LOW DIMENSIONAL VISUALISATION OF FOLK MUSIC SYSTEMS USING THE SELF ORGANISING CLOUD

Zoltán Juhász
Res Inst. For Technical Physics and Materials Sciences, H-1525 Budapest, Hungary
P.O.B. 49.
juhasz@mfa.kfki.hu

ABSTRACT

We describe a computational method derived from self organizing mapping and multidimensional scaling algorithms for automatic classification and visual clustering of large vector databases. Testing the method on a large corpus of folksongs we have found that the performance of the classification and topological clustering was significantly improved compared to current techniques. Applying the method to an analysis of the connections of 31 Eurasian and North-American folk music cultures, a clearly interpretable system of musical connections was revealed. The results show the relevance of the musical language groups in the oral tradition of the humanity.

1. INTRODUCTION

The comparative study of different folk music cultures goes back to the early 20th century [1-2]. Although ethnomusicologists seemed to gradually forget the conception of the classical structural analysis and classification, the development of the computation tools led to a renaissance of the idea in recent years [3-4]. At the same time, the number of representative national/regional digital folksong databases is also increasing rapidly. Therefore, a computer aided comparison of different musical cultures in order to reveal hidden contacts of different musical cultures became very topical.

Current interdisciplinary research, based on the cooperation of musicology, artificial intelligence research and data mining, focuses on automatic similarity measurement, segmentation, contour analysis and classification using different statistical characteristics, e.g. pitch-interval or rhythm distribution. A very widely used kind of artificial neural networks, the self organising map (SOM) proved to be a very versatile tool of computing musicology [5]. SOM-based systems have been elaborated for simultaneous analysis of the contour as well as the pitch, interval and duration distributions, based on the symbolic representation of the music [6]. A cross-cultural study of different musical cultures was also based on SOM technique [7].

The operation of a SOM can be summarised for our case as follows: Our input data to be classified are contour vectors, containing subsequent pitch values of melodies of a folksong database. The main goal of self organising mapping is to characterise the multidimensional point system constructed by the set of these melody contour vectors by a significantly smaller set of “contour type vectors” describing the average contours in the local condensations of the input contour vectors. Although the details of the calculations are different, this goal essentially corresponds to that of the so-called K-means algorithm [8]. However, the SOM produces something more: it assigns the resulting contour type vectors to the lattice points of a grid topographically. The topographic structure of the resulting map is provided by a cooperative learning, modifying the contour type vectors located in neighbouring lattice points in parallel. As a result of this local cooperation, similar contour type vectors are located in neighbouring lattice points after learning.

Due to the topographic lattice, the SOM allows us to describe the inherent relations of a melody collection in two levels. Similar melodies are classified as variants of a common contour type in the first level, while the relations of the classes represented by the contour types themselves are mapped into the topographic lattice in the second one.

The overall relations in a data set can be excellently represented on a SOM, providing that these relations can be well approximated by a two-dimensional structure. However, stretching a more complicated structure into a plain lattice results in a significant loss of the accuracy of the classification on one hand, and a non-perspicuous map on the other hand. In principle, it is possible to extend the map dimension, but the resulting exponential increase in the number of lattice points dramatically increases the computing time and the memory demand. Therefore, we need some other technique to increase the degree of freedom of the points in the map.

Therefore, we elaborated a system combining the SOM technique with a special version of the multidimensional scaling (MDS) algorithm [9]. In MDS technique, the input data to be visualised are presented in a quadratic matrix containing some distance-like or similarity-like values between some objects. (For instance, the matrix can contain geographical distances between towns, or dissimilarity ratings of melodies, etc.) The aim of the algorithm is to represent the objects (towns or melodies) in a low dimensional space (often in a plane) with the requirement that the distances of the low dimensional points must optimally correspond to the input values.

In the present work, firstly we describe a method constructed by two independent stages corresponding to the above-mentioned two-level characterisation of melody corpora. The first stage is a simplified, non-cooperative – and therefore non-topographic - version of SOM learning. In the second stage, the topographic low-dimensional mapping of the resulting contour type vectors is accomplished by a variant of the MDS algorithm. This allows us to project the spatial regularities of the multidimensional input vector system to a continuous low-dimensional space without the restrictions of the planar grid structure of the SOM. In order to express the contact to the original SOM principle and to emphasize the increased degree of freedom of the low dimensional mapping, we call this technique "self organising cloud" (SOC).

