A MEASURE OF MELODIC SIMILARITY BASED ON A GRAPH REPRESENTATION OF THE MUSIC STRUCTURE

Nicola Orio
Department of Information Engineering
University of Padova
nicola.orio@dei.unipd.it

Antonio Rodà
Lab. AVIRES
University of Udine
antonio.roda@uniud.it

ABSTRACT

Content-based music retrieval requires to define a similarity measure between music documents. In this paper, we propose a novel similarity measure between melodic content, as represented in symbolic notation, that takes into account musicological aspects on the structural function of the melodic elements. The approach is based on the representation of a collection of music scores with a graph structure, where terminal nodes directly describe the music content, internal nodes represent its incremental generalization, and arcs denote the relationships among them. The similarity between two melodies can be computed by analyzing the graph structure and finding the shortest path between the corresponding nodes inside the graph. Preliminary results in terms of music similarity are presented using a small test collection.

1. INTRODUCTION

One approach to content-based access to music documents is to provide users with tools to retrieve music documents that are *similar* to a set of one or more documents already known, which are the starting point of a query-by-example paradigm. The effectiveness of the results depends on the way a measure of music similarity is computed. This task is difficult to define, because the notion of music similarity is subjective and also because the role played by the different music dimensions – i.e., rhythm, melody, harmony, timbre, orchestration, tempo – is task dependent. For instance, the perceived similarity of two ballroom songs is mainly related to rhythm, while in jazz music it can depend on chord progressions.

In recent years, a major interest has been given to the retrieval of audio documents, typically in compressed formats such as MP3. This trend is explained by the increasing availability of large collections of audio files, and by the fact that users without a music training are usually not interested in accessing symbolic representations. For this reason, the notion of music similarity has been biased towards the music dimensions that, on the one hand, are more relevant for music listeners and, on the other hand, can be reliably extracted from audio files. A typical approach is to extract some timbre descriptors to address different tasks, such as genre and artist identification, automatic playlist generation, or music collection visualization [12].

Music documents can be represented also in symbolic forms, such as a notated digital score or a MIDI file. The access to these documents can be based on higher level features, such as melodic profile and harmonic progressions, which are not easily extracted from audio files. In particular, the melodic profile has been often used as the main dimension to compute the similarity between music documents [1, 7]. A typical task of melody-based retrieval is the automatic identification of a melody sung or hummed by the user. For this application, the similarity measure has to be robust to local mismatches due to imprecise recall from memory and to a lack of singing skills by the users, because it is assumed that the query and the relevant documents are representing the same information. Although the term *query-by-humming* was very popular in the early days of MIR research [3], recently it has been often replaced by the more general term query-by-example because it is assumed that users can easily record with a portable device an excerpt of the song they are interested to retrieve and are not willing to hum a melody in front of an user interface. Approaches of this kind may be based on approximate matching to identify the music work corresponding to the recorded performance, a task that is usually defined as cover [5] or music [8] identification.

The computation of melodic similarity can be useful also for applications other than an identification task. For instance, in musicological analysis the study of the melodic material used by different composers, or consistently used by a given composer, is of particular interest. Also ethnomusicological studies can take advantage from melodic similarity in order to track the evolution of a given song over the centuries and its diffusion in different geographical regions. Melodic similarity can be exploited to retrieve music that "sounds like" other well known songs, for instance to find a suitable soundtrack for a TV show.

We propose a novel similarity measure computed between music content in symbolic format, that takes into account the musical structure of the composition through the application of an analytic method. Our approach aims at representing a music collection with a graph structure where terminal nodes directly describe the music content, internal nodes represent its incremental generalization, and arcs take track of the relationships among them. Results are presented using an excerpt of the dataset used for the melodic similarity task at the Music Information Retrieval Evaluation eXchange (MIREX) campaign of 2005 [17].

