Linking Sheet Music and Audio – Challenges and New Approaches

Verena Thomas[1], Christian Fremerey[*2], Meinard Müller[†3], and Michael Clausen[4]

1,2,4 University of Bonn, Department of Computer Science III
Römerstr. 164, 53117 Bonn, Germany
{thomas,fremerey,clausen}@cs.uni-bonn.de

3 Saarland University and MPI Informatik
Campus E1-4, 66123 Saarbrücken, Germany
meinard@mpi-inf.mpg.de

Abstract

Score and audio files are the two most important ways to represent, convey, record, store, and experience music. While score describes a piece of music on an abstract level using symbols such as notes, keys, and measures, audio files allow for reproducing a specific acoustic realization of the piece. Each of these representations reflects different facets of music yielding insights into aspects ranging from structural elements (e. g., motives, themes, musical form) to specific performance aspects (e. g., artistic shaping, sound). Therefore, the simultaneous access to score and audio representations is of great importance. In this paper, we address the problem of automatically generating musically relevant linking structures between the various data sources that are available for a given piece of music. In particular, we discuss the task of sheet music-audio synchronization[1] with the aim to link regions in images of scanned scores to musically corresponding sections in an audio recording of the same piece. Such linking structures form the basis for novel interfaces that allow users to access and explore multimodal sources of music within a single framework. As our main contributions, we give an overview of the state-of-the-art for this kind of synchronization task, we present some novel approaches, and indicate future research directions. In particular, we address problems that arise in the presence of structural differences and discuss challenges when applying optical music recognition to complex orchestral scores. Finally, potential applications of the synchronization results are presented.

1998 ACM Subject Classification H.5.1 Multimedia Information Systems, H.5.5 Sound and Music Computing, I.5 Pattern Recognition, J.5 Arts and Humanities–Music

Keywords and phrases Music signals, audio, sheet music, music synchronization, alignment, optical music recognition, user interfaces, multimodality

Digital Object Identifier 10.4230/DFU.Vol3.11041.1

1 Introduction

Significant advances in data storage, data acquisition, computing power, and the worldwide web are among the fundamental achievements of modern information technology. This

* Christian Fremerey is now with Steinberg Media Technologies GmbH, Germany.
† Meinard Müller has been supported by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI). He is now with Bonn University, Department of Computer Science III, Germany.
[1] We use the term *sheet music* as equivalent to scanned images of music notation while *score* refers to music notation itself or symbolic representations thereof.

Multimodal Music Processing. *Dagstuhl Follow-Ups, Vol. 3.* ISBN 978-3-939897-37-8.
Editors: Meinard Müller, Masataka Goto, and Markus Schedl; pp. 1–22

Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

technological progress opened up new ways towards solving problems that appeared nearly unsolvable fifty years ago. One such problem is the long-term preservation of our cultural heritage. Libraries, archives, and museums throughout the world have collected vast amounts of precious cultural material. The physical objects are not only difficult to access, but also threatened from decay. Therefore, numerous national and international digitization initiatives have been launched with the goal to create digital surrogates and to preserve our cultural heritage.[2] However, generating and collecting digitized surrogates represents only the beginning of an entire process chain that is needed to avoid digital graveyards. To make the digitized data accessible, one requires automated methods for processing, organizing, annotating, and linking the data. Furthermore, intuitive and flexible interfaces are needed that support a user in searching, browsing, navigating, and extracting useful information from a digital collection.

In this paper, we address this problem from the perspective of a digital music library project,[3] which has digitized and collected large amounts of Western classical music. Such collections typically contain different kinds of music-related documents of various formats including text, symbolic, audio, image, and video data. Three prominent examples of such data types are sheet music, symbolic score data, and audio recordings. Music data is often digitized in some systematic fashion using (semi-)automatic methods. For example, entire sheet music books can be digitized in a bulk process by using scanners with automatic page turners. This typically results in huge amounts of high-resolution digital images stored in formats such as TIFF or PDF. One can further process the image data to obtain symbolic music representations that can be exported into formats such as MusicXML,[4] LilyPond,[5] or MIDI.[6] This is done by using *optical music recognition* (OMR), the musical equivalent to optical character recognition (OCR) as used in text processing. Symbolic music representations and MIDI files are also obtained from music notation software or from electronic instruments. Last but not least, modern digital music libraries contain more and more digitized audio material in form of WAV or MP3 files. Such files are obtained by systematically ripping available CD collections, converting tape recordings, or digitizing old vinyl recordings.

As a result of such systematic digitization efforts, one often obtains data sets that contain items of a single type,[7] see also Figure 1. For example, scanning entire sheet music books results in a collection of image files, where each file corresponds to a specific page. Or, ripping a data set of CD recordings, one obtains a collection of audio files, where each file corresponds to an audio track. In the case of digitizing a vinyl recording, a track covers an entire side of the recording that may comprise several pieces of music.[8] In order to make the data accessible in a user-friendly and consistent way, various postprocessing steps are required. For example, the scanned pages of sheet music need to be pooled, cut, or combined to form musically meaningful units such as movements or songs. Furthermore, these units

[2] For example, the project *Presto Space* (`http://www.prestospace.org`) or the internet portal *Europeana* (`http://www.europeana.eu`).
[3] PROBADO, for more information we refer to `http://www.probado.de/en_home.html`.
[4] `http://www.recordare.com/musicxml`
[5] `http://lilypond.org`
[6] `http://www.midi.org`
[7] For example, the *Archival Sound Recordings* of the British Library (`http://sounds.bl.uk`), the Chopin Early Editions (`http://chopin.lib.uchicago.edu`), or the *Munich Digitization Center* of the Bavarian State Library (`http://bsb-mdz12-spiegel.bsb.lrz.de/~mdz`).
[8] The notion of a piece of music usually refers to individual movements or songs within bigger compositions. However, the particular segmentation applied by music libraries can vary.

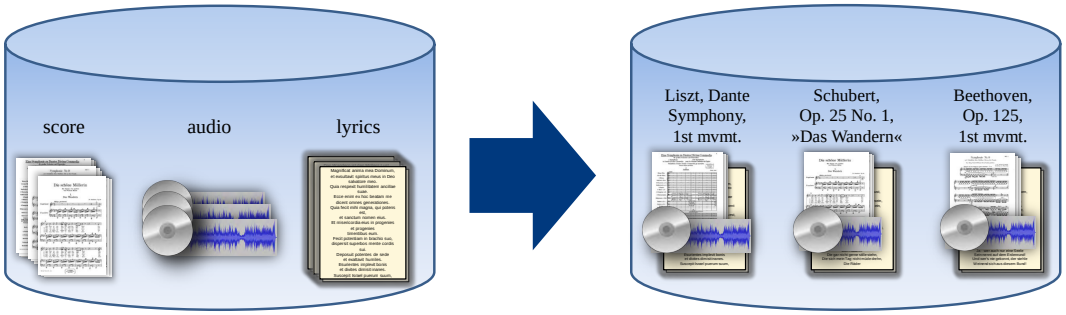
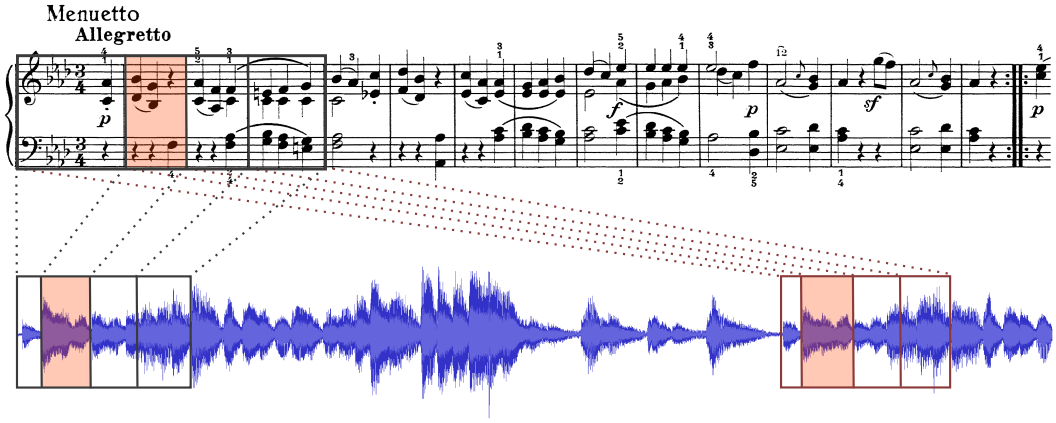


Figure 1 Change from a document and document type centered data collection (**left**) to an arrangement focusing on pieces of music (**right**).

