INTERDISCIPLINARY SCIENCE REVIEWS, Vol. 35 No. 2, June, 2010, 138–53

A Multimodal Way of Experiencing and Exploring Music

Meinard Müller and Verena Konz

Saarland University and MPI Informatik, Saarbrücken, Germany

Michael Clausen, Sebastian Ewert and Christian Fremerey

Department of Computer Science, Bonn University, Germany

Significant digitization efforts have resulted in large multimodal music collections, which comprise music-related documents of various types and formats including text, symbolic data, audio, image, and video. The challenge is to organize, understand, and search musical content in a robust, efficient, and intelligent manner. Key issues concern the development of methods for analysing, correlating, and annotating the available multimodal material, thus identifying and establishing semantic relationships across various music representations and formats. Here, one important task is referred to as music synchronization, which aims at identifying and linking semantically corresponding events present in different versions of the same underlying musical work. In this paper, we give an introduction to music synchronization and show how synchronization techniques can be integrated into novel user interfaces that allow music lovers and researchers to access and explore music in all its different facets thus enhancing human involvement with music and deepening music understanding.

KEYWORDS Synchronization, Alignment, Multimodality, User interfaces

Introduction

The last years have seen increasing efforts in building up comprehensive digital music collections, which contain large amounts of textual, visual, and audio data as well as a variety of associated data representations. In particular for Western classical music, three prominent examples of digitally available types of music representations are *sheet music* (available as digital images), *symbolic score data* (e.g., in the MusicXML, LilyPond, or MIDI format), and *audio recordings* (e.g., given as WAV or MP3). These three classes of representations complement each other by describing different characteristics

DOI 10.1179/030801810X12723585301110

of music. Sheet music, which in our context denotes a printable form of musical score notation, is used to visually describe a piece of music in a compact and human-readable form. This representation allows musicians to create a performance and musicologists to study structural, harmonic, or melodic aspects of the music that may not be obvious from mere listening. Symbolic score data, which can be parsed by computers, is typically used for triggering synthesizers to create music performances or for supporting larger scale analysis tasks that could not be done manually. Finally, an audio recording encodes the sound wave of an acoustic realisation, which allows the listener to playback a specific interpretation.

In the field of *music information retrieval* (MIR), great efforts are directed towards the development of techniques that allow users to access, explore, and experience music in all its different facets. For example, we can imagine that during playback of some CD recording, a digital music player of the future presents the corresponding musical score while highlighting the current playback position within the score. On demand, additional information about melodic and harmonic progression or rhythm and tempo is automatically presented to the listener. A suitable user interface displays the musical form of the current piece of music and allows the user to directly jump to any key part within the recording without tedious fast-forwarding and rewinding. Furthermore, the listener is equipped with a Google-like search engine that enables him or her to explore the entire music collection in various ways: the user creates a query by specifying a certain note constellation or some harmonic or rhythmic pattern by whistling a melody or simply by selecting a short passage from a CD recording; the system then provides the user with a ranked list of available music excerpts from the collection that are musically related to the query.

To make the various music sources accessible in a convenient, intuitive, and user-friendly way, various alignment and synchronization procedures have been proposed with the common goal to automatically unfold musically meaningful relations between various types of music representations. Here, the idea is to exploit the fact that there is an increasing number of relevant digital documents even for a single musical work. These documents may comprise various audio recordings, MIDI files, digitized sheet music, or symbolic score representations. In the musical context, *music synchronization* denotes a procedure which, for a given position in one representation of a piece of music, determines the corresponding position within another representation, thus coordinating the multiple information sources related to a given musical work (see Figure 1 for an illustration). Depending upon the respective data formats, one distinguishes between various synchronization tasks (Arifi *et al.* 2004, 9; Müller 2007). For example, *Audio-Audio* synchronization (Dixon and Widmer 2005, 492; Müller *et al.* 2006, 192; Turetsky and Ellis 2003) refers to the task of time-aligning two different audio recordings of a piece of music. These alignments can be used to jump freely between different interpretations, thus affording efficient and convenient audio browsing. The goal of *Score-Audio* and *MIDI-Audio* synchronization (Arifi *et al.* 2004, 9; Dannenberg and Hu 2003, 27; Müller *et al.* 2004, 365; Raphael 2004, 387; Soulez *et al.* 2003) is to coordinate note and MIDI events with audio data.