2. MEASURING MELODIC SIMILARITY

A common approach to the computation of melodic similarity is to make some assumptions on the perception of pure tones and melodic intervals and to simplify the melodic representation accordingly. This process can be considered as a variant of stemming applied to the music domain, because it aims at conflating into a single stem all the melodic variants that are musically or perceptually similar. A simple example is the representation of the melodic profile using only three classes of pitch intervals – ascending, descending, and same nome – as proposed initially in [3]. Clearly, more complex representations are possible using either a finer quantization of the intervals or the analysis of the harmonic role of melodic intervals. In all these cases, it is assumed that similar melodic excerpts share the same representation. Segmentation can be applied to melodic information [10], where the similarity between melodies can be computed as a weighted sum of the similarities between pairs of segments. This latter approach aims also at efficiency, because retrieval can be based on indexing [2].

Another approach is to exploit the properties of well known distance measures, such as the edit distance using approximate string matching techniques [4] or the earth mover's distance [16], in order to deal with variants in music content. An alternative is to apply statistical modeling, such as Markov chains described in [14], to cope with local variations. The general idea is that the melody used to query the system is transformed in order to be matched with the melodies in the collection, assuming that the cost of the transformation is related to the melodic similarity. One limitation of these approaches is that they make little use of structural information and musicological analysis.

The similarity measure proposed in this paper is based on a different approach. The basic idea is that all melodies – the ones in the music collection and the query – undergo a process of generalization (or simplification). The idea is motivated by the results of musicological studies, such as the Generative Theory of Tonal Music [6], the Implication-Realization Model [9], and Schenkerian analysis [13]. In particular, we aim at finding structural dependencies among the notes of a composition in order to organize them into a coherent hierarchy. This task can be achieved by means of a series of simplifications of the melodic content of a piece, assuming that these simplifications correspond to a generalization of the melodic profile.

The central part of our approach is the determination of which notes in a music passage are more structurally significant than others. We use this property to build a hierarchical representation of a single music document and, incrementally, of a collection of documents. At each step, the melodic content is analyzed and transformed in a melody with fewer notes and with a simpler profile that should represent the most musicologically relevant content information. The procedure ends when the melodic segments are represented by a single note. We called this representation Pseudo-Structural Representation (PSR), because it is inspired by structural analysis, yet the algorithm implementing it exploits some simplifications. The PSR is a graph-based representation, in which the terminal nodes are related to music surface and the internal nodes are progressive generalizations of the surface.

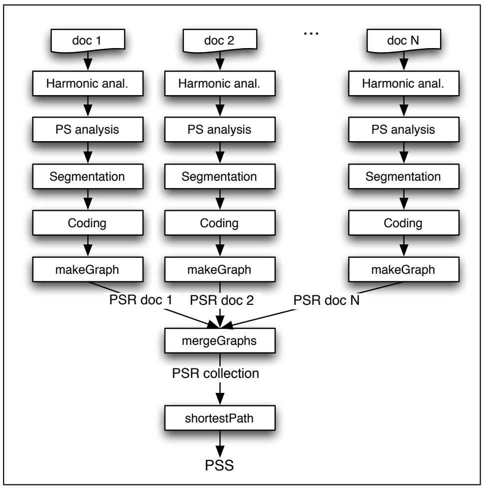

Figure 1. From the music document to the Pseudo-Structural Representation.

The steps to build the PSR of a collection of documents are depicted in Figure 1. First, each music document undergoes harmonic analysis, which highlights the harmonic function – i.e., tonic, dominant, subdominant – of each chord. Although a number of automatic routines is available for the computation of the harmonic function, in our experiments we choose to manually annotate the chords. This task, which is the only manual intervention, is in general not required for polyphonic documents for which reliable systems for inferring the chord progression already exists. Given that the evaluation has been carried out using the MIREX 2005 dataset for the Symbolic Melodic Similarity task, which documents contain only the main melody, we preferred to manually annotate the chords to not introduce possible sources of mismatch while evaluating a novel approach. Automatic chord annotation will be addressed in future work, with the aim to create a totally automatic procedure. At this stage we prefer to not introduce possible sources of preprocessing errors that are not dependent on the approach.