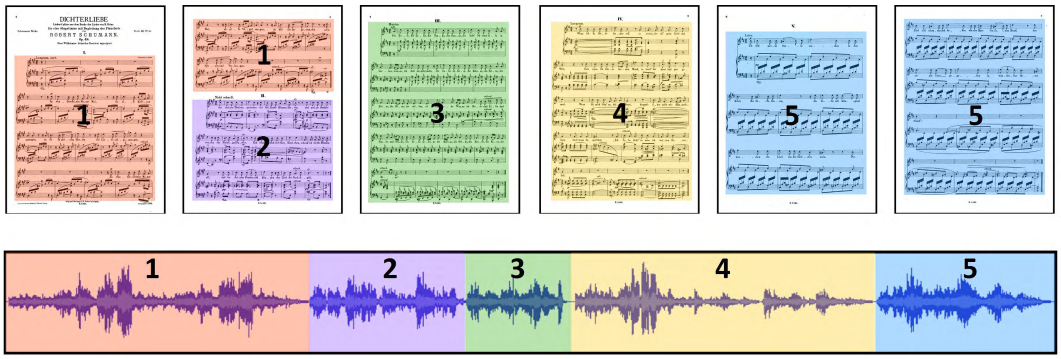
need to be assigned to the respective musical work and annotated accordingly. Similarly, audio tracks are to be identified and trimmed to meet certain standards and conventions. Finally, suitable metadata needs to be attached to the digitized documents. When trying to automate the stated postprocessing steps for real-world music collections, they become challenging research problems. The main issues are the inconsistency and the complexity of the given data. For instance, sheet music contains a lot of textual metadata but its extraction and proper interpretation are non-trivial tasks (e.g., *Allegro* can likewise constitute a tempo instruction or the name of a piece of music, see [25] for further details).

The availability of accurate metadata is essential for organizing and indexing huge music collections. For example, searching for the keywords “Beethoven” and “Op. 125”, one should be able to retrieve all documents that refer to Beethoven Symphony No. 9. In this way, suitable metadata information allows for re-arranging the music documents to obtain a data collection, where all versions that refer to the same piece of music are compiled irrespective of their format or modality, see Figure 1. However, such a document-level compilation of musically related versions constitutes only the first step towards a comprehensive system for multimodal music navigation and browsing. In the next step, one requires linking structures that reveal the musical relations within and across the various documents at a lower hierarchical level. For example, such a linking structure may reveal the musical correspondence between notes depicted in a scanned sheet music document and time positions in an audio recording of the same piece of music. Such links would then allow for a synchronous display of the audible measure in the sheet music representation during the playback of a music recording. Similarly, in a retrieval scenario, a musical theme or passage could be marked in the image domain to retrieve all available music recordings where this theme or passage is played.

In this paper, we address the problem of how suitable linking structures between different versions of the same piece of music can be computed in a fully automated fashion. In particular, we focus on the multimodal scenario of linking sheet music representations with corresponding audio representations, a task we also refer to as *sheet music-audio synchronization*, see Figure 2a. In Section 2, we give an overview of an automated synchronization procedure and discuss various challenges that arise in the processing pipeline. One major step in this pipeline is to extract explicit note events from the digitized sheet music images by using OMR. Even though there is already various commercial OMR software on the market for many years, the robust extraction of symbolic information is still problematic for complex scores. Some of these challenges are discussed in Section 3. In particular, certain extraction errors have severe consequences, which may lead to erroneous assignments of entire instrument tracks or



(a) Score-audio synchronization on the measure-level. Time segments in the audio stream are mapped to individual measures in the score representation. The depicted audio track contains a repetition. Therefore, the according score measures have to be mapped to both audio segments.



(b) Score-audio mapping on the detail level of pieces of music. The score and the audio data are segmented into individual pieces of music. Afterwards, the correct score-audio pairs have to be determined.

Figure 2 Examples for score-audio synchronization on different detail levels.

to deviations in the global music structure. In Section 4, we discuss common computational approaches to sheet music-audio synchronization and present various strategies how the resulting global differences between documents can be handled within the synchronization pipeline. Finally, in Section 5, we describe some applications and novel interfaces that are based on synchronization results. We conclude the paper with an outlook on future work. A discussion of relevant work can be found in the respective sections.

2 Task Specification

The goal of music synchronization is the generation of semantically meaningful bidirectional mappings between two music documents representing the same piece of music. Those documents can be of the same data type (e.g., audio-audio synchronization) or of different data types (e.g., score-audio synchronization or lyrics-audio synchronization). In the case of score-audio synchronization the created linking structures map regions in a musical score, e.g., pages or measures, to semantically corresponding sections in an audio stream (see Figure 2).

Although the task of score-audio synchronization appears to be straightforward, there exist several aspects along which the task and its realization can vary (see Figure 3). The

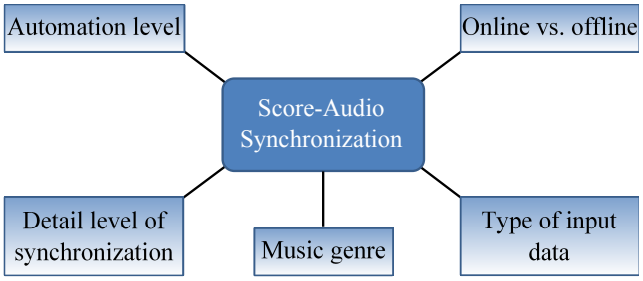


Figure 3 Aspects of score-audio synchronization.

particular choice of settings with respect to these aspects is always influenced by the intended application of the synchronization results.

The first choice concerns the sought detail level or granularity of the synchronization. A very coarse synchronization level would be a mapping between score and audio sections representing the same piece of music, see Figure 2b (e.g., *Neue Mozart-Ausgabe*[9]). This type of alignment is also referred to as *score-audio mapping*. Finer detail levels include page-wise [2, 21], system-wise, measure-wise [34], or note-wise [8, 46] linking structures between two music documents. The choice of granularity can in turn affect the level of automation. The manual annotation of the linking structure might be achievable for page-wise synchronizations. However, for finer granularities semi-automated or automated synchronization algorithms would be preferable. While automatic approaches do not need (and also not allow) any user interaction, in semi-automatic approaches some user interaction is required. However, the extent of the manual interaction can vary between manually correcting a proposed alignment on the selected detail level and correcting high-level aspects (e.g., the repeat structure) before recalculating the alignment. The selected automation level obviously also depends on the amount of data to be processed. For a single piece of music given only one score and one audio interpretation, a full-fledged synchronization algorithm might not be required. But, for the digitized music collection of a library, manual alignment becomes impossible. Finally, reliability or accuracy requirements also take part in the automation decision.

Another huge differentiation concerns the runtime scenario. In *online* synchronization, the audio stream is only given up to the current playback position and the synchronization should produce an estimation of the current score position in real-time. There exist two important applications of online score-audio synchronization techniques, namely *score following* and *automated accompaniment* [13, 17, 36, 37, 46, 48, 54]. The real-time requirements of this task turn local deviations between the score and the audio into a hard problem. Furthermore, recovery from local synchronization errors is problematic. In contrast, in *offline* synchronization the complete audio recording and the complete score data are accessible throughout the entire synchronization process [34, 42]. Also, the computation is not required to run in real-time. Due to the loosened calculation time requirements and the availability of the entire audio and score data during calculation, offline synchronization algorithms usually achieve higher accuracies and are more robust with regard to local deviations in the input data. The calculated linking structures can afterwards be accessed to allow for, e.g., score-based navigation in audio files.

The genre/style of the music to be synchronized also influences the task of score-audio synchronization. While Western classical music and most popular music feature strong

[9] `http://www.nma.at`

melodic/harmonic components other music styles, like African music, may mainly feature rhythmic drumming sounds. Obviously, using harmonic information for the synchronization of rhythmic music will prove ineffective and therefore different approaches have to be employed.

The type of input data—more precisely the score representation—constitutes the last aspect of score-audio synchronization. The score data can either be available as scanned images of music notation (i.e., sheet music) or as symbolic score (e.g., MIDI or MusicXML). Obviously, the choice of score input affects the type of challenges to be mastered during synchronization. While symbolic score representations are usually of reasonable quality and the extraction of the individual music events is straightforward, some sort of rendering is required to present the score data. In contrast, sheet music already provides a visualization. But the music information needs to be reconstructed from the image data before the linking structures can be calculated. OMR systems approach this task and achieve high reconstruction rates for printed Western music. Nevertheless, the inclusion of OMR into the synchronization process may result in defective symbolic score data (see Section 3). Usually, the errors are of mainly local nature. Thus, by choosing a slightly coarser detail level (e.g., measure level) sound synchronization results can be achieved. For a differentiation between these two types of input data, the term *sheet music-audio synchronization* is often utilized if scanned images are given as score input.

Various researchers are active in the field of score-audio synchronization and work on all settings of the listed aspects has been reported. Considering all aspects and their specific challenges would go beyond the scope of this paper. Instead, we focus on the task of automated offline sheet music-audio synchronization for Western classical music producing linking structures on the measure level. Furthermore, the processing of large music collections should be possible.