As a generalisation of the original SOM principle, we also present the cooperative version of the above learning system, where the topographic arrangement is improved by a feedback between the multidimensional learning and the low dimensional mapping functions.

We describe the results of a cross-cultural study of 31 representative Eurasian and North-American folksong collections, based on the modelling by "self organising cloud" technique. The studied cultures are as follows: Chinese, Mongolian, Kyrgyz, Mari-Chuvash-Tatar-Votiac (Volga Region), Sicilian, Bulgarian, Azeri, Anatolian, Karachay, Hungarian, Slovak, Moravian, Romanian, Cassubian (North-Poland), Warmian (East-Poland), Great-Polish (Southern-Central Poland), Finnish, Norwegian, German, Luxembourgish, French, Dutch, Irish-Scottish-English (mainly Appalachian), Spanish, Dakota, Komi, Chanty, Serbian-Croatian (Balkan), Kurd, Russian (Pskov). Our database contains digital notations of nearly 32000 folk songs arising from different written sources. All of these sources apply the Western notation, thus, the microtonal phenomena of the different cultures were eliminated by the authors themselves. The time duration and musical structure of the melodies is very variable, therefore we normalized the length of the melody contours as follows.

2. THE MELODY CONTOUR VECTORS

The generation of vectors from melodies is summarised in Figure 1, showing the first section of a Hungarian folksong as an example. The continuous pitch-time function derived from the score is represented by the thick line in Figure 1.

There, the pitch is characterised by integer numbers, increasing 1 step by one semitone, with the zero level of the pitch corresponding to the C tone. (In order to assure uniform conditions, each melody was transposed to the final tone G.)

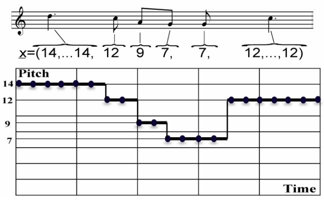

Figure 1. The generation of the melody contour vectors $\underline{x}$.

One can see in the figure that the duration of the temporal intervals of the pitch-time function is determined by the rhythmic value of the corresponding note. Thus, the main rhythmic information is also encoded. For sampling, the total length of the pitch-time function was divided into D portions. Then, the "melody vector" $\underline{x}_k = [x_{1,k}, x_{2,k} \ldots x_{D,k}]^T$ was constructed from the sequence of the pitch-time samples of the k th melody (See Figure 1.).

Since D was uniform for the whole set, melodies could be compared to each other using a distance function defined in the D -dimensional melody space, independently of their individual length. Due to this normalisation, melody contours can be compared independently of their measure, tempo and syllabic structure. We studied the melody vectors of the entire songs in the analysis, and we have found that a choice of $D = 64$ resulted in an appropriate accuracy for each melody.

3. DETERMINATION OF THE CONTOUR TYPE VECTORS

In the first phase of the process, we determined N D=64 dimensional "contour type" vectors $\underline{c}_i$, characterising the most important melody forms in a database containing M melodies. In a training step, the distances between a randomly selected melody contour $\underline{x}_k$ and the contour type vectors are determined, and the contour type of minimal distance $\underline{c}_i$ is considered as the "winner". The winner contour type is moved closer to the melody contour.

In the initial state, the vectors $\underline{c}_i$ were filled by randomly selected melodies of the database. The size of the contour type sets varied between 400 and 576. The algorithm consists of the following steps.

1. A melody of the database was selected randomly and its melody vector $\underline{x}_k$ was compared to the contour type vectors $\underline{c}_i$ using the Euclidean distance metric.

2. The contour type vector of the minimal distance $\underline{c}_i$ was determined as the “winner” and it was modified using

$$\underline{c}_i' = \underline{c}_i + \lambda(\underline{x}_k - \underline{c}_i) \ , \tag{1}$$

where λ is a scalar factor controlling the rate of convergence and the accuracy.
The above technique can be considered as a K-means algorithm [8], or equivalently, as a SOM with a learning radius of zero. This fact results in a remarkable simplification of the SOM algorithm and a significant improvement of the classification as we will illustrate it below. However, these advantages imply the disadvantage that the topographic arrangement of the contour types – being a natural consequence of the original SOM process - requires further computation. The algorithm producing a more comprehensive and adequate spatial arrangement of the contour type vectors is a version of the multidimensional scaling technique, and is described below.

