A Framework for Managing Multimodal Digitized Music Collections

Frank Kurth[1], David Damm[2], Christian Fremerey[2], Meinard Müller[3], and Michael Clausen[2]**

[1] Research Establishment for Applied Science (FGAN), FKIE-KOM
Neuenahrer Strasse 20, 53343 Wachtberg, Germany,
kurth@fgan.de
[2] Department of Computer Science III, University of Bonn,
Römerstraße 164, 53117 Bonn, Germany
{damm,fremerey,clausen}@iai.uni-bonn.de
[3] Max-Planck-Institut für Informatik, Department D4 - Computer Graphics,
66123 Saarbrücken, Germany,
mmueller@mpi-inf.mpg.de

Abstract. In this paper, we present a framework for managing heterogeneous, multimodal digitized music collections containing visual music representations (scanned sheet music) as well as acoustic music material (audio recordings). As a first contribution, we propose a preprocessing workflow comprising feature extraction, audio indexing, and music synchronization (linking the visual with the acoustic data). Then, as a second contribution, we introduce novel user interfaces for multimodal music presentation, navigation, and content-based retrieval. In particular, our system offers high quality audio playback with time-synchronous display of the digitized sheet music. Furthermore, our system allows a user to select regions within the scanned pages of a musical score in order to search for musically similar sections within the audio documents. Our novel user interfaces and search functionalities will be integrated into the library service system of the Bavarian State Library as part of the Probado project.

1 Introduction

Recently, significant digitization efforts have been carried out for large collections of books and other types of printed documents. These efforts naturally lead to the need for powerful tools that automatically process, analyze, and annotate the scanned documents, which provides the basis for efficient and effective content-based searching, navigation, and browsing in the digitized data. In the case of scanned text documents, various solutions for automated document processing have been proposed [1], which typically contain a component for optical character recognition (OCR) to extract the textual content from the images as well as

** This work was supported in part by Deutsche Forschungsgemeinschaft (DFG) under grant 554975 (1) Oldenburg BIB48 OLof 01-02.

a component for fault tolerant full-text indexing [2] and information retrieval [3]. Here, note that recent systems can cope with possible extraction errors, which are due to the error-prone OCR step by employing fault tolerant text-retrieval techniques [4, 5]. Furthermore, the systems present the user high-quality scans of the text passages of interest, while using the error-prone OCR extraction results only for the internal processing stages of analysis and retrieval. Here, the general idea is to suitably combine the strengths of both types of data representations (scan and text) for convenient navigation in the scanned documents. A well-known example is the Google Book Search project [6]. In spite of these advances in the textual domain, there is still a significant lack of corresponding solutions for handling general digitized material including images, videos, 3D graphical data, or audio data. Particularly, tools are needed to automatically extract semantically meaningful entities (or regions of interest) from the scanned documents and to create links between related entities.

In this paper, we focus on the particular scenario of heterogeneous music collections, which contain digitized sheet music (image data) as well as CD recordings (audio data). More precisely, for each musical work in our collection, we assume the availability of both a complete set of scanned pages of the musical score as well as at least one audio recording of the musical work. As those documents concern both the visual and the auditorial modalities, the document collection may be called *multimodal.* We now give an overview of our integrated framework for automatically managing such a multimodal music collection. Our framework comprises the three layers depicted in Fig. 1.

Repository Layer. This layer consists of the digitized music data including the scanned sheet music and the audio recordings. Furthermore, the layer comprises tools for automatically analyzing both types of data. In particular, these tools are used for feature extraction (including OMR and audio processing), for audio indexing, as well as for the synchronization (or *alignment*) of sheet music and audio data, where the objective is to link 2D regions (measured in pixels) within the scanned pages to semantically corresponding temporal segments (measured in seconds) within an audio recording [7].

User Interface Layer. This layer comprises various user interfaces for multimodal music access and interaction. Besides offering standard functionalities for acoustic audio playback, file handling, and playback controls, the *Score Viewer Interface* synchronously presents musically corresponding regions of the scanned sheet music during audio playback. The interface can be employed for marking regions within the digitized score by simply using the mouse pointer, which can then be used as query to start content-based music search. The retrieved audio documents are then presented in the *Result Viewer Interface*, which provides functionalities for navigating to and within these documents.

