Content-based Retrieval in Digital Music Libraries

Michael Clausen, Frank Kurth, Meinard Müller, and Andreas Ribbrock

Department of Computer Science III, University of Bonn,
Römerstraße 164, 53117 Bonn, Germany
{clausen,frank,meinard,ribbrock}@iai.uni-bonn.de,
WWW home page: http://www-mmdb.iai.uni-bonn.de

Abstract. MiDiLiB is a six year research project on digital music libraries funded by the German Research Foundation (DFG) as a part of the *Distributed Processing and Delivery of Digital Documents* (V^3D^2) research initiative. MiDiLiB's main focus is the development of content-based retrieval algorithms for both score- and waveform-based music. In this paper we give an overview of our research results, describe several prototypical systems for content-based music retrieval which have been developed during the project, and discuss applications of the presented techniques in the context of today's and future digital music libraries.

1 Introduction

During the course of the development of digital libraries for non-textual or *non-standard* document types, the last five years have seen increasing efforts in the field of music libraries[1]. Problems arising from the task of handling non-standard document types in digital libraries are manifold. A rough and non-exhaustive life-cycle of a non-standard document within a digital library may include the stages of digitization, choice of a suitable data format, transfer to and registration with the library. Furthermore, one has to deal with the issues of content-analysis (and annotation) as well as the generation of classical and content-based index-structures for efficient document access and retrieval. Finally, the creation of (multimodal) user interfaces, usage of system independent mechanisms for long-term document storage, and development of novel services for end-users to access the documents are of fundamental importance within a library scenario.

Among those tasks, content-based document analysis and retrieval is one of the most challenging problems. In *content-based* document processing, raw data contained in a document are processed directly, rather than relying on secondary document descriptions such as annotated metadata related to the document. With regard to the huge existing collections of digital documents, efficient mechanisms for content-based document analysis are of fundamental

[1] To avoid confusion, in this paper we shall use *music* as a general term comprising score- and digital waveform-based data. The term *audio* denotes digital waveform-based data like CD-audio or radio broadcast signals.

importance, in particular as generally the manual creation of secondary document descriptions is unfeasible. In content-based retrieval, queries to document collections are processed based on suitable index structures derived from an automatic content analysis.

One of the earliest and most intuitive tasks in content-based music retrieval is the *name-that-tune* application, where a user is interested in finding the title of a tune or a song which has been broadcast on the radio. Frequently, a listener is familiar with parts of the main theme or the hook line of the song although he might not remember the composer or interpreter. In order to find out about such kind of information (metadata), one could call the radio station or, provided availability, investigate the station's playlist on the internet. In case none of these alternatives is available, a comfortable solution could be to hum or whistle the tune in question into a microphone and let a computer do the work of finding the desired information. For this purpose, the whistled tune is converted into a suitable sequence q of notes and then compared to a collection $m_1, \ldots, m_N$ of melodies which are used as a reference database. All melodies which are close to q with respect to a suitable distance measure are returned as query results. Music search based on note-representations is commonly refered to as *score-based retrieval*. One of the pioneering works in this field [1] is based on transforming query and database melodies into so called down-up-repeat (DUR, also known as *Parson's code*) sequences, for roughly representing pitch intervals between subsequent notes. Such sequences provide a certain robustness against query errors with respect to musical intervals and absolute pitches of queried notes. The book of Barlow and Morgenstern [2] is one of the first content-based music dictionaries for manual search. It contains key-normalized pitch sequences representing the introductory parts of a large number of classical pieces.

Although the previous example as well as the sketched solution sound intuitive, real score-based retrieval scenarios impose several fundamental problems such as developing methods and user interfaces for query formulation, facilitating fault-tolerant queries, or devising index structures for efficient query processing. While the scenario of melody-based search allows us to use well-known data structures and algorithms form the field of text retrieval, the search in collections of complex, polyphonic musical scores requires data models and retrieval algorithms which are better adapted to music data. When the underlying music documents consist of audio signals as, e.g., CD-audio, content-based retrieval requires methods from digital signal processing. Important research topics in the field of content-based audio retrieval are audio identification, genre classification, and recommendation ("customers who bought song a also bought song b").