The basic idea in most score-audio synchronization scenarios is to transform both input data types into a common mid-level representation. These data streams can then be synchronized by applying standard alignment techniques, see Section 4 for an overview. Independent of the selected approach, one has to cope with the following problems to get reasonable synchronization results:

- **Differences in structure:** A score can contain a variety of symbols representing jump instructions (e.g., repeat marks, segno signs, or keywords such as *da capo*, *Coda*, or *Fine*, see Figure 4). While OMR systems are capable of detecting repeat marks, they often fail to reliably detect most other textual jump instructions in the score. Therefore, the correct playback sequence of the measures cannot be reconstructed. However, even if all jump instructions are correctly recognized, the audio recording may reveal additional repeats or omissions of entire passages notated in the score. Again, the given sequence of measures does not coincide with the one actually played in the audio recording. Such structural differences lead to major challenges in score-audio synchronization.

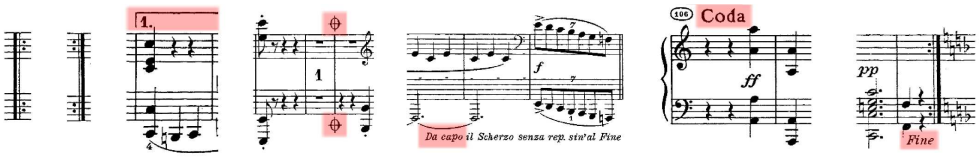


Figure 4 Examples of jump indicators used in music notation (adapted from [25]).

- **Differences between music representations:** Score pages and audio recordings represent a piece of music on different levels of abstraction and capture different facets of the music. One example is the tempo. Music notation may provide some written

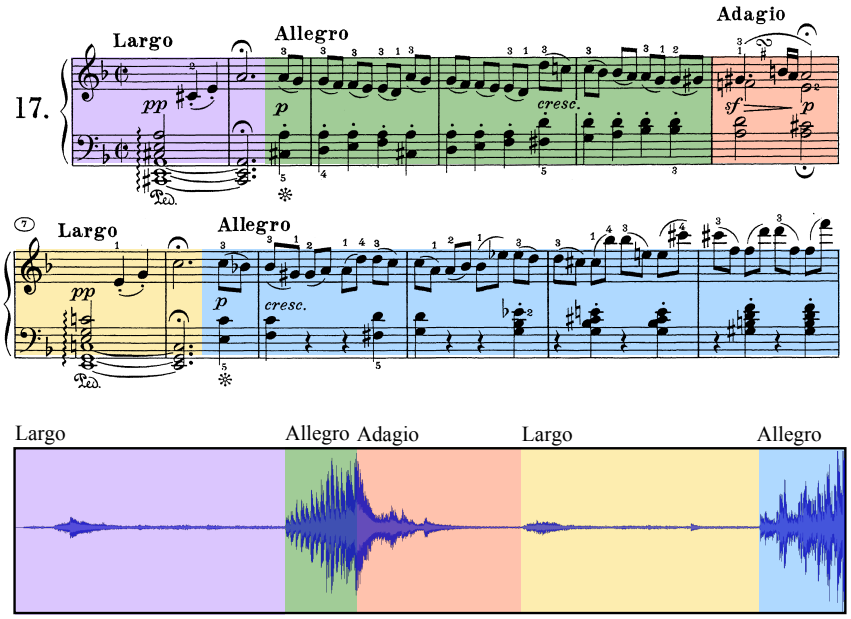


Figure 5 Extract of Beethoven's *Piano Sonata No. 17* (publisher: *Henle Verlag*, pianist: V. Ashkenazy). In the first nine measures alone four substantial tempo changes are performed. Thus, the duration of the measures in the audio recording varies significantly. However, in the score only vague instructions are available that result at best in an approximation of the intended tempo changes.

information on the intended tempo of a piece of music and tempo changes therein (e.g., instructions such as *Allegro* or *Ritardando*). However, those instructions provide only a rough specification of the tempo and leave a lot of space for interpretation. Therefore, different performers might deviate significantly in their specific tempo choices. In addition, most musicians even add tempo changes that are not specified by the score to emphasize certain musical passages. For an example we refer to Figure 5.

The differences in the loudness of instruments and the loudness variations during the progression of a piece of music are further important characteristics of a given performance. Just like tempo, loudness is notated only in a very vague way and OMR systems often fail to detect the few available instructions. Similarly, music notation only provides timbre information through instrument labels. Therefore, timbre-related sound properties such as instrument-dependent overtone energy distributions are not explicitly captured by the score.

In conclusion, in view of practicability, score-audio synchronization techniques need to be robust towards variations in tempo, loudness, and timbre to deal with the mentioned document type related differences.

- **Errors in the input data:** As already mentioned, OMR is not capable of reconstructing the score information perfectly. The errors introduced by OMR can be divided into local and global ones. Local errors concern, e.g., misidentifications of accidentals, missed notes, or wrong note durations. In contrast, examples for global errors are errors in the detection of the musical key or the ignorance of transposing instruments. Further details will be presented in Section 3. While for sheet music, errors are introduced during the reconstruction from the scanned images, the audio recordings themselves can be erroneous. The performer(s) may locally play some wrong notes or a global detuning

occurred. For Western classical music a tuning of 440 Hz for the note $A4$ was defined as standard. However, most orchestras slightly deviate from this tuning.[10] Furthermore, for Baroque music a deviation by a whole semitone is common.

- **Sheet music-audio mapping:** Especially in library scenarios, the goal is not the synchronization of one piece of music. Usually, the input consists of whole sheet music books and whole CD collections. Therefore, before calculating the linking structures, the score and the audio data need to be segmented into individual pieces of music. As the order in the sheet music books and on the CDs might differ, a mapping on this granularity level needs to be created before the actual synchronizations can be calculated.

Although we focus on sheet music-audio synchronization in this contribution, most of the mentioned problems also exist for other score-audio synchronization variants.

3 Optical Music Recognition

Similarly to optical character recognition (OCR) with the goal to reconstruct the textual information given on scanned text pages, optical music recognition (OMR) aims at restoring musical information from scanned images of sheet music. But, the automatic reconstruction of music notation from scanned images has to be considered much harder than OCR. Music notation is two-dimensional, contains more symbols, and those symbols mostly overlap with the staves. A large number of approaches to OMR has been proposed and several commercial and non-commercial OMR systems are available today. Three more popular commercial systems are SharpEye,[11] SmartScore,[12] and PhotoScore.[13] All of them operate on common Western classical music. While the former two only work for printed sheet music, PhotoScore also offers the recognition of handwritten scores. Two prominent examples for non-commercial OMR systems are Gamera[14] and Audiveris.[15] While Audiveris is not competitive in terms of recognition rates, Gamera is actually a more general tool for image analysis. Therefore, Gamera requires training on the data to be recognized to yield adequate recognition results. Since the introduction of OMR in the late 1960s [45] many researchers worked in the field and relevant work on the improvement of the recognition techniques has been reported. For further information, we refer to the comprehensive OMR bibliography by Fujinaga [29].

As with score-audio synchronization, there are three factors that affect the difficulty of the OMR task and the selection of the pursued approach. First, there exist different types of scores (e.g., medieval notation, modern notation or lute tablatures) that differ significantly in their symbol selection and their basic layout. Therefore, the type of music notation present on the images has to be considered. Second, the transcription format is of influence. Printed score is regular and usually well formatted while handwritten score can be rather unsteady and scrawly. Additionally, crossing outs, corrections, and marginal notes make the interpretation of handwritten scores even more challenging. Finally, the envisioned application of the resulting symbolic representation influences the required precision. OMR results intended for playback or score rendering have to present a much higher accuracy on the note level than a reconstruction serving as score representation during sheet music-audio synchronization on the measure level (see Section 4). In the first scenario, most OMR systems

[10] List of standard pitches in international orchestras: `http://members.aon.at/fnistl/`
[11] `http://www.music-scanning.com`
[12] `http://www.musitek.com`
[13] `http://www.sibelius.at/photoscore.htm`
[14] `http://gamera.informatik.hsnr.de`
[15] `http://audiveris.kenai.com`

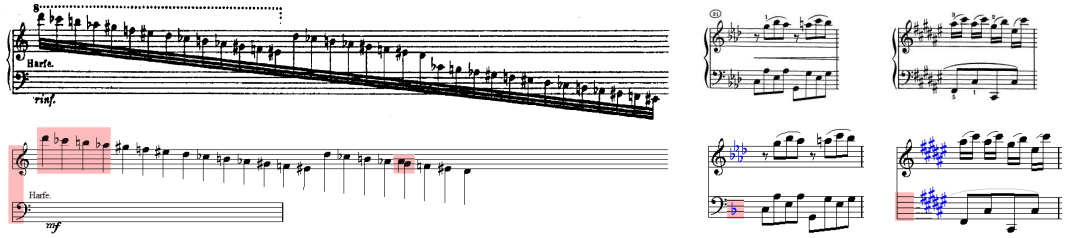


Figure 6 Examples of common OMR errors. **Left:** Besides wrong note durations and an accidental mistaken for a note, the staff system was split into two systems. **Middle:** The key signature was not correctly recognized for the lower staff. **Right:** In the lower staff, the clef was not detected.

support the creation of an initial approximation of a symbolic representation and provide user interfaces for manual correction.