Server Layer. This layer connects the Repository- and the User Interface Layers. On the one hand, the Server Layer has direct access to all data contained and generated in the Repository Layer. On the other hand, it receives and handles the various requests from the user interfaces. The core component of the

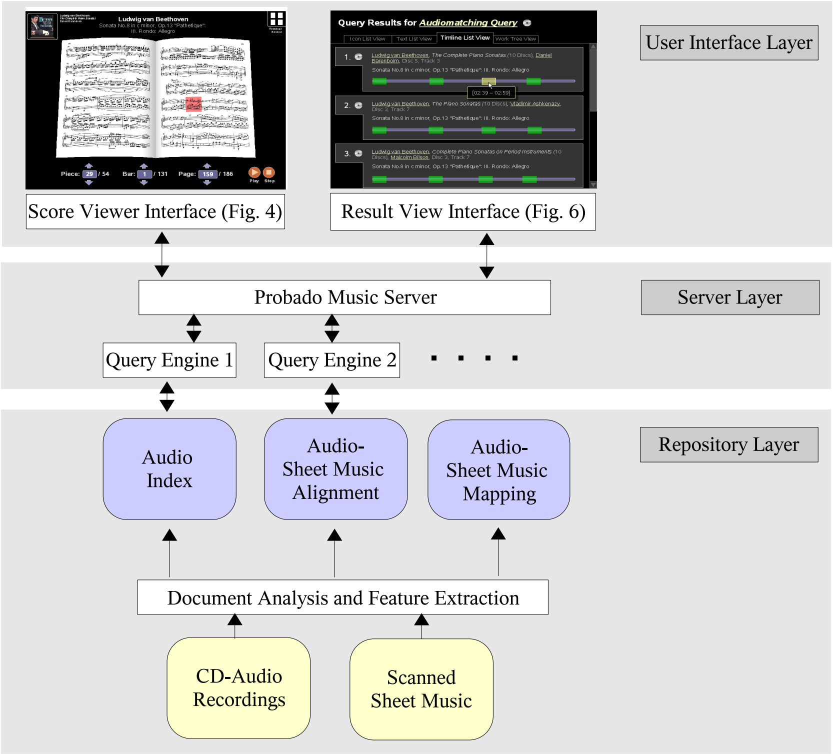

Ludwig van Beethoven
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro
Piece: 29 / 54
Bar: 1 / 131
Page: 159 / 186
Play
Stop
Query Results for Audiomatching Query
Icon List View
Text List View
Timline List View
Work Tree View
1. Ludwig van Beethoven, The Complete Piano Sonatas (10 Discs), Daniel Barenboim, Disc 5, Track 3
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro
[02:39 – 02:59]
2. Ludwig van Beethoven, The Piano Sonatas (10 Discs), Vladimir Ashkenazy, Disc 3, Track 7
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro
3. Ludwig van Beethoven, Complete Piano Sonatas on Period Instruments (10 Discs), Malcolm Bilson, Disc 3, Track 7
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro
Score Viewer Interface (Fig. 4)
Result View Interface (Fig. 6)
User Interface Layer
Probado Music Server
Server Layer
Query Engine 1
Query Engine 2
. . . .
Audio Index
Audio-Sheet Music Alignment
Audio-Sheet Music Mapping
Repository Layer
Document Analysis and Feature Extraction
CD-Audio Recordings
Scanned Sheet Music

Fig. 1. Digital library framework for managing multimodal music collections containing scanned sheet music and audio recordings. The *User Interface Layer* (top) connects to the *Server Layer* (middle), which in turn accesses the *Repository Layer* (bottom) via suitable query engines. The Repository Layer consists of both the digitized documents and supplementary information including previously generated audio features, music alignments, annotations, and indexes.

Server Layer is the Probado Music Server. Its task is to handle the communication with the User Interface Layer and to schedule incoming queries to a set of available query engines. Each query engine offers a particular functionality, e.g. content-based retrieval using audio matching or delivery of data via streaming.

Note that the processing tasks within the Repository Layer are carried out *offline*, while the tasks within the Sever Layer and User Interface Layer are performed *online*.

