MOTIVE IDENTIFICATION IN 22 FOLKSONG CORPORA USING DYNAMIC TIME WARPING AND SELF ORGANIZING MAPS

Zoltán Juhász
Research Institute for Technical Physics
and Materials Science.
P.O.B 49. Budapest H-1525.

juhasz@mfa.kfki.

ABSTRACT

A system for automatic motive identification of large folksong corpora is described in this article. The method is based on a dynamic time warping algorithm determining inherent repeating elements of the melodies and a self-organizing map that learns the most typical motive contours. Using this system, the typical motive collections of 22 cultures in Eurasia have been determined, and another great common self organising map has been trained by the unified collection of the national/areal motive collections. The analysis of the overlaps of the national-areal excitations on the common map allowed us to draw a graph of connections, which shows two main distinct groups, according to the geographical distribution.

1. INTRODUCTION

In order to study interethnic and historical relations, Bartók and Kodály compared different layers of Hungarian folk music to those of other nations living in the neighborhood of Hungarians. Later, they extended the study to Anatolian, Mari and Chuvash folk music [1-2]. These exciting results raise the question, whether it is possible to describe a whole and clear system of musical contacts in Eurasia by a systematic comparison of a sufficient number of national or regional cultures.

A further question, raised by the classical results mentioned above, refers just to the method of the analysis. The aim of these classical works was to find parallelism of entire melody structures. The similarity of whole melody contours seems to be really a sufficient condition to find genetic musical relations [1-3]. However, the question rises: do less rigorous requirements also exist? Instead of comparing the complete melody structures, our aim was to find and analyse the smallest independent melodic units. It is well known that folksongs can usually be divided into certain phrases on the basis of musical and textual regularities. In a previous work, we have shown some results comparing individual phrases, as well as whole melodies of 6 European cultures [4].

The idea of a motive identification algorithm can be derived from the recognition that phrases are not necessarily the smallest intelligible units in folk music. We want to find the most frequently appearing motive types in a well defined melody corpus, with the assumption that each motive type may have several variants. However, the repetition inside a melody can also be considered as an indication of a motive. Therefore, we suppose two possible detections of the motives. In addition to the "culture-defined" motive identification, based on the frequent appearance in different songs, we also suppose the existence of the "melody-defined" identification which is based on the repetitions inside the melodies.

The central problem of algorithmic melody pattern identification is the musical relevance of the results [5]. The most frequently applied melody segmentation techniques can be divided into two main groups. In the first group, segmenting is based on pre-defined and data-independent rules [6-8]. Using such rules, the so-called Local Boundary Detection Model (LBDM) determines a boundary strength value between each couple of notes, and determines the segment boundaries at the maximal strength values [6,9]. Due to the requirement of pre-defined rules, such methods are not available for the sake of a learning system. The second group of segmenting techniques is based on a learning process to determine the regularities of a given melody corpus. Such regularities can be characterised by the frequencies or conditional probabilities of the motives [10-12]. The so called Markov technique operating with conditional probabilities has already been applied to folk songs [13-14]. A further data-based self learning method for segmenting a large corpus of folksongs has been also described, which determines the conditional entropy of the motives and defines an average entropy increment value for a given segmentation [15]. A method based on knowledge representation has been elaborated for identifying recurrent melody parts in large folksong corpora [16].

The learning unit of the system described in this paper is a self organising map (SOM), trained by the contour functions of the motives [17-18]. The motive identification in a given melody is accomplished in two steps. Firstly we determine the repeating elements of the melody by an algorithm based on dynamic time warping (DTW). After that, the remaining melody parts are analysed using a self organizing map, which learns and identifies the most frequently appearing patterns as "culture-defined" motives.

