

Musical Information Retrieval using Melodic Surface

Massimo Melucci and Nicola Orio

Padua University

Department of Electronics and Computing Science

Via Gradenigo, 6/a - 35131 - Padova - Italy

{melo,orio}@dei.unipd.it

Abstract

The automatic best-match and content-based retrieval of musical documents against musical queries is addressed in this paper. By "musical documents" we mean scores or performances, while musical queries are supposed to be inserted by final users using a musical interface (GUI or MIDI keyboard). Musical documents lack of separators necessary to detect "lexical units" like text words. Moreover there are many variants of a musical phrase between different works. The paper presents a technique to automatically detect musical phrases to be used as content descriptors, and conflate musical phrase variants by extracting a common stem. An experimental study reports on the results of indexing and retrieval tests using the vector-space model. The technique can complement catalogue-based access whenever the user is unable to use fixed values, or he would find performances or scores being "similar" in content to known ones.

KEYWORDS: Information Retrieval, Computer Music, Musical Digital Libraries, Automatic Indexing, Automatic Melodic Segmentation.

1 INTRODUCTION

The research projects in digital libraries, and specifically those carried out in cultural heritage domain have shown that the integrated management of diverse media - text, audio, image, video - is necessary [7, 12, 14]. As stressed in [18], the problem with content-based access to multimedia data is twofold. On the one hand, each media requires specific techniques that cannot be directly employed for other media. On the other hand, these specific techniques should be integrated whenever different media are present in a indi-

vidual item. The core information retrieval (IR) techniques based on statistics and probability theory may be more generally employed outside the textual case and within specific non-textual application domains. This is because the underlying models, such as the vector-space and the probabilistic models, are likely to describe fundamental characteristics being shared by different media, languages and application domains [18].

As music is one of the most important means of expression of cultural heritage, the organisation, the integration with other media, and the access to the digitized version of music documents becomes an important multimedia digital library component. We can thus speak about *content-based musical* IR. Specific and effective techniques being capable of indexing and retrieving such multimedia documents as the musical ones need to be designed and implemented. The requirement for a musical content-based IR has been stressed within the research area of musical information systems as well. The developments in the representation of music "suggest a need for an information retrieval philosophy directed toward non-text searching and eventual expansion to a system that encompasses the full range of information found in multimedia documents", as stressed by McLane in [13]. As IR has dealt with the representation and the disclosure of content from its early days [5, 11, 17, 21], it is natural to think that IR techniques should be investigated to evaluate their application in music retrieval. By concluding his survey, McLane stressed that "what has been left out of this discussion, and will no doubt be a topic for future study, is the potential for applying some of the standard principles of text information retrieval to music representations" [13].

The use of standard principles of text information retrieval to index and retrieve musical documents requires the design of segmentation algorithms to produce musical phrases like words in textual documents. Like textual words, musical phrases occur in documents or queries with many variants. For example, the same melodic pattern may occur in more than one musical work, perhaps composed by different authors. It is therefore necessary to detect these variants and conflate all the different phrases into a common stem.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DL 99 Berkeley CA USA

Copyright ACM 1999 1-58113-145-3/99/08...\$5.00

2 ISSUES OF CONTENT-BASED INDEXING AND RETRIEVAL OF MUSICAL DATA

The musical data, in their different representation forms, can be considered as another media together with text, image, video, and speech. There are some issues that make music different from other IR application domains.

The same entity, i.e. a musical work, can be represented in two different main *forms*: the notated and the acoustic form, respectively corresponding to score and performance. Hence the communication in music is performed at two levels: (i) the composer translates his emotions in a musical structure (music as a composing art), and (ii) the musician translates the written score into sounds (music as a performing art). Also users may have different needs, in particular the music scholar may look for a given composition, while the melomane may look for a particular performance.

Each musical work may have different *instantiations*. As musicians can interpret scores, the resulting performances may differ and therefore more than one performance correspond to an individual score. Moreover, the same musical work may be transcribed into different scores, depending on the revisers's choices. As a consequence, the same musical work may be represented by different scores and performances.

Different *dimensions* characterise the information conveyed by a musical work. Melody, harmony, rhythm, and structure are dimensions, carried by the written score, that may be all or in part of interest for the final user. In the case of a musical performance other dimensions (like timbre, articulation, and timing) should be added. It is likely that the dimensions of interest vary with the level of user's expertise and the specific user's search task.