Several studies on the performance of OMR systems and the types of errors that occur were conducted [10, 11, 12, 25]. Those studies show that OMR systems vary with regard to their strengths and weaknesses. However, the types or classes of recognition errors are the same for all systems. Some examples of common errors are given in Figure 6. Most of those errors are of a local nature and concern individual music symbols or small groups thereof. Examples are articulation marks, ornaments, accidentals, dynamics, and note durations that are mistaken for some other symbol or missed altogether. In the context of sheet music-audio synchronization those local errors are less severe because the applied synchronization methods are capable of managing local deviations between the two sequences to be aligned. In contrast, several types of recognition errors, influencing larger areas of the score, exist. Those might include wrong time signatures, missed clefs, wrong key signatures, staff systems being split up (e.g., due to arpeggios traveling through several staves or due to textual annotations disrupting the vertical measure lines), or missed repetition instructions. While the time signature is of little importance for sheet music-audio synchronization, the other error types can have a strong impact on the alignment result. To achieve high quality alignments, these kinds of errors should be corrected, either by offering user interfaces for manual intervention or by developing new OMR techniques improving on those specific deficits.

Another shortcoming of most OMR systems is the interpretation of textual information in the score. While some systems are capable of determining text such as lyrics correctly, text-based instructions on dynamics, title headings, and instruments are often recognized without associating their (musical) meaning or are not detected at all. For sheet music-audio synchronization the most significant textual information is the one on transposing instruments.[16] If transposing instruments are part of the orchestra and their specific transposition is not considered during the reconstruction, their voices will be shifted with respect to the remaining score, see Figure 7. However, to the best of our knowledge, no OMR system considers this type of information and attempts its detection.

[16] For transposing instruments, the sounding pitches are several semitones higher/lower than the notes written in the score.

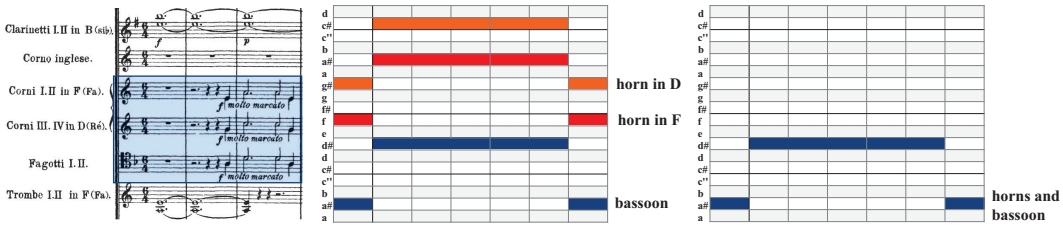


Figure 7 Voices of transposing instruments are shifted with respect to other voices if their transpositions are not known. **Middle:** Erroneous reconstruction in absence of transposition information. **Right:** Correct symbolic representation of the highlighted score extract.

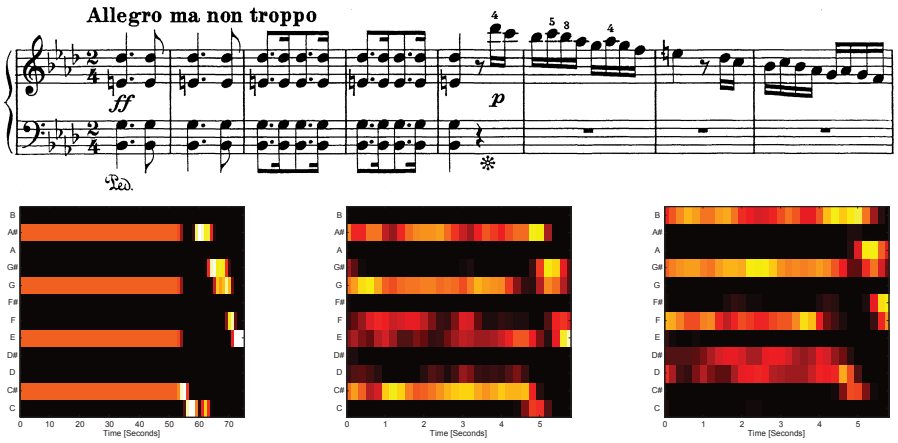


Figure 8 Illustration of chroma features for the first few measures from the third movement of Beethoven's *Piano Sonata No. 23*. The color values represent the intensity of a chroma at a given position (black: low intensity, red: medium intensity and yellow/white: high intensity). The left diagram shows a chroma sequence created from the depicted sheet music extract. The middle and the right diagram show the chroma features for two audio interpretations of the same music extract. The chroma features clearly capture the higher tuning (by one semitone) of the second recordings.

4 Sheet Music-Audio Synchronization

The goal of sheet music-audio synchronization is to link regions in two-dimensional score images to semantically corresponding temporal sections in audio recordings. Therefore, the two data sources need to be made comparable by transforming them into a common mid-level representation. In the synchronization context, chroma-based music features turned out to be a powerful and robust mid-level representation [7, 31]. A chroma vector represents the energy distribution among the 12 pitch classes of the equal-tempered chromatic scale (C, C$^{\#}$, D, ..., B) for a given temporal section of the data, see Figure 8. Chroma features have the property of eliminating differences in timbre and loudness to a certain extend while preserving the harmonic progression in the music. Therefore, their application is most reasonable for music with a clear harmonic progression, like most Western classical music. In addition, by choosing the size of the sections represented by individual chroma vectors appropriately, local errors in the input data can be canceled out for the most part. To transform sheet music into chroma, OMR is performed on the score scans. Afterwards, a MIDI file is created from this data assuming a fixed tempo and standard tuning (see [34] for more information).

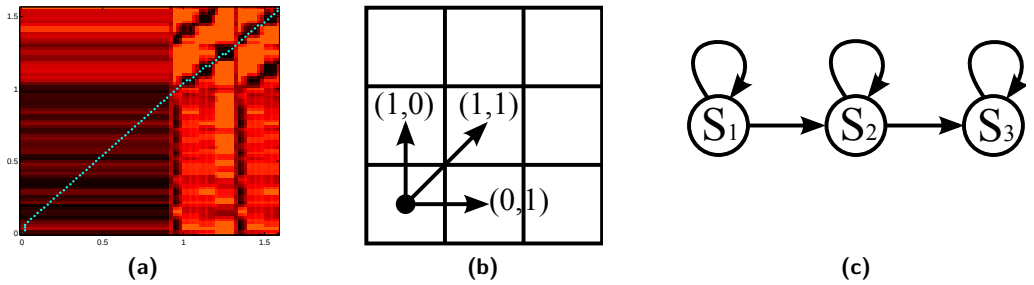


Figure 9 **(a)** Local cost matrix for the score chromagram and one of the audio chromagrams depicted in Figure 8. The optimal alignment path is highlighted in light blue. **(b)** Example of allowed steps during DTW based synchronization. **(c)** Illustration of a left-to-right connected HMM with 3 states.

At the moment, we assume that we are given one sheet music representation and one audio interpretation of the same piece of music. We will address the issue of sheet music-audio mapping in Section 4.1. Furthermore, we will for now assume that the structure of the score and the audio recording coincide. Some ideas on how to handle structural differences will be presented in Section 4.2. After calculating chroma features for both music representations, a local cost matrix can be constructed by pair-wise measuring the similarity between the vectors of the two chroma sequences. Then, the goal is the identification of a path through this matrix that is connecting the two beginnings and endings of the feature sequences and is optimal with respect to the local costs along the path (*optimal alignment path*). See Figure 9a for an example.

There exist two commonly used computational approaches to this task. The first approach is called *dynamic time warping* (DTW) and is based on dynamic programming techniques [1, 18, 24, 31, 42, 43]. After creating the local cost matrix using an appropriate cost measure an accumulated cost matrix is constructed. In this matrix the entry at position (n, m) contains the minimal cost of any alignment path starting at $(1, 1)$ and ending at (n, m). However, during the creation of the alignment path only a certain set of steps is allowed to move through the matrix, e.g., $\{(1, 0), (0, 1), (1, 1)\}$, see Figure 9b. The optimal alignment path is then constructed by backtracking through the matrix using the allowed steps. At each point we chose the predecessor with the lowest accumulated costs. The second approach applies *Hidden Markov Models* (HMM) to determine the optimal alignment path [36, 37, 46, 48]. In this scenario one of the feature sequences is used as hidden states of the HMM and the other sequence forms the set of observations. Usually a left-to-right connected HMM structure is used for score-audio synchronization, see Figure 9c.