4. LOW DIMENSIONAL MAPPING OF THE CONTOUR TYPE VECTORS

The basic idea of the multidimensional scaling algorithm can be formulated for our problem as follows: We have a set of *N* pieces of *D=64* dimensional contour type vectors $\underline{c}_i$, and we can calculate the *N*N* dimensional quadratic, symmetric matrix $\underline{Q}$ containing the squared Euclidean distances $q_{i,j}$ of them. (The advantage of squaring will be explained below.) We want to represent the *N* contour types by *N* vectors $\underline{v}_i$ of a low dimensional point system, so that the distances $d_{i,j}$ between these points converge to the best low-dimensional approximations of the

$$q_{i,j} = \sum_{k=1}^{D} (c_{i,k} - c_{j,k})^2 \tag{2}$$

values in the sense of

$$S = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{i,j} (d_{i,j} - q_{i,j})^2 = \min \, , \tag{3}$$

where S is the stress function to be minimised, and $w_{i,j} = w_{j,i}$ are weights expressing the importance of the distance of the corresponding points in the stress function. (For instance, the exact distance of very dissimilar vectors may not be important in certain cases. Thus, the weight values can be defined as functions of the input distances $q_{i,j}$.)
The minimum of the stress function is searched by a gradient algorithm. For sake of simplicity, we consider the case when the low dimensional space is a plane, but the results can be easily generalised to higher dimensions. At the beginning, the *N* points are randomly located in the plane with the coordinates $(v_{m,1}, v_{m,2})$, where m denotes the serial number of the points. The gradient components of the stress function in the 2*N* dimensional space of the point coordinates are the partial derivatives

$$\frac{\partial S}{\partial v_{m,k}} = \sum_{i=1}^{N} 2 \sum_{j=1}^{N} w_{i,j} (d_{i,j} - q_{i,j}) \frac{\partial d_{i,j}}{\partial v_{m,k}} \ ,$$
$$k = 1,2 \quad m = 1 \ldots N \ . \tag{4}$$

Let the “distance” of the *i*th and *j*th points in the plane be defined as

$$d_{i,j} = \frac{1}{2} \sum_{k=1}^{2} (v_{i,k} - v_{j,k})^2$$
$$k = 1,2 \ \ i = 1 \ldots N \, , \ \ j = 1 \ldots N \ . \tag{5}$$

This definition yields a very simple expression for $\frac{\partial d_{i,j}}{\partial v_{m,k}}$, and the gradient components of the stress function in Equation (4) become finally:

$$\frac{\partial S}{\partial v_{m,k}} = 2 \sum_{i=1}^{N} w_{i,m} (v_{m,k} - v_{i,k})(d_{m,i} + d_{i,m} - q_{m,i} - q_{i,m})$$
$$k = 1,2 \ m = 1 \ldots N \tag{6}$$

According to the gradient search principle, the new estimates of the optimal point co-ordinates are determined as

$$v_{m,k}' = v_{m,k} - \mu \frac{\partial S}{\partial v_{m,k}} \, , \tag{7}$$

where the small scalar value μ determines the rate and the accuracy of the convergence.
In the subsequent steps of the algorithm, the gradient components of the stress function are re-calculated in the new

point locations using Equations (5) and (6), and the points are replaced using Equation (7) again. The algorithm can be easily generalised to 3 or more dimensional point systems.

Comparing the above algorithm to the self organising map (SOM), an important difference lies in the fact that the low dimensional vectors $\underline{v}_i$ are not fixed to lattice points, so they are allowed to roam in the low dimensional space, in search of their own optimal position. In order to express this free roaming of the point system during learning, and to distinguish between the original SOM and the above described algorithm, we call it "self organising cloud" (SOC).