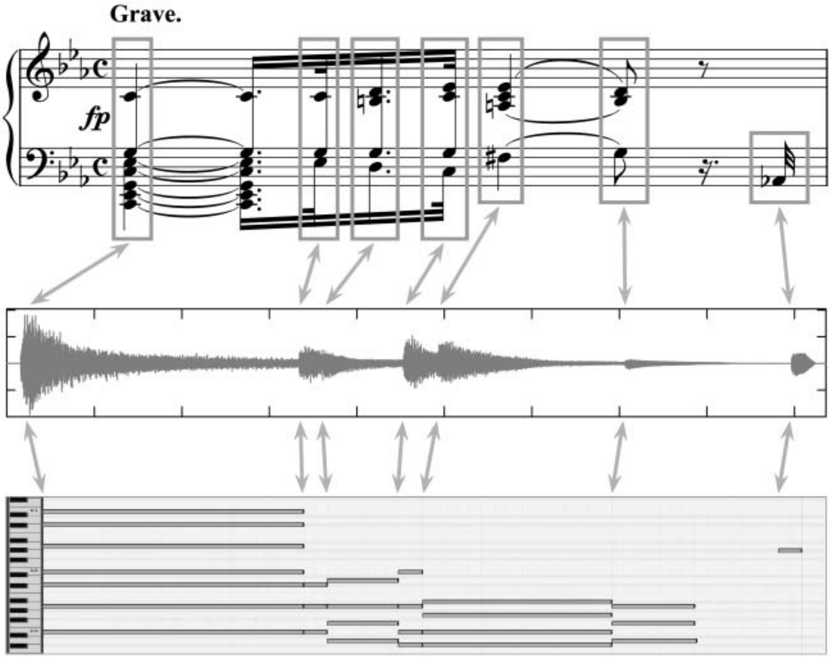


FIGURE 1 First measure of Beethoven's Piano Sonata Op. 13 (Pathétique). A scanned musical score, a waveform of a recording, as well as a MIDI file (as piano roll) are shown. The results of a SheetMusic-Audio and of a MIDI-Audio synchronization, respectively, are indicated by grey bidirectional arrows.

The result can be regarded as an automated annotation of the audio recording with available score and MIDI data. A recently studied problem is referred to as *SheetMusic-Audio* synchronization (Kurth *et al.* 2007, 261), where the objective is to link regions (given as pixel coordinates) within the scanned images of given sheet music to semantically corresponding physical time positions within an audio recording. Such linking structures can be used to highlight the current position in the scanned score during playback of the recording. Similarly, the goal of *Lyrics-Audio* synchronization (Fujihara *et al.* 2006, 257; Müller *et al.* 2007, 112; Wang *et al.* 2004, 212) is to align given lyrics to an audio recording of the underlying song, which is useful not only for retrieval but also for karaoke applications. For an overview of related alignment and synchronization problems, we also refer to Dannenberg and Raphael (2006, 39) and Müller (2007).

Our goal in this paper is to indicate how music synchronization techniques can be applied to enhance the human involvement with music. In particular, we discuss three different case studies where automated synchronization methods play an important role for supporting the user in experiencing and exploring music. In the first study, we introduce a novel user interface for multimodal (audio-visual) music presentation and navigation. Our system offers high-quality audio playback with time-synchronous display of the digitized sheet music associated with a musical work. In our second case study, we report on an experiment conducted at the University of Music Saarbrücken with the goal of introducing a novel MIR user interface to music education and to get feedback from music experts. To this end, we asked several music students to play the same piece of music. The different interpretations were recorded, aligned, and integrated in a user interface, which allows for synchronous playback of the different performances. Upon using this interface, the music students were then asked to analyse the performances according to a well designed questionnaire. Doing so, we not only tested and evaluated our interface in a setting of practical relevance, but also indicated the potential of MIR methods in music education. Finally, in our third study, we show how synchronization techniques can be used for extracting temporal information from a music recording in a fully automated fashion. This information is given in the form of a tempo curve that reveals the relative tempo difference between an actual performance and some reference representation of the underlying musical piece. Before describing these three case studies in more detail, we give a brief introduction to music synchronization. Further problems and prospects on future work are discussed in the conclusions of this paper.

Music synchronization

As was already mentioned in the introduction, a musical work is far from simple or singular. In particular, there may exist various audio recordings, MIDI files, digitized sheet music, and other symbolic representations. The general goal of *music synchronization* is to automatically link the various data streams thus interrelating the multiple information sets related to a given musical work (Müller 2007). More precisely, *synchronization* is taken to mean a procedure which, for a given position in one representation of a piece of music, determines the corresponding position within another representation. The result of a synchronization process is illustrated by Figure 1 in the form of grey bidirectional arrows. Automated music synchronization constitutes a challenging research field since one has to account for a multitude of aspects such as the data format, the genre, the instrumentation, or differences in parameters such as tempo, articulation and dynamics that result from expressiveness in performances. In the design of synchronization algorithms, one has to deal with a delicate trade-off between robustness, temporal resolution, alignment quality, and computational complexity. In order to synchronize two different music representations, one typically proceeds in two steps, which are explained next. For details, we refer to Müller (2007).