In combination with chroma features these alignment techniques allow for some variations in timbre, loudness, and tempo. In addition, small deviations in the data streams (due to errors) can be handled. In contrast, tuning differences are not considered by the presented approaches. Here, the feature sequences show significant differences that can result in a poor synchronization quality (see Figure 8). To suitably adjust the chroma features, a tuning estimation step can be included in the feature calculation process [19]. Instead, one may also apply brute-force techniques such as trying out all possible cyclic shifts of the chroma features [30, 39]. Thus, the presented approaches already cope with some of the problems mentioned in Section 2. In the remainder of this section we want to introduce approaches tackling some of the remaining unsolved problems (i.e., structural differences, certain types of errors, and sheet music-audio mapping).

4.1 Sheet Music-Audio Mapping

Arranging the music data in a digital library in a work-centered way or, more precisely, piece of music-wise has proven beneficial. Thus in the context of a digitization project to build up a large digital music library, one important task is to group all documents that belong to the same piece of music, see Figure 1. Note that in this scenario, the music documents that are to be organized are not given as individual songs or movements, but rather as complete sheet music books or audio CD collections that usually contain several pieces of music.[17] In addition, we typically have to deal with numerous versions of audio recordings of one and the same piece of music,[18] and also with a number of different score versions (different publishers, piano reductions, orchestra parts, transcriptions, etc.) of that piece. Thus, the final goal at this level of detail is to segment both the score books and the audio recordings in such a way that each segment corresponds to one piece of music. Furthermore, each segment should be provided with the appropriate metadata. This segmentation and annotation process, called *sheet music-audio mapping*, is a crucial prerequisite for the sheet music-audio synchronization described in the previous section. One possibility to solve this task is to manually perform this segmentation and annotation. However, for large collections this would be an endless undertaking. Thus semi-automatic or even fully automatic mapping techniques should be developed.

For audio recordings and short audio extracts, music identification services like *Shazam*[19] can provide a user with metadata. Furthermore, ID3 tags, CD covers, or annotation databases such as Gracenote[20] and DE-PARCON[21] can contain information on the recorded piece of music. However, their automated interpretation can quickly become a challenging task. To name just two prominent issues, the opus numbers given by the different sources might not use the same catalogue or the titles might be given in different spellings or different languages. Furthermore, the mentioned services do not provide information for public domain recordings. Another issue can be introduced by audio tracks containing several pieces of music. Here, the exact start and end positions of the individual pieces of music have to be determined.[22] However, this information is usually not provided on CD covers or in metadata databases. Still, the mentioned information sources can be used to support the manual segmentation and annotation process. The automatic extraction and analysis of textual information on scanned score images has to be considered at least equally challenging.

Given one annotated audio recording of all the pieces contained in a score book, Fremerey et al. [25, 27] propose an automatic identification and annotation approach for sheet music that is based on content-based matching. One key strategy of the proposed procedure is to reduce the two different types of music data, the audio recordings as well as the scanned sheet music, to sequences of chroma features, which then allow for a direct comparison across the two domains using a variant of efficient index-based audio matching, see [33]. To this end, the scan feature sequence is compared to the audio feature sequence using subsequence dynamic time warping. The resulting matching curve combined with the information on the

[17] In the context of the PROBADO project, the Bavarian State Library in Munich digitized more than 900 sheet music books (approx. 72,000 score pages) and about 800 audio CDs.
[18] For example, the British Library Sounds include recordings of about 750 performances of Beethoven String Quartets, as played by 90 ensembles, see `http://sounds.bl.uk/Classical-music/Beethoven`
[19] `http://www.shazam.com`
[20] `www.gracenote.com`
[21] `http://www.de-parcon.de/mid/index.html`
[22] Usually, longer periods of silence can hint at the beginning of a new piece. However, the direction *attacca* resulting in two successive movements played without a pause, can prevent this clue from existing.

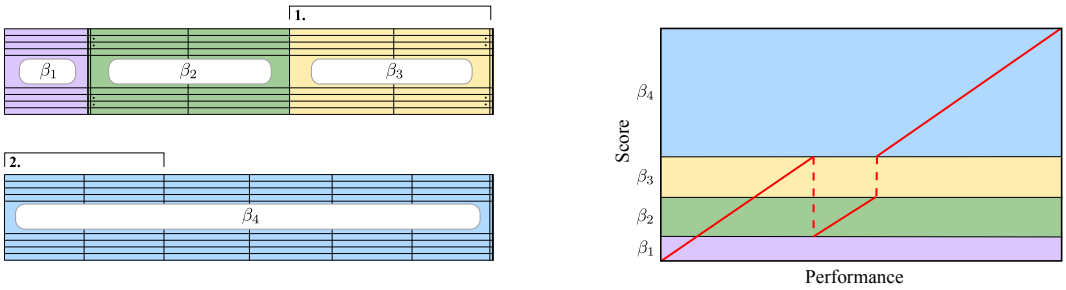


Figure 10 Score block sequence $\beta_1\beta_2\beta_3\beta_4$ created from notated score jumps and alignment path for an audio with block structure $\beta_1\beta_2\beta_3\beta_2\beta_4$ (adapted from [25]).

audio segmentation finally gives both the segmentation and the annotation of the scanned sheet music.

In the same manner, additional audio recordings of already known pieces can be segmented and annotated. Therefore, through the presented approach the manual processing of only one manifestation of each piece of music is required.

4.2 Dealing with Structural Differences

When comparing and synchronizing scores and performances, it may happen that their global musical structures disagree due to repeats and jumps performed differently than suggested in the score. These structural differences have to be resolved to achieve meaningful synchronizations. In the scenario of online score-audio synchronization this issue has already been addressed [2, 35, 44, 51]. Pardo and Birmingham [44] and Arzt et al. [2] both use structural information available in the score data to determine music segments where no jumps can occur. In the first publication an extended HMM is used to allow for jumps between the known segment boundaries. In the second approach an extension of the DTW approach to music synchronization is used to tackle structural differences. At each ending of a section, three hypotheses are pursued in parallel. First, the performance continues on to the next section. Second, the current section is repeated. Third, the subsequent section is skipped. After enough time has passed in the performance the most likely hypothesis is kept and followed. Besides approaches exploiting structural information available from the score, Müller et al. [38, 40] approached a more general case where two data sources (e.g., two audio recordings) are given but no information on allowed repeats or jumps is available. In this case, only partial alignments of possibly large portions of the two documents to be synchronized are computed.

Fremerey et al. [25, 26] presented a method for offline sheet music-audio synchronization in the presence of structural differences, called *JumpDTW*. Here, jump information is derived from the sheet music reconstruction thus creating a block segmentation of the piece of music (see Figure 10). As already mentioned, OMR systems may not recognize all types of jump instructions (especially, textual instructions are often missed). Therefore, bold double bar lines are used as block boundaries. At the end of each block the performance can then either continue to the next block or jump to the beginning of any other block in the piece, including the current one (in contrast to [2] where only forward jumps skipping at most one block are considered). To allow for jumps at block endings, the set of DTW steps is modified. For all block endings, transitions to all block starts in the score are added to the usual steps. By calculating an optimal alignment path using a thus modified accumulated cost matrix,

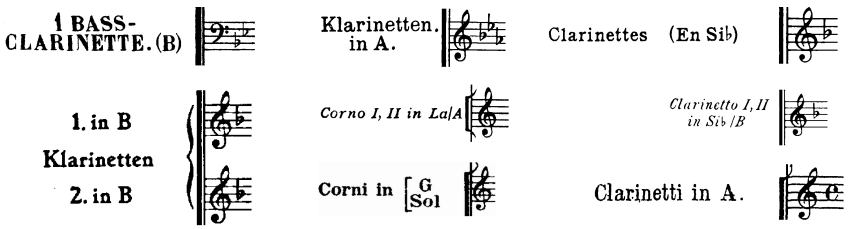


Figure 11 Examples of transposition labels applied by different editors.

possible jumps in the performance can be detected and considered during the synchronization process.

4.3 Dealing with Orchestral Music

Because of the large number of instruments in orchestral music, the score notation inevitably becomes more complex. Typically, this results in a decreased OMR accuracy. Furthermore, orchestral scores contain information commonly neglected by OMR systems. One very important example is the transposition information. The specific transposition of an instrument is usually marked in the score by textual information such as "Clarinet in E", see Figure 11. Obviously, by disregarding this information during the OMR reconstruction, the pitch information for transposing instruments will be incorrect. In the context of sheet music-audio synchronization such global errors in the reconstructed symbolic score data can result in a significant accuracy loss [52, 53].

