Multimodal Presentation and Browsing of Music

David Damm*, Christian Fremerey*, Frank Kurth†, Meinard Müller‡, Michael Clausen*

*Department of Computer Science III
University of Bonn
53117 Bonn, Germany
{damm,fremerey,clausen}@iai.uni-bonn.de

†Research Establishment
for Applied Science
53343 Wachtberg, Germany
kurth@fgan.de

‡Max-Planck-Institut
für Informatik
66123 Saarbrücken, Germany
meinard@mpi-inf.mpg.de

ABSTRACT

Recent digitization efforts have led to large music collections, which contain music documents of various modes comprising textual, visual and acoustic data. In this paper, we present a multimodal music player for presenting and browsing digitized music collections consisting of heterogeneous document types. In particular, we concentrate on music documents of two widely used types for representing a musical work, namely visual music representation (scanned images of sheet music) and associated interpretations (audio recordings). We introduce novel user interfaces for multimodal (audio-visual) music presentation as well as intuitive navigation and browsing. Our system offers high quality audio playback with time-synchronous display of the digitized sheet music associated to a musical work. Furthermore, our system enables a user to seamlessly crossfade between various interpretations belonging to the currently selected musical work.

Categories and Subject Descriptors

H.5.5 [**Computer Applications**]: Information Interfaces and Presentation—*Sound and Music Computing*

General Terms

Design

1. INTRODUCTION

Recent advances in Music Information Retrieval (MIR) have brought up new possibilities for accessing, browsing and navigating in large collections of digital music. In this paper, we employ recently introduced MIR-techniques for music synchronization to develop novel user interfaces for multimodal presentation, navigation and inter-document browsing of music documents belonging to a musical work based on music synchronization.

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.

A musical work can be described by various formats and document types. Throughout the paper, we distinguish between a *musical work* (in an abstract sense), its *sheet music* representation and a particular *audio recording* (a concrete interpretation). In our scenario, for a musical work we have a sheet music representation as well as associated audio recordings representing different interpretations by various musicians. While sheet music contains musical symbols, which describe the musical content *visually* and abstracts from a concrete realization, i.e. interpretation as well as instrumentation, an audio recording retains a concrete realization of music, which describes the musical content *acoustically.* Note that both sheet music and audio recordings, respectively, may be considered as two natural forms of music representation as they explicitly address the visual and auditorial modalities respectively. For this reason, both of those representations are most widely employed by users for accessing music, and corresponding multimodal user interfaces are of high importance. It turns out that the key challenge in designing such interfaces and in suitably preprocessing the underlying music documents is to find an appropriate common representation for both music modes coming from different domains in order to compare and relate the musical content. Therefore the various document types have to be reducible to the same representation.

In the context of MIR, *music synchronization* denotes a procedure which, for a given position in one representation of a musical work, determines the corresponding position within another representation, e.g. the coordination of score symbols with audio data. Various synchronization tasks have been described in [1, 5, 6, 8]. In the following, we focus on the special scenario of *SheetMusic-Audio synchronization* as well as *Audio-Audio synchronization.* SheetMusic-Audio synchronization is used to align corresponding note events within a sheet music representation of a musical work and time regions within an associated interpretation. Audio-Audio synchronization is used to time-align corresponding note events in various interpretations, i.e. different audio recordings, of a musical work.

In this paper, we present novel user interfaces utilizing these alignments to support a high-quality audio-visual presentation of a musical work and an intuitive navigation by choosing a particular measure within the sheet. Furthermore the user has the option to seamlessly crossfade between different interpretations, which allows for a convenient comparison of two or more interpretations. The *Sheet Music Interface* (see Fig. 2) enables the user enjoying music in a multimodal way. On the one hand, it provides an automatic score tracking while listening to a musical work. Furthermore, the user has the option to navigate within the sheet music and choose a particular measure to change the playback position. The *Interpretation Switcher* (see Fig. 4) allows the user to switch between several interpretations (audio recordings) of the current musical work. A user may listen to one of the selected interpretations and then, at any time during playback, switch to another interpretation. The playback in the target interpretation will start at the position that musically corresponds to the position inside the current interpretation. In other words, a seamless transition from one interpretation to another is performed while maintaining audio playback.

