JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY

JASIST VOLUME 55 NUMBER 12 2004

CONTENTS

Papers published online in Wiley InterScience, 13 May 2004 through 25 May 2004

WILEY
Coden AISJB6 ISSN 1532-2882

Journal of the American Society for Information Science and Technology

JASIST

VOLUME 55 NUMBER 12 OCTOBER 2004

CONTENTS

PERSPECTIVES ON...
MUSIC INFORMATION RETRIEVAL

(*Continued on next page*)

On the Cover: Taking Note, designed by Lois Lunin and Mike Rutkowski

Volume 55, Number 12 was mailed the week of September 6, 2004

Combining Melody Processing and Information Retrieval Techniques: Methodology, Evaluation, and System Implementation

Massimo Melucci and Nicola Orio
Department of Information Engineering, University of Padova, Via Gradenigo, 6/A, 35131 Padova, Italy.
E-mail: {melo, orio}@dei.unipd.it
IRCAM—Centre Pompidou, 1, Place Igor Stravinsky, 75004 Paris, France

The article describes the project on music information retrieval that has been carried out at the University of Padova, Italy. The research work has been characterized by the synergy of the modular integration of sound techniques of melody processing and of statistical information retrieval. After illustrating the background from which the project has originated, we describe the complete process, from methodology design through evaluation and system implementation. Conclusions, impacts on research in music information retrieval, and future directions are also described.

Introduction

The research project in music information retrieval (MIR) that has been carried out at the University of Padova and reported in this article originated from a feasibility study for the ADMV (*Archivio digitale della musica veneta del XVIII secolo*, Digital Archive for the Venetian Music of the 18th Century) digital library (DL) project (Agosti, Bombi, Melucci, & Mian, 2000). The study defined architecture, technology, and search functions for a music DL, and it investigated the interrelationships of all the typical aspects of a real, effective DL—distributed databases, preservation, wide-area networking, security, image and sound representation and storage, and data management. Like other music archives, ADMV provides music search functionalities based on fixed attributes, which have some drawbacks if end users want to access the music collections on the basis of document content, as explained in section on data-based approaches.

The need for providing content-based access has been studied in nontextual or multimedia information retrieval (IR) domains and is specifically addressed in cultural heritage domain, where different media and search functions live together (Gladney, Mintzer, Schiattarella, Bescós, & Treu, 1998; Lesk, 1997, Moen, 1998). As music is one of the most important repositories of cultural heritage, content-based search functions allow for the access to music documents that are stored in DLs with different formats—from notation to recording—which meet different aims as explained in the section on Music Representation, Format, and Processing. *Incipites* can sometimes be used as source from which content descriptors are extracted, yet they are a nonrepresentative sample of complete music works, which are often arranged in many different parts. Therefore, the representation of the full music document content is an important requirement. Moreover, inexperienced users may want to access the collections without knowing composer names, work titles, or classification codes. A MIR system should allow for browsing among the music collection, navigation across hyperlinks, and friendly querying.

Content-based access can be accomplished through the construction of indexes mapping content descriptors to documents. This is the core task of IR because it gives content descriptors the representational power to discriminate between relevant and nonrelevant documents. Manual indexing effectively describes documents, but it is clearly inefficient for large music collections, which hence should be handled by automatic indexing techniques. In our research project, we hypothesized the possibility of automatically indexing and retrieving full music documents at an acceptable degree of effectiveness and at a very high level of efficiency. We designed and implemented techniques for melody surface segmentation to extract perceptually meaningful phrases from documents and queries, for indexing music documents using the extracted phrases, for retrieving documents matching queries, and for generating hyperlinks among documents and phrases to enable navigation. Normalization is applied to cope with melodic variations of phrases, analogously to stemming algorithms that are applied to cope with word variants. Efficiency is guaranteed

 • Published online 25 May 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20058

because indexing algorithms are derived from the textual IR domain. Melody segmentation algorithms give automatic indexing levels of effectiveness comparable to manual indexing because they are based on human perception and coupled with statistical weighting schemes.

Related Work

Approaches to Music Information Retrieval

There is a variety of approaches to MIR and many related disciplines are involved. For this reason, we cite only the work we believe being most relevant to our research aims. Current approaches to MIR can broadly be classified into content-based and data-based approaches.

Content-based approaches. Content-based approaches take into account the music document content and automatically extract some features, which are used as content descriptors. Among the different dimensions of music documents, a main focus is given to the melody. The typical methodology extracts note strings from the full-score music document. If arbitrarily extracted, note strings may be musically meaningless because no music information is exploited to detect them, yet they may allow for a good coverage of all the possible features to be extracted. Content-based approaches can sometimes be oriented to disclosing music document semantic content, under the hypothesis that music documents hold a meaning and that some fragments can effectively convey it. In this case, some information on music theory is exploited to detect the strings. For instance, fragments may be considered relevant if they musically make sense when they are played individually.