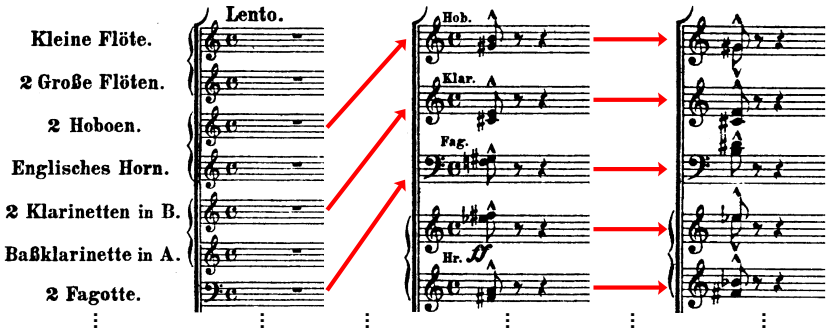


Figure 12 Extracts from Franz Liszt: *Eine Sinfonie nach Dantes Divina Commedia* using compressed notation (publisher: *Breitkopf & Härtel*).

In Western classical music, the score notation usually obeys some common typesetting conventions. Examples are the textual transposition information but also the introduction of all instruments playing in a piece of music by labeling the staves of the first system. Furthermore, a fixed instrument order and the usage of braces and accolades help in reading the score [49]. But despite of all these rules, the task of determining which instrument is supposed to play in a given staff (instrument-staff mapping) and whether or not it is a transposing instrument can be challenging. For most scores the number of staves remains constant throughout the entire piece of music. Therefore the instrument names and transposition information are often omitted after the first system and the information given in the first system needs to be passed on to the remaining systems. The task of determining the instrument of a staff and its transposition becomes even more complicated for compressed score notations where staves of pausing instruments are removed (see Figure 12). Here, the

instrument order is still valid, but some of the instruments introduced in the first system may be missing. To clarify the instrument-staff mapping in these cases, textual information is given. However, in these cases the instrument names are usually abbreviated and therefore more difficult to recognize. Furthermore, transposition information is often only provided in the first system of a piece or in the case that the transposition changes. The textual information might be omitted altogether if the instrument-staff mapping is obvious for a human reader (e.g., strings are always the last instrument group in a system).

Although a great deal of research on OMR has been conducted (see, e.g., [4, 29]), the particular challenges of orchestral scores have not yet been addressed properly. A first approach for the reconstruction of the transposition information was presented in [53]. The instrument-staff mapping as well as the transposition information are reconstructed during three distinct processing steps.

In the first step, the textual information available on the score scans is recovered and interpreted to regain as many instrument labels and transposition labels as possible, see Figure 13. Using the staff location information available in the OMR result, image regions that possibly contain text/words naming an instrument or a transposition are detected and processed by an OCR engine. Subsequently, the detected instruments are mapped to the according staves. To account for different spellings and abbreviations, a library of all possible textual representations of the instruments is used as additional knowledge. Transpositions are recognized by searching for the keyword "in" followed by a valid transposition information.

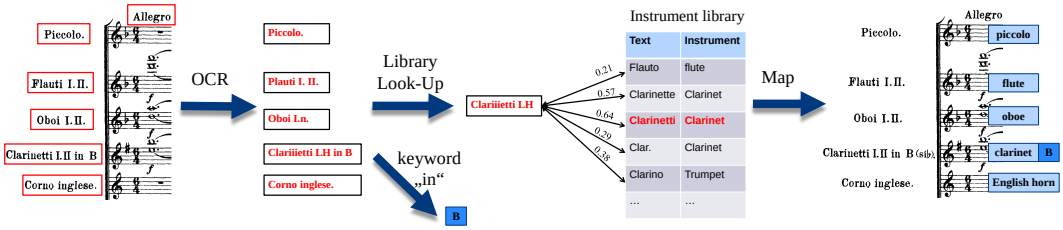


Figure 13 Overview: Reconstruction of instrument and transposition labels from the textual information in the score.

In the second step, the reconstruction from the previous step is used as initialization of an iterative process, see Figure 14. To this end, musical knowledge and common notation conventions are employed. As both the OCR-reconstruction and all information deduced through musical knowledge are uncertain, all instrument-staff mappings are equipped with plausibility values. Besides filling missing mappings, the following iterative update process also strengthens/weakens existing plausibilities. Each iteration of step two can again be divided into three parts. First, the already detected instrument information is successively propagated between consecutive systems by employing the convention that the initially established instrument order is not altered. If two instruments occur in both systems and the number of intermediate staves between these instruments coincides, the instrument information of the intermediate staves of the first system is propagated to the according staves in the subsequent system. Second, musical properties such as "trombone and tuba play in subsequent staves and are grouped by an accolade" are deduced from the score and employed to determine the instrumentation. In the third and final part, the instrument order established in the first system is used again. For all subsequent systems deviations from this order are determined and the according instrument-staff mappings are weakened.

In the last step of the proposed method, the transposition labels given in the first system (and reconstructed in step one) are transferred to the remaining systems. Thereby a

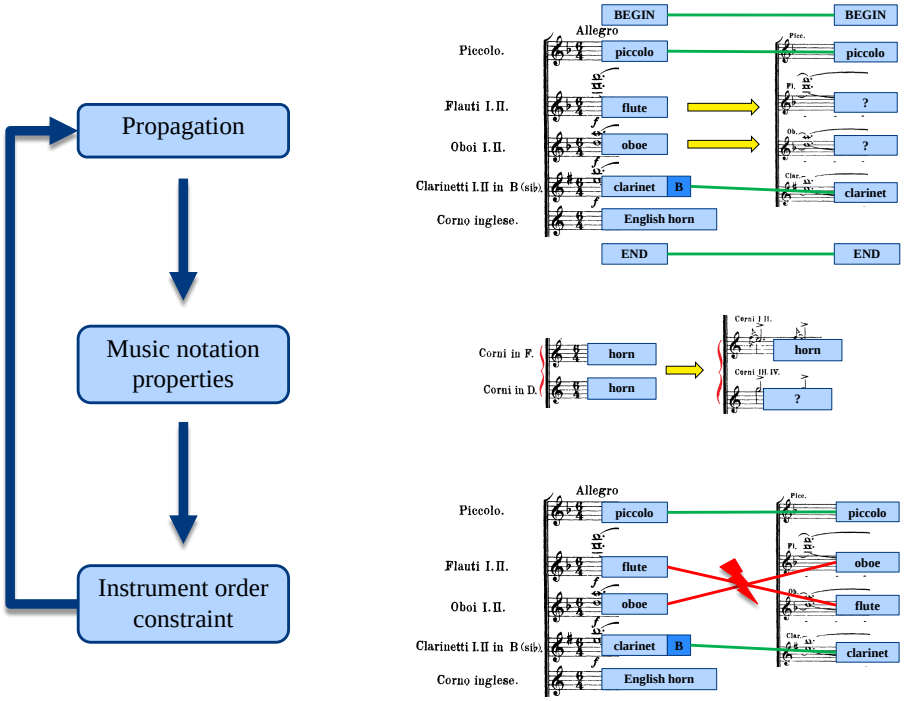


Figure 14 Overview: Recursive approach to the reconstruction of missing instrument and transposition labels.

global correction of the transposition information is achieved even if textual transposition information is only available in the first system.

5 Applications of Sheet Music-Audio Synchronization

In Section 1 we already touched upon possible applications of sheet music-audio synchronization. In this section we first give a more detailed overview of existing user interfaces that employ synchronization techniques (using sheet music or symbolic score data). Then, we focus on current issues in *music information retrieval* (MIR) and show how to incorporate sheet music-audio synchronization to solve specific MIR tasks.

5.1 User Interfaces

The *Laboratorio di Informatica Musicale* at the University of Milan developed the IEEE 1599 standard for the comprehensive description of music content. The proposed XML-format can handle and relate information of various kinds including music symbols, printed scores, audio recordings, text, and images. In addition, music analysis results and synchronization information can be stored as well. Based on this IEEE standard, user interfaces for the simultaneous presentation of multiple music documents have been proposed [3, 5, 6]. To this end, the synchronization results are used for enhanced, multimodal music navigation. At the moment, the synchronization information is created manually but work towards automated score-audio synchronization has been reported [14]. Another project that uses manually

created alignment information is the *Variations* project [21].[23] The goal of *Variations* is the development of a digital music library system to be used in the education context. The system offers music analysis and annotation tools (e.g., structure analysis, time stretching) and page-wise score-audio synchronization. Work on automated synchronization has been described in [47].