Many are the *formats* that may be used whenever scores or performances are stored into a digital music collection. It is important to stress that a given format can be able to capture only a reduced number of dimensions. In particular the storage of performances can be done by recording the sound, but at the state of the art is almost impossible to retrieve information about, for instance, melody and harmony; tracing the musical events, like the beginning and the end of each note, can usually be done using the MIDI format, but in this case there is no information about timbre and note envelopes. Therefore, the choice of a representation format has a direct impact on the degree to which a music retrieval system can describe each dimension.

While text, image, video, or speech-based documents convey some information that form their content, it is still unclear what type of *content*, if any, musical works do convey. In principle, music language does not convey information as, for instance, text does. Many composers wrote music to stir up emotions, and in general they aimed to communicate no specific information to the listener. The final user feels emotions on listening to the music, and he interprets some information independently from the composer's and performer's thought and differently from the other users. There is a particular kind of musical works, called *musica a programma*, in which the title (like Vivaldi's "The Tempest") or a lyric (like Debussy's "Prélude à l'après-midi d'un

faune") suggests a meaning to the listener; this sort of textual data would be better managed using a database system rather than a IR system. Moreover in sung music, such as Cantatas, the accompanied text gives the work some meaning, yet that sort of text would require ad-hoc IR techniques to be effectively managed. In general the availability of textual material together with musical documents is insufficient.

As always happens in IR, the effectiveness of techniques does strongly depend on the final *user*. IR systems does indeed interact with final users of very diverse types and with different levels of expertise in the use of the system itself and in the application domain. For instance, the layman may find querying by musical content, such the melodic incipit, more effective than by bibliographic values, since he may find playing some notes easier than knowing the necessary data to retrieve the searched works.

3 A TECHNIQUE FOR AUTOMATIC INDEXING OF MUSICAL DATA

Given the peculiar characteristics of the musical media, highlighted in Section 2, the first task in the development of a system for content-based IR of musical data is the identification of the *dimensions* that have to be used to index and retrieve. Dimensions are related to the form, i.e. notated or acoustic, chosen to store the musical data into the digital library. The acoustic form presents some problems for automatic indexing purposes. In fact, the automatic extraction of melody and harmony from polyphonic performances is not well developed yet. Moreover, only a few parameters of the acoustic signal, such as brightness and roughness were found perceptually significant for the listener, however they can be used to describe a single sound, not a complete performance where the musician continuously changes his performing parameters.

For these reasons we have chosen to use the notated form, which carries information on melody, harmony, rhythm and structure. Among these different dimensions, the melody is the one which is more easily recognizable for a final user who does not have a specific expertise in music. Hence we have chosen to use melody to index musical data. In order to apply the algorithms developed for textual IR to this dimension, we need to develop a segmentation of the melody in musically relevant phrases. The problem of musical segmentation is well known in musicological literature, and some models were proposed to automatically perform a segmentation of musical works in short musical phrases called *melodic surfaces*.

3.1 Detection of Melodic Surfaces

The algorithm used to perform the segmentation algorithm is based on a model due to Cambouropoulos [2], who proposed the Local Boundaries Detection Model (LBDM). The basic idea of the model is that the listener perceives the presence of boundaries in a melodic surface whenever there are some changes in the relationships among the notes. In particular, these changes may regard the musical intervals

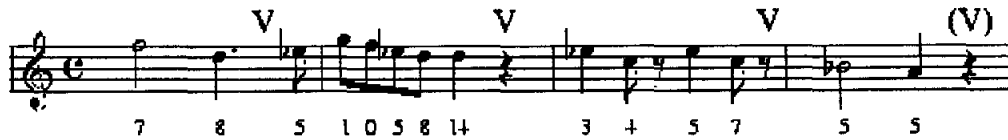


Figure 1: Mozart’s Concert for Clarinet K622, for each note the respective weights are quoted

and the note duration. That is, the listeners perceive the presence of a melodic surface whenever there are changes in the melodic structure. The LBDM detects the boundaries of a melodic surface by giving a weight to all the possible places in which the boundaries may occur, that is between two subsequent notes of the melody. The weight values are evaluated depending on:

- the relationship among the *musical intervals* that each note forms with the previous and the subsequent notes. If intervals are different (e.g. a fifth versus a third), then a given value is added to the two weights (2 for the larger interval and 1 for the smaller one);
- the relationship among *note durations*. If durations are different a given value is added to both intervals (respectively, 4 and 1 if the first note is longer than the second one, or 3 and 2 if the first note is shorter);
- the presence of *musical rests*. In this case a value of 4 is added to the weight.

The boundaries of melodic surfaces can be detected by analyzing the weights trend: Cambouropoulos proposed that the boundaries are associated to the presence of maxima in the weight function.

