Music Information Retrieval Trough Melodic Similarity Using Hanson Intervallic Analysis

Johnny F. Serrano [α,β] *and José M. Iñesta* [α]

[α] Departamento. de Lenguajes y Sistemas Informáticos
Universidad de Alicante, España
[β] Centro de Investigación en Computación e Informática
Universidad Latina, Costa Rica
`johsersa@gmail.com, inesta@dlsi.ua.es`

Abstract. In music information retrieval, the melodic similarity is used as main feature for the detection of relevant information. Among the possible applications there is a detection of plagiarism of ideas exposed by an artist in the past, the payment for copyrights using detection of musical pieces in radio transmissions, the computer-aided composition, etc. There exist several techniques exposed in melodic similarity that use diverse statistical and probabilistic analysis. The objective in this work is to establish a text words equivalent to musical notation using a representation based on intervallic relationships and duration, and to evaluate three textual information retrieval techniques applied to this representation, as well as to propose changes to improve the system's performance.

1 Introduction

The music representation through descriptors such as duration, pitch, contour, intervals, etc., is common in the system analysis and musical information retrieval [1,2]. The two dimensions in which a score is codified (vertical for pitch and horizontal for duration) like symbolic registration of the music have been represented in very different ways, in order to carry out analysis as probabilistic techniques like hidden Markov chains [3], entropy analysis [4] and statistical correlations [4], among other [5].

The intervallic analysis will be the fundamental unit for the melodic similarity task, where the musical fragments are analyzed in their intervallic relationships in a tempered chromatic scale [6]. The interval is the basic unit of analysis, putting aside the pitch and the tonality as direct descriptors. Some previous works have used this concept making a cartesian notation of the intervals [7] or determining the correlation of the intervallic progressions [8]. A well-known work that makes special emphasis in the relationships among contiguous intervals may be the one of Narmour [9]. The author establishes a series of groups of contiguous intervals (implication patterns - resolution) that allow analyzes melodies through their transformation in symbols. An important disadvantage of this focus is that it ignores the relationships of durations, which is an inconvenience for the information retrieval.

© A. Gelbukh, S. Torres, I. López (Eds.)
Advances in Artificial Intelligence
Research in Computing Science 20, 2006, pp. 131-142

This work explored the conversion to text words from the extracted intervals of MIDI documents. This focus is not completely new in the sense that other authors have previously used techniques coming from the analysis of texts to solve problems related with the processing of the music information.

One possible encoding is the one proposed by Downie [10] or Doraisami and Rüger [11] that have used models based in n-grams for melody identification. On the other hand, in [12] the authors use words models to identify musical styles.

Three techniques of textual information retrieval have been used: Cosine [13], Kaszkiel [14] and Pivoted Cosine [15] to check the effectiveness in the determination of the retrieval for relevant information through melodic similarity. Also, this work proposes some modifications in the techniques mentioned above to improve the performance.

2 Methodology

2.1 Conversion to Text Words

Starting from a musical fragment (see Fig. 1), it takes the first note and the distance is calculated in semitones with regard to the following note. This distance (interval) is represented with the letters d-s-n-m-p-t of the Hanson system [6] plus the five vowels a-e-i-o-u to form the intervals of two octaves and one half[1] [16] (see Fig. 2). Thus, the intervals are coded into characters in this form.

The original work of Hanson was addressed to the modern music analysis. Because of that, it is convenient to introduce some modifications to adapt the code to this work. A third letter is added to the representation as follows: "l" to represent a descendent interval and "r" for an upward interval. The representation of an interval of zero semitones was not considered originally in the Hanson's analysis but it has been considered in this work to obtain a more complete representation of the melodies. For it the word "loo" has been used. This example shows some intervals and its corresponding textual codes.

Also, the duration of the two notes that conform it is considered as the duration of the interval representing units according to their musical figure, where the value 1 is given to a black note. This way, the duration can take fractional values. A couple of examples of Hanson intervallic representation of the melody can be appreciated in Fig. 3. This representation is a simple and comprehensible musical notation of a fragment. Through the definition of any initial note, one melody can be reproduced in different tonalities keeping its similarity [16]. Therefore, when untying the tonality of the melody the analysis is simplified and it is focused to the relationship between the intervals and their durations.

[1] These syllables have been used for many years in teaching systematic intervallic solfege.