By now, the community of researchers specializing in music information retrieval (MIR) has grown to a considerable size, which is documented by the success of the the annual *International Conferences on Music Information Retrieval (ISMIR)* bringing together music researchers, audio engineers, computer scientists, librarians, and music industry [3]. The recently finished MiDiLiB-project, which has been funded by the German Research Foundation (DFG) as a part of the *Distributed Processing and Delivery of Digital Documents* (V^3D^2) research

initiative, has contributed to several of the recent advances in content-based music retrieval. In this paper, we outline the technical concepts on music retrieval developed by the MiDiLiB-project. Besides developing techniques for content-based indexing and search of music documents this includes concepts for the important tasks of audio monitoring and synchronization of musical documents in different formats. We present several of our prototypical systems for efficient music retrieval and give an overview on our test results. Taking into account related work, we outline current issues in music retrieval and discuss future trends in the area of digital music libraries.

Concerning content-based music retrieval, MiDiLiB's main contributions are the development of data structures and efficient, fault tolerant techniques for polyphonic search in collections of polyphonic scores [4]. Besides for the first time allowing efficient search in polyphonic scores, a particular strength of the proposed techniques is a natural mechanism for finding and precisely localizing partial matches, the latter accounting for possibly incomplete user knowledge. Extensions of our technique have been successfully applied to the problems of fast audio identification [5]. As a generalization, we developed a technique for content-based search in large classes of multimedia documents including digital 2D-images, shapes, and 3D-models [6].

The paper is organized as follows. In Sections 2–4, we give an overview on our techniques for content-based music retrieval on different types of music documents. Section 2 deals with the task of searching in scores of polyphonic music. As a second task, in Section 3 we discuss melody-based retrieval, which may be considered as the monophonic version of the latter (general) score-based retrieval. However, when allowing vague user queries such as melodies whistled into a microphone, special care has to be taken to develop suitable mechanisms for incorporating fault tolerance. Sections 4 sketches how the developed techniques may be extended to search in large collections of audio signals. In particular, we describe a technique for identifying short excerpts of audio signals and present some recent applications. Finally, Section 5 discusses techniques for synchronizing music documents given in different data formats (like, e.g., score- or signal-based formats) and outlines the importance of such algorithms in digital library scenarios. Concluding, Section 6 discusses possible applications of the retrieval techniques in the context of future digital music libraries.

2 Score-based Retrieval

One of the key problems in searching scores of polyphonic music by content is to find appropriate models for representing the score data. In a classical approach, which has already been sketched in the introduction, a sequence of notes is simply modeled as a string of symbols representing the notes' pitches. For example, the melody *Twinkle twinkle little star* would be represented by the string `ccggaagffeedc`. In a score-based retrieval scenario, a collection of N melodies would be modeled by a database of strings $m_1, \ldots, m_N$. Likewise, a user's query is represented by a string q. For query processing, q would be compared to all

melody strings m_i using a suitable similarity measure d. A simple similarity measure is given by the *edit distance* $d(A, B)$ between two strings A and B, i.e., the minimum number of *edit operations* required to transform string A into string B. The three classical edit operations are insertion, deletion, and replacement of individual symbols.