Using this model, it was developed an algorithm that evaluates all the weights after reading the score from a MIDI file, and extracting all the melodies. A segmentation is then performed. We have chosen to use two kinds of musical segments, which we called *phrases* and *periods*. Phrases are the melodic surfaces given by the LBDM model, while periods are groups of subsequent phrases: the beginning and the end of each period is related to the presence of a maximum with a higher weight than the surrounding maxima (hence a maximum among maxima). The introduction of periods is due to a twofold requirement: (i) we were interested in testing if the length of the melodic surface had a significant impact on the retrieval effectiveness; (ii) we intended to use the periods together with the phrases to index documents and queries. Since a period is formed with phrases, it does not introduce any new information, but it takes into account the temporal relationship among subsequent phrases. In this way it is expected that, between two documents with the same phrases matching the query phrase, the one with the same temporal occurrence of phrases will have a higher score than the other one.

3.2 Data Normalisation

Once phrases and periods are translated in a textual notation, four normalizations regarding both pitch and duration of notes are applied.

- *Pitch Transposition (PT)*: the first note of each phrase or period is forced to be a C4, and the pitch of the other notes is calculated in order to maintain the same musical intervals. This normalisation is proposed to consider the relationship among notes rather than their exact pitch. Notes length are expressed in multiples of a given duration, common for all the documents.
- *Pitch Transposition and Duration Normalisation (PTDN)*: pitches are expressed as in PT, while note durations are expressed in multiples of their Greatest Common Divisor. In this way phrases are independent on the particular Tempo and what is important is just the relationship among durations.
- *Pitch Normalisation and Duration Normalisation (PNDN)*: pitches are quantized in a number of different levels, related to the musical intervals. In particular a level corresponds to the unison, a second level to a small interval (from minor second to major third), a third one to an average interval (from perfect fourth to a major sixth) and a fourth one to a big interval (above a minor seven). In this way the trend of the melodic surfaces becomes significant, not the exact sequence of intervals. The durations are the same of PTDN.
- *Pitch Normalisation and Duration Removal (PNDR)*: pitches are the same as in PNDN, while no information about duration is maintained. This kind of massive normalisation was adopted to test the important of timing versus interval sequence.

3.3 An Example of Segmentation

In order to illustrate the performance of the segmentation algorithm, we can take as an example the segmentation of the first four measures of Mozart’s Concert for Clarinet K622, reported in Figure 1. At first a given weight is assigned to each interval between two subsequent notes, depending on the rules explained in Section 3.1; the weight of each interval is quoted under the previous note. Then the local maxima are calculated to detect the boundaries of the melodic surfaces: in Figure 1 the boundaries are shown by a “V”. In this example four phrases are detected, considering that the

last phrase is forced by the length of the query. Moreover it can be seen that two periods are detected, both made with two phrases.

Then the algorithm can produce four different text files, depending on the kind of normalization to be applied. In Table 1 are quoted the different outputs (only the phrases are reported): as it can be seen the musical information on melodic surfaces is now transformed in textual information, which can then be treated with classical textual information retrieval algorithms.

Normalization	Phrases
PT	C524A418 C56E56D56C56B46B424 C512A412C512A412 C524B412
PTDN	C54A43 C51E51D51C51B41B44 C51A41C51A41 C52B41
PNDN	O4N3 O1P1N1N1N1O4 O1N1P1N1 O2N1
PNDR	ON OPNNNO ONPN ON

Table 1: The four different output formats of the segmentation algorithm. The following convention is applied; in PT and [PTDN]: note + octave + [normalized] duration; in PNDN and [PNDR]: interval (O=unison, N=descending small, P=ascending small) + [normalized duration].

4 EXPERIMENTAL STUDY

The experimental study we carried out aimed to describe the impact of two factors – the normalization technique and the use of musical periods together with musical phrases, illustrated in Section 3 – on the results of musical document retrieval. We expect that the degree of normalization, and the use of either musical phrases or periods at indexing-time are means to tune the degree of exhaustivity and specificity of musical document retrieval. If the hypothesis is true, we can therefore use those means to give the user the capability at query formulation time of deciding the quantity of extraneous musical material that is likely to be retrieved, and the quantity of pertinent musical material that is likely to be missed.