The subsequent sections of this paper are organized as follows. In Sect. 2, we introduce our workflow for automatically processing multimodal music collections. In particular, we describe how to temporally align and match scanned pages of sheet music with corresponding audio material. As a further important contribution of this paper, we introduce two novel user interfaces (Score Viewer

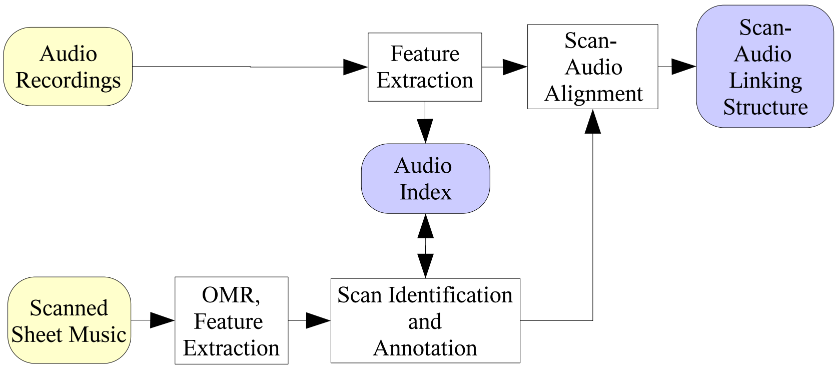


Fig. 2. Overview on the workflow of automatic document processing for the two given collections of scanned sheet music and audio recordings.

Interface, Result Viewer Interface) that facilitate multimodal music access and interaction (Sect. 3). In particular, we discuss in detail the functionalities of the User Interface Layer as mentioned above. Our framework will be integrated in the ongoing Probado digital library initiative [8] and is put into practice at the Bavarian State Library (BSB) in Munich to supplement the existing library services. In Sect. 4, we give details on the current real-world test repository provided by the BSB. Furthermore, we discuss open problems, prospects on future work, and further extensions of our framework.

2 Automatic Document Processing

In this section, we describe the underlying methods that are needed to process, match, and align the various types of music data, see Fig. 2 for an overview. The basic idea is to transform both the scanned images as well as the audio recordings into a common feature representation, which then allows for a direct comparison of the two different types of data. In this context, chroma-based music features have turned out to be a powerful mid-level music representation [9–11]. In Sect. 2.1, we describe the steps required for transforming the audio documents as well as the sheet music documents into chroma representations. In our system, the extracted features are organized by means of a suitable index structure, which can then be used for efficient music matching tasks (Sect. 2.2). Furthermore, we introduce a novel mechanism that allows for identifying and annotating scanned pages of sheet music by means of available annotated audio material (Sect. 2.3). Finally, in Sect. 2.4, we summarize our music synchronization procedure for generating the alignments that are needed for visualization and synchronous playback in the user interface layer.

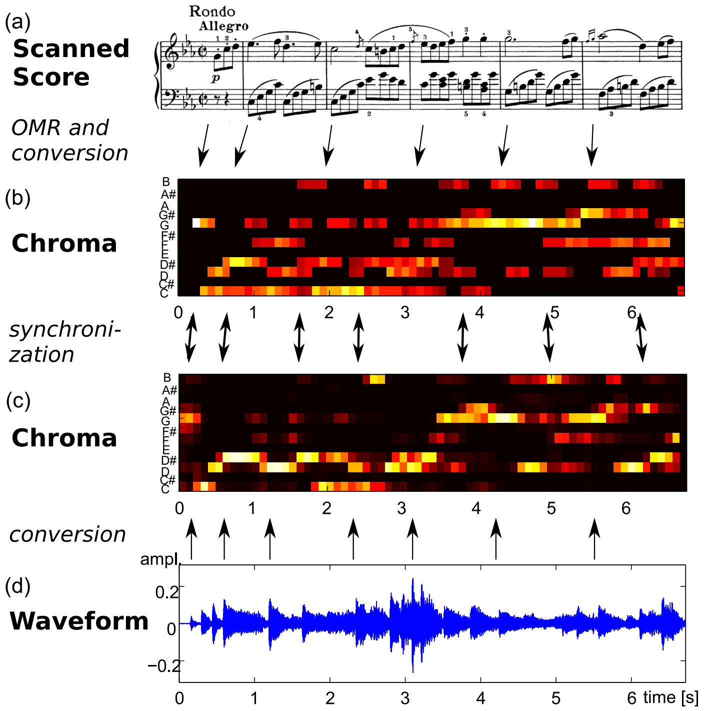


Fig. 3. Data types involved in automatic document processing for the first few measures of Beethoven's Piano Sonata No. 8, Op. 13 "Pathethique", Rondo (3rd movement). **(a)** Scanned sheet music. **(b)** Sheet music chromagram. **(c)** Audio chromagram. **(d)** Audio recording (waveform). The scan-audio linking structure (double-headed arrows) is obtained by aligning the two chromagrams, see Sect. 2.4.