Query processing can significantly differ within content-based approaches. After a query has been created by playing a melody, the system can represent it either as a single note string or as a sequence of melodic fragments. The latter can be either arbitrary, such as *n*-grams, or suitably extracted using melody information. Regarding the query as a single note string makes content-based retrieval very difficult, because it would be similar to retrieving textual files using Unix grep-like commands, which is known to provide poor results. On the contrary, extracting fragments using melody information can result in a more effective query description and can be integrated with partial match retrieval techniques. Examples of content-based approaches to MIR are described in Ghias, Logan, Chamberlin, and Smith (1995), Blackburn and DeRoure (1998), Tseng (1999), Hsu, Liu, and Chen (1998), and Uitdenbogerd and Zobel (1998).

Data-based approaches. Data-based MIR systems allow users for searching databases by specifying exact values for predefined fields, such as composer name, title, and type of work. These approaches make content-based IR almost impossible because the music content cannot easily be conveyed by bibliographic values, which are usually related only to general features as the musical form, the tonality, and the solo instrument. Moreover, descriptive titles, such as Tchaikovsky's "Pathétique," are insufficient to express a user's information need whenever he would find them not being a good description of the music work.

From an IR point of view, data-based approaches are quite effective if the user can exhaustively and precisely use the available search fields. However, bibliographic values may not always be able to describe the content of music works. For example, the terms "Sonata" and "C Major" as value of the type of work cannot sufficiently discriminate all the existing sonatas in that tonality. The additional use of cataloging number, like K279 for Mozart's "Sonata in C Major," will be effective only if the user has complete information on the musical work, and in this case a database system will suffice. Searching by composer name can be effective. However, works of less known composers may not be retrieved if only because their names are little known, while for a prolific composer an extremely high number of documents, unbearable to the final user, may be retrieved. Examples of data-based approaches to MIR are reported in CANTATE (1999), HARMONICA (1999), Harvell and Clark (1996), Musica (1999), and RISM (1999).

Music Digital Library Systems

DL projects have recently been conducted for designing, implementing, and testing real MIR systems. Some of them implement data-based, content-based, or both approaches. We cite some of the projects most relevant to our research aims. The reader can access the cited papers to have a complete description of methods and systems. The VARIATIONS DL (Dunn & Mayer, 1999) and MELDEX project (Bainbridge, Nevill-Manning, Witten, Smith, & McNab, 1999) are two interesting examples of music DLs. A project involved the University of Milan and the Teatro alla Scala (Ferrari & Haus, 1999) to implement a multimedia object-relational database storing the music contents of the archive, as well as catalogue data about the nights at the Teatro alla Scala.

Music Representation, Format, and Processing

The first problem that arises in the automatic music processing is that a music work may be digitally stored in different formats. A work can be represented by a reproduction of the manuscript, a symbolic notation of the score, a sequence of time-stamped events corresponding to notes, or a digital recording of a performance. Each format carries different information on the document content. For instance, it is impossible to recover the written score from a polyphonic audio signal while, on the other hand, the score carries little information about timbre, expressive timing, and other performing parameters. Hence, document format has to be carefully chosen depending on the aims of the DL, which may encompass preservation, display, listening, indexing, and retrieval.

A number of standard formats exist for the representation of music information, each one more suitable for a particular class of users. A standard created for musicians is the Musical Instrument Digital Interface (MIDI) protocol, which was proposed in 1982 for data exchange among digital instruments (Rothstein, 1991). Even if MIDI is somehow obsolete, it is still widely used because it carries information about the musical events from which it is possible to both reconstruct the score and listen to a performance played, for instance, by the internal synthesizer of a computer. The fortune of MIDI is fading because of the development of new formats for compression, storage, and distribution on the Internet of digitized performances, like MP3 (MPEG, 2001). Unlike MIDI, MP3 is totally oriented to listeners, because it is not possible to recover the music score. With the aim of taking into account all the variety in which music information can be represented, the Standard Music Description Language (SMDL) (ISO, 1995) has been proposed as an application of the Standard ISO/IEC Hypermedia/Time-based Structuring Language. In SMDL, a music work is divided into different domains, each dealing with different aspects, from visual to gestural and analytical. SMDL then provides a linking mechanism to external, preexisting formats for visual representation or storage of performances. Hence SMDL may be a useful way for music representation standardization, but the solution is just to collect different formats rather than proposing a new one able to deal with all the aspects of the communication in music.

Approach

The key idea of our approach is the synergic combination of melody processing and information retrieval methods, where the former produces the output being processed by the latter as input. Our approach hinges on three fully automated processes: Melody processing to extract melody phrases that are meaningful from a perceptual point of view; indexing to produce melody phrases describing music content; automatic hyperlink generation to allow end users to navigate a space of music documents and phrases.

Melody Processing

Melody processing allows for the automatic extraction of music content descriptors of both documents and queries respectively at indexing and retrieval time. Documents and queries have a number of differences that affect the choice of the content descriptors. Typical end users are likely to express their information needs by humming or playing on a simple digital interface. Their queries will then be monophonic, played at a random tonality, probably affected by errors, and without any information on music structure. On the other hand, music documents may give a richer and more reliable information, for instance on harmony and rhythm structure.