Our current possibilities allowed us to set up 22 folksong corpora, each of them consisting of 600-2400 melodies, representing Hungarian, Slovak, Moravian, Chinese, Mongolian, Kyrgyz, Mari-Chuvash-Tatar, Karachay-Balkar, Anatolian Turkish, Azeri, Sicilian, Spanish, Rumanian, Bulgarian, Polish - Cassubian, Finnish, Norwegian, German, Luxembourgian-Lotharingian, French, Dutch and Irish-Scottish-English musical traditions. In order to make an unbiased and general analysis, these nearly 40 000 melodies were transposed to the common final tone G automatically in the analysis.

2. DETERMINATION OF MOTIVES DEFINED BY REPETITION WITHIN MELODIES

To search for essentially identical, but not completely uniform motives inside melodies, we developed an algorithm based on dynamic time warping technique [17]. The operation of the algorithm is illustrated in Figures 1 and 2.

In the first step, the contour vectors $\underline{v}$ of the melodies are generated in the way demonstrated in Figure 1. The time duration of the *k*th melody is divided into small units according to the rhythmic value of 1/16, and the pitch values belonging to these subsequent small time intervals are stored in a multidimensional vector $\underline{v}_k$.

Figure 1. Generation of the contour vector $\underline{v}$.

The original aim of a DTW process is to determine a non-negative scalar number characterising the difference of two vectors. In order to calculate this DTW-distance between melody contours $\underline{v}$ and $\underline{u}$, the matrix $\underline{\mathrm{P}}$ is generated containing the deviations of the *n*th and *m*th pitch samples of the vectors $\underline{v}$ and $\underline{u}$:

$$\delta_{n,m} = \left| v_n - u_m \right| \quad n = 1...D_v, m = 1...D_u, \qquad (1)$$

where D_v and D_u are the dimensions of $\underline{v}$ and $\underline{u}$ respectively. Figure 2 shows an example of the above calculation for the contour vectors demonstrated by the diagrams on the left side and the bottom of the matrix.

The zero elements of the matrix $\underline{\mathrm{P}}$ marked by bold italic characters indicate local warping curves assigning similar parts of the two vectors to each other. Our algorithm is based right on this recognition: instead of determining the total DTW distance of the vectors, we search for such local warping paths in matrix $\underline{\mathrm{P}}$. To do this, the partial time warping distances $\Delta_{n,m}$ are calculated, according to the dynamic time warping process:

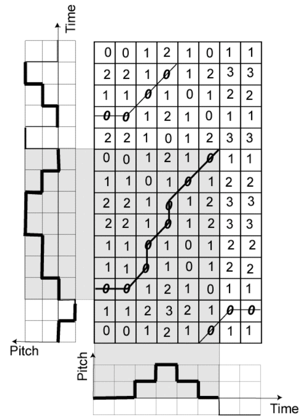


Figure 2. Generation of the partial deviation matrix $\underline{\mathrm{P}}$, and the path of 0 elements indicating the relation between corresponding motives.

$$\Delta_{n,m} = \min\begin{pmatrix} \Delta_{n-1,m} + \delta_{n,m} \\ \Delta_{n-1,m-1} + \delta_{n,m} \\ \Delta_{n,m-1} + \delta_{n,m} \end{pmatrix} \rightarrow \begin{pmatrix} v_{n,m} = 1 \\ v_{n,m} = 2 \\ v_{n,m} = 3 \end{pmatrix} \qquad n = 2..D_v, m = 2...D_u$$

(2)

The original DTW algorithm produces the final distance at the end of the above recursive calculation as $\Delta_{D_v D_u}$. The local warping paths can be determined using the $D_v * D_u$ dimensional matrix $\underline{\mathrm{V}}$. Since the elements of the matrix $\underline{\Delta}$ cannot be defined for negative indices, the algorithm starts with the values of $(n = 2, m = 2, v_{1,1} = 2)$, and the initial values of $\underline{\Delta}$ are $\Delta_{1,m} = \delta_{1,m}$ and $\Delta_{n,1} = \delta_{n,1}$.

The overall similarity of the vectors can be characterised by the summed length of the similar sub-sequences compared to the sum of the total length of the vectors $D_v + D_u$. Thus, our technique can characterise the similarity of two different contour vectors by a scalar number ranging between 0 and 1. This similarity measure ignores the order of the motives, in contrast to the original DTW and the Euclidean distances. Therefore it is able to detect the relationship even if the successions of the characteristic melody parts are different in the compared melodies.

