PARATAXIS: MORPHOLOGICAL SIMILARITY IN TRADITIONAL MUSIC

Andre Holzapfel
Institute of Computer Science, FORTH and
University of Crete
hannover@csd.uoc.gr

Yannis Stylianou
Institute of Computer Science, FORTH and
University of Crete
yannis@csd.uoc.gr

ABSTRACT

In this paper an automatic system for the detection of similar phrases in music of the Eastern Mediterranean is proposed. This music follows a specific structure, which is referred to as *parataxis*. The proposed system can be applied to audio signals of complex mixtures that contain the lead melody together with instrumental accompaniment. It is shown that including a lead melody estimation into a state-of-the-art system for cover song detection leads to promising results on a dataset of transcribed traditional dances from the island of Crete in Greece. Furthermore, a general framework that includes also rhythmic aspects is proposed. The proposed method represents a simple framework for the support of ethnomusicological studies on related forms of traditional music.

1. INTRODUCTION

In the field of ethnomusicology, computer based methods are adequate for simplifying musicological studies. Useful methods can be the recognition of intervals played by an instrument, or determining the meter structure of a signal. Using such methods, a search engine can be developed that can detect similarities between different pieces. Such a tool is valuable for research in ethnomusicology, because it enables to get a faster access to pieces that are interesting for a comparison. In this paper, a general framework for the morphological analysis of the Eastern Mediterranean traditional music is proposed and the parts related to melodic characteristics are presented and evaluated.

In general, morphology of music is defined as the methodical description of the structure of the form of musical works [12]. The elements of this organization are themes, phrases and motives, which themselves are made up of sound characteristics like tonal height, duration, intensity and timbre. The analysis aims at the discovery of the sentence structure (Periodenbau) and the transformative structure of these elements. This discovery is the core of morphological analysis.

Recently, the research presented in Sarris *et al.* [11] shed light on the difficulty of understanding traditional music in the eastern Mediterranean area: to a great extent, it is following a different kind of morphology, the logic of *parataxis*. The term *parataxis* stems from the field of linguistics, where it denotes a way of forming phrases using short sentences, without the use of coordinating or subordinating conjunctions [10]. In music following this logic, the tunes are built from small melodic phrases which do not follow a specific morphologic structure. This means, that there is no composed elaboration of a theme like for example in a fuga, neither there is a clear periodic structure, according to which a musical theme is repeated, like the repeating element of a chorus in popular western music. As mentioned in Theodosopoulou [13], it is a major effort to transcribe and analyze a big number of pieces. In this paper, the goal is to derive at least some conclusions about the content and similarity between pieces in an automatic way. Thus, a concept is presented that is aimed to discover recurring elements in a musical signal. These recurring elements are the melodic phrases that are the characteristic themes of the music following the logic of *parataxis*. The recognition of these phrases and their assignment to a specific dance appears to be a complex task even for a human being. In interviews the author conducted with local musicians, repeatedly the recognition of a dance was connected with the recognition of a specific melodic phrase. This process was also described in Tsouchlarakis [15]. Also, in listening test conducted *e.g.* in [6], it was observed that dancing teachers had memorized almost all melodies they have been presented with. With this knowledge they were able to conduct assignments to a class of dance much faster and with higher accuracy than their students. It is apparent that the similarity estimation between the used motifs is the key to a concept for a search engine for this music.

Recently, similarity in folk song and traditional melodies has drawn increasing attention of the Music Information Retrieval research community. Most of the related publications investigate symbolic transcriptions of melodies [7, 14,16]. For audio signals, Moelants *et al.* [9] and Bozkurt [1] derive pitch histograms from monophonic recordings, the former using African music and the latter in the context of Turkish music. Both methods are aimed towards the recognition of underlying tonal concepts (*i.e.* scales or *makams*, respectively), and stress the importance of a finer frequency solution than the one provided by the chroma features. Cabrera *et al.* [2] investigate the estimation of melodic similarity on a set of mainly monophonic vocal Flamenco recordings.