Since content descriptors should be coherent between documents and queries, we decided to reduce the music dimensions to be used for content extraction to only the melodic information, where melody is considered as a continuous flow of notes. To this end, the content descriptors that we chose are the *musical phrases*. A musical phrase is a melody excerpt that is perceived by listeners as a *unit* of the music language. Informally, a musical phrase can be played alone preserving its musical meaning, even if we are aware that the context is very important in music. We believe that the choice of a content descriptor that is musically meaningful, and that can efficiently be indexed using, for example, inverted files, will improve the system performance in terms of response time while preserving good accuracy and precision. The first step in melody processing is the automatic *segmentation* of the melody.

When indexing music documents and queries, variants in the musical phrases can occur. Indeed, musical phrases may not be equal but may be perceived as similar. To make the variants of a phrase comparable to one another, we propose to perform two kinds of normalization in the representation of musical phrases, respectively on pitch and duration of the notes. This process can be considered as an analogous of stemming in text document indexing. The second step in melody processing is the automatic *normalization* of the musical phrases.

Segmentation. The method for automatic segmentation is based on the Local Boundaries Detection Model (LBDM) (Cambouropoulos, 1997). The basic idea of the model is that listeners perceive the presence of boundaries in a melody whenever there are changes affecting musical intervals and note duration. Hence a musical phrase can be bounded before or after a note that has a duration or that forms an interval different from the surrounding notes.

Melodic boundaries are uncertain events because there is no explicit separator in the melody (as spaces or commas in text) and their presence depends on listeners' perception. Because of this uncertainty, the LBDM assigns a weight to all possible places in which a boundary may occur, that is, between two subsequent notes of the melody. A weight represents the probability of a boundary at the corresponding position. The actual boundaries can be detected by analyzing the weight trend. Cambouropoulos (1997) proposed that the boundaries are associated with the presence of a local maxima in the weight function. We have developed an algorithm by implementing the LBDM. The weight values are computed using the following data and rules:

- If *intervals* between two notes are different (e.g., a perfect fifth versus a third but not a minor versus a major third), a value is added to the two weights—2 for the larger interval and 1 for the smaller one.
- If *durations* are different (e.g., a quarter note versus a whole note), a value is added to both weights—respectively 4 and 1 if the first note is longer, 3 and 2 if it is shorter.

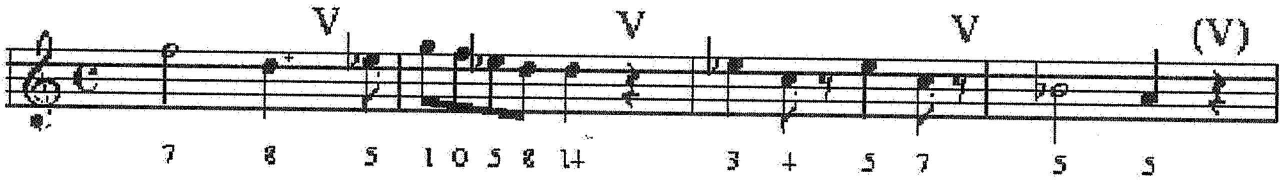

FIG. 1. Example of segmentation: For each note the respective weights are quoted and four **V** markers are inserted to highlight the boundaries.

- If a *long duration* or a *rest* occurs, a value is added after the long note or to the rest position—the value is 4 in both cases.

Let us take as an example the segmentation of the first four measures of Mozart's Concerto for Clarinet K622, reported in Figure 1. At first a weight is given to each interval between two subsequent notes, depending on the rules previously explained (weights are quoted under the previous note). Then the local maxima are calculated for detecting the actual boundaries. In Figure 1 boundaries highlighting four musical phrases are shown by a "V".

Normalization. The concept of perceptual similarity between musical phrases is widely used by composers, and the continuous variations of the same melodic material is a typical compositional technique (e.g., the subject in a fugue, the *leitmotif*, the musical form Theme and Variations). Moreover, variants may be randomly and involuntarily made by the final user when querying the system. Errors in intonation and timing are typical for musically untrained users when they are using a query-by-humming system.

Different levels of normalization, which may affect both pitch and duration, can be applied. First of all, we propose a *Pitch Transposition* (PT): All the notes are transposed having each musical phrase beginning with the same note. This makes the indexing robust to possible differences in tonality. Then we introduce a *Duration Normalization* (DN): Note durations are expressed in multiples of their Greatest Common Divisor, so that what becomes important is the relationship among durations. This may overcome the problem of differences in tempo between queries and documents and among documents. Since variants of musical phrases may regard small changes in pitch contour, due, for instance, to modulations or to wrong notes, we also introduce a *Pitch Normalization* (PN): Pitches are quantized in a number of different levels, related to the musical intervals (i.e., unison, from minor second to major third, from perfect fourth to major sixth, and so on). Finally, we propose to consider only pitches, when *Duration Removal* (DR) is performed. We chose to combine them to obtain an increased generality of the musical phrases. The user may choose which kind of normalization has to be applied by the system, with the following conventions:

- *Very specific, Not exhaustive* = Pitch Transposition with duration unchanged.
- *Specific, Little exhaustive* = Pitch Transposition and Duration Normalization.
- *Little specific, Exhaustive* = Pitch Normalization and Duration Normalization.
- *Not specific, Very exhaustive* = Pitch Normalization and Duration Removal.