Example 1 shows two couples of melodies arising from different cultures, with a significant amount of similar parts found by the above described method. For instance, the first, second and fourth phrases of the Hungarian song in the first example are practically identical to the second and fourth phrases of the corresponding Appalachian melody, and the third phrase of the Appalachian song appears as a dominant part of the corresponding Hungarian phrase, too. Due to these local correspondences, the melodies are found to be similar, in spite of the difference between the domed, as well as descending character of the two melodies.

The above technique can be applied also to identical vectors (i.e. $\underline{v} = \underline{u}$). In such cases, the trivial result that the whole melody is identical to itself is indicated by the zero elements of the diagonal, but the partial warping paths marked by zero matrix elements indicate the similar subsequences. Therefore, our technique is also able to find similar parts within one given melody (see Figure 3).

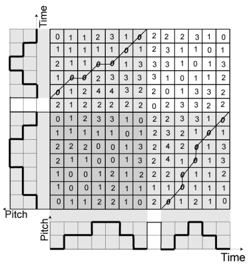
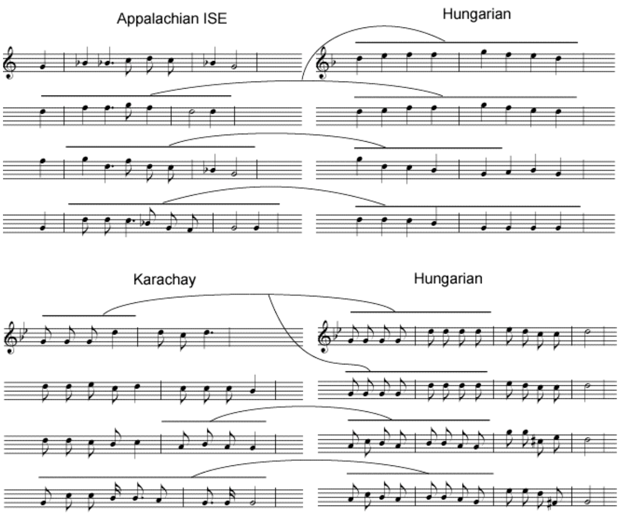


Figure 3. Application of the DTW technique to search for repeated parts in a melody.

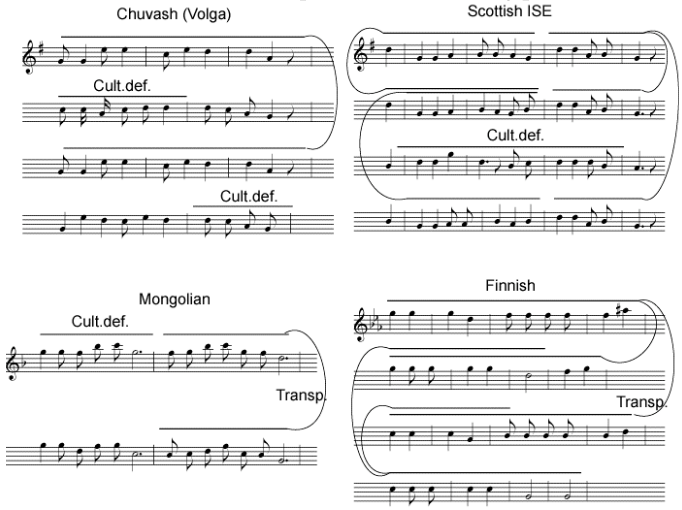
The technique can be generalized for not exactly identical pitch values, too, by the extension of the search for paths of small elements in matrix $\underline{\mathrm{P}}$. Some results of the method are shown in Example 2.



Example 1. Common motives of related melodies arising from different cultures.