It turns out that when considering polyphonic music, the classical string-based approach is no longer feasible. One of the main reasons is that simultaneous notes cannot be modeled appropriately. Also note durations and rhythmic behaviour are not modeled by the above string-based representation. A main problem with the string-based approach is the treatment of the notes' onset positions which are represented only implicitly, i.e., by their respective position within the melody string. In [4] we present a framework for modelling polyphonic music where note onset positions are made *explicit*. In this, a note is a pair $[t, p]$ consisting of a pitch p and an onset position t. Then, a score-based music document may be easily modelled as a *set* of notes. For example $D_1 := \{[10, c^1], [10, e^1], [14, c^2]\}$ is a music document consisting of three notes, two simultaneous notes of pitches c^1 and e^1 played at time position 10 and one note of pitch c^2 played at time 14. The set of all possible notes will hence be denoted by $U := \mathbb{Z} \times \{c^1, d^1, e^1, \ldots\}$, i.e., each note has an integer time-component ($\mathbb{Z}$) and a certain pitch ($c^1, d^1, e^1, \ldots$). We consider a database $D_1, \ldots, D_N$ where each document D_i is simply a set of notes, i.e., $D_i \subseteq U$. *Exact matches* for a query $Q \subseteq U$ may then be easily defined by requiring that a time-shifted version $Q + \tau$ of Q occurs in a document D_i. As a toy example, consider the query $Q := \{[5, c^1], [9, c^2]\}$. As $Q + 5 := \{[5 + 5, c^1], [9 + 5, c^2]\} = \{[10, c^1], [14, c^2]\}$ is contained in the above document D_1, the pair $(5, 1)$ will be called a match, the number 1 specifying the matching document and the number 5 representing the time-lag required to shift the query to the matching position.

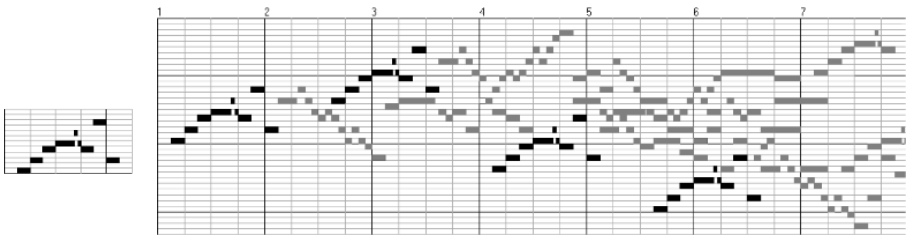

Fig. 1. Query to a database in the piano roll notation (left) and excerpt J.S. Bach's Fugue in C major, BWV 846, where all occurrences of the query are highlighted (right).

The proposed idea may be easily extended to other types of matches. As an example, assume that instead of using symbols $c^1, d^1, e^1, \ldots$, pitches are modelled using integers $0, 1, 2, \ldots$. Then we could define a match to be a triple (τ, ρ, i) satistfying $Q + (\tau, \pi) \subseteq D_i$, i.e., the query Q time-shifted by τ *and* pitch-shifted

by π semitones occurs in document D_i. Score-based music is typically visualized using the so called *piano-roll* notation, where each note is represented by a rectangle located at position $[t, p]$ corresponding to its pitch p and onset-time t. The width of a rectangle is proportional to the corresponding note's duration. As an example, Fig. 1 (left) shows the piano-roll representation of a small query document. To the right, all pitch- and time-shifted occurrences of the query within an excerpt of J.S. Bach's Fugue in C major, BWV 846, are highlighted. Further types of matches are necessary when considering fault-tolerant search. An important example is that a few, say k, notes of a query Q do not match a target position $Q + t$ within document D_i. To account for this, we say that (t, i) is a hit with k *mismatches*. Further types of fault tolerance including a concept of *fuzzy notes* as well as a mechanism for incorporating a user's prior knowledge on certain aspects of the desired document are discussed in [4].

Efficient index-based algorithms for the proposed types of matches have been devised and successfully tested on a database of 12,000 pieces of music containing about 33 million notes. The algorithms are based on a modified version of inverted files, which are well-known from classical full-text retrieval. Intuitively, for a given database $\mathcal{D} = (D_1, \ldots, D_N)$, one inverted file $H_{\mathcal{D}}(p)$ is created for each pitch p. If a note $[t, p]$ occurs in piece D_i, an object (t, i) is included in the inverted file $H_{\mathcal{D}}(p)$. Together, the inverted files form an *inverted index*. In our PROMS system [4], exact queries consisting of some 10–100 notes can be answered in about 50 milliseconds on a Pentium II, 300 MHz PC. The underlying database consists of music given in the popular score-like MIDI format [7]. Queries may be specified, e.g., by using an integrated piano-roll editor or by recording a piece of music using a MIDI-piano connected to the system.