Once musical phrases are normalized, it is possible to assign them a textual representation. Textual representations of the music excerpt shown in Figure 1, for each of the four levels of normalization, are quoted in Table 1.

Indexing and Retrieval

Indexing takes place on the phrase-based representation of music documents and queries given by segmentation and normalization shown in Table 1. The IR model employed to construct indexes and retrieve documents is the vector-space model. The weight w_{ij} of descriptor j within document

TABLE 1. The four different output formats of the segmentation algorithm.

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

Note. 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].

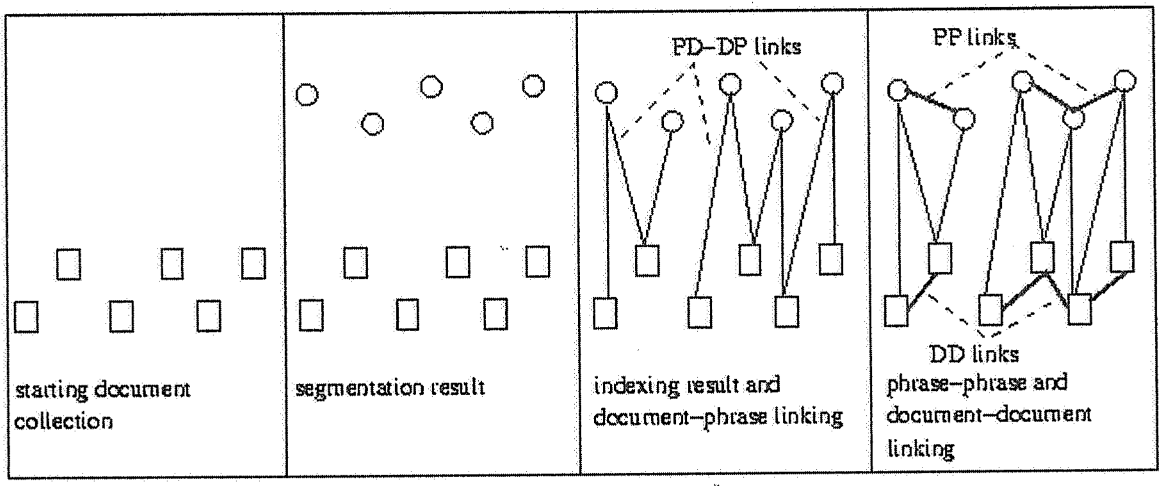

FIG. 2. The process of hyper-music construction: Documents are represented by "squares" and phrases are represented by "circles."

i belongs to the $tf \times idf$ family, and it is expressed as $tf_{ij} \times idf_j$, where tf_{ij} is the frequency of occurrence of descriptor 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 descriptor j.

As the cosine function normalizes the retrieval status value to query and document lengths, their sizes have been controlled, so long documents have the same chance of being retrieved as short ones. The use of a $tf \times idf$ weight implicitly assumes that documents with high intradocument frequency (tf_{ij}) and low intracollection frequency (n_j) are more likely to be relevant than documents with low tf or low idf. Of course, the direct correlation between $tf \times idf$ and probability of relevance needs to be tested, like many other aspects of MIR do. However, it is likely that the statistical properties observed in the textual domain still hold in non-textual or multimedia domain (Sparck Jones & Willett, 1997.

Link Generation and Navigation

The methods for segmentation-normalization and indexing can be combined to automatically build a *hyper-music* (by "hyper-music" we mean a hyper-media of which nodes are pieces of music). In fact, phrases extracted from documents can be used as content descriptors, like keywords do for textual documents, while indexes can be used to build links using co-occurrence information, based on the number of phrases shared by two documents. Figure 2 illustrates the whole automatic hyper-music construction process in four steps.

a. The starting point is a "flat" collection of "flat" music documents. They can be complete, or parts of documents, and form the D level of the hyper-music.
b. The next step is automatic segmentation as described earlier in the section on segmentation.
c. Indexing produces an index of musical phrases being associated with the music documents that occur within. Standard algorithms have been chosen (Frakes & Baeza-Yates, 1992).
d. Automatic link construction involves four types of link and differs whether a link is among nodes of the same level, i.e., DD or PP links, or among nodes of different levels, i.e., D or DP links. Details follow.

DD (PP) links connect documents (phrases). A numerical measure of similarity is computed between documents (phrases) i and j to estimate the closeness between the respective semantic content. The measure is the cosine of the angle between the document (phrase) vectors. For each document (phrase) i, a ranked list of similarity can then be computed. A DD (PP) link is inserted between documents (phrases) i and j if the measure of similarity is over a stated threshold. DP links connect a document with phrases, while PD links connect a phrase with documents. The weight of phrase within document can be used as a numerical measure of link strength. The measure can vary according to the weighting scheme. For example, if the $tf \times idf$ scheme is adopted, the phrase j occurring within document i is weighted with $tf_{ij} \log N/n_j$, where tf_{ij} is the frequency of phrase j within document i, n_j is the number of documents indexed by phrase j, and N is the total number of documents. A DP (PD) link is inserted between documents and phrases if the weight is over a stated threshold.