This non-cooperative form of the SOC algorithm accomplishes a two-level systematisation of melody collections. In the first step, the contour type vectors $\underline{c}_i$ are determined, representing the centres of local clusters of the melody contour vectors in the D=64 dimensional melody space. Thus, the first level of the systematisation is assigning the melodies to the most similar contour type vectors. Having accomplished this classification process, the connections of the melodies can be described, the higher-level connections of the resulting melody classes, however, remain unrevealed. These latter relations are described by mapping the D=64 dimensional contour type vectors to a low dimensional space. Thus, the second level of the systematisation is the low dimensional representation and visualisation of the relations between the melody classes having been determined in the first level.

5. COOPERATIVE LEARNING

Up to this point, we have emphasized the advantages of the independence of the non-topographic learning- and the topographic visualising parts of the SOC technique. However, the system can easily be modified to learn the contour types in a cooperative way. In this case, all of the contour type vectors located in the surroundings of the winner are modified by the current training vector, and their new low dimensional coordinates are re-calculated simultaneously with the contour type learning steps, using Equations (5), (6) and (7). Since the vectors $\underline{v}_i$ can freely move in the low dimensional space during the process, this cooperative learning approaches similar vectors to each other, resulting in a more articulated system of the low dimensional clusters. However, an uncontrolled cooperative process can lead to an accelerated approach of neighbouring vectors, resulting in a total collapse of the whole system into one point. This principal problem can be solved by the prohibition of the cooperative training within a critical radius around the winner. Although this version produces a suboptimal contour type estimation - similarly to the SOM algorithm -, it may significantly improve the visual representation of the clusters.

6. CROSS-CULTURAL ANALYSIS OF 31 MUSICAL CULTURES USING THE SOC ALGORITHM

As an application of the SOC algorithm, we summarise the procedure and the results of a cross-cultural study of 31 folk music cultures in this chapter. The cultures were represented by 31 databases containing 1000 – 2500 melodies by culture. The first step of the analysis was the determination of the contour type collections of the 31 cultures, using non-cooperative SOC mapping of the databases one by one. In the second phase, we unified the resulting 31 contour type collections into one training set, and trained a two-dimensional "common" SOC having 1000 contour type vectors. After training by the nearly 12000 contour type vectors arising from the 31 collections (400-500 vectors by culture), the resulting 1000 common vectors represent the most characteristic melody contours appearing in the 31 cultures. Figure 2 shows the resulting common musical maps generated by non-cooperative, as well as cooperative training of the SOC. The figure verifies that the cooperative learning yields a much more arranged "musical map". The musical meaning of the main areas of this map is demonstrated by the contour type examples in Figure 3.

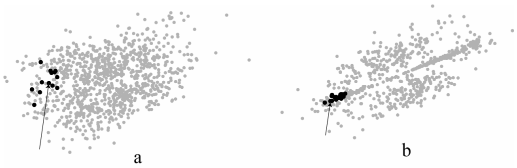

Figure 2. Self organising clouds of the common contour type collection using non-cooperative (a), and cooperative (b) learning.

At this point, we have to define the concept of "activation" of the common contour type vectors as follows: a contour type vector of the common SOC is "activated" by a training vector when the distance between them is less than a threshold value (see Equation 2). For example, the black points in Figure 2 correspond to the contour types activated by the Hungarian melody of Figure 4. The distribution of the points illustrates that the cooperative learning moves similar contour types into a more compact cluster. Extending this concept to national/areal sets of training vectors, we can say that the 31 contour type collections activate different subsets of the 1000 common vectors.

Figure 3 shows the common SOC with 6 different national activations and some contour type examples being very characteristic in the given cultures. Since the arrangement of the SOC reflects purely musical conditions, it is not a trivial result that the different cultures are located in more or less continuous areas. This fact refers to different musical styles dominating in different cultures. Some of these very characteristic melody forms are also indicated in Figure 3.

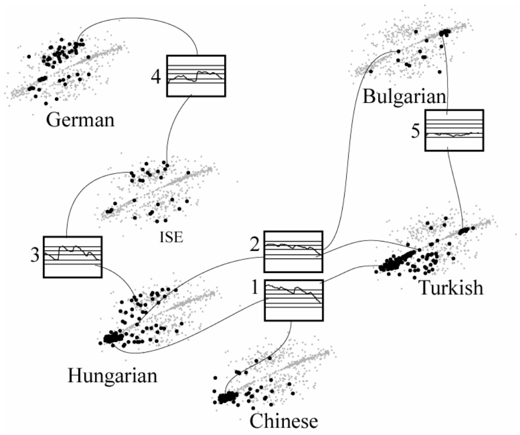

Figure 3. Activated area of the common contour type cloud by contour type collections of 6 different cultures.