3. THE COMPLEX MOTIVE IDENTIFICATION ALGORITHM

In addition to the melody-based motive identification, we also need a technique for the culture-defined identification which was defined as the determination of those melody parts which frequently appear in a whole national/areal database. While the melody-based technique needs the analysis of one given melody, the culture-based identification requires a self learning process



Example 2. Melody-defined and culture-defined motives in 4 folksongs.

analysing the whole database simultaneously. In order to solve this problem, i.e. to identify the most frequent melody parts automatically, we developed a system based on a self organising map, as it is shown in Figure 4.

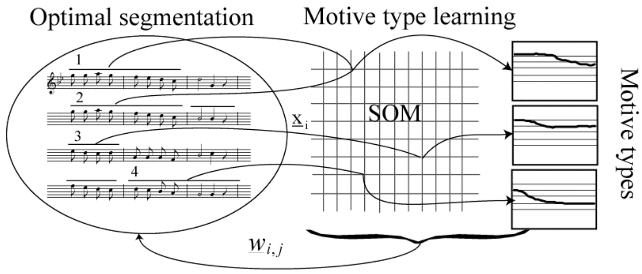


Figure 4. The complex motive identification system.

The input to the algorithm is a melody selected randomly from the database. At the beginning of the process, the D dimensional motive type contour vectors assigned to the lattice points of the SOM, $\underline{w}_{i,j}$ are filled by random numbers. The choice of $D = 32$ proved to be sufficient for our database.

The processing is done by the following steps:

1. In the first step, all melody-defined motives of a melody are determined, using the melody-based identification algorithm.

2. All possible motives of the remaining parts of the melody are determined. The time duration of each possible motive is divided into D parts, and the pitch values belonging to the subsequent time intervals are stored in a vector $\underline{x}$ of dimensionality D. This operation has been discussed in reference to melody contour generation (see Figure 1), but it is worth mentioning here an important

difference: When generating motive contour vectors, the vector dimension D is a pre-defined constant, while it is variable for melody vector generation, because the sampling time unit is pre-defined in this latter case.

3. The optimal motives are identified on the basis of the current estimates of the most typical motive types assigned to the lattice points of the SOM. Let $\underline{x}_k$ denote the contour vector belonging to the kth possible motive, and $\underline{w}_{i,j}$ the current estimate of the motive contour type belonging to the lattice point with the coordinates (i, j). The motive contour vector $\underline{x}_k$ is assigned to the most similar motive contour type vector of the SOM:

$$d_k = \min_{(i,j)}(\delta_k(i,j)) \quad , \quad (3)$$

where the similarity measure $\delta_k(i, j)$ is the Euclidean distance between the $\underline{x}_k$ and $\underline{w}_{i,j}$.

$$\delta_k(i,j) = \sqrt{(\underline{x}_k - \underline{w}_{i,j})^T(\underline{x}_k - \underline{w}_{i,j})} \quad (4)$$

Finally, the culture-based motives are defined using the following constraints:

- The distance of the motive and the corresponding motive type must be less than a critical value.
- The culture-defined motives are defined as the longest melody parts satisfying the above requirement.
- The culture-defined motives should not overlap with melody-defined motives. Melody-defined motives have priority.

4. The SOM is trained with the resulting set of culture-defined and melody-defined motives, using the well known algorithm. Each $\underline{x}_k$ vector determines a "winner" motive type contour on the SOM according to Equation 3, and the winner vectors $\underline{w}_{m,n}$ are modified towards the corresponding motive contour (denoting a winner position by (m,n) on the SOM). The motive type vectors located in the surroundings of a winner are also modified, while the radius defining the surroundings decreases during the training steps [17].

The input data vectors are usually invariable during the training process of self organising maps. In our system, however, they are variable, because the optimal culture-based motive identification depends on the current state of the motive type vectors $\underline{w}_{i,j}$ (see Equations 3 and 4). Since $\underline{w}_{i,j}$ are modified during the learning process, the optimal segmentation itself depends on the current state of the SOM. In other words, there exists a feedback between the segmentation and the learning algorithm, thus, our system converges to an optimal training- and feature vector set in parallel. The results of many independent training processes verified that all of the characteristic motive contour types have been learned consistently and independently of the starting conditions of the SOM-s.