Browsing through automatically constructed links can integrate querying. It can produce higher recall because the user can select additional documents through DD and PD links, and additional phrases through PP and DP links. Browsing can also produce higher precision because the user can select documents or phrases with different weights. Browsing steps are depicted in Figure 3, which illustrates the different actions the user can take after having queried the system.

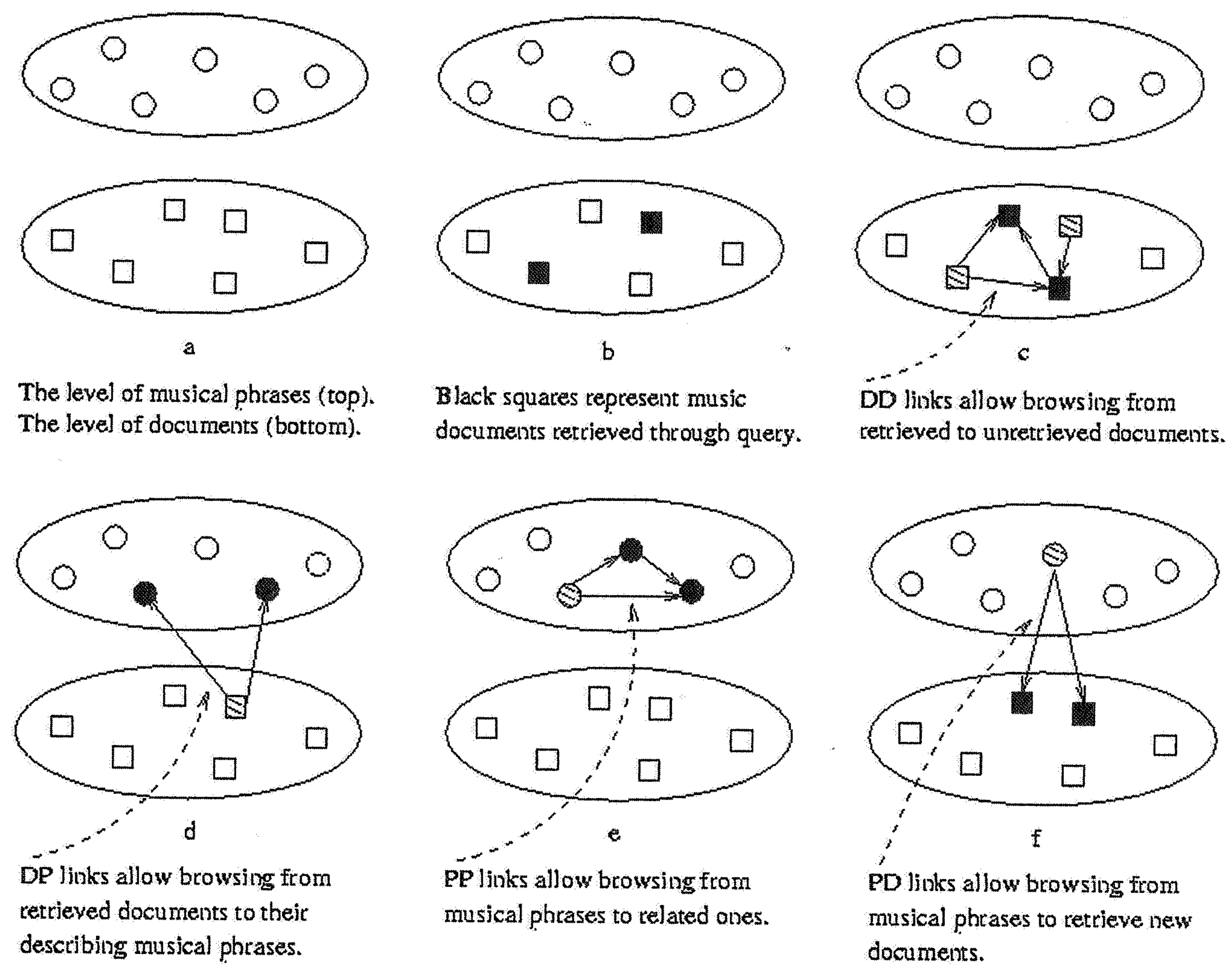

FIG. 3. The process of hyper-music browsing: Black squares or circles indicate retrieved objects through querying or browsing, while gray ones indicate starting objects.

Experiments and Results

We carried out an experimental study to assess the degree to which normalization technique affects retrieval and the pertinence of queries to retrieved documents. Experiments have been carried out in a laboratory setting using a collection of 419 music documents, each work stored in a MIDI file. Only music data were used to index documents, while textual information possibly present in the files were discarded. Documents were complete works of Tonal Western Music of various lengths—from 3 up to 35 minutes—and of six composers—Handel, Bach, Mozart, Beethoven, Liszt, and Tchaikovsky.