In this paper, the goal is to estimate similarities between

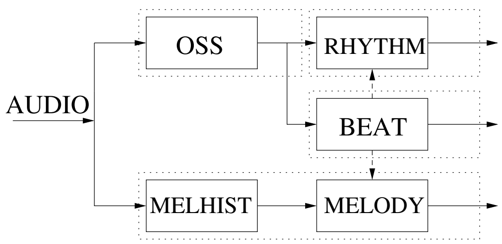

Figure 1. Block diagram of the proposed morphological analysis system

lead melodies in polyphonic mixtures. The focus lies upon the melodic aspects of the morphological analysis system depicted in Figure 1. This is achieved by the beat synchronous computation of melody histograms (MELHIST), as detailed in Section 4. The rhythmic aspects, such as the computation of onset strength signals (OSS), rhythmic similarity and beat tracking on this kind of music has been the subject of investigation in other publications [4, 6]. An integration of rhythmic and melodic similarity as depicted in Figure 1 for such a task is meaningful: It has been repeatedly confirmed by local musicians, that not only the melody is of importance for recognizing a specific dance, but also the way the instrument player puts emphasis on particular notes of the melody. However, the type of available data made it necessary to concentrate on the melodic aspect, as will be detailed in the following Section.

2. DATASET

A small dataset of polyphonic samples has been collected that enables for a preliminary evaluation of a system for the detection of morphological similarity. For this, samples from the Crinno data of the Institute of Mediterranean Studies [1] have been used. In the Crinno collection for some samples of the dance *Sousta* the lead melodies have been transcribed by musicologists and then analyzed for their morphology. All encountered phrases have been indexed, and using the list of these indexes it is feasible to locate the morphologically identical phrases in different pieces. The way to index the phrases follows the method described in Theodosopoulou [13]: the phrases have a length of either one or two bars as shown in Figures 2 and 3. Based on Theodosopoulou [13], during the analysis the first encountered two bar phrase will be titled `1a1b`. If, for example, the next encountered two bar phrase contains the second part of the first phrase in its second measure, while its first measure is an unknown phrase, then it will be titled `2a1b`, denoting the partial relation with the first pattern. In Figures 2 and 3 the titles of the depicted melodic phrases are denoted above the score. It is obvious that an exact partial or complete matching can be localized by using this way of indexing the phrases. However, no conclusions can be drawn about the similarity of phrases with different titles. As the amount of transcribed data is rather small

[1] http://gaia.ims.forth.gr/portal/

Figure 2. Example of a one measure melodic phrase

Figure 3. Example of a two measure melodic phrase

(20 pieces), there are not many phrases that appear several times in various pieces. It has been achieved to compile a data set of 40 sound samples, each containing a complex musical mixture signal with the instruments Cretan *laouto* and *lyra* and sometimes singing voice. Each sample contains several repetitions of melodic phrase of two measures length. Each of the 40 pieces has a "partner" within the dataset that contains a similar or equal musical phrase played by the *lyra*, according to the analysis of musicologists. Thus, in this dataset exist 20 pairs of samples that contain similar phrases. Please note that according to the musicological analysis these phrases are exactly the same. However, the audio files differ because they are performed by different artists and vary due to their different playing style.

3. MELODIC PATTERN SIMILARITY

For the computation of similarity, a baseline system as presented in Ellis and Poliner [3] will be used. This system uses beat synchronous chroma features to describe the melodic content. It was proposed for the detection of coversongs in western pop music, and it will serve as a starting point for the studies of detecting morphological similarity in traditional music. The first computational step in this approach is a beat tracking that uses a spectral flux like OSS as an input, and derives the beat time instances using dynamic programming. Then, for each beat time a 12-dimensional chroma feature is computed. These chroma features record the intensity associated with each of the 12 semi-tones of the equal-tempered tonal system. In order to determine, how well two songs match, the cross-correlations between two feature matrices are computed for each possible transposition. In the following, this system will be referred to as BASE-SYS.

As the sound files are complex mixtures, melodic similarity is degraded by the other instruments contained in the mixture, which play to some extent a similar accompaniment in all examples that is characteristic for this type of dance. Thus, a lead melody extraction using a method as the one proposed in Klapuri [8] could be included as a pre-processing. Furthermore, instead of using chroma features, in the context of traditional music melodic histograms of a finer resolution have been found useful for the classification of melodic content [1]. In order to determine if such approaches can be adapted to the beat synchronous melody description framework, the lead melody will be estimated using the algorithm presented in Klapuri [8], which was provided by the author of the paper. The parameters given as input to the algorithm are the desired number of fundamental frequency tracks to be estimated from the signal (set to 1), and the fundamental frequency range of the desired F0 tracks. This range was set to $60Hz...480Hz$, after an analysis of the available scores of the recordings. The next step is the computation of beat synchronous melody histograms. Motivated by the work presented in Bozkurt [1], the frequency resolution of these histograms is set higher than necessary for music using scales of the equal-tempered system. This is because in Greek traditional music many modal scales are encountered which make use of tonal steps different from the half tone of the equal-tempered system. For example, some of these scales have their roots in the scales investigated in Bozkurt [1]. Scales like *Hidzaz* and *Kurdi* are examples for this case, and because these scales are also used in Cretan music the finer resolution of the histograms is theoretically justified. Thus, for a song a matrix is obtained with one column for a beat instance which contains the melody histogram for this beat. Again, for matching two samples the method proposed in Ellis and Poliner [3] has been used in the same way as for the chroma features. The system that uses this kind of melody histograms will be referred to as HIST-SYS.