The experiments have been conducted in a laboratory setting using a collection of musical documents and queries as testbed. As we lack of test collections being usually available in textual IR, we had to construct an ad-hoc set of experimental documents and queries. The construction of a set of test relevance judgements or the employment of a large number of human judges were too expensive tasks and out of the scope of this research. We consider these laboratory-based experiments as the first step towards an experimental

environment where many “real” final users shall be involved. The experiments are supposed to give us some useful and necessary insights in order to design and implement a prototype musical IR system which can be employed within user-based experimentations.

The set of test documents consists of 419 musical works, each work is stored into a MIDI file. Many works include textual descriptions about, for example, time, tonality, or title. While the full musical data has been used to index documents, the textual data have not been used as we are interested on studying musical content-based retrieval on the basis of musical data only. Documents are pieces of Tonal Western Music (TWM), and specifically complete musical works of Baroque, Classic and Romantic periods, such as movements of Concertos or Sonatas of various lengths – from 3 up to 35 minutes – and of six different composers¹.

The choice of TWM is due to a number of reasons: (i) bibliographic catalogs usually concern this music genre, (ii) the system developed for segmentation is based on melodic surface and in this repertoire melodies are more complex and structured than in other music genres (e.g. jazz, pop, rock, or ethnic music), hence the TWM is a good testbed for the system, (iii) most of the studies on music structures are based on the TWM language, (iv) in other music genres, especially jazz and pop music, the final user is usually interested about the performer rather than the composer, and so bibliographic data normally suffices.

We created $8 = 4 \times 2$ different versions of the same document set corresponding to the 4 normalization criteria and to 2 types of token used to index the documents, i.e. the musical phrase and the musical period. Table 2 summarizes the main numerical parameters about the document set. For each normalization criterion and type of token, the average number of unique tokens per document, and the average number of documents being indexed by an individual token are reported.

	Type of Token	
	Only Phrase	Phrase & Period
Av. document length	713.7	969.6
Standard deviation	822.5	1106.5
<i>PT</i>		
Av. tokens/document	366.0	518.0
Av. documents/token	1.40	1.26
<i>PTDN</i>		
Av. tokens/document	344.6	496.5
Av. documents/token	1.53	1.33
<i>PNDN</i>		
Av. tokens/document	293.6	440.8
Av. documents/token	1.68	1.39
<i>PNDR</i>		
Av. tokens/document	134.1	264.0
Av. documents/token	4.47	2.05

Table 2: Main numerical parameters about documents.

On setting queries up, we kept two objectives in mind.

¹G.F. Händel, J.S. Bach, W.A. Mozart, L. van Beethoven, F. Liszt, P.I. Tchaikovskij.

The first objective has been to describe the statistical behaviour of retrieval results, and therefore we needed a quite large number of queries. The second objective has been to understand the behaviour of retrieval results from a musical point of view. We did therefore perform a quantitative analysis to attain the former, and perform a qualitative analysis to attain the latter. The quantitative analysis has been performed by computing numerical measures, while the qualitative analysis has been performed by employing a musician as expert final user. The aim of the quantitative analysis has been to study the relationships between the quantity of retrieved musical documents as some independent factors change. The aim of the qualitative analysis was to study the variations of genre and composer of the retrieved musical documents as some independent factors change.

According to the two different types of query, we set up two types of query sets – set A and set B – used to perform the quantitative and the qualitative analysis respectively. Set A consists of different versions of a set of 419 document *incipites* which have been extracted automatically from the document set. One version of set A has been created for each of six different query lengths corresponding to six numbers of starting tokens – i.e. 419 queries have automatically been created using x starting phrases, provided $x = 1, 2, 3, 4, 5, 10$. One version has also been created for each of the four normalization criteria, and for each of the two types of token used at indexing-time. We were then provided with $(2 \times 4 \times 6)$ different versions of set A. Set B includes 15 short musical pieces of around eight phrases being manually played by the musician. These queries have been chosen in order to test the behaviour of the indexing and retrieval techniques as the musical style changes. As consequence, the following types of query have been chosen: Four queries were about composers – Vivaldi and Grieg – being absent from the document set, and six queries were about composers being already stored into the document set. Two queries out of the latter are about works being already stored into the document set. The other four queries were about different works. Five of the total fifteen queries of set B are similar to other five queries, but “biased” through the manual insertion of errors. Errors are a sample of common performance mistakes. The queries of set B have then been normalized to produce one version of the query set for each normalization criterion and for each type of token used at indexing-time. Specifically, either phrase- or period-based versions of query set B have been considered in order to study the impact of the use of period on retrieval results. We are then provided with (2×4) different versions of set B.