Interval in semitones: 1 2 3 4 5 6 7 8 9 10 11 12 . .
Corresponding codification: da sa na ma pa ta pe me ne se de te . . .

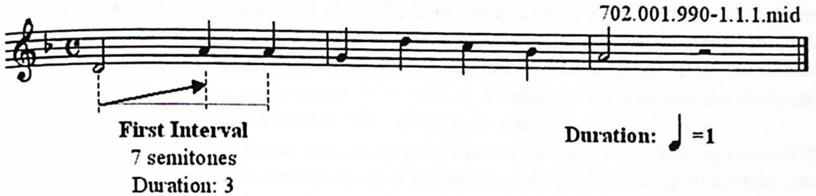


Fig. 1. Interval analysis.

Fig. 2. Syllables' examples of the systematic-intervallic solfege. In the superior part of the staff, the syllables corresponding to the code of the interval are indicated between the intermediate note and the superior one of each chord, and the lower code corresponds to the intermediate note in relation to the lowest one.

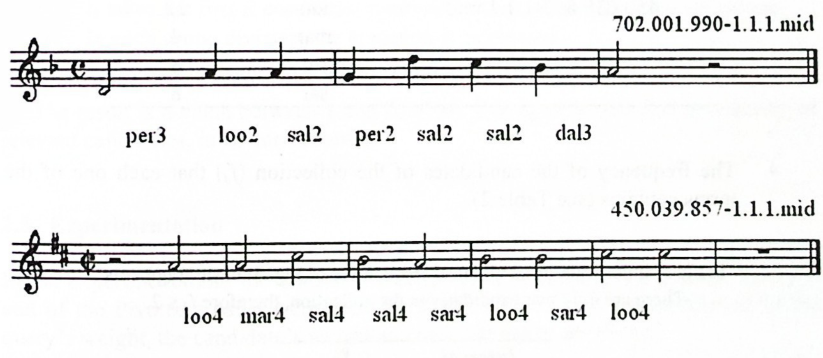


Fig. 3. Hanson intervallic representation for two music fragments. The code showed over the scores is the key corresponding to the RISM database (see section 2.3).

2.2 The Information Retrieval System

One of the objectives of the presented work is to export techniques, commonly used in textual information retrieval, to the field of music information. In this way, we will consider that the documents we work with are monophonic melodic fragments, and the terms of these candidates are composed in textual codes for each one of the intervals present in such candidates.

Given a database of melodic fragments, the indexation will made through the following procedure. First, the musical fragments are converted to text according to the methodology described. Then, an index is extracted out of each fragment, which is addition of durations for all the intervals of the same type found in this fragment. In this way, the concept of "frequency" which is characteristic of the field from the textual information retrieval is assimilated as the duration of the notes in each melodic fragment. For it is determined as follows:

- The frequency (i.e., the duration) of the terms (intervals of one type) in the document (f_t) (see Table 1). For example, in the first candidate, the interval "sal" appears 3 times with duration of 2 units in each interval, because of that, the value of the f_t of the interval "sal" is 6.

Table 1. Indexation for frequencies of terms corresponding to Fig. 3 music fragments.

Music Fragment	*Interval*	f_t
702.001.990-1.1.1.mid	dal	3
	loo	2
	per	5
	sal	6
450.039.857-1.1.1.mid	loo	12
	mar	4
	sal	8
	sar	8

- The frequency of the candidates of the collection (f_d) that each one of the terms contains (see Table 2).

Table 2. Indexation by candidates' frequencies of Fig. 3 music fragments. There are only two candidates in the collection, therefore $f_d \leq 2$.

Interval	f_d
dal	1
loo	2
mar	1
per	1
sal	2
sar	1

2.3 Data Set

The corpus used for the experimentation has 581 melodic fragments with an average of 14 intervals for melody. The format of the melodies is MIDI. These melodies conform the answers to 11 queries generated by the team of Rainer Typke of the Utrecht University [17], being the most similar (relevant) candidates evaluated and ordered by 35 experts in the musical field starting from the collection RISM A/II out of more than 456.000 melodies.

The 11 consultations have around 50 similar melodies. These are also ordered according to a degree of similarity and in groups of same similarity. In the Fig. 4, a query example and an extract of the melodies that are more relevant to this query are visualized

2.4 Evaluation

In this section the measures proposed are presented to evaluate the precision, recall and capacity of the algorithms used in this work to retrieval the candidates in an appropriate order of relevance. Two methods will be used for this evaluation:

- Evaluation according to the TREC[2] [18] competitions that considers all the answers with the same degree of relevance, where the (not interpolated) average precision, R-precision, the precision at 5 candidates and the charts of precision versus recall interpolated will be mentioned.
- ADR (average dynamic recall) evaluation [19] also takes into consideration a degree of relevance of each candidate, where groups determine the degrees of relevance.