In Ellis and Poliner [3], the features are computed beat synchronous. This means that a beat tracking is necessary as a pre-processing step. For this purpose, OSS derived from amplitude are used to perform the beat tracking [3]. However, results described in Holzapfel and Stylianou [5] indicate that for the investigated type of music a beat tracking based on phase characteristics gives more accurate results. Thus, it should be evaluated as well if the accuracy of the beat tracking has some impact on the results of the matching experiments.

4. EVALUATION METHODS

Two different evaluation methods are suggested. In the first one, only the 40 short samples containing the melodic phrases are used to compute their mutual similarity regarding melodic content. The quality of the obtained similarity measure can be evaluated using the Mean Reciprocal Rank (MRR)

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{Q}\frac{1}{rank_i} \tag{1}$$

where $|Q|$ is the number of queries. For our data set this means that each sample is used as a query once, *i.e.* $|Q| = 40$. If *e.g.* the correct partner is found on place 3 of the most similar samples, the reciprocal rank is $\frac{1}{3}$. This means that the closer the MRR is to the value 1, the better the similarity measurement.

In the second evaluation method, a sample from the dataset is used as a query and similarities are computed for the whole duration of the piece that contains its partner motif at some time instance. If this similarity measure shows a peak at the position of the true partner, the goal of locating it in a continuous piece is achieved.

5. EXPERIMENTS

5.1 Setup 1: Matching pairs

In the first experiment, the BASE-SYS system was applied to the data set of 40 song excerpts. Each song was used as a query and the mean reciprocal rank as defined in (1) was computed, which resulted in a value of $MRR_{BASE-SYS} = 0.38$, as shown in Table 1.

In the following we will show that the performance in terms of the mean reciprocal rank of the BASE-SYS system can be improved by involving an estimation of the main melody from the polyphonic samples and the usage of high resolution histograms in the HIST-SYS system. In Bozkurt [1], a resolution of one Holdrian comma (Hc) was referred to as the smallest interval considered in Turkish music theory, and the authors used a resolution of $\frac{1}{3}$ Hc for their histograms. One Holdrian comma is equal to 22.6415 cents, and the octave interval can be divided into 53 Hc or 1200 cents. Various resolutions have been tried, but no clear result regarding the optimum value could be obtained on the limited sized dataset. For that reason, the resolution was set to 2 Hc, or about 2.25 times higher than the resolution of equal-tempered scales (about 4.5 Hc). As it can be seen from the second row in Table 1, the obtained mean reciprocal rank (0.58) is improved compared to the BASE-SYS system. This improvement is present almost independently of the histogram resolution, which indicates that the sensitivity to microtonal changes is not of importance at least for the present dataset. We acknowledge, however, that bigger and more diverse datasets have to be obtained to achieve more insight into the parameter settings.

Table 1. Mean reciprocal rates (MRR)

BASE-SYS	0.38
HIST-SYS	0.58

5.2 Setup 2: Matching queries in whole songs

As described in Section 4, the second evaluation method is using one of the short samples contained in the dataset as a query. For this experiment 10 phrases of two measures length have been selected as depicted in the first column of Table 2. For example, the query file `13b42b:234` is the phrase 13*b*42*b* taken from the recording number 234 in the collection. The target file is the whole piece which contains the partner of the query at some time instance (*i.e.*