2.1 Chroma-based Music Features

In order to compare and relate music data of various types and formats, one needs to find a suitable feature representation satisfying several critical requirements. One the one hand, the feature representation has to be robust to semantic variations and transformation errors. Furthermore, the various types of data should be reducible to the same representation. On the other hand, the feature representation has to be characteristic enough to capture distinctive musical aspects of the underlying piece of music. Chroma-based features constitute a good tradeoff between these—to some extent conflicting—requirements [9–11]. The *chroma* correspond to the twelve traditional pitch classes of the equal-tempered scale and are commonly indicated by the twelve pitch spelling attributes $\mathrm{C}, \mathrm{C}^\sharp, \mathrm{D}, \ldots, \mathrm{B}$ as used in Western music notation. Chroma-based features account for the well-known phenomenon that human perception of pitch is periodic in the sense that two pitches are perceived as similar in "color" if they differ by an octave [9].

In the case of CD audio recordings, normalized chroma-based features indicate the short-time energy distribution among the twelve chroma and closely correlate to the harmonic progression of the underlying piece. Based on signal processing techniques, the transformation of an audio recording into a chroma representation (or chromagram) can be done either by using short-time Fourier transforms in combination with binning strategies [9] or by employing suitable multirate filter banks [11]. For the technical details, we refer to the literature. Fig. 3 (c) shows an audio chromagram for the first few measures of a recording (d) of the 3rd movement of Beethoven's Piano Sonata No. 8, Op. 13 ("Pathethique"),

The transformation of scanned sheet music into a chromagram requires several steps, see [7]. First, each scanned page is analyzed using optical music recognition (OMR) [12, 13]. In our system, we use the commercially available SharpEye software [14] to extract musical note parameters (onset times, pitches, durations) along with 2D position parameters as well as bar line information from the scanned image. Then, using this explicit pitch and timing information, a chromagram can be computed essentially by identifying pitches that belong to the same chroma class, see [10] for details. Fig. 3 (b) shows a chromagram obtained from a scanned score (a) of the "Pathethique". Note that a sheet music chromagram is, in general, much "cleaner" than an audio chromagram, see Fig. 3. However, the OMR software often produces serious note extraction errors, which are only partially absorbed by the chroma features. For the test collection of piano sonatas considered in our experiments (see Sect. 4) it turns out that the OMR quality is in most cases good enough to obtain reasonable matching and synchronization results in the subsequent processing stages.

2.2 Audio Indexing and Audio Matching

The key idea we exploit for automatic document analysis is that reducing the two different types of data (visual and acoustic music data) to the same type of representation (chromagram) allows for a *direct* comparison on the feature level *across* the two domains. To also allow for an *efficient* comparison, we further process the chroma features by quantizing the chroma vectors using semantically meaningful codebook vectors, see [15] for details. According to the assigned codebook vectors, the features can then be stored in some inverted file index, which is a well-known index structure that is frequently used in standard text retrieval [2].

In our system, we employ audio matching as described in [15] as an underlying engine for the various music retrieval and identification tasks. The basic matching approach works as follows. Each music document of the repository is converted into a sequence of 12-dimensional chroma vectors. In our implementation, we use a feature sampling rate of 1 Hz. While keeping book on document boundaries, all these chroma sequences are concatenated into a single sequence $(d_0, \ldots, d_{K-1})$ of chroma features. Similarly, a given query music clip is also transformed into a sequence $(q_0, \ldots, q_{L-1})$ of chroma features. This query sequence is then compared with all subsequences $(d_k, d_{k+1}, \ldots, d_{k+L-1})$, $k \in [0 : K - L]$, consisting of L consecutive vectors of the database sequence. Here, we use the distance measure $\Delta(k) := 1 - \frac{1}{L}\sum_{\ell=0}^{L-1}\langle d_{k+\ell}, q_\ell\rangle$, where the brackets denote the inner vector product. The resulting curve Δ is referred to as *matching curve*. Note that the local minima of Δ close to zero correspond to database subsequences that are similar to the query sequence. Those subsequences will constitute the desired *matches* for content-based retrieval as described in Sect. 3. Because of the bookkeeping, document numbers and positions of matches within each document can be recovered easily. To account for possible temporal differences between the query clip and corresponding temporal regions within the documents (e.g., think of tempo differences between different interpretations of the same piece of music), we employ the technique of multiple querying with various chromagrams at different sampling rates. Another important point is that the matches can be computed efficiently using standard text retrieval algorithms based on the above mentioned index structure. For the technical details, we refer to [15].