We summarize some related work in the field of polyphonic score retrieval. Lemström et al. recently considered several retrieval tasks using a data model which is very similar to our approach [8]. Doraisamy et al. [9] model polyphonic scores based on n-grams and use standard database techniques for query processing. Pickens et al. [10] consider polyphonic score-based retrieval where the query consists of an *audio signal*. For this purpose, the query is first transformed into an approximate score version and then processed further. A very difficult and yet unsolved aspect of score-based retrieval is *similarity-based* search. Typke et al. use the Earth Movers Distance in combination with a set-based data model to incorporate a similarity measure within an efficient retrieval algorithm [11].

3 Melody-based Retrieval

In a melody-based retrieval scenario we assume that a melody, i.e., a monophonic sequence of notes, is used for querying a database of melodies. Typically, a melody-based retrieval system allows queries to be formulated by humming, singing, or whistling a tune into a microphone. Hence such a system is targeted to a much broader class of users than a system for polyphonic retrieval.

The above set-based approach may be naturally extended to handle melody queries. However, due to the differences in the general query scenario, much more

fault-tolerance is needed to successfully process a query. First, the hummed or whistled query is is transformed into a sequence of notes using a suitable extraction algorithm. Besides note extraction being an error-prone task in itself, the users's input may contain wrong notes, missing notes, deviations in rhythm and tempo, or the query may be formulated in the wrong key. Taking into account such sources of errors, our set-based approach to score-retrieval has been extended by a tempo-tracking mechanism as well as a technique to account for missing notes. The tempo-tracker is used to account for the typical tempo variations occurring in many hummed or whistled queries. For this purpose, during query processing each match candidate is assigned a tempo-tracking parameter. In each step of the retrieval algorithm, a limited variation of this parameter (w.r.t. the queries' tempo curve) is allowed, otherwise a match candidate is excluded from further processing. In [12], we present the NWO-system (notify!-

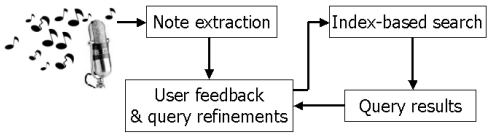

Fig. 2. Overview of the NWO system architecture

by-whistling online) for recognizing tunes whistled into a microphone. Fig. 2 shows an overview of the NWO system's architecture. Following the extraction step, the extracted notes are presented to the user in a piano-roll representation. The user then may verify his query by acoustic playback and make corrections to the extracted notes. Before actually issuing a query, the user is allowed to incorporate his prior knowledge about the desired query result by specifying parameters like rhythm tolerance or the maximum number of missing notes. After querying the user is allowed to copy melody fragments of particular query results for reusing them as new queries. Hence a special type of relevance-feedback is possible. An online version of NWO working on a manually compiled (and hence limited) database of about 2,000 melodies is currently made available.

It turns out [12] that our set-based approach for tempo-tracking shows significant advantages when users are able to remember rythmic and harmonic details of the query item. In such a case our search algorithms, in contrast to the classical edit-distance based approaches, only retrieve the few relevant database items as query results. This holds even if a query consist of only a few, say 4–6, notes. If, on the other hand, queries are of low quality, an edit distance- (or generally string-) based retrieval approach usually yields better results. However, this comes at the expense of longer result lists and longer required query lengths.

In the field of melody-based retrieval, a significant amount of research has been done during the last decade. Besides the pioneering work carried out in the

New Zealand Digital Library project [1] we only mention two recent contributions. First, the Cuby-Hum system incorporates many of the essential technical aspects of a state-of-the-art query-by-humming system. The paper [13] is thus a good starting point for further reading. As a second aspect, a robust extraction of note events from user queries is crucial for obtaining high quality queries. In connection with the musicline.de database project, a query-by-humming technique has been developed which uses a physiological model for pitch extraction [14].