Considering R as the number of relevant candidates for a query:

- It takes the first *R* candidates returned by the information retrieval system.
- In each group division, the precision is calculated.
- The precision curve is integrated and is divided by *R*.

The result is a value between 1 and 0, where 1 is equal to retrieval the entirety of relevant candidates, in its correct order.

2.5 Experimentation

In the experimentation, the information retrieval techniques of the Cosine, Kaszkiel and of the Pivoted Cosine were used. Next, the formulas for the calculation of the query's weight, the candidate's weight and their similarity are shown.

[2] TREC: Text Retrieval Conference.

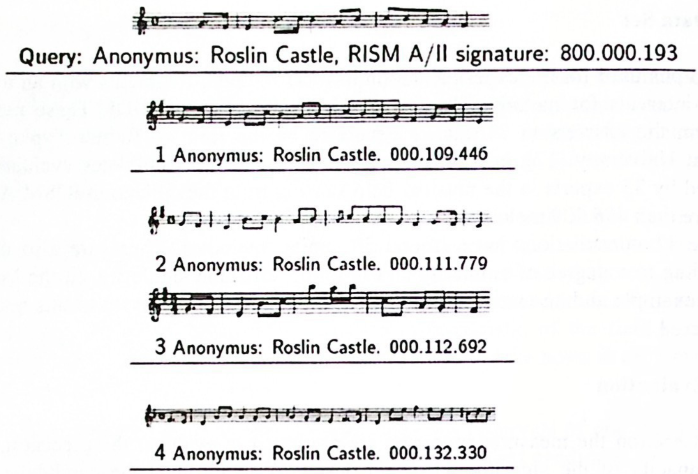


Fig. 4. Query example (above) with the 4 more relevant candidates in order of relevance.

2.5.1 Cosine

Given two fragments, one for query, Q, and other the candidate, D, selected in a database where there are N candidates, the similarity is defined between both as:

$$sims(Q,D) = \frac{\sum_{i=1}^{k} q_i \cdot d_i}{\|Q\| \cdot \|D\|}, \tag{1}$$

where $\|Q\| = \sqrt{\sum_{i=1}^{k} q_i^2}$, $\|D\| = \sqrt{\sum_{i=1}^{k} d_i^2}$ and k is the number of terms in Q.

The weight of each term is defined as $q_i = ft_{q,i} \cdot \log_e(\frac{N}{fd_i})$ for the terms of the query and $d_i = ft_{d,i} \cdot \log_e(\frac{N}{fd_i})$ for the each candidate. **(2)**

These expressions $ft_{d,i}$ represent the frequency of the term *i-th* of the candidate and $ft_{q,i}$ the frequency of the term *i-th* of the query.

2.5.2 Kaszkiel:

In an equivalent way, this measure of similarity among the consultation Q and the candidates D is defined as:

$$sims(Q,D) = \frac{\sum_{i=1}^{k} q_i \cdot d_i}{\|Q\| \cdot \|D\|}, \tag{3}$$

where $\|Q\| = \sqrt{\sum_{i=1}^{k} q_i^2}$, $\|D\| = \sqrt{\sum_{i=1}^{k} d_i^2}$, and the weight of the terms:

$$d_i = \log_e(ft_{d,i} + 1) \ , \ q_i = \log_e(ft_{q,i} + 1) \cdot \log_e(\frac{N}{fd_i} + 1) \tag{4}$$

The initial evaluation was made using these techniques shown above. The results were not more favorable than that technique of Cosine. It was determined that the normalization of the natural logarithm affected negatively because it attenuated in an excessive way the frequencies, therefore it was replaced by the square root of the frequencies. In this way, the weights of the terms of the equation (4) were finally substituted for:

$$d_i = \sqrt{ft_{d,i}} \ , \ q_i = \sqrt{ft_{q,i}} \cdot \log_e(\frac{N}{fd_i} + 1) \tag{5}$$