To study the impact of normalization, we created a versions of the same document set for each normalization criteria. Table 2, which summarizes the main numerical parameters about the document set, shows that the strongest impact occurs when duration is removed. After PNDR normalization, the average document size is less than the average document size after PNDN normalization, while the average number of documents indexed by a phrase is doubled. Furthermore, we created a set of queries by automatically extracting the *incipites* of the documents in the set. Normalization was applied also to the queries. As reported in a previous paper (Melucci & Orio, 1999), test retrieval results showed that PNDR gives the highest increase of number of retrieved documents and of different composers. In particular, documents of around three different composers are retrieved on average and the number of different composers increases as normalization becomes stronger.

TABLE 2. Main numerical parameters about documents.

	Type of Normalization			
	PT	PTDN	PNDN	PNDR
Av. phrases/document	366.0	344.6	293.6	134.1
Av. documents/phrase	1.40	1.53	1.68	4.47

Note. On average, documents are 713.7 ± 822.5 phrases long. For each normalization criterion, the average number of unique phrases per document, and the average number of documents being indexed by an individual phrase are reported.

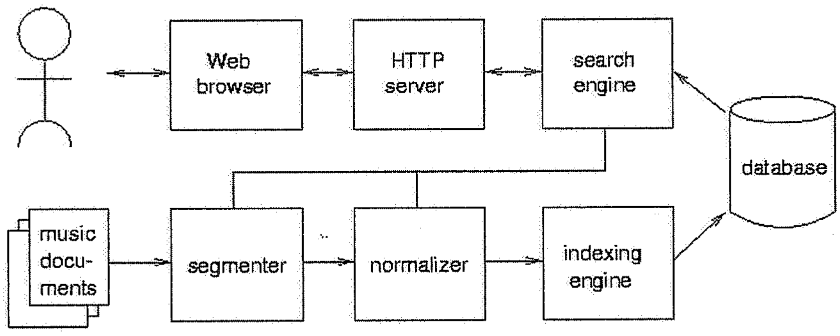

FIG. 4. The overall SMILE architecture.

With the aim of studying the behavior of the indexing and retrieval techniques as musical style changes, we developed a user-centered test. A musician prepared 15 short musical excerpts from known works to be used as queries. Queries varied depending on normalization level and whether they were about authors or documents present in the document set. Some queries were "biased" through the manual insertion of common performance mistakes. The first interesting result was that, independently from the applied normalization, queries using works present in the set always gave that same work with the highest retrieval status value (RSV). Moreover, when the author was present in the set, his pieces were retrieved with a high RSV. The distribution of RSV among authors become more homogeneous when the degree of specificity decreased. An analysis of the top ranked documents revealed that most of the retrieved documents had a melodic structure similar to the corresponding query. When PT was applied, the retrieved works had a tempo similar to the query, even if this behavior was also slightly maintained when durations were normalized. In general the user-centered analysis gave satisfying results, meaning that the system was able to assign high ranks to documents musically similar to queries.

A System for Music Information Retrieval Environments

We developed a working MIR system called SMILE, which stands for a System for Music Information retrievaL Environments. SMILE has been thought of as one of the components of a system providing access to music DLs. Thus, SMILE implements three requirements: (i) It allows for full content-based music retrieval; (ii) it relies on fully automatic processes—indexing, retrieval, and link construction; (iii) it integrates navigation and retrieval within a single methodological and technological framework. SMILE runs on a client-server architecture, where the client side consists of a standard Java-enabled Web browser running a Java applet, while the server side consists of a music indexing and search engine, the segmenter, the normalizer, a music database system, and the usual HTTP server. The overall architecture is depicted in Figure 4.

A Java applet is used to create interactively the musical queries, and to tune the retrieval. A virtual keyboard can be played through the mouse, and queries can be created both recording a short music performance and inserting the notes one by one. The applet then converts the music score to a text notation that is used to invoke the search engine. The segmenter takes the representation of both music documents and queries and creates a list of their musical phrases by using the methodology described in the section on segmentation. The normalizer creates four different textual representations, according to the ones proposed in the section on normalization, of the musical phrases produced by the segmenter. The indexing engine takes the textual representations as input and produces the indexes being used at search time; indexes have been implemented as inverted files (Frakes & Baeza-Yates, 1992). The search engine takes the query being played using the Java applet, runs the segmenter and the normalizer taking the query as input, and accesses the indexes to retrieve the matching musical documents; the vector-space model has been employed to match and rank documents. The database made available is the one used in our experiments. SMILE, which user interface is shown in Figure 5, is available at http://www.dei.unipd.it/~ims.

Conclusions and Future Directions

We encountered many exciting and hard problems, the most important being the notion of "relevance" in music. The latter can be solved but in a very long time: We think current methods and tools can in principle support end users in retrieving music information relevant to their requests, but the notion of relevance and music query in MIR is still unclear. The major obstacle in designing a MIR system is caused by the fundamental differences between music and text queries, and more generally between music and text documents. Music data are basically different from text data because the latter are unidimensional and their own content is independent on the format used to represent them. In contrast music data have different dimensions and do not clearly carry a semantic that could be used to express their content. Moreover, music data have different representation