Table 2. Results of matching patterns from MS1 in whole song files

QUERY FILE	$max(R_{neg})$	R_{source}	$max(R_{pos})$	MATCH
(1) 13b42b:234	0.5796	0.9200	0.6403	EXACT
(2) 4a31b:217	0.3602	0.9301	0.6741	EXACT
(3) 3a3b:027	0.5059	0.9297	0.6238	CORRECT
(4) 35a35b:196	0.5482	0.9416	0.6866	CORRECT
(5) 3a21b:051	0.4511	0.8549	0.7040	EXACT
(6) 89a46b:143	0.4881	0.6571	0.5451	EXACT
(7) 31a31b:035	0.4830	0.8989	0.6351	WRONG
(8) 6a72a:167	0.5535	0.8778	0.6578	EXACT
(9) 7a6b:008	0.5073	0.8242	0.5870	EXACT
(10) 62a62b:249	0.4484	0.8333	0.5869	EXACT

a different interpretation of the same phrase). It has been tried to locate the appearance of the phrase in the target file using the HIST-SYS method, which lead to the best pattern matching results as shown in Table 1. The highest correlation measures in these files are depicted in the column titled $max(R_{pos})$ in Table 2. In the column titled `MATCH` the success of this matching is judged. If the position connected to this highest correlation measure is exactly the position where the partner file has been extracted from, then the label `EXACT` is assigned. If the position of the correlation maximum is related to another appearance of the same pattern in the file, it is labeled as `CORRECT`. Finally, when a different pattern from the query pattern is located at the position of the correlation peak, the label `WRONG` is assigned. This evaluation has been performed entirely by hand, by locating the time instance of the correlation maximum of the melody histogram in the related musical score. It can be seen that only in one case the matching gave a wrong result, while all the other 9 matches were related to an appearance of the same melodic phrase in the target file.
Let us stress again that all the target files are different from the file that the query was taken from: The target files used in the column titled $max(R_{pos})$ are different recordings than the recordings which contain the query at some time instance. They have been recorded by different players, but they contain at one or more time instances a melodic phrase that has been judged to be identical with the query by an analysis conducted by musicologists.
The correlation between the F0 histogram of a query sample and the histogram of the whole recording it has been extracted from has been computed as well. This means that at some time instance exactly the same pattern is encountered, without the variation introduced by a different interpretation. This enables to determine how good the matching works in the perfect case, when the pattern we are looking for, is indeed contained in the file exactly as found in the query. The resulting correlations are depicted in the column entitled R_{source} in Table 2. It can be seen that the correlations shown in R_{source} are always larger than the correlation depicted in $max(R_{pos})$, but never equal to 1. This is likely to be caused by slightly differing beat tracking and F0 estimation results on the small query samples and on the whole file.

Furthermore, each query has been applied also to a file, where according to the annotation the phrase is not contained neither as a whole nor half of it. The correlation maxima are depicted in the column titled $max(R_{neg})$, and these values are always smaller than the correlation values computed in the other columns. This supports the assumption that the proposed method is able to separate similar phrases from those that do not share a large similarity with the query phrase.
In Figures 4 and 5, all R_{pos} vectors of the 10 queries shown in Table 2 are plotted. These vectors have been obtained by computing the two dimensional correlations between the query and the target histogram matrices, and the choosing the row (*i.e.* the tonal transposition between the files) in the correlation matrix, that contains the maximum value. In all plots, maxima have been chosen and it has be evaluated if at the related measures in the score indeed the query phrase is found. For these cases, maxima are shown with dashed boxes, while maxima which are not related to the query pattern have been marked with dotted boxes.
A first and important result of this analysis is that in none of the cases an occurrence of the query pattern in the investigated audio file has been missed, which means that in every case the occurrence of the pattern was related to a maximum in R_{pos}. Also the overall number of true positives (dashed boxes) is 21 while the number of false positives (dotted boxes) is only 7. However, these false positives do not imply that there is no similarity between the query and the target at the time instance of the false positive. The false positive only indicates that at this position the phrase played by the lead instrument does not have exactly the same label. Taking a closer look at the false positives reveals that for example all wrong detections for query (3) are phrases which contain the pattern `3a` which is also contained in the query sample (`3a3b`). A closer look has been taken at the only case, where the maximum in R_{pos} is connected to a false positive (query (7)). The query phrase and the phrases found in the dotted boxes in Figure 5.(7) are depicted in Figure 6. It is apparent that at least the first parts of the two phrases share a big amount of similarity. Thus, at least in this case, the false positive is related to a similar melodic phrase.