2.5.3 Pivoted cosine

The third technique used to evaluate the similarity between the query and the candidates, is defined as:

$$sims(Q,D) = \sum_{i=1}^{k} \frac{q_i \cdot d_i}{W_d}$$, and the weights:

$$d_i = 1 + \log_e(ft_{d,i} + 1) \ , \ q_i = 1 + \log_e(ft_{q,i} + 1) \cdot \log_e(\frac{N+1}{fd_i}) \tag{6}$$

In the same way as for Kaszkiel's case and since the Pivoted Cosine neither improved the technique of the Cosine, the normalization of the natural logarithm of the equation was replaced (6) for the square root of the frequencies:

$$d_i = 1 + \sqrt{ft_{d,i}} \ , \ q_i = 1 + \sqrt{ft_{q,i}} \cdot \log_e(\frac{N+1}{fd_i}) \tag{7}$$

The normalization factor that appears in the equation (6) would be:

$$W_d = (1 - slope) + slope \cdot \frac{L_B}{\overline{L_B}}, \tag{8}$$

being *slope* the factor of participation of the normalization, L_B the candidate longitude in Bytes, and $\overline{L_B}$ the average longitude of the all candidates in Bytes. But in the results presented in section 3.1 of this work, modification of this normalization factor has been used, because the norm has been used with regard to the average duration of all candidates' intervals. As follows:

$$W_d = (1 - slope) + slope \cdot \frac{L_I}{\overline{L_I}} \quad , \qquad (9)$$

where L_I is the total duration of the candidate's intervals and $\overline{L_I}$ is the average of the total duration of intervals for the candidates.

3 Results

3.1 Preliminary Results

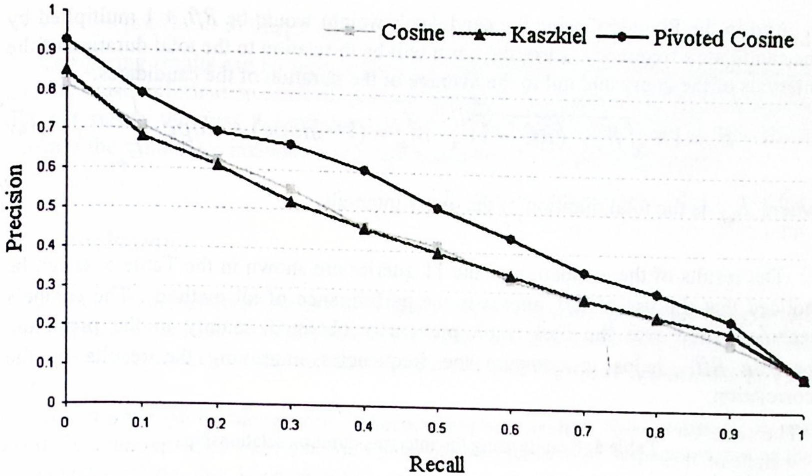
The first test was to evaluate the performance applying the three methods, using as weight of the query and weight of the candidate, the respective frequencies. In the case of the Cosine, the frequencies were normalized between 0 and 1 with regard to the maximum frequency found inside the analyzed candidate. And in the case of the Pivoted Cosine, the norm was used with regard to the average of candidates' intervals duration (23.55) with a *slope* of 0.4. The modifications proposed for the normalization of the weights in the cases of Kaszkiel and pivoted cosine improved the results regarding those obtained in the preliminary tests with the original equations. The results of the evaluation of the 11 queries can be seen in Graph 1 and in Table 3.

It can be seen that the technique of the Pivoted Cosine provides better results with 55.18% of ADR for the average of the 11 queries. Observing the results individually for each question, it can be noticed how it affects the Pivoted Cosine since the result of some questions favors small candidates, but in other ones the result is good, close or above 70%.

Table 3. Preliminary Results

Technique	*AP not int.*[3]	*Prec. at 5*[4]	*R-prec.*[5]	*ADR*[6]
Cosine	0.4062	0.6364	0.4245	0.4569
Kaszkiel	0.3968	0.5455	0.4144	0.4436
Pivoted Cosine	0.4900	0.7091	0.4805	0.5518

[3] AP not Int. : Not interpolated average precision
[4] Prec. at 5: Precision at 5 retrieval candidates
[5] R-prec.: R-precision.
[6] ADR: Average Dynamic Recall