2.3 Scan Identification and Annotation

After the digitization process, the digitized documents need to be suitably annotated before they can be integrated into the holding of a digital library. In the case of digitized audio recordings, one has to assign metadata such as *title*, *artist*, or *lyrics* to each individual recording. Besides the labor and cost intensive option of manual annotation, one may exploit several available databases that specialize on various types of metadata such as Gracenote [16] or DE-PARCON [17].

For annotating the scanned sheet music, we now introduce a novel automated procedure that, to the best of our knowledge, has not yet been described in the literature before. In our scenario, we assume the existence of an audio database containing annotated digitized audio recordings for all pieces to be considered in the sheet music digitization process. We then automate the annotation of the scanned pages as follows (see also Fig. 2). In a preprocessing step, we transform the audio documents into sequences of chroma vectors and build up an audio index structure. Then, in the annotation step, each scanned page is converted into a sequence of chroma vectors. Using this sequence as a query, we compute the top match within the audio documents as described in Sect. 2.2. Assuming that each page is contained in a single musical work, the top match may usually be expected to lie within a musically corresponding audio recording. As first experiments show, this particularly holds in case that there are no severe OMR errors. In other words, the scanned page can be identified by the top match and can then be automatically annotated by the metadata already attached to the corresponding audio recording. In the presence of severe OMR errors or in the case a page does not correspond to a single piece of music (occasionally, a single page contains both the end and the beginning of two consecutive movements), this procedure frequently fails. To overcome this problem and nevertheless obtain correct matching results, one can exploit the fact that subsequently scanned pages most likely result in subsequent matching regions within an audio recording. We also expect that the improvement of the OMR results and the simultaneous usage of different OMR procedures will significantly

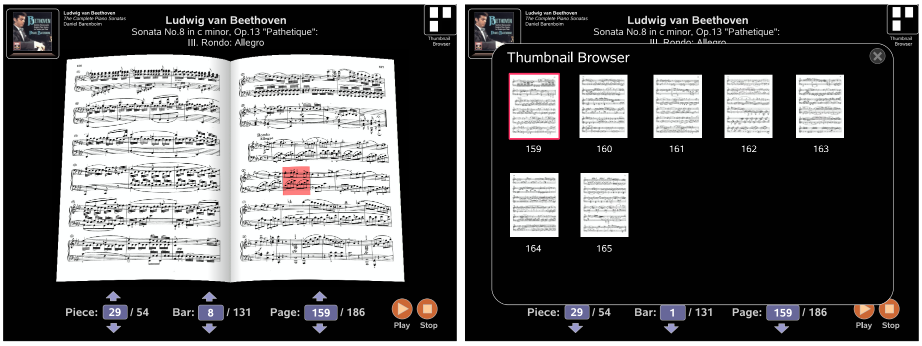


Fig. 4. The Score Viewer Interface for multimodal music presentation and navigation. Synchronously to audio playback, corresponding musical measures within the sheet music are highlighted (left). The Thumbnail Browser (right) allows to conveniently nagivate through the currently selected score.

improve the matching quality [13]. Technical details on this novel mapping and annotation procedure will be reported elsewhere.

2.4 Scan-Audio Alignment

Once having identified scanned pages of sheet music and corresponding audio recordings, one can automatically link semantically related events across the two types of music representation. To this end, we employ music synchronization techniques [10, 11, 18, 19] to link regions (given as pixel coordinates) within the scanned images of given sheet music to semantically corresponding time positions within an audio recording. Such a procedure has been described in [7]. The basic idea is to convert both the scanned pages as well as the corresponding audio recording into sequences of chroma features which can then be synchronized based on standard alignment techniques such as dynamic time warping [11]. In case of the scanned pages, we exploit the identification results described in Sect. 2.3 to construct the feature sequence for the underlying music document by appropriately grouping the features obtained from the individual pages. An example of the discussed scan-audio synchronization is shown in Fig. 3, where the resulting linking structure is indicated by the double-headed arrows. The importance of such linking structures has been emphasized in the literature [19]. In Sect. 3, we will introduce user interfaces that exploit the latter scan-audio alignments in order to facilitate multimodal music navigation and offer a suitable music presentation.