Table 3 reports the numerical parameters about queries of set A being segmented, normalized using only phrases. In Table 3, the average number of unique phrases per query and the average number of queries being indexed by phrase are reported for each normalization level, and for each *incipit* query length.

The IR model employed to construct indexes and retrieve documents was the vector-space model (VSM). Accordingly to the VSM model, both documents and queries are represented as k -variate vectors of token weights w_{ij} , where

Only Phrases	Incipit Query Length					
	1	2	3	4	5	10
<i>PT</i>						
Av. phrases/query	1.0	2.0	3.0	3.9	4.8	9.2
Av. queries/phrase	1.1	1.1	1.2	1.2	1.2	1.2
<i>PTDN</i>						
Av. phrases/query	1.0	2.0	3.0	3.9	4.8	9.1
Av. queries/phrase	1.1	1.2	1.3	1.3	1.3	1.4
<i>PNDN</i>						
Av. phrases/query	1.0	2.0	2.9	3.8	4.6	8.6
Av. queries/phrase	1.2	1.4	1.5	1.5	1.5	1.6
<i>PNDR</i>						
Av. phrases/query	1.0	1.9	2.8	3.6	4.4	7.8
Av. queries/phrase	2.1	2.6	2.9	3.1	3.3	3.5

Table 3: Main numerical parameters about queries of set A.

a token can be either phrase or period, and k is the total number of unique tokens. The weight w_{ij} of token j within document i can be expressed as $tf_{ij} \times idf_j$, where tf_{ij} is the frequency of occurrence of token j within document i , $idf_j = \log N/n_j$, N is the total number of documents, and n_j is the number of documents indexed by token j . The retrieval status value (RSV) is the usual cosine of the angle between the query vector and the document vector. As the cosine function normalizes the RSV to the query and document lengths, the sizes of document and query have been controlled. The $tf_{ij} \times idf_j$ weighting scheme gives higher RSVs to documents within which query tokens occur with high intra-document frequency (tf_{ij}) and low intra-collection frequency (n_j).

In the following, the analysis of test results has been split into quantitative and qualitative analysis. For each type of analysis, some measures and discussions are presented.

4.1 Quantitative analysis

Table 4 summarizes test results for the quantitative analysis performed using queries of set A against the whole document set. As queries of set A have been set up using both phrases and periods, results refer to documents retrieved and matched using either phrase or period. The first part of the Table refers to phrase-based indexing and retrieval, while the second one refers to period-based indexing and retrieval.

As regards to the difference between the use of musical phrases and the use of musical periods, it is apparent that the absolute quantity of retrieved documents using phrases is higher than the quantity using periods. Therefore, periods are more selective than phrases, as we expected on designing musical period-based indexing. As consequence, musical phrases are able to detect more pieces of different composers than musical periods.

The reported values are directly correlated to the *incipit* query length, i.e. the longer the *incipit* query, the higher the number of retrieved documents and the number of different composers. This can be explained by the higher chance that a long query token occurs within a document than a short query token. This may also be explained by the fact that

Only Phrases	Incipit Query Length					
	1	2	3	4	5	10
<i>PT</i>	1.2	2.0	2.6	2.9	3.1	3.5
	8.0	13.2	17.9	20.7	22.3	26.1
<i>PTDN</i>	1.4	2.0	2.7	3.1	3.3	3.8
	9.1	13.4	18.1	21.3	23.3	27.4
<i>PNDN</i>	1.8	2.5	2.9	3.3	3.5	3.9
	11.4	16.0	19.5	21.9	23.7	27.6
<i>PNDR</i>	3.0	3.3	3.6	3.7	3.8	3.9
	21.0	23.9	25.6	26.8	27.5	29.2

Only Periods	Incipit Query Length					
	1	2	3	4	5	10
<i>PT</i>	0.0	0.2	0.3	0.4	0.5	1.0
	1.1	2.0	2.4	3.0	3.7	6.7
<i>PTDN</i>	0.0	0.4	0.8	0.8	0.9	1.7
	1.0	3.1	4.2	5.0	6.1	10.4
<i>PNDN</i>	0.1	0.6	1.0	1.3	1.5	2.5
	1.2	4.1	6.2	7.9	9.2	15.3
<i>PNDR</i>	0.5	1.3	2.0	2.4	2.7	3.6
	3.6	8.7	13.1	16.2	17.8	23.9

Table 4: Retrieval results using either phrase or period at indexing and retrieval time. For each normalization level, and for each *incipit* query length, two values are reported – the average number of retrieved document of which composer is different from the query author (first row) and average number of retrieved documents (second row).

the number of documents indexed by an individual token is rather low, and that the individual tokens index different sets of documents.