Graph 1. Precision versus recall for the preliminary results

3.2 Normalization Change

As the objective is finding the most relevant candidates according to the query, in the case of melodic similarity it is important to make a hard analysis of the intervals duration. In order to do this, the factor Rft_i is established as the relationship between the durations of the candidate and the durations of the query. This value is calculated dividing the minor value by the major, and a numeric factor is obtained between 1 and 0 indicating the degree of similarity among the duration of the intervals of the same candidate's type with regard the query:

$$\text{if } ft_{q,i} > ft_{d,i}\ ,\ Rft_i = \frac{ft_{d,i}}{ft_{q,i}} \qquad (10)$$

$$\text{if } ft_{q,i} < ft_{d,i}\ ,\ Rft_i = \frac{ft_{q,i}}{ft_{d,i}} \qquad (11)$$

Next, the Cosine is evaluated using the Rft_i as the candidate's weight:

$$d_i = Rft_i \cdot \log_e\left(\frac{N}{fd_i}\right) \qquad (12)$$

In the case of Kaszkiel, the candidate's weight would be Rft_i multiplied by the frequency of the query term:

$$d_i = \sqrt{Rft_i \cdot ft_{q,i}} \qquad (13)$$

And in the Pivoted Cosine the candidate's weight would be $Rft_i + 1$ multiplied by the candidate's frequency. Also, the norm will be in relation to the total duration of the intervals of the query and not to the average of the duration of the candidates:

$$d_i = 1 + \sqrt{ft_{d,i} \cdot (Rft_i + 1)}\ ,\quad W_d = (1 - slope) + slope \cdot \frac{L_I}{L_{qI}}\ , \qquad (14)$$

where L_{qI} is the total duration of the query intervals.

The results of the evaluation of the 11 queries are shown in the Table 5. It can be noticed that the use of Rft_i improves the performance of all methods. The cosine's technique improves the best result previously obtained, mainly in the precision, because Rft_i helps to compare the frequencies improving the results of the correlation.

Table 5. Results using the intervals duration relationships[7].

Technique	*AP not int.*	*Prec. at 5*	*R-prec.*	*ADR*
Cosine	0.4645	0.7818	0.4703	0.5495
Kaszkiel	0.4798	0.8364	0.4653	0.5684
Pivoted Cosine	0.5314	0.8000	0.5112	0.6303

The Pivoted Cosine generates the best result with 63.03% of ADR for all 11 queries. The improvement obtained through Rfti is very significant (9% in precision and 8% in ADR).

Observing the results one by one for each query, it can be noticed how the result affects the Pivoted Cosine of some queries, since it favors small candidates, but in other, the result exceed 75%.

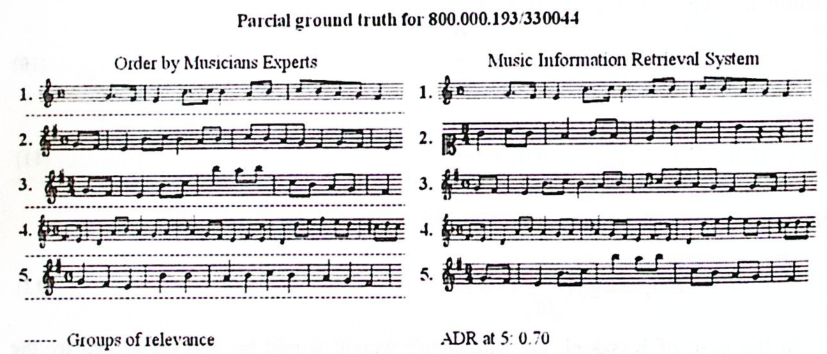


Fig. 5. Comparison results for a query between the music experts and the music information retrieval system.

[7] $slope = 0.3$

Comparing results can be seen in the Fig. 5 for a query between the music experts and the music information retrieval system. Although the information retrieval system did not return the first 5 candidates in the proper order, the order is acceptable because the candidates maintain a relevant similarity in relation to the query.

4 Conclusions

It has been demonstrated that the intervallic relationships are important elements to take into consideration for the melodic similarity task and that it is possible to make an equivalent to text words and use the current techniques of textual information retrieval.

In the case of Kaszkiel and Pivoted Cosine, a change was necessary in the normalization replacing the function of the natural logarithm for the square root of the frequency. Also, the improvements introduced to the three methods by the relationship of frequencies (Rft_i) demonstrated to be of great help, and they improved much results in the evaluation of the TREC as well as the ADR.