WEDELMUSIC is one of the first systems presenting sheet music and audio data simultaneously [8]. During playback a marker moves through the sheet music to identify the currently audible musical position. In addition, page turning is performed automatically by gradually replacing the current sheet/system with the next one. However, the employed automatic synchronization approach was rather simple. Using the start and end points in the sheet music and the audio as anchor points, linear interpolation was applied. As local tempo deviations may result in alignment errors, a user interface for the manual rework of the proposed synchronization was available. Xia et al. [55] present a rehearsal management tool for musicians that exploits semi-automated score-audio synchronization. Here, recordings of various rehearsals are clustered and aligned to a score representation of the piece of music. Additional challenges are introduced by the fact that the recordings can differ in length and may cover different parts of the piece. In the PROBADO project, a digital music library system for the management of large document collections was developed (see Figure 15). The most prominent features are content-based retrieval techniques and a multimodal music presentation implemented by sheet music-audio synchronization [15, 16]. The alignment structures are calculated nearly automatically in this system.

Another application designed to support musicians is automated accompaniment. To this end, online score-audio synchronization determines the current position in the score as well as the current tempo to replay a time-stretched audio recording. Two well known accompaniment systems are *Music Plus One* by Raphael [46, 48] and ANTESCOFO by Cont [13].

5.2 MIR Research

There are various MIR tasks that exploit score information as additional knowledge. For example, in score-informed source separation one assumes that along with the audio recording a synchronized MIDI file is given. Through this file the occurring note events along with their position and duration in the audio are specified. We refer to Ewert and Müller [23] for an extensive overview. At the moment, all approaches use symbolic score data (e.g., MIDI) but sheet music may be applicable as well. However, in this case recognition errors need to be considered by the source separation method. A similar task is the estimation of note intensities in an audio recording where the notes are specified by a symbolic representation [22]. Again, to avoid the manual creation of a MIDI file, the exploitation of score scans together with sheet music-audio synchronization techniques, seems reasonable.

Another important research topic is lyrics-audio synchronization [28, 32]. Instead of using the commonly employed speech analysis techniques, sheet music can be added as additional information. Thereby, the lyrics can be derived from the OMR results. Afterwards, the lyrics-audio alignment can be calculated by means of the sheet music-audio synchronization [15, 41, 50].

[23] http://www.dlib.indiana.edu/projects/variations3

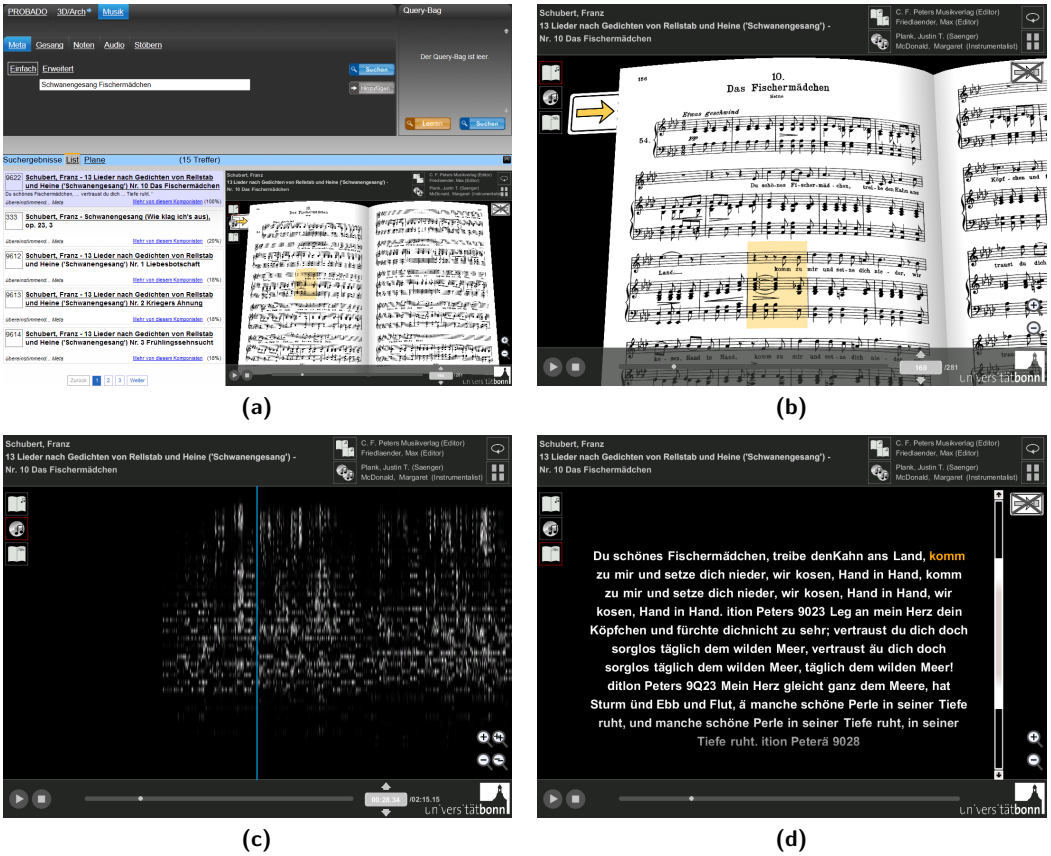


Figure 15 The PROBADO music user interface. **(a)** Search interface with a result presentation on the piece of music level. On the bottom right, access to all documents containing the selected piece is provided. Besides sheet music **(b)**, the interface offers visualizations of audio recordings **(c)** and lyrics **(d)**. Sheet music-audio synchronization results allow for the currently audible measure to be highlighted. Equally, the sung word is marked in the lyrics [50]. Different sheet music edition or other audio recordings can easily be selected. During a document change, the linking structures help to preserve the musical position and playback continues smoothly.

There are several other tasks where score-audio synchronization might help reducing the complexity of the problem. Some examples are structure analysis, chord recognition, and melody extraction.

6 Outlook

Although, current sheet music-audio synchronization algorithms perform quite well, there still exist some open issues. First, to allow for a higher level of detail, the input data has to become more reliable. In particular, the OMR accuracy needs to be improved. After achieving a high-resolution synchronization, e.g., on the note level, the question of how to present this alignment structure arises. For orchestral music, highlighting the currently audible notes in all voices would result in a very nervous visualization. At the moment only printed sheet music of reasonable quality is being used. However, huge amounts of old sheet music volumes exist that are heavily yellowed and stained. In addition, large collections of handwritten scores and hand-annotated printed sheet music are available. Some OMR

systems are capable of dealing with those types of sheet music but the applicability of the resulting symbolic representation (in terms of recognition accuracy) to the synchronization task would have to be investigated.

In Section 4.1, we discussed the task of sheet music-audio mapping and presented a method for segmenting and identifying score data using already segmented and identified audio documents. With this approach, at least one version of a piece of music has to be manually annotated. For large music databases a full automation or at least some support in the unavoidable manual tasks is highly desired. Looking at sheet music and CD booklets, they contain a wealth of textual information (composer, title, opus number, etc.). Automatically detecting and interpreting this information constitutes an important future step.

One can think of a variety of applications that would benefit from the presented synchronization techniques. The method could be extended to allow for online score following and live accompaniment of musicians using a scanned score. In [20, 44] the synchronization of lead sheets with a fully instrumented audio recording was suggested. In a similar manner, the sheet music of individual voices could be synchronized to an orchestra recording. These linking structures could for example be of use in the context of digital orchestra stands [9]. All parts are synchronized to the conductors score and upon selecting a position in the conductors score the position in all score visualization changes accordingly.

7 Acknowledgment

This work was supported by the German Research Foundation DFG (grants CL 64/7-2 and CL 64/6-2). Meinard Müller is funded by the Cluster of Excellence on Multimodal Computing and Interaction (MMCI). We would like to express our gratitude to Maarten Grachten, Masataka Goto, and Markus Schedl for their helpful and constructive feedback.

References

1 Vlora Arifi, Michael Clausen, Frank Kurth, and Meinard Müller. Automatic synchronization of musical data: A mathematical approach. *Computing in Musicology*, 13:9–33, 2004.

2 Andreas Arzt, Gerhard Widmer, and Simon Dixon. Automatic page turning for musicians via real-time machine listening. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 241–245, Patras, Greece, 2008.

3 Denis Baggi, Adriano Baratè, Goffredo Haus, and Luca Andrea Ludovico. NINA—navigating and interacting with notation and audio. In *Proceedings of the International Workshop on Semantic Media Adaptation and Personalization (SMAP)*, pages 134–139, Washington, DC, USA, 2007. IEEE Computer Society.

4 David Bainbridge and Tim Bell. The challenge of optical music recognition. *Computers and the Humanities*, 35(2):95–121, 2001.

5 Adriano Baratè, Goffredo Haus, and Luca A. Ludovico. IEEE 1599: a new standard for music education. In *Proceedings of the International Conference on Electronic Publishing (ELPUB)*, pages 29–45, Milan, Italy, 2009.

6 Adriano Baratè, Luca A. Ludovico, and Alberto Pinto. An IEEE 1599-based interface for score analysis. In *Computer Music Modeling and Retrieval (CMMR)*, Copenhagen, Denmark, 2008.