Another observation from Figures 4 and 5 is that max-

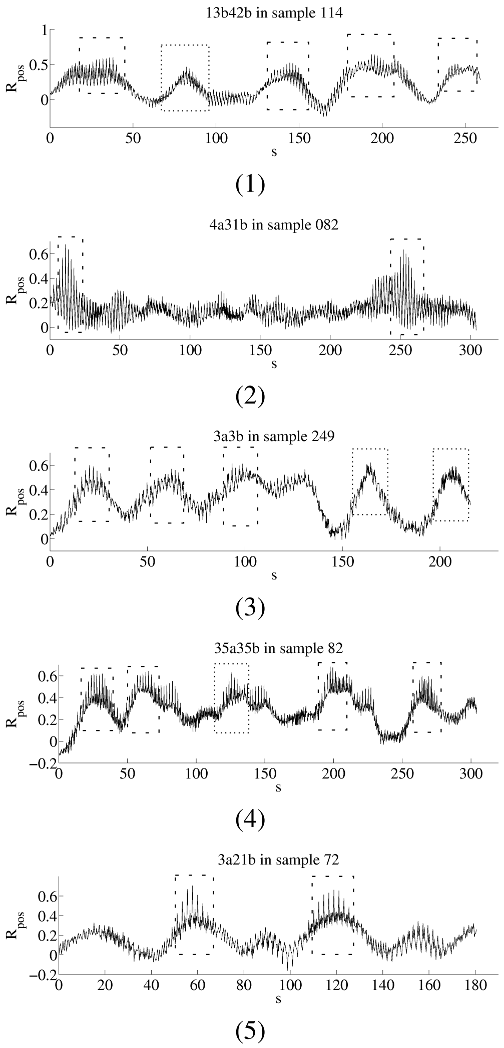

Figure 4. Complete R_{pos} obtained for queries 1-5 in Table 2, positive matches in dashed boxes, negative matches in dotted boxes

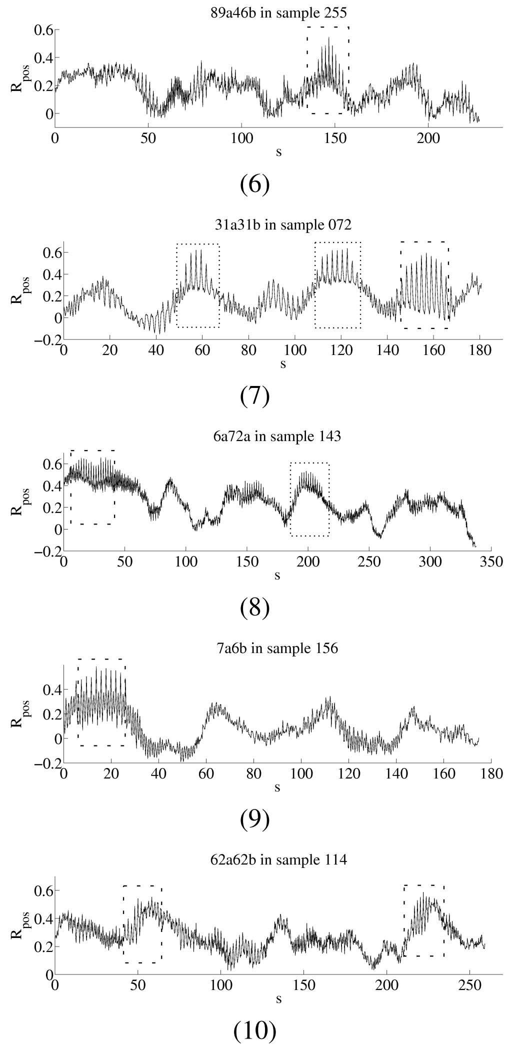

Figure 5. Complete R_{pos} obtained for queries 6-10 in Table 2, positive matches in dashed boxes, negative matches in dotted boxes

ima related to true positives seem to be characterized by a strong oscillation. This oscillation has been observed to have the frequency of exactly two measures. This means that the correlation shows a strong peak whenever the beginnings of the query phrase and the related phrase in the investigated file are aligned. This effect should be further investigated when a larger dataset is available, and it is possible that a detection of such oscillations, besides high correlation envelopes, further improves the result of the pattern retrieval.

Finally, the impact of the beat tracker has been evaluated. In order to determine how large the change in the matching procedures would be if the beat tracking and hence the synchronization is optimized, all samples in the dataset and all complete samples used for the computation of R_{pos} in Table 2 have been beat annotated by the author. However, rerunning all experiments in the experimental setups 1 and 2 using these ground truth beat annotations did not qualitatively change the results. Since the original beat tracker used in Ellis and Poliner [3] lead mainly to local misalignments with the beat annotation, and it has to be concluded that these misalignments have no impact on the systems used in this work, at least when applied to the limited size of data that is currently available.