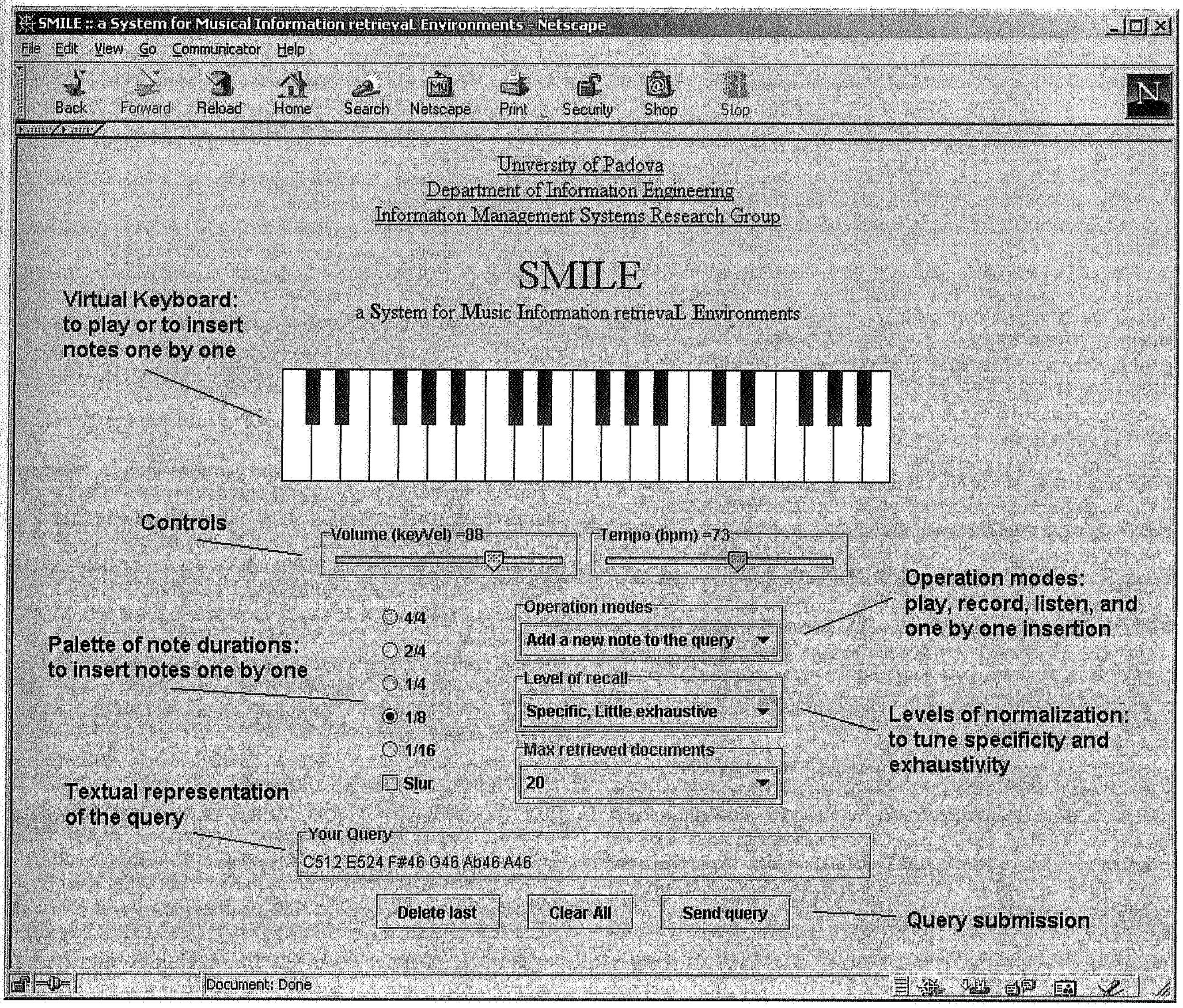

FIG. 5. The user interface of SMILE.

power depending on the format—sound represents information that cannot be represented by notation, and vice versa (Selfridge-Field, 2000). On the other hand, McLane (1996) stressed that "what . . . 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." Specifically, it is interesting to assess the degree to which some aspects of text analysis, which are used in textual IR, "apply equally well to musical analysis, where the concept of a lexical unit corresponding to a word may have more meaning in some musical works than others" (McLane, 1996)

The inadequacy of MIR systems in providing efficient ways to express music queries leads end users to ineffectively express their own information needs; for example, end users cannot express polyphonic queries if the system does not provide such a dimension with its query language. The integration of diverse search strategies may be a means to overcome the difficulty in expressing music information need. If the end user is unable to satisfactorily express his information need through a query, the system may provide him with alternative ways to access the music data. Using a search strategy the end user may be able to express what could not be expressed through another search strategy. For example, navigation of links among data is commonly used to integrate traditional query-based search strategies. Navigation is of course possible if links among data are provided, but the large size of actual music data collection makes manual link generation infeasible. Therefore, our approach integrates link navigation and query-based search strategies in a single model and system by using fully automated processes.

References

Agosti, M., Bombi, F., Melucci, M., & Mian, G. (2000). Towards a digital library for the Venetian music of the eighteenth century. In J. Anderson,

M. Deegan, S. Ross, & S. Harold (Eds.), DRH 98: Selected Papers from Digital Resources for the Humanities, Glasgow, UK (pp. 1–16). Office for Humanities Communication.