4 Audio Retrieval

In audio retrieval, rather than working on high level musical features such as notes, retrieval is performed based on the digital waveform signals of the underlying music. An early approach for classifying sounds according to characteristic acoustic and perceptual features has been proposed by Wold et al. [15]. In this approach, short acoustic fragments constituting similar sounds are grouped to clusters, examples being clusters for laughter, scratchy sounds, or barking dogs.

An important problem in audio retrieval is the *identification* of audio signals. Given a large database of known audio signals, the latter may be regarded as a retrieval problem. Typically, a query signal will be a short and probably very noisy excerpt of an original signal. A recently very popular application scenario consists of recording a part of an unknown song using a mobile phone. Such a scenario could take place in a car, a restaurant, or some other noisy environment. The recorded song is then transmitted to an identification agency. After successful identification, the user is provided with the song's title, composer, interpreter, and possibly ordering information for the corresponding CD.

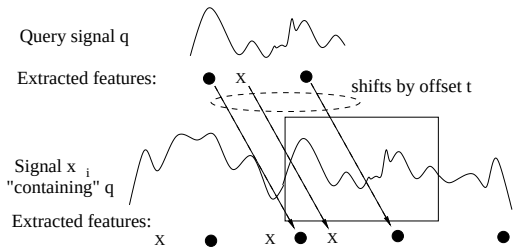

Fig. 3. Waveforms and feature representations extracted from a query signal q (top) and a database signal x_i (bottom). A t-shifted version of the query signals' feature representation $F[q]$ occurs in $F[x_i]$.

A main problem in audio retrieval are the large volumes of data which have to be handled. However, it turns out that the set-based approach to score-retrieval presented in Section 2 may also be used to obtain efficient algorithms for audio identification. Fig. 3 gives an overview on the underlying concepts, where the

basic idea consists of converting the huge number of sample values constituting an audio signal to a so-called *feature representation*. Indexing and searching is then performed based on those feature representations. Feature representations are obtained using a so-called *feature extractor* F, which assigns class labels c from a set $\mathcal{X}$ of available feature classes to signal positions t within an audio signal. A *feature* $f = [t, c] \in \mathbb{Z} \times \mathcal{X}$ is a pair consisting of a time position t and a class label c assigned to this position. Fig. 3, shows a feature extractor F extracting significant local maxima and minima from an input signal. In this case, $\mathcal{X} = \{\mathrm{x}, \bullet\}$, where feature class $\bullet$ denotes essential local maxima and class x denotes essential local minima. The figure shows a query signal q and a database signal x_i which are processed by F. The upper part shows the waveform signal q and the feature representation $F[q]$. Note that the the extracted features are plotted at their respective sample positions. Hence, for the given signal q, the feature representation $F[q]$ contains three elements. The lower part of Fig. 3 shows the corresponding data for the signal x_i. To illustrate our concept of feature-based identification, we note that a t-shifted version of the query signal q occurs in the database signal x_i. This in turn is reflected by the t-shifted feature representation $F[q]$ matching a subset of the feature representation $F[x_i]$. As compared to the above technique for score-based search, the notes $[t, p]$ are just replaced by features $[t, c]$, hence making the full index-based retrieval technique available. Thus for audio indexing, instead of creating inverted files for each pitch, the search index consists of one inverted file for each feature class.

To give an impression of resulting index sizes, we mention one of our test collections consisting of 15 GB of uncompressed audio which has been indexed using several different feature extractors. The resulting index sizes range from 33 to 128 MB amounting to averages of 85–420 features per second. For more test results and detailed information on the various feature- and application-settings as well as our prototypical *audentify!*-system, we refer to [16, 17].