The second step consists in the progressive simplification of the melodic profile, by means of an algorithm inspired by musicological analyzes. First of all, the surface melody is processed to assign three weight coefficients to each note. These coefficients are related to: the underlying harmonic function (harmonic weight), the metric position (metric weight), and the pitch interval between the tone of the melody and the root of the underlying chord

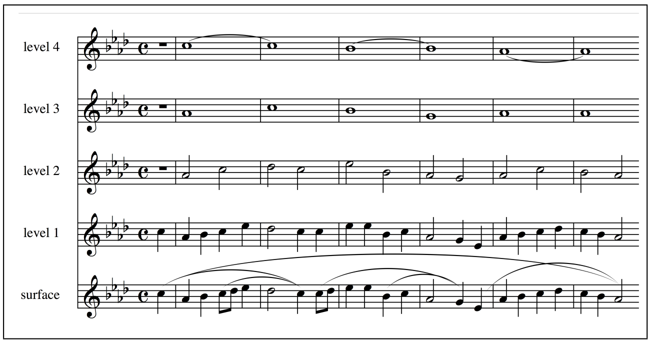

Figure 2. Example of Pseudo-Structural analysis on a Bach's Choral.

(melodic weight). A number of weighting schemes have been devised and evaluated experimentally, as presented in Section 4. The basic idea is to assign higher weights to relevant notes. For instance, a note with tonic function has a higher harmonic weight than a note with a subdominant function, the same applies to notes in strong beats in respect to notes in weak beats for the metric weight; as regards the melodic weight, a perfect fifth with the chord root has higher weight than a perfect four or a major second.

The algorithm works locally on a sliding window, progressively eliminating notes with smallest weights within a window. In the particular implementation, window length has been set equal to the double of the minimum duration of the melody, thus a window contains at most two notes. For instance, the surface melody of Figure 2 has a minimum duration of an eight-note and consequently the window length is a quarter-note. In case the window contains two notes with the same weight, the algorithm applies additional heuristics that take into account (in descending order of relevance): only melodic weight, only metric weight, only harmonic weight, and finally the relative position. The less relevant note is removed and the other one is prolonged to cover the duration of the removed note.

When the sliding window reaches the end of the melody, the first level of abstraction is completely calculated. Then, the algorithm is applied iteratively to calculate the higher levels. The algorithm stops when the highest level has one or two notes. Figure 2 shows an example of Pseudo-Structural analysis applied to the first six bars of the Bach's Choral BWV345, with four levels of progressive generalization. The higher the level the more general representation of the melodic profile.

The analyzed documents are then segmented in musical phrases, a task that can be carried out using one of the different algorithms that have been presented in the literature [11]. With the aim of separately measuring the effect of all the components, we perform segmentation both manually and automatically, using the algorithms provided by the Miditoolbox. All the approaches have been evaluated on the same test collection. Segmentation is carried out at the surface level and inherited by the higher levels. In general segments can overlap, although automatic segmentation algorithms usually provide non overlapping segments.

The subsequent step regards the coding of the pitch sequences, one for each segment. Because duration is used to perform the Pseudo-Structural analysis, it is implicitly modeled at the higher levels of generalization and not directly represented in the PSR. Pitch information is represented in the form of melodic intervals, that is the difference between two subsequent notes. Apart from the segments representing the melodic surface, pitch information undergoes different levels of quantization, from a coarse representation using three symbols (0 unison; ± 1 up/down tone) to the distance in semitones.

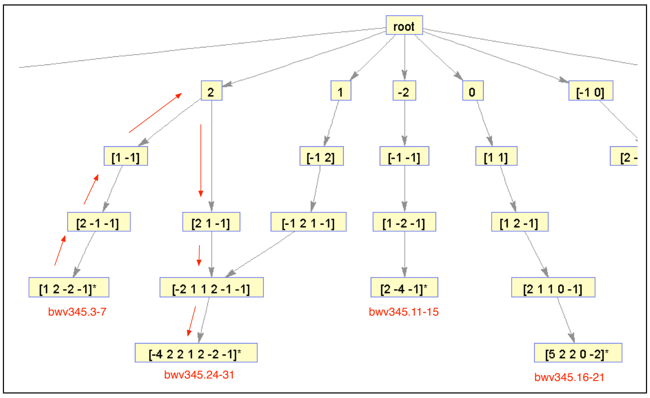

Figure 3. Pseudo-Structural Representation (partial view) of the document analyzed in Figure 2.