The rest of this paper is organized as follows. In Sect. 2, we give an outline of the involved MIR techniques used for SheetMusic-Audio synchronization as well as Audio-Audio synchronization. Subsequently, in Sect. 3 and 4, we will introduce our new, multimodal user interfaces. In particular, in Sect. 3 we will present the user interface exploiting the SheetMusic-Audio synchronization in order to facilitate multimodal music presentation of a musical work as well as an intuitive navigation. In Sect. 4 we will present the inter-document browser interface utilizing the Audio-Audio synchronization to switch between various interpretations of the current musical work. Conclusions are given in Sect. 5.

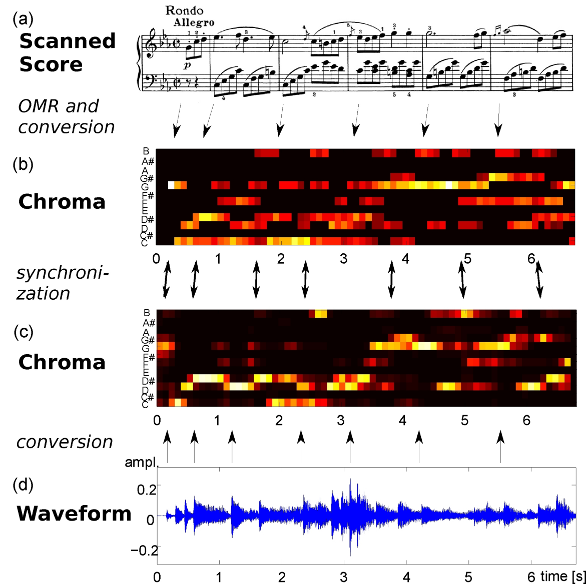

Figure 1: Data types involved in SheetMusic-Audio synchronization for the first few measures of Beethoven's piano sonata no. 8, op. 13 ("Pathethique"), Rondo (3rd movement). (a) Sheet music. (b) Sheet music chromagram. (c) Audio chromagram. (d) Audio recording. The SheetMusic-Audio linking structure (double-headed arrows) is obtained by aligning the two chromagrams.

2. UNDERLYING TECHNIQUES

In this section, we describe our approach to synchronize a given sheet music representation of a musical work with one or more associated audio recordings. The approach is based on the background of techniques from MIR and audio signal processing. In particular, we use chroma-based features [2, 6, 7] as mid-level representation for both, the sheet music as well as the audio recordings, which is then used for the alignment of sheet music to audio as well as audio to audio.

The mid-level feature representation has to satisfy several critical requirements. On the one hand, it has to be robust to semantic variations and transformation errors. On the other hand, it has to be characteristic enough to capture distinctive musical aspects of the underlying musical work. It has been carried out that chroma-based features do achieve these requirements. The chroma correspond to the twelve traditional pitch classes of the equal-tempered scale [2]. In Western music notation, the chroma are commonly indicated by the set $\{C, C^{\#}, \ldots, B\}$ consisting of the twelve pitch spelling attributes. Chroma-based features are well-known to reflect the phenomenon that human perception of pitch is periodic in the sense that two pitches are perceived as similar to the human auditory system if they differ by one or more octaves [2].

For an audio recording, the sample sequence is transformed into a sequence of normalized 12-dimensional chroma vectors. Here, each vector embodies the local energy distribution among the twelve pitch classes. Based on signal processing techniques, a chroma representation can be obtained using either short-time Fourier analysis in combination with binning strategies [2] or multirate filter banks [7]. Fig. 2 (c) shows an audio chromagram for the first few measures of an audio recording (d) of the 3rd movement of Beethoven's piano sonata no. 8, op. 13 ("Pathethique"). Chroma-based features absorb variations in parameters such as dynamics, timbre and articulation and describe the coarse harmony progression of the underlying audio signal. Note that this is needed to compare an audio recording with a sheet music representation due to the absence of these parameters in the latter.