4. ANALYSYS OF THE CULTURAL CONTACTS AMONG 22 CORPORA

Let suppose that we can create a whole collection of motive contour types, containing all the significant contours that appear in any of the 22 cultures. It is obvious that the national/areal sets of motive types can be considered as different subsets of this great common collection, therefore the study of musical connection between different cultures can be determined by the analysis of the intersections of these subsets.

Being in possession of the size of the great common motive contour type collection (N), the sizes of its two national/areal subsets (A and B), as well as the size of their intersection (X), the measure of the relationship between these cultures can be expressed by a probability as follows.

As a first step we compute the probability of the event that a random choice of two subsets with sizes A and B from the set of size N results in an intersection of size x, as

$$p(x) = \binom{N}{x}\frac{\binom{N-x}{A-x}\binom{N-A}{B-x}}{\binom{N}{A}\binom{N}{B}} \quad . \quad (5)$$

Using this probability density function, the probability of the event that the size of the intersection is less than X, is expressed as

$$P(X) = \sum_{x=1}^{X-1} p(x) \quad (6).$$

A high value of this probability indicates that the number of common contour types in the two corpora is much higher than the expected value in case of random correlations. Consequently the similarity, manifested by such high intersection of two corpora, cannot be a product of occasional coincidences of independent musical evolutions. It can be stated in such cases of similarity that the common musical characteristics implicate a historical or present, immediate or intermediate cultural interaction, that is, the established relationship is necessarily deterministic.

To construct the above mentioned sets, we first had to deduce the characteristic motive contour type collections for eash of the 22, by training 22 SOM-s of size 20*20 lattice points separately. After determining the 22 national/areal motive contour type collections, a new large self organizing map of size 30*30 was trained by the united set of them, in order to determine the set of all possible motive contour types appearing anywhere in the 22 cultures.

This common SOM allows us to classify all motive types of a given national/areal collection on it using Equations 3 and 4. We call this process "excitation of the common map by a culture". The values *A, B* and *X* can be determined for any selected two cultures by counting up the lattice points excited in the great common SOM. With these quantities, the calculation of the probability $P(X)$ can be carried out using Equations 5 and 6, knowing that *N* is equal to the total number of the common contour types. It is worth mentioning here that this calculation avoids the problems arising from the different sizes of the corpora, since the expected intersection decreases with decreasing subset sizes *A* and *B*.

The graph of the system of closest relationships is summarized in Figure 5, where a connection line indicates a high probability ($P(x) > 0.999$) of deterministic contact between the nodes of musical cultures. The Figure shows two main sub-graphs containing an "Eastern" - Mongolian, Chinese, Volga, Hungarian, Slovak, Moravian, Spanish, Kyrgyz, Romanian, Bulgarian, Azeri, Sicilian, Turkish and Karachay-Balkar, as well as a "Western" - Finnish, Norwegian, Irish-Scottish-English, French, German, Dutch, Luxembourgish and Cassubian – group of nodes. There are some interconnections between these two large sets due to the close connections of the Hungarian – Slovak – Finnish – (Irish-Scottish-English), and the Moravian - Norwegian corpora. Besides these close contacts of the Carpathian Basin to the Scandinavian and Irish-Scottish-English cultures, the Irish-Scottish-English and Norwegian corpora have certain further Eastern contacts to the Volga-region and Kyrgyzstan. Anyhow, the connection of the two main subsystems indicates a special role of the above mentioned cultures inside their main groups and also in the whole system.

The structure of the graph indicates certain smaller groups inside the great "Eastern" system. The majority of the motives belonging to the large pattern excited by the Mongolian, Chinese and Volga group on the common SOM move in the highest regions of the melodies – they start or end at the octave or higher notes (See the Mongolian motive contour type in Figure 4). The visible overlaps of the patterns of the Hungarian, Slovak, Karachay-Balkar, Turkish and Sicilian excitations with the above mentioned triad are based mainly on the above mentioned motives in the highest region of the melodies.