As a second retrieval task we briefly discuss *audio monitoring*. Monitoring applications generally deal with the detection or inference of certain events from particular real-time data streams and have recently gained a great deal of attention, particularly in the fields of databases and information retrieval. When monitoring real time audio streams such as broadcast channels (e.g., radio or TV), one is interested in detecting occurrences of known audio fragments or, more generally, particular acoustic events within those streams. To illustrate how audio monitoring may be performed using our retrieval technique, we assume that a collection of radio commercials is given as a database. In a preprocessing step, a search index is built from this collection as described above. During the process of monitoring, a computer receives a radio program via cable or network connection. Subsequently, the incoming audio signal is transformed into a feature representation. After a fixed number of features have been extracted, we use this so-called *feature segment* as a query to the search index. By storing the query results of each feature segment in an appropriate data structure, it is possible to precisely transcribe which commercial of the database has been broadcast at what time. Applications of such a scenario include automatical creation of ad-

vertising statistics, or, using a search index built from a large collection of music signals, automatic generation of playlists. For a description of our prototypical monitoring systems *audentify!-live* and Sentinel we refer to [5, 18].

We briefly summarize related work on audio retrieval. An early approach to recognize distorted musical recordings was proposed in [19]. In the context of large data collections, algorithms for robust audio identification include audio hashing [20], geometric hashing-based approaches as proposed by Wang et al. [21], hidden markov models (HMMs) [22], or clustering-based approaches [23]. An overview on the proposed techniques may be found in [24]. Only recently, audio identification services for the above mobile phone scenario have been launched in several parts of Europe as, e.g., the service offered by Shazam Entertainment. Future work will also be concerned with using compact feature representations of audio signals for exchanging content-based information [25].

5 Music Synchronization

Modern digital music libraries contain textual, visual, and audio data. Among this multi-media based information, musical data poses many problems, for musical information is represented in many different data formats which, depending upon the application, fundamentally differ in their respective structure and content. So far we have encountered two such data formats: the score data format and the digital waveform-based data format which we simply referred to as audio. Score data roughly describes music in a formal language depicted in a graphical-textual form, whereas audio data encodes all information needed to reproduce an acoustic realization of a specific musical interpretation. Other data formats such as MIDI may be thought of as a hybrid of the score and audio data format. In MIDI, relevant content-based information such as the notes of a score as well as agogic and dynamic niceties of a specific interpretation can be encoded.

Hence, a musical work in the digital context is far from being unique since it can have several different realizations in several different formats. This heterogeneity makes content-based browsing and retrieval in digital musical libraries a challenging tasks. For example, one may think of a user who tries to find a specific passage in some audio CD but only roughly knows the melody or only remembers some score-based information such as a configuration of certain notes.

One important step towards a solution is the synchronization of multiple information sets related to a single piece of music. In the audio framework, by *synchronization* we denote a procedure which, for a given position in one representation of a piece of music, determines the corresponding position within another representation (e.g., the coordination of score symbols with audio data). Such linking structures could extend score-based music-retrieval to facilitate access to a suitable audio CD and could assist content-based retrieval in heterogenous digital music libraries — for example, allowing melody-based retrieval in the audio scenario and vice versa. Furthermore, linking of score and audio data could be useful for automatic tracking of the score position in a performance or for the investigation of tempo studies.

Within the MiDiLiB-project, we designed and implemented algorithms for the automatic synchronization of score-, MIDI- and audio-data streams representing the same piece of music [26]. To align, for example, an audio data stream with a score data stream, we first extract score-like parameters such as onset times and pitches from the audio data stream. Then the actual alignment is computed based on the score-parameters by a technique similar to the classical dynamic time warping (DTW) approach. Only recently, two similar DTW-based synchronization algorithms have been proposed: Turetsky et al. [27] first convert the score-data stream into an audio-data stream using a suitable synthesizer and perform the alignment in the audio domain. Soulez et al. [28] use the score data to design a sequence of suitable filter models which can then be compared with the audio data stream. In contrast to these two approaches, we perform the synchronization purely in the score-like domain which has advantages in view of both efficiency and accuracy.