The transformation of a sheet music representation into a chromagram consists of several steps. In a first step, the score data such as note events, i.e. onset times, pitches and durations, the key and time signatures and other musical symbols, are extracted using optical music recognition (OMR) [3, 4]. This process is similar to the well-known optical character recognition (OCR), where textual content is extracted from an image. Note that this process is error-prone and the recognition accuracy of the OMR process strongly depends on the quality of the input image data as well as the complexity of the underlying score. In the context of this paper, we consider high quality scans of piano music at a resolution of 600 dpi and 1 bit color depth (b/w). In addition to the musical score data, the OMR process provides us with spatial information. In particular, we get the exact pixel coordinates of the extracted data as well as bar line information. This allows us to localize all musical symbols within the sheet music. In a second step, a sequence of normalized 12-dimensional chroma vectors is synthesized from the OMR output, which consists of a sequence of the extracted note events, encoded by parameters for pitch, onset time and duration. Fig. 2 (b) shows a chromagram obtained from a sheet music of the "Pathethique". The chromagram derived from a sheet music is computed by sliding across the time axis with a temporal window while adding energy to the chroma bands that correspond to pitches that are active during the current temporal window. A single temporal window equals a single chroma vector. For timing informa-

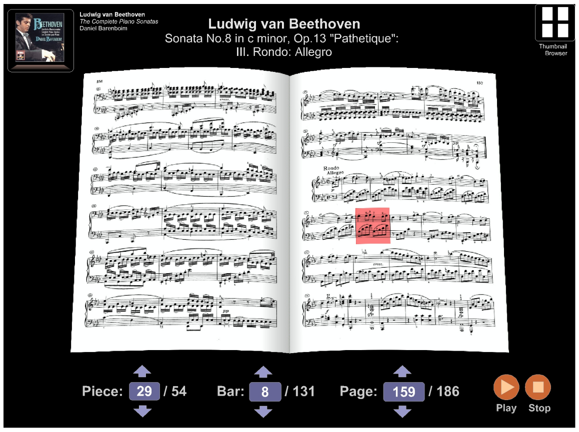

Figure 2: The Sheet Music Interface for multimodal music presentation and navigation. Synchronously to audio playback, corresponding musical measures within the sheet music are highlighted. A click on a measure changes the playback position. Additional control elements on the bottom can be used to navigate through the currently selected musical work.

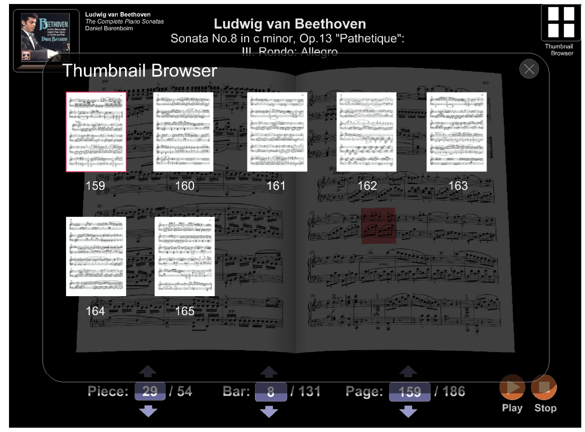

Figure 3: The Thumbnail Browser allows the user to conveniently navigate through the currently selected musical work.

tion we assume a constant tempo of 100 bpm. Note that this particular choice of tempo is not crucial, because differences in tempo will be compensated in the subsequent synchronization step.