Each coded segment is then inserted into the PSR graph structure. In particular, *terminal nodes* contain segments of the melodic surface and *internal nodes* contain segments which are the results of the generalization process. As a result of the Pseudo-Structural analysis, an internal node is connected with a directed arc to all the nodes, either internal or terminal, that correspond to the immediate lower level of generalization. Nodes that hold the same content are joined together in a single one, that inherits the ancestors and the descendants of the starting nodes, obtaining a direct acyclic graph. An example of the PSR of a music document is shown in Figure 3. The PSR of a complete collection can be built by iterating the process for all the segments and the documents in the collection.

Given a PSR, we can define the distance between two melodic segments $s1$ and $s2$, represented by the terminal nodes $n1$ and $n2$, as the length of shortest path from $n1$ and $n2$ considering PSR as an undirect graph where all the arcs have the weight set to 1. It can be noted that PSR could be also a weighted graph, where each arc has a weight that takes into account the kind of applied simplification or the relative frequency nodes appear in a music work and in a collection. This aspect will be investigated in future work. The way the distance is computed can be described through an example. With reference to the music document shown in Figure 3, the distance between the melody segment from beat 3 to beat 7 (coded by [1 2 -2 -1]) and the melody segment from beat 24 to beat 31 (coded by [-1 2 2 1 2 -2 -1]) is equal to 6. This distance is a metric, because it is clearly reflexive and symmetric. Moreover, being based on the shortest path inside an undirect graph, it is easy to show that the triangular inequality holds.

3. APPLICATIONS

The first application of the similarity measure is a musicologically grounded approach to content-based retrieval of music documents using a query by example paradigm. For this application, retrieval can be carried out by performing the same Pseudo-Structural analysis also to the query, which is then (temporarily) added to the PSR graph, in order to compute the similarity. As for other approaches, melodic similarity of the complete documents can be computed as a weighted sum of the melodic similarity of their segments. In particular, the distance between two documents $d(c_i, c_j)$ is defined as the mean of the PSD between all the segments of c_i and c_j. The similarity $s(c_i, q)$ between c_i belonging to a collection of N documents and the query q is calculated through equation

$$s(c_i, q) = \left(1 + \frac{d(c_i, q)}{\sum_{j=1}^{N} \frac{d(c_i, c_j)}{N-1}}\right)^{-1}, \qquad (1)$$

where the similarity is proportional to the reciprocal of the distance $d(c_i, q)$ between the document c_i and the query q divided by a normalizing factor, which is the mean distance between c_i and all the other documents in the collection. The normalizing factors can be computed off-line in order to speed up retrieval.

The graph representation provides a simple way to define the maximum allowable distance between two segments. For instance, the user may choose to limit the length of the paths across the graph or to define the maximum allowed level of generalization, that is the number of times a path can jump to a higher level. Moreover, the user can be presented with a list of documents, their relevant segments, and a representation of the internal nodes that shows which is the path across the PSR that transforms the query into the retrieved segments. It is interesting to note that through this approach, the user can modify its personal view of the PSR, because past queries can be stored in the graph and eventually affect the results of the current retrieval session.

The analysis of the structure of the PSR can provide novel tools for exploring a music collection. For instance, the user can choose a branch of the PSR and explore the melodic excerpt that are represented by internal or by terminal nodes and, in the latter case, to listen to the compositions they belong to. Thus, the user can navigate inside the music segments of a collection, and their generalization based on musicological properties. To this end, informal tests showed that PSR tends to group similar composing styles in close regions of the graph. This ability of the proposed approach will be explored in future work.