3 Score- and Result Viewer Interfaces

The central component for presenting sheet music and associated audio recordings to the user is the *Score Viewer Interface* depicted in Fig. 4. To the left, the

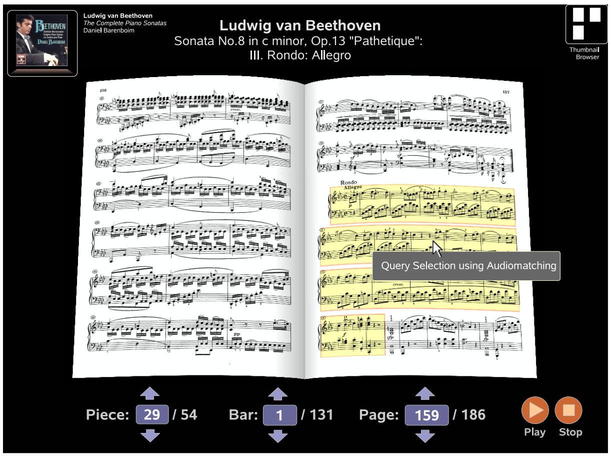
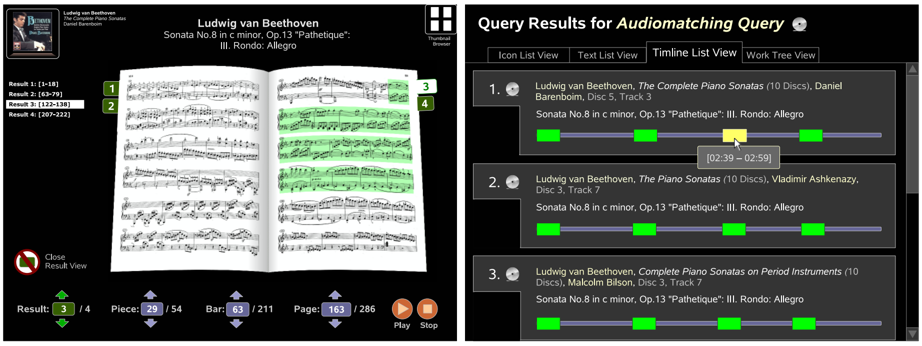


Fig. 5. Using the Score Viewer Interface for content-based retrieval. A query is selected by marking a region of measures within the score.

main visualization mode is illustrated for two scanned pages of the above example, Beethovens Piano Sonata No. 8, Op. 13 (“Pathethique”). When starting audio playback, corresponding measures within the sheet music are synchronously highlighted based on the linking structure generated by the scan-audio alignment described in Sect. 2.4. In Fig. 4, a region in the center of the right page, corresponding to the eighth measure of the 3rd movement (Rondo), is currently highlighted by a surrounding box. When reaching the end of odd-numbered pages during playback, pages are turned over automatically. Additional control elements allow the user to switch between measures of the currently selected piece of music. The Score Viewer Interface manages entire scanned scorebooks and hence also allows to navigate through those books using piece- or page numbers that are located below the scanned pages. Using the *Thumbnail Browser* shown on the right of Fig. 4, a local context of pages around the current playback position is displayed and may be used for navigation. An icon in the top left corner indicates which CD is currently used for audio playback. If more than one recording is available for the currently active piece of music, the user may switch between those using an icon list that is available by clicking on the current icon.

A user friendly functionality for music retrieval based on the query-by-example paradigm has been integrated in the Score Viewer Interface. More precisely, the user is enabled to select specific regions within the sheet music using the mouse pointer. By right-clicking on the selected region, a query may be issued. As an example, consider Fig. 5 where the first 17 measures of the Beethoven Rondo are selected as a query. Exploiting the linking structure generated by the scan-audio alignment, the selected sheet music region is assigned to the corresponding time interval of the audio recording identified in the preprocessing stage. In our example, the first 17 measures of the Rondo correspond to seconds 1-20 of an



Ludwig van Beethoven
The Complete Piano Sonatas
Daniel Barenboim

Ludwig van Beethoven
Sonata No.8 in c minor, Op.13 "Pathetique":
III. Rondo: Allegro