However, due to the complexity and diversity of music data the problem of automatic music alignment is still far from being solved — not only concerning the data format but also concerning the genre (e.g., pop music, classical music, jazz), the instrumentation (e.g., orchestra, piano, drums, voice), and many other parameters (e.g., dynamics, tempo, or timbre). For the future it seems promising to devise a system incorporating multiple competing strategies (instead of relying on one single strategy) in combination with statistical methods as well as explicit instrument models in order to cope with the richness and variety of music.

6 Conclusions and Future Work

In this paper, we discussed recent advances in the field of digital music libraries. We focused on the important aspect of content-based retrieval and gave an overview on some of the techniques which have been developed in our MiDiLiB-project. More precisely, we described a set-based technique for searching scores of polyphonic music by content. Subsequently, we showed how this technique may be exploited in melody-based retrieval, e.g., name-that-tune applications, as well as efficient audio identification and monitoring. The underlying general technique is not restricted to the field of music retrieval, but may be used for searching in general collections of multimedia documents by content [6] Another important aspect of the MiDiLiB-project are algorithms for synchronizing music given in different formats. We sketched an underlying technique and pointed to several applications in the context of digital music libraries.

The last years have seen significant technological progress in content-based music retrieval, where several retrieval tasks such as polyphonic score search and audio identification for the first time became manageable on large scale document collections. Whereas complex score-based search by now is mostly restricted to music experts, existing prototypes for melody search offer a sufficient degree of fault tolerance to make them suitable for a broader class of users.

Based on the technological advances and the large existing collections of music data, the next years will probably see an increasing number of publicly

available (online-) music services. To conclude our paper we briefly sketch such a scenario of a client-server based service for real-time exchange of music-related information. In our scenario, a modified audio player (serving as a client application) during acoustic playback of an audio track receives the track's lyrics from a server application, possibly located within some library. The lyrics may then be displayed synchronously to the actual acoustic playback. Such a service may be realized using techniques for audio indexing, identification and audio-to-text snychronization. For this, an index of audio fingerprints is generated in a proprocessing step. The audio fingerprints (consisting of suitable feature sets) of a certain track are then suitably linked to the corresponding lyrics. During playback of an audio track, fingerprints are extracted from that track. Those are transmitted to the server and then used to determine the actual song and playback position. Subsequently, the corresponding lyrics are transmitted to the client application. Note that in the proposed approach one is not required to make the actual audio tracks publicly available, but works on extracted fingerprints only. Besides efficiency issues this has significant advantages concerning legal issues such as copyright- and content-protection.

Future work will be more and more concerned with developing the latter type of applications. However, in spite of the significant recent advances in the underlying technologies for content-based document processing, there is still much fundamental research work to be done in the field of semantic content analysis.

References

1. Rodger J. McNab et al.: The New Zealand Digital Library MELody inDEX. D-Lib Magazine (1997)
2. Barlow, H., Morgenstern, S.: A Dictionary of Musical Themes. Faber and Faber, London (1991)
3. ISMIR: International Conference on Music Information Retrieval (2001) `http://www.ismir.net/`.
4. Clausen, M., Engelbrecht, R., Meyer, D., Schmitz, J.: PROMS: A Web-based Tool for Searching in Polyphonic Music. In: Proceedings Intl. Symp. on Music Information Retrieval 2000, Plymouth, M.A., USA. (2000)
5. Clausen, M., Kurth, F.: A Unified Approach to Content-Based and Fault Tolerant Music Recognition (2003) IEEE Transactions on Multimedia, Accepted for Publication.
6. Clausen, M., Körner, H., Kurth, F.: An Efficient Indexing and Search Technique for Multimedia Databases. In: SIGIR Workshop on Multimedia Retrieval, Toronto, Canada. (2003)
7. Selfridge-Field, E., ed.: Beyond MIDI: The Handbook of Musical Codes. MIT Press (1997)
8. Ukkonen, E., Lemström, K., Mäkinen, V.: Geometric Algorithms for Transposition Invariant Content-Based Music Retrieval. In: International Conference on Music Information Retrieval, Baltimore. (2003)
9. Doraisamy, S., Rüger, S.: Robust Polyphonic Music Retrieval with N-grams. Journal of Intelligent Information Systems **21** (2003) 53–70