4. EXPERIMENTAL EVALUATION

The methodology has been experimentally evaluated using the dataset provided for the Symbolic Melodic Similarity task at MIREX 2005 [17]. The dataset is based on the RISM collection of incipits, where relevance judgments on melodic similarity have been provided by a pool of expert musicologists. We present three different measures of retrieval effectiveness. The common measures *average precision* (AP) and *R-precision* (RP), and the *Average Dynamic Recall* (ADR) which takes into account that relevance judgments are not binary [15] and has been used as the main parameter at MIREX 2005.

# symbols	ADR	AP	R-P
3	0.65	0.60	0.54
5	0.66	0.60	0.52
7	0.65	0.59	0.51
no quantization	0.67	0.64	0.56

Table 1. Results with manual segmentation and using different levels of quantization.

segmentation	ADR	AP	R-P
manual	0.67	0.64	0.56
gestalt	0.69	0.64	0.55
probabilistic	0.67	0.61	0.53
LBDM	0.61	0.53	0.50

Table 2. Results with no quantization and using different approaches to segmentation.

Results on a subset of 110 incipits using 11 queries are reported in Table 1, showing the effect of different levels of quantization when manual segmentation is applied. The reduced size of the collection is due to the fact that, for this initial evaluation, part of the process shown in Figure 1 – annotation of the harmonic function and segmentation – has been carried out manually. As it can be seen, results are similar, with slightly better performances when no quantization is applied, although the differences are not statistically significative. This aspect should be investigated in more detail with a larger collection because a coarse quantization allows us to reduce the size of the PSR improving efficiency, yet a fine quantization preserves more information about the melodic content.

We carried out an experiment using three automatic approaches to segmentation when no quantization was applied. Segmentation algorithms are the ones provided by the Miditoolbox, which are based on gestalt concepts, on probabilistic model, and on the Local Boundaries Detection Model respectively. Results are reported in Table 2, showing that the gestalt-based approach gives results completely comparable with manual segmentation. A test on the statistical significance of the differences between these results showed that none of the differences reached the significance, thus this step can be carried out automatically without affecting retrieval effectiveness.

Other experiments have been carried out to evaluate different weighting schemes for the Pseudo-Structural analysis. As regards the harmonic weight, we tested the effectiveness of grouping the harmonic functions in 3, 4, or 7 classes (denoted with letter H). For instance, for $3H$ we grouped the harmonic functions depending on their degree on the scale, namely I and VI had the highest weight, IV and V had a intermediate weight and other degrees had the

weighting scheme	ADR	AP	R-P
$3H, MH, 3M$	0.67	0.64	0.56
$4H, MH, 3M$	0.65	0.63	0.56
$7H, MH, 3M$	0.65	0.64	0.56
$3H, MH, 4M$	0.67	0.64	0.55
$3H, MH, 7M$	0.61	0.60	0.51
$3H, MS, 3M$	0.66	0.63	0.52

Table 3. Results using different weighting schemes.

smallest weight. The metric weight was varied considering either a simple subdivision (MS) in strong and weak beats or a hierarchical organization (MH) depending on the position in the measure. Finally, the melodic weight has been tested in a similar fashion of the harmonic weight, with 3, 4, and 7 classes (denoted with letter M) where, for example, in $3M$ notes forming an interval of a unison/octave, third or fifth from the fundamental had the highest weight, a seventh had an intermediate weight, and other intervals had the smallest weight. Weighting schemes with more classes, such as $4H$ and $7H$ for harmonic weight, simply introduce new classes either taking into account new harmonic functions or splitting an existing class in two or more smaller classes. Similar considerations apply to the melodic weights.

Results are reported in Table 3, for some combination of weighting schemes, showing that the use of three classes for both the harmonic ($3H$) and the melodic ($3M$) weights with a hierarchical metric (MH) weight gave the best performances, although differences are minimal and not statistically significative. Considering that in all the experiments the three effectiveness measures are considerably higher than the ones obtained at MIREX, we can assume that with a larger collection the performances will be at least comparable with other approaches.