For instance, contour example 1 shows that descending melodies with a high range are simultaneously dominating in the Chinese, Hungarian and Turkish activation area. An example for such melodies with Hungarian, Chinese, Anatolian and Dakota parallels is shown in Figure 4.

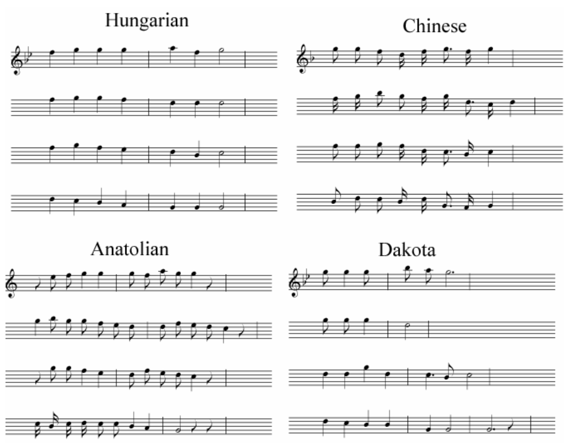

Figure 4. Melody examples of type 1 in Figure 3.

Contour example 2 and 5, representing melodies with low range demonstrate the musical background of the definite overlap between Anatolian and Bulgarian cultures.
The Hungarian area shows a significant overlap with the Chinese and Anatolian ones, but contour example 3 also demonstrates a significant common musical style of domed melody forms with the Irish-Scottish-English culture.

At the same time, the Irish-Scottish-English corpus has also a significant overlap with the German one in the area of ascending forms moving beyond the final tone (see contour example 4).

The sizes of the overlaps benchmarked against the total sizes of the activated area refer to the intensity of the relations of musical cultures [7]. We considered these relative overlap sizes as similarity ratings of musical cultures, and represented the resulting system of musical language groups using the MDS algorithm described above. The two-dimensional MDS plot of the connections is shown in Figure 5. The edges indicate pairs of cultures with the largest overlaps. We also indicated some sub-graphs where the nodes mutually are in close musical contacts with each other. The graph shows a very clear structure with seven musically well interpretable clusters. The right branch of the system contains the mutually very closely related {Chinese – Volga – Mongolian}, {Hungarian – Slovak} and {Turkish – Karachay – Sicilian – Dakota} groups. The left branch is constructed by the {Finnish – Norwegian – ISE} and {German – Luxembourgian – French – Holland} clusters, whereas the {Bulgarian – Balkan - Kurdish – Azeri} and {Russian – Komi - Warmian (East-Poland)} groups construct clearly separate clusters.
The close contacts of the above discussed seven "musical language groups" can be traced back to certain musical styles being simultaneously present in more cultures. Comparing Figure 5 to Figure 3, one can recognise that the six activator cultures of the common musical map can be considered as representatives of the above mentioned "musical language groups". Therefore, contour examples 1-5 in Figure 3 represent right the most characteristic common musical forms contacting the musical language groups as well.

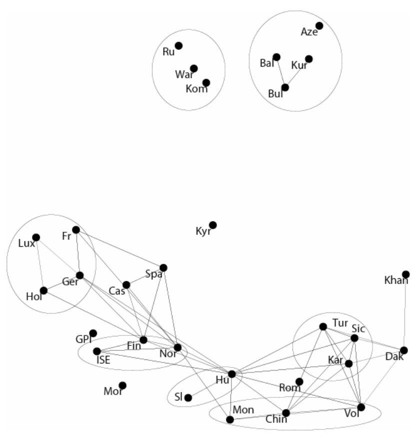

Figure 5. MDS plot of the connections of 31 folk music cultures. Connecting lines indicate the mutually largest relative overlaps.