A direct correlation can also be observed between the reported values and the level of normalization. Normalized tokens allows for the retrieval of higher quantity of documents, perhaps of different composers. Specifically, *PNDR* gives the highest increase of number of retrieved documents and of different composers. Actually, normalization have been designed to conflate similar phrases and periods together in order to have larger sets of documents being indexed by each token. The combination of two facts – the low decrease of the average number of unique tokens per query as normalization proceeds (Table 3), and the increase of the average number of documents being indexed by an individual token – can explain the order of magnitude of the reported values.

In particular, it is interesting to note that documents of around three different composers, i.e. around one tenth of the total retrieved quantity, can be retrieved on average. It is also worth of mentioning that the two tested devices – the level of normalization, and the use of either phrase or period – allow for the retrieval of documents at different degrees of specificity and exhaustivity at every *incipit* query length.

4.2 Qualitative analysis

The first interesting result was that, independently from the kind of normalisation, queries using pieces that were present in the collection always gave that same piece with the highest RSV. More generally, when the author of the melody was included in the collection, his pieces were retrieved with an high RSV. This is particularly true when *PT* was applied: queries made with Bach's pieces retrieved a large number of Bach's works with the highest RSV; the same applied

to Mozart's queries, even if in this case also other authors were retrieved. When *PTDN* was applied this trend was maintained, even though also pieces written by composers who had a small amount of musical works in the collection (e.g. Tchaikovskij) were retrieved with an high RSV. When *PNDN* is applied, the distribution of the RSV among composers is more homogeneous; Bach and Mozart's works were usually the most retrieved. The *PNDR* stressed this behaviour: the RSV is homogeneously distributed among the different composers.

The same musician who compiled the queries of set B was asked to judge the results of the queries from a musicological point of view. In particular we were interested to test if the retrieved works had some analogies with the melodies used to create the queries, that is if a final user would find them close enough to the works he was searching. The analysis was developed only for the pieces with a high RSV. Results are related to the kind of normalization applied to the segmented melodies, with some general trends. Most of the retrieved melodies shared the same *temper* of the queries, even if this *temper* was related to the musical style of the different composers. For queries using Bach's, Mozart's and Grieg's melodies, in most of the case it was possible to recognize a similar structure of the melody, that is a similar organization of the melodic surfaces. This was not the case of Vivaldi's, even if it was possible to recognize the melodic segments shared by the query and the retrieved works. When *PT* was applied, the retrieved works had a Tempo similar to the queries: that is, fast melodies retrieved works with a fast Tempo. It is interesting to note that this behaviour was maintained also when durations were normalized.

In general the qualitative analysis gave satisfying results. Retrieved works shared some musical properties with the test queries, for example a query made with a Bach's "Invention" retrieved other inventions and Grieg's "Morning"

retrieved works with plain melodies built on major chords. Moreover a musicologist may find interesting to analyze works where the same melodic expert (i.e. the melodic surface related to our concept of phrase) appears in different musical context, while a common user may find melodies that simply *sound like* the one used for the query.

5 A MUSICAL INFORMATION RETRIEVAL SYSTEM

Figure 2 depicts the skeleton of the architecture of a possible musical IR system implementing the technique based on melodic surface and presented in previous sections. The architecture may take into account other components, such as Z39.50 interface to provide wider accessibility and interoperability. The system components are:

- a Web browser-based user interface – the browser would be Java-enabled to permit the processing of MIDI objects;
- the segmentation routine transforming MIDI files into segmented textual files to be indexed by the indexing component;
- the indexing component generates inverted files to be used as indexes at retrieval-time;
- the search engines retrieves musical documents matching the user's query;
- the digital object server, such as database server manages digitized manuscripts, scores and performances at different levels of quality.

The flow of interaction can be described in the following steps being numbered consistently to the Figure:

1. This step is not really part of the interaction flow, but it is included since it is what converts MIDI files into a form that can be indexed.
2. The user expresses an information need using a musical interface based on either a graphical Web form or a MIDI keyboard. In both cases, the query is a MIDI file being converted and segmented into a form that can be indexed.
3. The search engine takes as input the indexed query and the document indexes, and produces a list of anchors to the digitized versions of the retrieved documents. The list of retrieved documents is sorted by RSV.
4. The user selects one of the retrieved documents and asks the system to retrieve one of the digitized objects. Different types of object may exist: Manuscripts, high-quality performances, scored notations, and MIDI files. The choice of the object type may depend, for instance, on network bandwidth or user privileges.
5. The user access the digitized object database and delivers to the user the selected one.