An important characteristic of the PSR is the relationship between the number of documents in the collection and the number of different nodes in its graph representation. It is expected that the size of the graph will increase with sublinear trend when documents are added to the PSR. Figure 4 shows this trend, where the number of new nodes – both internal and terminal – that are added with each new document decreases with the number of documents, although local variations can still be seen due to the reduced size of the test collection.

Figure 5 shows the results of a nearest neighbor query on the collection, with the aim of highlighting which features are captured by the proposed similarity measure and whether this definition can take into account for progressive differences among melodies. The searched pattern is the surface melody of the first segment of $000.109.406 - 1.1.1$ in the RISM collection. Pseudo-Structural analysis of this composition highlights that the notes at positions 2, 4, and 6 (notes B, B, and G respectively) are less relevant than the others. Indeed, all these notes are passing notes on a weak metric position. At an higher level of generalization in the PSR, these notes are therefore omitted, and the

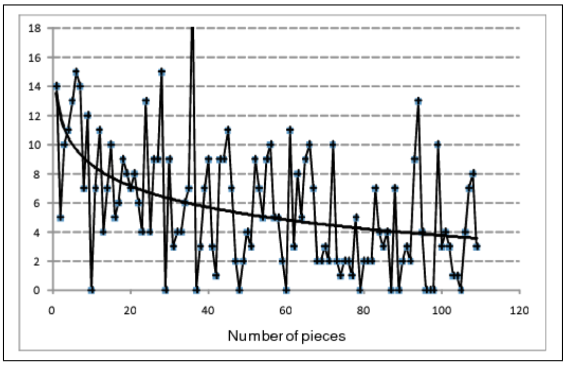

Figure 4. Number of new nodes added to the PSR graph when new music documents are added. The bold line is a logaritmic aproximation in a least square error sense.

segment is represented by an ascending melodic interval (A-C), followed by two descending melodic intervals (C-A and A-F), which correspond to the code [+2 -2 -2] using the coarsest quantization. At a distance $d = 1$, the result reports a segment with identical pitch, but with augmented duration values. It can be noted that the ratio among durations are unchanged in respect to the query segment. At a distance $d = 2$, there are segments with similar but not equal pitches, and with minor metric variations. Finally, at distance $d = 3$ and $d = 4$ there are melodic segments composed by notes that at least share with the query segment the same higher level code [+2 -2 -2].

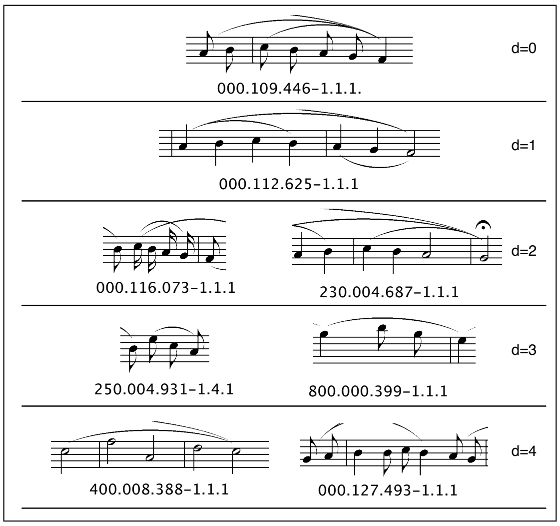

Figure 5. Results of a nearest neighbor query.

5. CONCLUSIONS

The proposed approach aims at representing the structural relationships of a music collection with an undirected graph, which is built from the analysis of the melodic content of music documents. Terminal nodes represent melodic segments of the documents, while internal nodes represent a progressive simplification/generalization of their content. Music similarity is then measured by the length of the shortest path between terminal nodes. This representation allows us to retrieve music documents that are relevant at least from a musicological point of view. Moreover, the proposed similarity is a metric because it is based on a topological distance, allowing us to efficiently carry out a number of retrieval tasks, such as range queries, k-nearest neighbor, and document clustering.