Given two chroma sequences extracted from a sheet music and an associated audio recording, respectively, one can use standard algorithms based on dynamic time warping (DTW) to align the two sequences. There have been proposed several synchronization strategies, see, e.g. [1, 6, 8] and the references therein. Most of these approaches rely on some variant of DTW. The main idea is to build up a cross-similarity matrix by computing the pairwise distance between each sheet music chroma vector and each audio chroma vector. Here, we use the inner product for the comparison. An optimum-cost alignment path is determined from this matrix via dynamic programming. The resulting path through the matrix encodes a spatial-temporal alignment of the sheet music and an associated audio recording. The spatial information of the OMR output allows for a localization of chroma vectors within the sheet music image, thus allowing a linkage between groups of note events and corresponding image regions. Combining the spatial information with the synchronization result of a sheet music chroma sequence and an audio chroma sequence, we have all linking information needed to track and highlight note events in a sheet music while playing an associated audio recording.

The same approach used for SheetMusic-Audio synchronization is used to align two or more interpretations of a musical work, which may differ in considerable deviations in tempo, note realization, dynamics and instrumentation. Here, the chroma sequences to align are obtained from the various audio recordings. For a more detailed description, see, e.g. [1, 6, 8].

3. SHEET MUSIC INTERFACE

In this section, we present our new Sheet Music Interface for presenting sheet music while playing back associated audio recordings, depicted in Fig. 2. Here, the main visualization mode is illustrated for two scanned pages of the above example, Beethoven's piano sonata no. 8, op. 13 ("Pathethique"). When starting audio playback, corresponding measures within the sheet music are synchronously highlighted based on the linking information generated by the SheetMusic-Audio alignment described in Sect. 2. In Fig. 2, a region in the center of the right page, corresponding to the eight measure of the 3rd movement (Rondo), is currently highlighted by a surrounding box. When reaching the end of an odd-numbered page during playback, pages are turned over automatically. Additional control elements allow the user to switch between measures of the currently selected musical work. The Sheet Music Interface allows to navigate through the sheets of music using piece- or page-numbers that are located below the scanned pages. By clicking on a measure, the playback position is changed and the audio recording is resumed at the appropriate time position. An icon in the top left corner indicates which interpretation is currently used for audio playback. If more than one associated audio recording is available for the currently active musical work, the user may switch between those using an icon list that is shown by clicking on the current icon. A more detailed description of this feature is given in Sect. 4. Clicking on the icon in the top right corner opens the *Thumbnail Browser*. Using the Thumbnail Browser shown in Fig. 3, a local context of pages around the current playback position is displayed and may be used for navigation.

The technical background underlying the Sheet Music Interface's presentation and navigation capabilities is summarized as follows. The Sheet Music Interface utilizes synchronization data which links image regions within the sheet music and time regions within the selected audio recording. For highlighting regions in a sheet music while playing an associated audio recording, one has to choose a suitable spatial-temporal granularity. To be more robust against local OMR errors we choose a granularity where the displayed region within the sheet music corresponds to one musical measure. This leads to a stable and accurate synchronization of sheet music and an associated audio recording even in the case of typical local OMR errors.

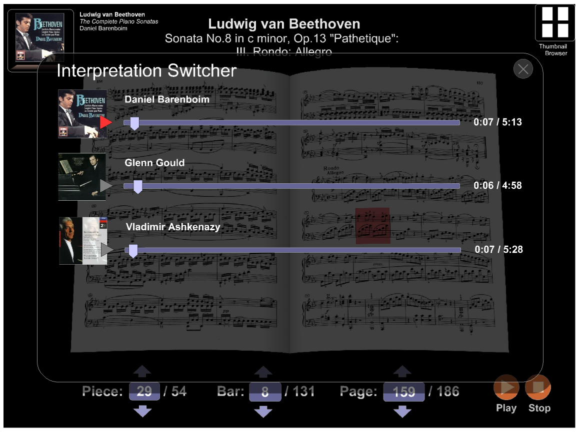

Figure 4: The Interpretation Switcher for a seamless crossfade from one interpretation to another. Synchronously to audio playback, the slider knobs of the various interpretations run along the time line bars. A click on a time line bar's play symbol (left) changes the interpretation used for audio playback.