The patterns of the Irish-Scottish-English, Finnish and Norwegian excitations also indicate an important role of such motives, resulting in the deterministic contacts of these cultures to the Carpathian Basin and the Volga-region. However, this "Eastern" part of the common motive type map empties in the further Western patterns.

The French and Dutch contour examples show that the most common Western motive types move in the lowest

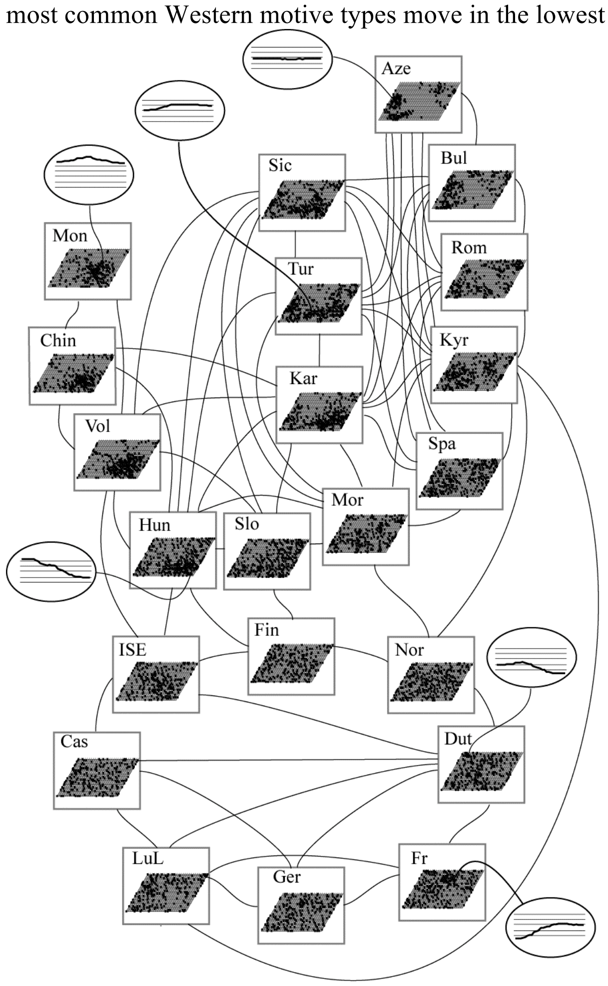


Figure 5. The graph of deterministic relations of 22 musical cultures in Eurasia.

ranges of the melodies, starting or ending at a fourth or fifth below the ending note.

The cloud of the high motives also disappears gradually along the branch of the Spanish – Kyrgyz – Romanian – Bulgarian – Azeri excitations, while the pattern on the left side of the motive type map becomes more and more emphasized. The Azeri motive example illustrates that the motive types belonging to this part of the map are of low ambit, ranging between the fourth, third or the second. The Sicilian, Turkish and Karachay-Balkar excitations show that these cultures also frequently apply such motive types, (beneath the above mentioned group of motives in high), indicating deterministic cultural contacts between the two branches. However, the group of these low-ambit motives practically misses in the Mongolian-Chinese-Volga branch, and it is also rather rare in the Hungarian, Slovak and Moravian melodies. Therefore, these cultures have no direct connections to the Spanish-Kyrgyz-Romanian-Bulgarian-Azeri branch.

SUMMARY

The very clear connections between the patterns of the different national/regional excitations on the common motive type map allowed us to analyze the musical structures of different cultures as different manifestations of a common motive set, and led to the conclusion that the main contacts between the cultures can be explained by the dominance/lack of a few motive types. This analysis clarified that "Eastern" cultures prefer motives in high regions of the melody, generally moving between the octave and the fifth as well as fourth, while the "Western" melodies prefer motives connecting the tonic to a fifth or a fourth beyond the tonic. The combined analysis of the contact probabilities and the overlaps of the national/areal patterns indicated several distinguishable branches among the Eastern cultures. The Mongolian-Chinese-Volga branch highly prefers motives in high, while the Sicilian-Turkish-Karachay branch evaluates a balance between these high motives and those of an explicitly low ambit. The close contacts of Hungarian, Slovak and Moravian cultures to these two distinguishable branches are based mainly on the high motive types. At the same time, the high motive types gradually disappear in the Spanish-Kyrgyz-Romanian-Bulgarian-Azeri branch, while the dominance of motives of low ambit connects them to the Sicilian-Turkish-Karachay branch.