Bainbridge, D., Nevill-Manning, C., Witten, I., Smith, L., & McNab, R. (1999). Towards a digital library of popular music. In E.A. Fox & N. Rowe (Eds.), Proceedings of ACM Digital Libraries (DL) Conference, Berkeley, CA (pp. 161–169). New York: ACM Press.

Blackburn, S., & DeRoure, D. (1998). A tool for content based navigation of music. In R. Nikatsu, E. Altman, & C. Pinhanez (Eds.), Proceedings of ACM Multimedia Conference, Bristol, UK (pp. 361–368). New York: ACM Press.

Cambouropoulos, E. (1997). Musical rhythm: A formal model for determining local boundaries. In E. Leman (Ed.), Music, gestalt, and computing (pp. 277–293). Berlin: Springer-Verlag.

CANTATE. (1999). Computer access to notation and text in music libraries. Retrieved April 15, 2004, from http://projects.fnb.nl/cantate/

Dunn, J., & Mayer, C. (1999). VARIATIONS: A digital music library system at Indiana University. In E.A. Fox & N. Rowe (Eds.), Proceedings of ACM Digital Libraries (DL) Conference, Berkeley, CA (pp. 12–19). New York: ACM Press.

Ferrari, E., & Haus, G. (1999). The musical archive information system at Teatro alla Scala. In Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS '99), Florence, Italy (Vol. II, pp. 817–821).

Frakes, W., & Baeza-Yates, R. (1992). Information Retrieval: Data structures and algorithms. Englewood Cliffs, NJ: Prentice Hall.

Ghias, A., Logan, J., Chamberlin, D., & Smith, B.C. (1995). Query by humming: Musical information retrieval in an audio database. In F.M. Shipman, R. Furuta, & D.M. Levy (Eds.), Proceedings of ACM Digital Libraries (DL) Conference, New York (pp. 231–236). New York: ACM Press.

Gladney, H., Mintzer, F., Schiattarella, F., Bescós, J., & Treu, M. (1998). Digital access to antiquities. Communications of the ACM, 41(4), 49–57.

HARMONICA. (1999). Accompanying action on music information in libraries. Retrieved April 15, 2004, from http://projects.fnb.nl/harmonica

Harvell, J., & Clark, C. (1996). Analysis of the quantitative data of system performance. Deliverable 7c, LIB-JUKEBOX/4-1049: Music across borders. Retrieved April 15, 2004, from http://statsbiblioteket.dk/Jukebox/edit-report-1.html

Hsu, J., Liu, C., & Chen, A. (1998). Efficient repeating pattern finding in music databases. In G. Gardarin, J.C. French, N. Pissinou, K. Makki, & L. Bouganim (Eds.), Proceedings of the ACM Conference on Information and Knowledge Management (CIKM), Bethesda, MD (pp. 281–288). New York: ACM Press.

International Organization for Standardization (ISO). (1995). Information technology: Standard Music Description Language (SMDL) (Technical Report ISO/IECDIS 10743:1995). Geneva, Switzerland: ISO International Electrotechnical Commission. Retrieved, from ftp://ftp.techno.com/pub/SMDL/

Lesk, M. (1997). Practical digital libraries: Books, bytes, and bucks. San Francisco, CA: Morgan Kaufmann.

McLane, A. (1996). Music as information. In M. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 31, Chap. 6, pp. 225–262). Silver Spring, MD: American Society for Information Science.

Melucci, M., & Orio, N. (1999). Musical information retrieval using melodic surface. In E.A. Fox & N. Rowe (Eds.), Proceedings of ACM Digital Libraries (DL) Conference, Berkeley, CA (pp. 152–160). New York: ACM Press.

Moen, W. (1998). Accessing distributed cultural heritage information. Communications of the ACM, 41(4), 45–48.

MPEG. (2001). Home-page of the moving picture experts group. Retrieved April 15, 2004, from http://www.iso.ch/

Musica. (1999). The international database of choral repertoire. Retrieved April 15, 2004, from http://www.musicanet.org/

RISM. (1999). Répertoire international des sources musicales. Retrieved April 15, 2004, from http://www.rism.stub.uni-frankfurt.de/

Rothstein, J. (1991). MIDI: A comprehensive introduction. Madison, WI: A-R Editions.

Selfridge-Field, E. (2000). What motivates a musical query? In International Symposium on Music Information Retrieval, Plymouth, MA. Retrieved April 15, 2004, from http://ismir2000.ismir.net/papers/invites/selfridge_invite.pdf

Sparck Jones, K., & Willett, P. (1997). Readings in information retrieval. San Francisco, CA: Morgan Kaufmann.

Tseng, Y. (1999). Content-based retrieval for music collections. In M. Hearst, F. Gey, & R. Tong (Eds.), Proceedings of the ACM International Conference on Research and Development in Information Retrieval (SIGIR), Berkeley, CA (pp. 176–182). New York: ACM Press.

Uitdenbogerd, A., & Zobel, J. (1998). Manipulation of music for melody matching. In R. Nikatsu, E. Altman, & C. Pinhanez (Eds.), Proceedings of ACM Multimedia Conference, Bristol, UK (pp. 235–240). New York: ACM Press.