7 Mark A. Bartsch and Gregory H. Wakefield. Audio thumbnailing of popular music using chroma-based representations. *IEEE Transactions on Multimedia*, 7(1):96–104, 2005.

8 Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, and Marius B. Spinu. Execution and synchronisation of music score pages and real performance audios. In *Proceedings of the IEEE Inter-*

national Conference on Multimedia and Expo (ICME), pages 125–128, Lausanne, Switzerland, 2002.

9 Pierfrancesco Bellini, Fabrizio Fioravanti, and Paolo Nesi. Managing music in orchestras. *Computer*, 32:26–34, 1999.

10 Esben Paul Bugge, Kim Lundsteen Juncher, Brian Søborg Mathiasen, and Jakob Grue Simonsen. Using sequence alignment and voting to improve optical music recognition from multiple recognizers. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 405–410, 2011.

11 Donald Byrd, William Guerin, Megan Schindele, and Ian Knopke. OMR evaluation and prospects for improved OMR via multiple recognizers. Technical report, Indiana University, Bloomington, Indiana, USA, 2010.

12 Donald Byrd and Megan Schindele. Prospects for improving OMR with multiple recognizers. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 41–46, Victoria, Canada, 2006.

13 Arshia Cont. A coupled duration-focused architecture for real-time music-to-score alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):974–987, 2010.

14 Antonello D'Aguanno and Giancarlo Vercellesi. Automatic music synchronization using partial score representation based on IEEE 1599. *Journal of Multimedia*, 4(1):19–24, 2009.

15 David Damm. *A Digital Library Framework for Heterogeneous Music Collections—from Document Acquisition to Cross-modal Interaction.* PhD thesis, University of Bonn (in preparation), 2012.

16 David Damm, Christian Fremerey, Verena Thomas, Michael Clausen, Frank Kurth, and Meinard Müller. A digital library framework for heterogeneous music collections—from document acquisition to cross-modal interaction. *International Journal on Digital Libraries: Special Issue on Music Digital Libraries (to appear)*, 2012.

17 Roger B. Dannenberg. An on-line algorithm for real-time accompaniment. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 193–198, 1984.

18 Johanna Devaney, Michael I. Mandel, and Daniel P. W. Ellis. Improving MIDI-audio alignment with acoustic features. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 45–48, New Paltz, NY, USA, 2009.

19 Karin Dressler and Sebastian Streich. Tuning frequency estimation using circular statistics. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 357–360, Vienna, Austria, 2007.

20 Zhiyao Duan and Bryan Pardo. Aligning semi-improvised music audio with its lead sheet. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 513–518, Miami, FL, USA, 2011.

21 Jon W. Dunn, Donald Byrd, Mark Notess, Jenn Riley, and Ryan Scherle. Variations2: Retrieving and using music in an academic setting. *Communications of the ACM, Special Issue: Music information retrieval*, 49(8):53–48, 2006.

22 Sebastian Ewert and Meinard Müller. Estimating note intensities in music recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 385–388, Prague, Czech Republic, 2011.

23 Sebastian Ewert and Meinard Müller. Score-informed source separation. In Meinard Müller, Masataka Goto, and Markus Schedl, editors, *Multimodal Music Processing (to appear)*, Dagstuhl Follow-Ups. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2012.

24 Sebastian Ewert, Meinard Müller, and Peter Grosche. High resolution audio synchronization using chroma onset features. In *Proceedings of the IEEE International Conference*

on Acoustics, Speech, and Signal Processing (ICASSP), pages 1869–1872, Taipei, Taiwan, 2009.

25 Christian Fremerey. *Automatic Organization of Digital Music Documents – Sheet Music and Audio*. PhD thesis, University of Bonn, 2010.

26 Christian Fremerey, Meinard Müller, and Michael Clausen. Handling repeats and jumps in score-performance synchronization. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 243–248, Utrecht, The Netherlands, 2010.

27 Christian Fremerey, Meinard Müller, Frank Kurth, and Michael Clausen. Automatic mapping of scanned sheet music to audio recordings. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 413–418, Philadelphia, USA, 2008.

28 Hiromasa Fujihara, Masataka Goto, Jun Ogata, and Hiroshi G. Okuno. Lyricsynchronizer: Automatic synchronization system between musical audio signals and lyrics. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1252 –1261, 2011.

29 Ichiro Fujinaga. Optical music recognition bibliography. `http://ddmal.music.mcgill.ca/wiki/Optical_Music_Recognition_Bibliography`, 2000.

30 Masataka Goto. A chorus-section detecting method for musical audio signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 437–440, Hong Kong, China, 2003.

31 Ning Hu, Roger B. Dannenberg, and George Tzanetakis. Polyphonic audio matching and alignment for music retrieval. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, US, 2003.

32 Min-Yen Kan, Ye Wang, Denny Iskandar, Tin Lay New, and Arun Shenoy. LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(2):338 –349, 2008.

33 Frank Kurth and Meinard Müller. Efficient index-based audio matching. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):382–395, 2008.

34 Frank Kurth, Meinard Müller, Christian Fremerey, Yoon-ha Chang, and Michael Clausen. Automated synchronization of scanned sheet music with audio recordings. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 261–266, Vienna, Austria, 2007.

35 John Lawter and Barry Moon. Score following in open form compositions. In *Proceedings of the International Computer Music Conference (ICMC)*, Ann Arbor, MI, USA, 1998.

36 Nicola Montecchio and Arshia Cont. A unified approach to real time audio-to-score and audio-to-audio alignment using sequential montecarlo inference techniques. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 193–196, Prague, Czech Republic, 2011.

37 Nicola Montecchio and Nicola Orio. A discrete filter bank approach to audio to score matching for polyphonic music. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 495–500, Kobe, Japan, 2009.

38 Meinard Müller and Daniel Appelt. Path-constrained partial music synchronization. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 65–68, Las Vegas, Nevada, USA, 2008.

39 Meinard Müller and Michael Clausen. Transposition-invariant self-similarity matrices. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 47–50, Vienna, Austria, 2007.

40 Meinard Müller and Sebastian Ewert. Joint structure analysis with applications to music annotation and synchronization. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 389–394, Philadelphia, Pennsylvania, USA, 2008.

41 Meinard Müller, Frank Kurth, David Damm, Christian Fremerey, and Michael Clausen. Lyrics-based audio retrieval and multimodal navigation in music collections. In *Proceedings of the European Conference on Digital Libraries (ECDL)*, pages 112–123, Budapest, Hungary, 2007.

42 Bernhard Niedermayer. Improving accuracy of polyphonic music-to-score alignment. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 585–590, Kobe, Japan, 2009.

43 Nicola Orio and François Déchelle. Score following using spectral analysis and hidden Markov models. In *Proceedings of the International Computer Music Conference (ICMC)*, Havanna, Cuba, 2001.

44 Bryan Pardo and William P. Birmingham. Modeling form for on-line following of musical performances. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 1018–1023, Pittsburgh, PA, USA, 2005.

45 Dennis Howard Pruslin. *Automatic recognition of sheet music*. PhD thesis, Massachusetts Institute of Technology, 1966.

46 Christopher Raphael. Music Plus One: A system for flexible and expressive musical accompaniment. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba, 2001.

47 Christopher Raphael. A hybrid graphical model for aligning polyphonic audio with musical scores. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 387–394, Barcelona, Spain, 2004.

48 Christopher Raphael. Music Plus One and machine learning. In *Proceedings of the International Conference on Machine Learning (ICML)*, Haifa, Israel, 2010.

49 Stanley Sadie, editor. *The New Grove Dictionary of Music and Musicians (second edition)*. Macmillan, London, 2001.

50 Markus Schäfer. Hochwertige automatische Extraktion von Gesangstext aus Notenbänden und mediensynchrone Darstellung. Diploma thesis, University of Bonn (in preparation), 2012.

51 Mevlut Evren Tekin, Christina Anagnostopoulou, and Yo Tomita. Towards an intelligent score following system: Handling of mistakes and jumps encountered during piano practicing. In *Computer Music Modeling and Retrieval (CMMR)*, pages 211–219, Pisa, Italy, 2005.

52 Verena Thomas, Christian Fremerey, Sebastian Ewert, and Michael Clausen. Notenschrift-Audio Synchronisation komplexer Orchesterwerke mittels Klavierauszug. In *Proceedings of the Deutsche Jahrestagung für Akustik (DAGA)*, pages 191–192, Berlin, Germany, 2010.

53 Verena Thomas, Christian Wagner, and Michael Clausen. OCR-based post-processing of OMR for the recovery of transposing instruments in complex orchestral scores. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 411–416, Miami, FL, USA, 2011.

54 Barry Vercoe. The synthetic performer in the context of live performance. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 199–200, 1984.

55 Guangyu Xia, Dawen Liang, Roger B. Dannenberg, and Mark J. Harvilla. Segmentation, clustering, and display in a personal audio database for musicians. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 139–144, Miami, FL, USA, 2011.