6 APPROACHES TO MUSIC INDEXING AND RETRIEVAL

Some approaches to access to databases about music are based on textual bibliographic data records. The final user can query the databases by specifying exact values for pre-defined fields, such as composer's name, title, date of publication, type of work, etc. Examples of projects addressing the access to bibliographic databases as component of larger information systems about musical material are Cantate [3], Harmonica [8], Jukebox [9], Musica [15], and RISM [16]. Some of these, such as Cantate, implement links to multimedia objects, e.g. digitized manuscripts or performances, or provide with tools to manage multilinguality, such as Musica.

From an IR point of view, these approaches are quite effective whenever the user (i) can exhaustively use the available search fields, and (ii) is able to precisely use the available search fields. Bibliographic values are not always able to describe exhaustively and precisely the content of musical works. For example, the term "Sonata" as value of the type of work cannot sufficiently discriminate all the existing sonatas. Moreover, many known work titles, such as the Tchaikovskij's "Pathetic", are insufficient to express a final user's query whenever he would find the title not being a good description of the musical work. The use of cataloguing number, like K525 for Mozart's "Eine Kleine Nachtmusic", will be effective only if the user has a complete information on the music work, and in this case a database system will suffice. Searching by composer's name can be very effective. However, some less known composers and their works may not be retrieved if only because the authors are little known. On the other hand, for a prolific composer, just like Mozart, a simple query by composer's name will retrieve an extremely high number of documents, unbearable for the final user.

Some automatic techniques have been proposed to index and search music databases using alternative methods. Most of these techniques are based on the "string matching" concept, i.e. musical notation can be seen as a string and document-query matching can be designed as more or less sophisticated string matching algorithms [6, 1, 10]. String matching-based search techniques are very efficient in textual IR, but poorly effective because text is very complex to be represented and searched as simple strings or substrings. Searching by means of strings may make sense for a given class of users and for a specific class of musical works. If we have no evaluation frameworks to compare different musical IR techniques, no conclusions can be drawn. For example, some evaluation proposals have been done in [20, 19].

Recently, the interest in content-based music retrieval is growing. Contributions are given from different perspectives. Dunn and Mayer provides a detailed description of a real experience on designing and implementing the VARIATIONS digital library [4]. The evaluation being conducted with real users is a useful source of information for the future work. Tseng [19] addresses the problem we faced as well, i.e. the problem of mismatch between stored musical melodies and musical user's queries. Bainbridge *et al.* illustrate the

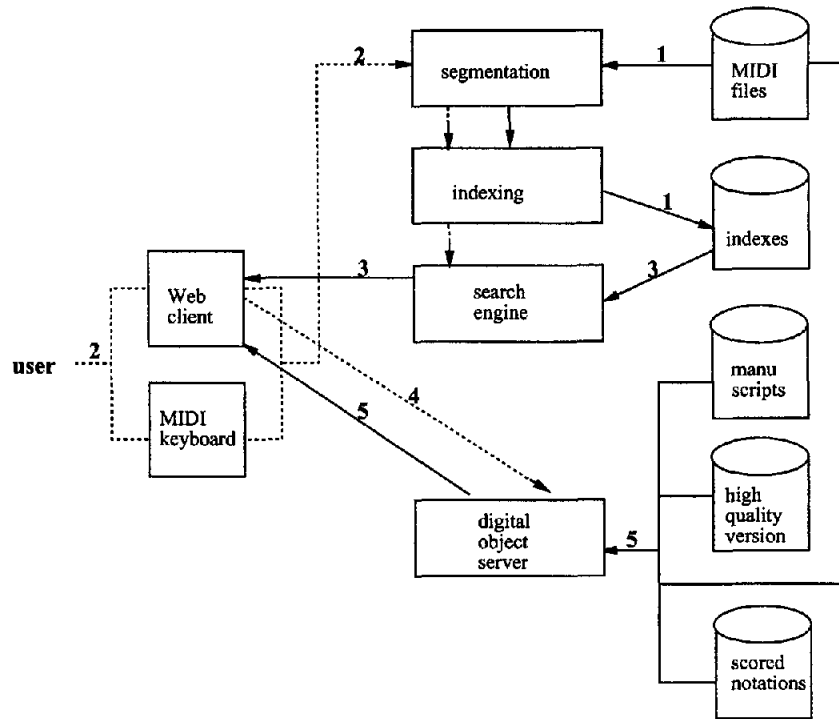


Figure 2: The architecture of a possible musical IR system.

state-of-the-art of a digital library including popular a large music document collection. That paper is interesting since describes a range of types of data and of functions to search the digital library.