The Sheet Music Interface described above is a convenient way to enjoy a musical work in a multimodal way. On the one hand, the user can see the sheet music along with the currently played measure highlighted while listening to the musical work. On the other hand, the navigation within the sheet music gives an intuitive modality to search for specific parts and change the playback position respectively.

4. INTERPRETATION SWITCHER

In this section, we present our new Interpretation Switcher for seamless transitions between different interpretations belonging to the currently active musical work, depicted in Fig. 4. In the Interpretation Switcher window, there are listed various available interpretations belonging to the musical work. Each of the interpretations is represented by a slider bar indicating the current playback position with respect to the recording's particular time scale. The interpretation that is currently used for audio playback, in the following referred to as reference recording, is indicated by a red playback symbol located to the left of the slider bar. The slider knob of the reference recording moves at constant speed while the slider knobs of the other recordings move accordingly to the relative tempo variations with respect to the reference. The reference recording may be changed at any time simply by clicking on the respective playback symbol located to the left of each slider bar. The playback of the new reference recording then starts at the time position that musically corresponds to the last playback position of the former reference. This has the effect of seamlessly crossfading from one interpretation to another while preserving the current playback position in a musical sense. One can also jump to any position within any of the recordings by directly selecting a position of the respective slider. This will automatically initiate a switch of reference to the respective recording.

The technical background underlying the Interpretation Switcher's switching capabilities is summarized as follows. The Interpretation Switcher exploits synchronization data which temporally links the selected audio recordings. This data is given as a table of time positions with each column representing one recording and each row representing mutually corresponding time positions. In a preprocessing step, the synchronization data is generated automatically from pairs of audio recordings using the algorithm outlined in Sect. 2.

The Interpretation Switcher described above assists the user in detecting and analyzing the differences between several interpretations of a single musical work. In the example depicted in Fig. 4, the user may switch between three interpretations of piano sonata no. 8, op. 13 ("Pathethique"). Besides the musical differences between the individual interpretations, also differences regarding acoustics, loudness or equalization become apparent.

5. CONCLUSIONS

In this paper, we have introduced a multimodal way for experiencing music. Given the associated sheet music as well as one or more associated interpretations of a musical work, the user has a very intuitive access to the musical content based on the synchronous presentation of two high quality audio-visual music representations. While one listens to an interpretation of the musical work, the listener can visually track the corresponding part within the sheet music representation. Furthermore the visual component offers an intuitive way to search for specific parts within the musical work. For a musical work there may exist more than one interpretation. Our system offers the user to seamlessly crossfade from one interpretation to another while maintaining the current playback position in a musical sense.

6. REFERENCES

[1] V. Arifi, M. Clausen, F. Kurth, and M. Müller. Synchronization of music data in score-, MIDI- and PCM-format. *Computing in Musicology*, 13, 2004.

[2] M. A. Bartsch and G. H. Wakefield. Audio thumbnailing of popular music using chroma-based representations. *IEEE Trans. on Multimedia*, 7(1):96–104, 2005.

[3] D. Byrd and M. Schindele. Prospects for improving OMR with multiple recognizers. In *Proc. ISMIR*, pages 41–46, 2006.

[4] G. Choudhury, T. DiLauro, M. Droettboom, I. Fujinaga, B. Harrington, and K. MacMillan. Optical music recognition system within a large-scale digitization project.

[5] R. Dannenberg and C. Raphael. Music score alignment and computer accompaniment. *Special Issue, Commun. ACM*, 49(8):39–43, 2006.

[6] N. Hu, R. Dannenberg, and G. Tzanetakis. Polyphonic audio matching and alignment for music retrieval. In *Proc. IEEE WASPAA*, 2003.

[7] M. Müller. *Information Retrieval for Music and Motion*. Springer, 2007.

[8] M. Müller, H. Mattes, and F. Kurth. An efficient multiscale approach to audio synchronization. In *Proc. ISMIR*, pages 192–197, 2006.