It is possible to increase the results using a balanced normalization of the musical fragments using the total duration of the intervals as the quantity of intervals that compose them.

Another solution could be the application of an initial filter for the recovery of most relevant candidates through these quick techniques, and then using a more precise algorithm, to obtain the final results.

Another pending task is studying the behavior of the system in function of the size of the musical fragments that are compared. The availability of a great database is necessary to carry out these tasks.

Acknowledgment

This work was possible thanks to the support of the Spanish Agency of International Cooperation (ref: AECI:B/2098/04). The Ministry of Science and Technology of Costa Rica and the CONICIT (ref: 84-2005), as well as the projects CICyT TIC 2003-08496-C04 and GV043-541.

References

1. Marsden, A. Pople (eds.) Computer Representations and Models in Music. Academic Press. 1992.
2. W.B. Hewlett, E. Selfridge-Field (eds.) The Virtual Score. Representation, Retrieval, Restoration. The MIT Press. 2001.
3. Ning Hu, Roger B. Dannenberg, Ann L. Lewis. A Probabilistic Model of Melodic Similarity. Proceedings of the International Computer Music Conference (ICMC), 2002.

4. Craig Stuart Sapp, Yi-Wen Liu, Eleanor Selfridge-Field. Searh-Effectiveness Measures for Symbolic Music Queries in Very Large Databases. Proceedings of 5th International Conference on Music Information Retrieval, ISMIR. 2004.
5. Rainer Typke, Frans Wiering, Remco C. Veltkamp. A survey of Music Information Retrieval Systems. Proceedings of 5th International Conference on Music Information Retrieval, ISMIR. 2004.
6. Howard Hanson. Music Theory. Harmonic Materials of Modern Music. ISBN: 0891972072 Publisher: Irvington Pub. 1960.
7. Michael Clausen, Frank Kurth, Roland Engelbrecht. Content-based Retrieval in MIDI and Audio. ECDL WS Generalized Documents. 2001.
8. Roland Engelbrecht. Statistical Comparison Measures for Searching in Melody Databases. Tech. Report. University of Bonn. 2002.
9. Eugene Narmour. The Analysis and Cognition of Basic Melodic Structures. Chicago University Press, 1990.
10. Stephen Downie. Evaluating a simple approach to music information retrieval: conceiving melodic n-grams as text. Tesis Doctoral. University of Western Notario. 1999.
11. S. Doraisami & S. Rüger. Robust polyphonic music retrievals with n-grams. Journal of Intelligent Information Systems, 21(1): 53-70.
12. C. Pérez-Sancho, José M. Iñesta, J. Calera-Rubio. Style recognition through statistical event models. Sound and Music Computing (SMC'04) Conference Proceedings, 135-139 2004.
13. W.B. Frakes, D.C. Baeza-Yates, editors. Information Retrieval: Data Structures and Algorithms. Prentice-Hall, 1992.
14. M. Kaszkiel, J. Zobel and R. Sacks-Davis. Efficient Passage Ranking for Document Databases. ACM Transactions on Information Systems, 17(4), pp 406-439. October 1999.
15. Singhal. A., Buckley, M. Mitra. Pivoted document length normalization. Proceedings of the 19th annual international ACM-SIGIR conference on research and development in information retrieval. 1996.
16. Bernal Flores. Music Theory. Systematically Orderly Sonorities. Universidad de Costa Rica, 1979.
17. Rainer Typke, Marc den Hoed, Justin de Nooijer, Frans Wiering, Remco C. Veltkamp. A Ground Truth For Half A Million Musical Incipits. Proceedings of the Fifth Dutch-Belgian Information Retrieval Workshop. Pp.63-70. 2005.
18. Edie Rasmussen. Evaluation in Information Retrieval. The MIR/MDL Evaluation Project White Paper Collection #3. Pp. 45-49. 2003.
19. Rainer Typke. Proposed measure for comparing ranked lists, for MIREX '05 (Symbolic Melodic Similarity). 5th Dutch-Belgian Information Retrieval Workshop. 2005.
20. Ning Hu, Roger B. Dannenberg and George Tzanetakis. Polyphonic Audio Matching and Alignment for Music Retrieval. 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. October 19-22, 2003, New Paltz, NY.
21. Emilios Cambouropoulos, Costas Tsougras. Influence of Musical Similarity on Melodic Segmentation and Algorithms. Proceedings of the International Conference on Sound and Music Computing (SMC04), Paris, France.