the smaller map, with a success classification rate of 82.9% for jazz melodies and of 60.1% for classical melodies. On the other hand the error rates are lower for the second map, and the difference is due to the higher *unclassified* rates for this second map. These results are probably due to the fact that in the second map the class clusters were more defined, leaving more space to unlabelled units. If we devise a way to assign this unlabelled units a class (based, for example, in taking into account the distances in the weight space for the trained map for assigning a label also to unlabelled units), probably the results would improve.

Table 1. Classification results (percentages) using melodic and harmonic descriptors

	Jazz	Classical
Map dimensions = 16 × 8		
Class success	82.9	60.1
Class error	17.1	34.6
Unclassified	0.0	5.3
One label	52.9	43.7
First label	30.0	16.4
Second label	12.9	18.9
Error	4.2	15.7
Map dimensions = 30 × 12		
Class success	72.9	51.2
Class error	10.0	23.6
Unclassified	17.1	25.2
One label	64.3	41.8
First label	8.6	9.4
Second label	2.9	5.0
Error	7.1	18.6

4 CONCLUSIONS AND FUTURE WORKS

We have shown the ability of SOM to map symbolic representations of melodies into a set of musical styles using their description in terms of melodic and harmonic features. The best recognition rate has been found with 21 descriptors that describe melodic and harmonic features of the melodies. The best recognition rate has not been achieved when the overlap was minimum, so the overlap ratio does not seem to be a key point when assessing the quality of a map.

Some of the misclassifications can be caused by the lack of a smart method for melody segmentation. The music samples have been arbitrarily restricted to 8 bars, getting just fragments with no relation to musical motives. This fact can introduce artifacts in the descriptors leading to less quality mappings. The main goal was to test the feasibility of the approach, dealing even with incomplete data. Nevertheless a best average recognition rate of 71.5% has been achieved, that is very encouraging keeping in mind these limitations and others like the lack of valuable information for this task, like timbre.

A number of possibilities are yet to be explored, like the development and study of new descriptors. A statistical multifactorial study of the whole set of descriptors can aid in the selection of a model that can achieve better results with a minimum subset of them. It is very likely that this subset is highly dependent on the styles to be discriminated.

To achieve this goal a large music database has to be compiled and put it to the test using our system. Different styles and more melodies are needed to draw significative conclusions.

Other future lines are based in the integration of time in the description process to capture the evolution of the whole melody. The map activations for a series of fragments of the same melody could be the input to other recognition algorithms in order to increase the classification power of the system, even with a higher number of music styles at the same time. Works in that direction are currently being developed.

References

1. T. Kohonen. Self-organizing map. *Proceedings IEEE*, 78(9):1464–1480, 1990.
2. A. Rauber and M. Frühwirth. *Automatically analyzing and organizing music archives*, pages 4–8. 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001). Springer, Darmstadt, Sep 2001.
3. Brian Whitman, Gary Flake, and Steve Lawrence. Artist detection in music with minnowmatch. In *Proceedings of the 2001 IEEE Workshop on Neural Networks for Signal Processing*, pages 559–568. Falmouth, Massachusetts, September 10–12 2001.
4. Hagen Soltau, Tanja Schultz, Martin Westphal, and Alex Waibel. Recognition of music types. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-1998)*. Seattle, Washington, May 1998.
5. O. A. S. Carpinteiro. A self-organizing map model for analysis of musical time series. In A. de Padua Braga and T. B. Ludermir, editors, *Proceedings 5th Brazilian Symposium on Neural Networks*, pages 140–5. IEEE Comput. Soc, Los Alamitos, CA, USA, 1998.
6. Belinda Thom. Unsupervised learning and interactive jazz/blues improvisation. In *Proceedings of the AAAI2000*, pages 652–657, 2000.
7. T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen. Som_pak, the self-organizing map program package, v:3.1. Lab. of Computer and Information Science, Helsinki University of Technology, Finland, April, 1995.

Self-organising maps for music style perception

Pedro J. Ponce de León, José M. Iñesta Departamento de
Lenguajes y Sistemas Informáticos
Universidad de Alicante
Ap. 99, E-03080 Alicante
inesta@dlsi.ua.es
Tf. 965903772

Resumen:

In this paper the capability of using self-organising neural maps (SOM) as music style classifiers from symbolic specifications of musical fragments is studied. From MIDI file sources, the monophonic melody track is extracted and cut into fragments of equal length. From these sequences, melodic and harmonic numerical descriptors are computed and presented to the SOM. The performance is analysed in terms of separability in different music classes from the activations of the map, obtaining different degrees of succeed. This scheme has a number of applications like indexing and selecting musical databases or the evaluation of style-specific automatic composition systems.

Keywords:

Perception, AI in Music, Signal analysis, Self-organising maps.

Consider it for *paper track*