10. Pickens, J., Bello, J.P., Monti, G., Crawford, T., Dovey, M., Sandler, M., Byrd, D.: Polyphonic Score Retrieval Using Polyphonic Audio. In: International Conference on Music Information Retrieval, Paris. (2002)
11. Typke, R., Giannopoulos, P., Veltkamp, R.C., Wiering, F., van Oostrum, R.: Using transportation distances for measuring melodic similarity. In: International Conference on Music Information Retrieval, Baltimore. (2003)
12. Kurth, F., Clausen, M., Ribbrock, A.: Efficient Fault Tolerant Search Techniques for Full-Text Audio Retrieval. In: Proc. 112th AES Convention, Munich, Germany. (2002)
13. Pauws, S.: CubyHum: a fully operational query by humming system. In: International Conference on Music Information Retrieval, Paris. (2002)
14. Heinz, T., Brückmann, A.: Using a Physiological Ear Model for Automatic Melody Transcription and Sound Source Recognition. In: Proc. 114th AES Convention, Amsterdam, Netherlands. (2003)
15. Wold, E., Blum, T., Kreislar, D., Wheaton, J.: Content-based classification, search, and retrieval of audio. IEEE Multimedia **3** (1996) 27–36
16. Ribbrock, A., Kurth, F.: A Full-Text Retrieval Approach to Content-Based Audio Identification. In: Proc. 5. IEEE Workshop on MMSP, St. Thomas, Virgin Islands, USA. (2002)
17. Kurth, F., Clausen, M., Ribbrock, A.: Identification of Highly Distorted Audio Material for Querying Large Scale Data Bases. In: Proc. 112th AES Convention, Munich, Germany. (2002)
18. Kurth, F., Scherzer, R.: Robust Real-Time Identification of PCM Audio Sources. In: Proc. 114th AES Convention, Amsterdam, Netherlands. (2003)
19. Fragoulis, D., Roussopoulos, G., Panagopoulos, T., Alexiou, C., Papaodysseus, C.: On the automated recognition of seriously distorted musical recordings. IEEE Trans. SP **49** (2001) 898–908
20. Haitsma, J., Kalker, T.: A Highly Robust Audio Fingerprinting System. In: Proc. ISMIR 2002. (2002)
21. Wang, A.: An Industrial Strength Audio Search Algorithm. In: International Conference on Music Information Retrieval, Baltimore. (2003)
22. Cano, P., Battle, E., Mayer, H., Neuschmied, H.: Robust Sound Modeling for Sound Identification in Broadcast Audio. In: Proc. 112th AES Convention, Munich, Germany. (2002)
23. Allamanche, E., Herre, J., Fröba, B., Cremer, M.: AudioID: Towards Content-Based Identification of Audio Material. In: Proc. 110th AES Convention, Amsterdam, NL. (2001)
24. Cano, P., Battle, E., Kalker, T., Haitsma, J.: A Review of Audio Fingerprinting. In: Proc. 5. IEEE Workshop on MMSP, St. Thomas, Virgin Islands, USA. (2002)
25. Tzanetakis, G., Gao, J., Steenkiste, P.: A Scalable Peer-to-Peer System for Music Content and Information Retrieval. In: International Conference on Music Information Retrieval, Baltimore. (2003)
26. Arifi, V., Clausen, M., Kurth, F., Müller, M.: Synchronization of Music Data in Score-, MIDI- and PCM-Format. In Hewlett, W.B., Selfridge-Fields, E., eds.: Computing in Musicology. MIT Press, accepted for publication (2003)
27. Turetsky, R.J., Ellis, D.P.: Force-Aligning MIDI Syntheses for Polyphonic Music Transcription Generation. In: International Conference on Music Information Retrieval, Baltimore, USA. (2003)
28. Soulez, F., Rodet, X., Schwarz, D.: Improving polyphonic and poly-instrumental music to score alignment. In: International Conference on Music Information Retrieval, Baltimore. (2003)