Figure 6. Two phrases found to be similar in query (7)

6. CONCLUSION

In this paper, methods have been evaluated that help to detect melodic similarity in polyphonic recordings following the logic of *parataxis*. It could be shown that a method based on histograms of the F0 estimation of the leading melody enables for an improvement compared to a baseline system that uses chroma features. Furthermore, it could be illustrated that the proposed method is capable of spotting appearances of small melodic patterns in a whole audio file, even when both files are polyphonic mixtures and the query pattern has been derived from a different recording. Such a method can be a valuable tool for research in the field of musicology, where similar phrases in a large collection could be located without the necessity of transcription, thus leading to a large saving of time.

Furthermore, the integration of melodic and rhythmic aspects is straight-forward, and it is likely to improve results for datasets in which different types of rhythms are contained. As features for melody and for rhythm can both be computed in a beat synchronous way, the correlation values obtained for a query from these two aspects could be simply added, or by using some weighting that favors either melody or rhythm derived correlations. However, the rhythmic similarity measure is quite questionable on the available dataset which is rhythmically very homogeneous, and for that reason it had to be postponed. As a future goal, the integration of rhythmic similarity as depicted in Figure 1 has to be evaluated on a more diverse dataset. However, for the compilation of such a dataset the support of experts in musicology is necessary.

7. REFERENCES

[1] Baris Bozkurt. An automatic pitch analysis method for turkish maqam music. *Journal of New Music Research*, 37(1):1–13, 2008.

[2] Juan J. Cabrera, Jose Miguel Díaz-Bánez, Francisco J. Escobar-Borrego, Emilia Gómez, Francisco Gómez, and Joaquín Mora. Comparative melodic analysis of a cappella flamenco cantes. In *Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08)*, 2008.

[3] Dan Ellis and G. Poliner. Identifying cover songs with chroma features and dynamic programming beat tracking. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP*, pages IV–1429–1432, 2007.

[4] Daniel P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[5] Andre Holzapfel and Yannis Stylianou. Beat tracking using group delay based onset detection. In *Proc. of ISMIR - International Conference on Music Information Retrieval*, pages 653–658, 2008.

[6] Andre Holzapfel and Yannis Stylianou. Scale transform in rhythmic similarity of music. *Accepted for publication in IEEE Transactions on Speech and Audio Processing*, 2010.

[7] Zoltán Juhász. Motive identification in 22 folksong corpora using dynamic time warping and self organizing maps. In *Proc. of ISMIR - International Conference on Music Information Retrieval*, pages 171–176, 2009.

[8] Anssi Klapuri. Multiple fundamental frequency estimation by summing harmonic amplitudes. In *Proc. of ISMIR - International Conference on Music Information Retrieval*, pages 216–221, 2006.

[9] Dirk Moelants, Olmo Cornelis, and Marc Leman. Exploring african tone scales. In *Proc. of ISMIR - International Conference on Music Information Retrieval*, pages 489–494, 2009.

[10] Edward P. Morris. *On Principles and Methods in Latin Syntax*. New York, C. Scribner's sons, 1901.

[11] Haris Sarris, Tassos Kolydas, and Panagiotis Tzevelekos. A framework of structure analysis for instrumental folk music. In *Proc. of CIM08, 4th Conference on Interdisciplinary Musicology*, Thessaloniki, Greece, 2008.

[12] Dimitris Themelis. *Morphology and analysis of music, (in Greek language)*. University Studio Press, Thessaloniki, 1994.

[13] Irini B. Theodosopoulou. *Methodology of morphological analysis and analytic data of small rhythmic patterns of cretan folk music, (in Greek Language)*. Athens: Kultura, 2004.

[14] Petri Toiviainen and Tuomas Eerola. Method for comparative analysis of folk music based on musical feature extraction and neural networks. In *In III International Conference on Cognitive Musicology*, pages 41–45, 2001.

[15] Ioannis Tsouchlarakis. *The dances of Crete - Legend, History, Tradition (in Greek language)*. Center of Cretan Culture Studies, Athens, 2000.

[16] Peter van Kranenburg, Anja Volk, Frans Wiering, and Remco C. Veltkamp. Musical models for folk-song melody alignment. In *Proc. of ISMIR - International Conference on Music Information Retrieval*, pages 507–512, 2009.