Not forgetting the simplifications made during the application of our technique, we can state that the motive analysis allowed us to draw a rather perspicuous picture of the cross-cultural connections of different folksong cultures. We hope that these results may demonstrate the feasibility of an extended research of "musical linguistics", and suggest an efficient and quantitative tool for "melody mining", using artificial intelligence and other mathematical tools.

Acknowledgement. The Author is grateful to Ewa Dahlig-Turek, Damien Sagrillo, Louis Grijp, Hans-Hinrich Thedens, János Sipos and Gergely Agócs for their altruistic help to extend the database with Cassubian, Luxembourgian, Lotharingian, Dutch, Norwegian, Anatolian, Kyrgyz, Mongolian, Azeri and Karachay-Balkar collections.

References

[1] B. Bartók: "On Collecting Folk Songs in Turkey" Tempo, *New Ser., No. 13, Bartok Number* (Autumn, 1949), pp. 15-19+38

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

[3] D. Huron: "The melodic arch in Western folksongs". *Computing in Musicology,* Vol. 10, pp. 3-23.

[4] Z. Juhász: "A systematic comparison of different European folk music traditions using self-organising maps". *Journal of New Music Research* 2006, Vol. 35, No. 2, pp 95-112.

[5] O. Lartillot and P. Toiviainen. (2007). "Motivic matching strategies for automated pattern extraction", *Musicae Scientiae*, Discussion Forum 4A, pp. 281-314.

[6] E. Cambouropoulos : « Musical Parallelism and Melodic Segmentation", *Proceedings XII Colloquium on Musical Informatics*, Gorizia, Italy

[7] D. Halperin: "A Segmentation Algorithm and its Application to Medieval Monophonic Music". *Musikometrika* 2 (1990)

[8] J. Singer: "Creating a nested melodic representation: competition and cooperation among bottom-up and top-down Gestalt principles". ISMIR 2004

[9] E. Cambouropoulos: "A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface". *Proceedings of the Troisiémes Journées d'Informatique Musicale (JIM-96),* Caen, France

[10] E. Charniak: "Tree-bank Grammars", *Pproceedings AAAI-96* Menlo Park, Ca

[11] E. Charniak: "A Maximum-Entropy-Inspired Parser". *Proceedings ANLP-NAACL'2000,* Seattle, Washington

[12] S. Seneff: "TINA: A Natural Language System for Spoken Language Applications". *Computational Linguistics 18(1), 61-86.*

[13] R. Bod: "Memory-Based Models of Melodic Analysis: Challenging the Gestalt Principles". *Journal of New Music Research* 31 (2002) 27-37.

[14] R. Bod: "Probabilistic Grammars for Music Proceedings*" BNAIC 2001*, Amsterdam

[15] Z. Juhász: "Segmentation of Hungarian Folk Songs Using an Entropy-Based Learning System". *Journal of New Music Research* 33 (2004) No 1, (pp 5-15).

[16] D. Conklin, Ch. Anagnostopoulou: "Segmental Pattern Discovery in Music", *Informs Journal on Computing*,
Vol. 18, No. 3, Summer 2006, pp. 285-293
DOI: 10.1287/ijoc.1040.0122

[17] T. Kohonen: „Self-organising Maps". Berlin:Springer-Verlag

[18] P. Toiviainen: "Symbolic AI Versus Connectionism in Music Research". *In E. Mirinda (Ed.), Readings in Music and Artificial Intelligence.* Amsterdam: Harwood Academic Publishers (2000).