The approach has been tested with a collection of incipits. Moreover, qualitative analysis have been carried out on the relationships between the graph structure and the melodic content of the documents. Results are encouraging, both in terms of average precision of the retrieval results and in terms of musicological significance. The approach can be further exploited for browsing a collection of music documents based on the traversal of the graph representing the music documents and their relationship.

The described similarity measure is tailored to music genres where harmony plays a functional role, like in Western tonal music, because the weighting schemes presented in Section 2 are mostly based on the harmonic content. The idea itself of generalizing the melodic content through structural analysis is motivated by musicological studies on Western music. Although we believe that it is difficult, if not impossible, to develop a general purpose approach to music similarity searches, it is likely that the idea of representing music content with a hierarchical graph, where levels are associated to an incremental simplification of the musical content, can be generalized to other music features and to other genres.

A major limitation is that, at the moment, the methodology is still partially based on manual annotation of the chord progressions of the musical documents. Given the encouraging results, future work will focus on the complete automatization of the analyzes.

6. ACKNOWLEDGMENTS

The work has been partially supported by a grant of the University of Padova for the project “Analysis, design, and development of novel methodologies for the study and the dissemination of music works”.

7. REFERENCES

[1] J.S. Downie. Music information retrieval. *Annual Review of Information Science and Technology*, 37:295–340, 2003.

[2] J.S. Downie and M. Nelson. Evaluation of a simple and effective music information retrieval method. In *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 73–80, 2000.

[3] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith. Query by humming: Musical information retrieval in an audio database. In *Proceedings of the ACM Conference on Digital Libraries*, pages 231–236, 1995.

[4] H.H. Hoos, K. Renz, and M. Görg. GUIDO/MIR – an experimental musical information retrieval system based on GUIDO music notation. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 41–50, 2001.

[5] P. Herrera J. Serrá, E. Gómez and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(6):1138–1151, 2008.

[6] F. Lerdhal and R. Jackendoff. *A Generative Theory of Tonal Music*. The MIT Press, Cambridge, MA, 1983.

[7] M. Melucci and N. Orio. Combining melody processing and information retrieval techniques: Methodology, evaluation, and system implementation. *Journal of the American Society for Information Science and Technology*, 55(12):1058–1066, 2004.

[8] R. Miotto and N. Orio. A music identification system based on chroma indexing and statistical modeling. In *Proceedings of the International Conference on Music Information Retrieval*, pages 301–306, 2008.

[9] E. Narmour. *The Analysis and Cognition of Basic Melodic Structures*. University of Chicago Press, Chicago, MI, 1990.

[10] G. Neve and N. Orio. A comparison of melodic segmentation techniques for music information retrieval. In *Proceedings of the European Conference on Digital Libraries*, pages 49–56, 2005.

[11] B.S. Ong. Structural analysis and segmentation of music signals. Master’s thesis, Universitat Pompeu Fabra, 2006.

[12] N. Orio. Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, 1(1):1–90, 2006.

[13] H. Schenker. *Der Freie Satz, Neue musikalische Theorien und Phantasien*. Universal Wien, O. Jonas, 1956 edition, 1935.

[14] J. Shifrin, B. Pardo, C. Meek, and W. Birmingham. HMM-based musical query retrieval. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, pages 295–300, 2002.

[15] R. Typke, R.C. Veltkamp, and F. Wiering. Evaluating retrieval techniques based on partially ordered ground truth lists. In *Proceedings of the International Conference of Multimedia and Expo*, 2006.

[16] R. Typke, F. Wiering, and R.C. Veltkamp. A search method for notated polyphonic music with pitch and tempo fluctuations. In *Proceedings of the International Conference of Music Information Retrieval*, pages 281–288, 2004.

[17] Mirex 2005 Wiki. First annual Music Information Retrieval Evaluation eXchange, July 2006. http://www.music-ir.org/mirex2005/.