Result 1: [1-18]
Result 2: [63-79]
Result 3: [122-138]
Result 4: [207-222]

Result: 3 / 4 Piece: 29 / 54 Bar: 63 / 211 Page: 163 / 286

Query Results for *Audiomatching Query*

1. Ludwig van Beethoven, *The Complete Piano Sonatas* (10 Discs), Daniel Barenboim, Disc 5, Track 3
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro
[02:39 – 02:59]

2. Ludwig van Beethoven, *The Piano Sonatas* (10 Discs), Vladimir Ashkenazy, Disc 3, Track 7
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro

3. Ludwig van Beethoven, *Complete Piano Sonatas on Period Instruments* (10 Discs), Malcolm Bilson, Disc 3, Track 7
Sonata No.8 in c minor, Op.13 "Pathetique": III. Rondo: Allegro

Fig. 6. Audio recordings containing matches to a query are listed in the Result Viewer Interface (right). In the Timeline List View, all matches within the same document are represented by markers on a horizontal bar, thus indicating their temporal positions. Upon selecting a match the correponding sheet music pages are displayed in the Score Viewer Interface and acoustic playback is started.

interpretation by Barenboim. A sequence of chroma features is then extracted from the audio recording. Subsequently, audio matching as described in Sect. 2.2 is used to query the feature sequence to the audio index. Note that as the query features are taken from an audio recording, they are not affected by possible OMR errors.

Following the audio matching step performed by the Server Layer, the documents containing matches to a query are listed in the Result Viewer Interface (Fig. 6, right). The Result Viewer offers several different display types. In the Timeline List View shown in Fig. 6, all matches within a single document are represented by small rectangles along a horizontal bar indicating their temporal positions within the document. The listed documents are ranked according to their respective highest-scoring matches. In the alternative Icon- and Text List Views, individual matches are listed as score thumbnails (icons) and in plain text format, respectively. The Work Tree View provides a tree-like representation of the matches in the context of a hierarchy of musical works [20] based on the FRBR-model (Functional Requirements for Bibliographic Records) [21].

When querying the above few measures of the Rondo, the top matches are contained in three different interpretations (by Barenboim, Ashkenazy, and Bilson) of the Rondo. The Timeline List View shown in Fig. 6, indicates that the 12 matches contained in these three pieces exactly correspond to the respective four occurences of the queried musical theme within the recordings.

Upon selecting an individual match, the corresponding sheet music region is highlighted in the Score Viewer, Fig. 6 (left) and synchronous acoustic playback is started. The Score Viewer Interface shows the other matches within the same audio recording in a list on the left of the document. In our running example, the four matches contained in the Barenboim interpretation are given. Index numbers displayed to the left and the right of the scanned pages may be used to select each individual match. In Fig. 6 (left), the third match that musically corresponds to the recapitulation of the query theme is currently selected. Forward and backward navigation through the matches is possible using additional controls in the bottom left corner.

4 Conclusions and Future Work

In this paper, we presented a digital library framework for managing collections of scanned sheet music and associated audio recordings. Starting with a workflow for document processing, our technical contributions concern methods for automatically identifying pages of scanned sheet music and subsequent alignment of sheet music to audio recordings. To facilitate multimodal music navigation, we presented the Score Viewer Interface for time-synchronous display and playback of the scanned sheet music and corresponding audio recordings. We additionally presented a novel mechanism for content-based music retrieval by directly selecting query regions from the sheet music. Query results are displayed in the Result Viewer Interface allowing for audio browsing and multimodal navigation using the Score Viewer.

Our current test collection consists of the 32 piano sonatas (101 audio files, mostly individual movements) by Ludwig van Beethoven. For each of those pieces, a scanned version of the corresponding sheet music taken from an edition by G. Henle Verlag is available, amounting to a total number of 604 scanned pages. For each movement, at least one audio recording is available. The identification rate for individual scanned pages using chroma-based music matching is 82.5%. As discussed in Sect. 2.3, we expect that one can significantly increase this rate by postprocessing the OMR data prior to the matching procedure.