7. CONCLUSIONS

We have described a technique which learns the group averages of the local condensations of multidimensional point systems on the one hand and represents the similarity conditions of the learned average vectors in a low dimensional point system on the other hand. Basically, the algorithm can operate in two modes: In the non-cooperative mode only one average vector is modified in one training step and the state of the other vectors is independent of this modification. In the cooperative mode the training is extended to a group of average vectors, and a feedback comes into existence between the learning of the multidimensional averages and the low dimensional arrangement.

The non-cooperative learning of the contour type vectors permits the convergence to the exact centres of the local condensations of the training vectors, therefore the SOC corresponds to the K-means algorithm in this case. The cooperative learning realises a compromise between the accuracy of the multidimensional learning and the low dimensional representation, therefore the system converges into a sub-optimal state in this case. However, the cooperativeness can be tuned by the learning radius parameters, and the benefit of a well accomplished cooperative training may be a more transparent low dimensional representation of the multidimensional clusters, whereas the accuracy of the learning also remains acceptable.

The low dimensional topographic representation of the contour type vectors is accomplished by a weighted MDS algorithm. This increases the degree of freedom of the mapping, because the locations of the low dimensional points are not bounded to a lattice, and their dimensionality can be optimised without a significant increase in the computing time.

We applied the method to an analysis of the connections of 31 Eurasian and North-American folk music cultures. We have found that the changeover to the continuous low dimensional space of the SOC from the plain lattice structure of the SOM yields a more articulated low dimensional data representation and a musically well interpretable systematisation of the melody contours.

Using the SOC technique, we have determined a conjugate musical map of the most important melody forms in the studied cultures, and have found that the different cultures occupy well defined continuous areas of this map. The technique allowed us to trace back this “musical geography” to the dominance of certain well distinguishable musical styles in different cultures. Exactly the close correlation of different cultures with certain areas of the musical map calls the attention to the overlaps, referring to significant interactions of the studied cultures. The analysis of these overlaps revealed a perspicuous system of cross-cultural connections, which was represented by an MDS plot of the probabilities of deterministic interactions. The common musical forms standing in the background of the most important cultural connections were also identified from the overlap areas. We hope that these results demonstrate the timeliness of an extensive study of musical language groups and call the attention to the importance of the oral musical tradition of the humanity.

This work was supported by the Hungarian National Research Found (grant no. K81954).

8. REFERENCES

[1] B. Bartók: “The Hungarian Folk Song by B. Bartók. “ Ed: B. Suchoff, Transl: M.D. Calvocoressi, Animations: Z. Kodály. State University of New York, 1981.

[2] Z. Kodály: “Folk Music of Hungary”. Budapest, Corvina.

[3] P. Kranenburg1, J. Garbers, A. Volk, F. Wiering, L. P. Grijp, and R. C. Veltkamp1 (2010): “*Collaboration Perspectives for Folk Song Research and Music Information Retrieval: The Indispensable Role of Computational Musicology”*. Journal of interdisciplinary music studies spring 2010, volume 4, issue 1, art. #10040102, pp. 17-43

[4] O. Cornelis, M. Lesaffre, D. Moelants: “*Access to ethnic music: Advances and perspectives in content-based music information retrieval*”**.** SIGNAL PROCESSING Volume**:** 90 Issue: 4 Pages: 1008-1031 Published: APR 2010

[5] T. Kohonen: „Self-organising Maps“. Berlin:Springer-Verlag

[6] P.Toiviainen, and T. Eerola: “*A computational Model of Melodic Similarity Based on Multiple Representations and Self-Organizing Maps*”. Proceedings of the 7th International Conference on Music Perception and Cognition, Sidney, 2002 C. Stevens, D. Burham, G. McPherson, E. Schubert, J. Rewick (eds.) Adelaide: Causal Productions. P236-239.

[7] Z. Juhász, J. Sipos: „*A comparative analysis of Eurasian folksong corpora, using self organising maps*”. Journal of Interdisciplinary Music Studies. (2009), doi: 10.4407/jims.2009.11.005

[8] T. Kanungo, D.M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, A. Y. Wu: "*An efficient k-means clustering algorithm: Analysis and implementation*". IEEE Trans. Pattern Analysis and Machine Intelligence 24: 881–892. doi:10.1109/TPAMI.2002.1017616.

[9] I. Borg, P. Groenen: „*Modern Multidimensional Scaling: theory and applications (2nd ed.)*”, Springer-Verlag New York, 2005.