7 CONCLUSIONS AND FUTURE WORK

Results show that the proposed indexing technique allow for the content-based retrieval of musical documents at different levels of specificity and exhaustivity. We believe that such a best-match and content-based indexing technique may be integrated with current exact-match musical document retrieval systems to improve their effectiveness. In fact, we observed that at the highest levels of indexing specificity the queries of set B were able to retrieve documents of the same author. This can be useful whenever the final user is unable to query by fixed values.

Future works shall concern with the design and implementation of an effective prototype to be used by "real" final users in an operational environment. The prototype will serve to build a test collection of musical documents and queries, together with relevance judgements, to be used as experimental testbed. Furthermore, the refinement of the segmentation and normalization techniques can be useful to

improve the effectiveness of the prototype that is going to be implemented.

8 Acknowledgements

Massimo Melucci was partially supported by the INTER-DATA project from the Italian Ministry of University and Scientific Research and University of Padova.

References

- [1] S. Blackburn and D. DeRoure. A tool for content based navigation of music. In *Proceedings of ACM Multimedia Conference*, pages 361–368, Bristol, UK, 1998.
- [2] E. Cambouropoulos. Musical rhythm: a formal model for determining local boundaries. In E. Leman, editor, *Music, Gestalt and Computing*, pages 277–293. Springer-Verlag, Berlin, 1997.
- [3] CANTATE. Computer Access to Notation and Text in Music Libraries. <http://www.svb.nl/project/cantate/cantate.htm>, 1998.

- [4] J.W. Dunn and C.A. Mayer. VARIATIONS: A Digital Music Library System at Indiana University. In *Proceedings of ACM Digital Libraries (DL) Conference*, Berkeley, CA, August 1999.
- [5] W.B. Frakes and R. Baeza-Yates, editors. *Information Retrieval: data structures and algorithms*. Prentice Hall, Englewood Cliffs, NJ, 1992.
- [6] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming: Musical information retrieval in an audio database. In *Proceedings of ACM Digital Libraries (DL) Conference*, pages 231–236, New York, NY, November 1995.
- [7] H.M. Gladney, F. Mintzer, F. Schiattarella, J. Bescós, and M. Treu. Digital access to antiquities. *Communications of the ACM*, 41(4):49–57, April 1998.
- [8] HARMONICA. Accompanying Action on Music Information in Libraries. <http://www.svb.nl/project/harmonica/harmonica.htm>, January 6 1999. Last update.
- [9] J. Harvell and C. Clark. Analysis of the quantitative data of system performance. Deliverable 7c, LIB-JUKEBOX/4-1049: Music Across Borders, 1996. See also <http://www.sb.aau.dk/Jukebox/edit-report-1.html>.
- [10] J.L. Hsu, C.C. Liu, and A.L.P. Chen. Efficient repeating pattern finding in music databases. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, pages 281–288, Bethesda, MD, November 1998.
- [11] F.W. Lancaster and A.J. Warner. *Information Retrieval Today*. Information Resources Press, Arlington, VA, 1993.
- [12] M. Lesk. *Practical Digital Libraries: Books, Bytes, and Bucks*. Morgan Kaufmann, San Francisco, CA, 1997.
- [13] A. McLane. Music as information. In M.E. Williams, editor, *Annual Review of Information Science and Technology (ARIST)*, volume 31, chapter 6, pages 225–262. American Society for Information Science, 1996.
- [14] W.E. Moen. Accessing distributed cultural heritage information. *Communications of the ACM*, 41(4):45–48, April 1998.
- [15] Musica. The International Database of Choral Repertoire. <http://www.MusicaNet.org/>.
- [16] RISM. Répertoire International des Sources Musicales. <http://www.rism.harvard.edu/rism/Welcome.html>.
- [17] G. Salton and M.J. McGill. *Introduction to modern Information Retrieval*. McGraw-Hill, New York, NY, 1983.
- [18] K. Sparck Jones and P. Willett. *Readings in Information Retrieval*. Morgan Kaufmann, San Francisco, CA, 1997.
- [19] Y.H. Tseng. Content-based retrieval for music collections. In *Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR)*, Berkeley, CA, August 1999.
- [20] A. Uitdenbogerd and J. Zobel. Manipulation of music for melody matching. In *Proceedings of ACM Multimedia Conference*, pages 235–240, Bristol, UK, 1998.
- [21] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, London, second edition, 1979.