The proposed workflow and user interfaces are part of the Probado music repository currently set up at Bavarian State Library in Munich, Germany. For this music repository, an even larger music collection of classical and romantic piano sonatas (Haydn, Mozart, Beethoven, Schubert, Schumann, Chopin, Liszt, Brahms) as well as a collection of German 19th centuries piano songs has been digitized, amounting to about 6.000 pages of scanned sheet music and 1.200 audio recordings [20].

Although the proposed technical workflow for automatic document processing is fully functional, there are yet restrictions regarding the underlying music material. Most important, we assume that the musical structure of the scanned sheet music is in perfect correspondence with the associated audio recordings. Particularly, we do not yet deal with repetitions of particular parts of a musical work which are present in one representation of the piece of music (e.g. the audio recording) but not in the other (e.g. the scanned sheet music). Methods for handling such types of structural differences involving *partial music synchronisation* are currently investigated and will be reported elsewhere.

References

1. Wang, P., Bunke, H.: Handbook on Optical Character Recognition and Document Image Analysis. World Scientific (1997)
2. Witten, I.H., Moffat, A., Bell, T.C.: Managing Gigabytes. 2nd edn. Van Nostrand Reinhold (1999)
3. Baeza-Yates, R.A., Ribeiro-Neto, B.A.: Modern Information Retrieval. ACM Press, Addison-Wesley (1999)
4. Ohta, M., Takasu, A., Adachi, J.: Retrieval methods for english-text with missrecognized ocr characters. In: ICDAR '97: Proceedings of the 4th International Conference on Document Analysis and Recognition, Washington, DC, USA, IEEE Computer Society (1997) 950–956
5. Harding, S.M., Croft, W.B., Weir, C.: Probabilistic retrieval of ocr degraded text using n-grams. In: ECDL '97: Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries, London, UK, Springer-Verlag (1997) 345–359
6. Google Inc.: Google Book Search (2007) http://books.google.com/.
7. Kurth, F., Müller, M., Fremerey, C., Chang, Y., Clausen, M.: Automated Synchronization of Scanned Sheet Music with Audio Recordings. In: Proc. ISMIR, Vienna, Austria. (September 2007) 261–266
8. Krottmaier, H., Kurth, F., Steenweg, T., Appelrath, H.J., Fellner, D.: PROBADO - A Generic Repository Integration Framework . In: Proceedings of the 11th European Conference on Digital Libraries. (September 2007)
9. Bartsch, M.A., Wakefield, G.H.: Audio thumbnailing of popular music using chroma-based representations. IEEE Trans. on Multimedia **7**(1) (2005) 96–104
10. Hu, N., Dannenberg, R., Tzanetakis, G.: Polyphonic audio matching and alignment for music retrieval. In: Proc. IEEE WASPAA, New Paltz, NY. (October 2003)
11. Müller, M.: Information Retrieval for Music and Motion. Springer (2007)
12. Choudhury, G., DiLauro, T., Droettboom, M., Fujinaga, I., Harrington, B., MacMillan, K.: Optical music recognition system within a large-scale digitization project. In: Proc. ISMIR, Plymouth, MA, USA. (2000)
13. Byrd, D., Schindele, M.: Prospects for improving OMR with multiple recognizers. In: Proc. ISMIR, Victoria, Canada. (2006) 41–46
14. Jones, G.: SharpEye Music Reader (2008) http://www.visiv.co.uk/.
15. Kurth, F., Müller, M.: Efficient Index-based Audio Matching. IEEE Transactions on Audio, Speech, and Language Processing **16**(2) (February 2008) 382–395
16. Gracenote: WWW (2008) http://www.gracenote.com/.
17. Krajewski, E.: DE-PARCON Softwaretechnologie (2008) http://www.de-parcon.de/.
18. Arifi, V., Clausen, M., Kurth, F., Müller, M.: Synchronization of music data in score-, MIDI- and PCM-format. Computing in Musicology **13** (2004)
19. Dunn, J.W., Byrd, D., Notess, M., Riley, J., Scherle, R.: Variations2: Retrieving and using music in an academic setting. Special Issue, Commun. ACM **49**(8) (2006) 53–48
20. Diet, J., Kurth, F.: The Probado Music Repository at the Bavarian State Library. In: Proc. ISMIR, Vienna, Austria. (September 2007) 501–504
21. IFLA Study Group on the Functional Requirements of Bibliographic Records: Functional Requirements for Bibliographic Records; Final Report. Saur, Munich (1998) available at www.ifla.org/VII/s13/frbr/frbr.pdf.