In the first step, the two music representations are transformed into sequences of suitable features. Here, on the one hand, the feature representations should show a large degree of robustness to variations that are to be left unconsidered in the comparison. On the other hand, the feature representations should capture characteristic information that suffices for synchronization. In this context, chroma-based music features have turned out to be a powerful tool for synchronizing harmony-based music (Bartsch and Wakefield 2005, 96). Here, the chroma refer to the twelve traditional pitch classes of the equal-tempered scale encoded by the attributes C, C$^{\#}$, D, . . ., B.

Representing the short-time content of a music representation in each of the twelve pitch classes, chroma features show a large degree of robustness to variations in timbre and dynamics, while keeping sufficient information to characterize harmony-based music.

In the second step, the derived feature sequences have to be brought into temporal correspondence to account for temporal variations in the two music representations to be synchronized. An important technique for computing such a correspondence is dynamic time warping (DTW), which is a well-known technique to find an optimal alignment between two given (time-dependent) sequences under certain restrictions. Intuitively, the alignment can be thought of as a linking structure indicated by bidirectional arrows, as shown in Figure 1. These arrows encode how the sequences are to be warped (in a non-linear fashion) to match each other. For a detailed account of music synchronization and further links to the literature, we refer to Müller (2007). In the next section, the concrete scenario of synchronizing scanned sheet music and audio recordings is discussed in more detail.

Multimodal music navigation

Sheet music and audio recordings represent and describe music on different levels of abstraction. Sheet music specifies high-level parameters such as notes, keys, measures, or repeats in a visual form. Because of its explicitness and compactness, most musicologists discuss and analyse the meaning of music on the basis of sheet music. On the other hand, most people enjoy music by listening to audio recordings, which represent music in an acoustic form. In particular, the nuances and subtleties of musical performances, which are generally not written down in the score, make the music come alive. In this section, we discuss the problem of *SheetMusic-Audio synchronization*, which is one way for automatically bridging the gap between the sheet music domain and the audio domain (Fremerey *et al.* 2008, 413; Kurth *et al.* 2007, 261). Here, the synchronization task is to link regions (given as pixel coordinates) within the scanned images of given sheet music to semantically corresponding physical time positions within an audio recording (see Figure 1). Such linking structures can be used to highlight the current position in the scanned score during playback of the recording, thus enhancing the listening experience as well as providing the user with tools for intuitive and multimodal music exploration. The importance of such a functionality has also been emphasized in the context of accessing and searching data in digital music libraries (Dunn *et al.* 2006, 53).

In the following, we consider two scenarios that are of great practical importance; we refer to Figure 2 for an overview. In the first scenario, one is given an audio recording and scanned images of the sheet music. Using optical music recognition (OMR), symbolic score data is generated from the sheet music. This process is similar to the well-known optical character recognition (OCR), where textual content is extracted from an image. In the second scenario, one is given an audio recording and symbolic score data (e.g., in the MusicXML or LilyPond format). In this case, a sheet music image has to be rendered from the symbolic score data. In both scenarios, the

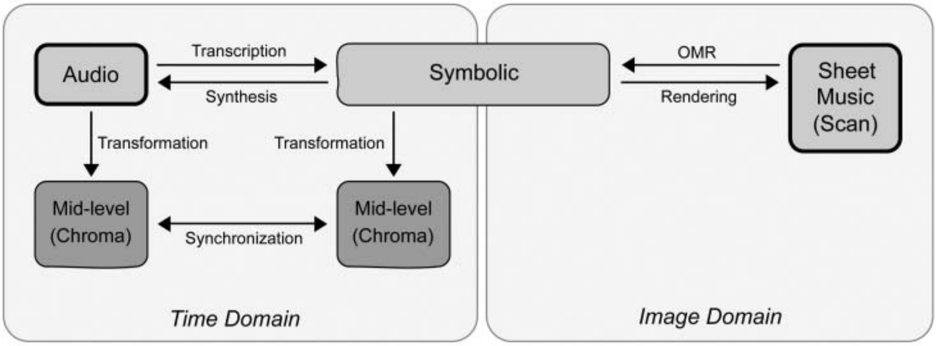
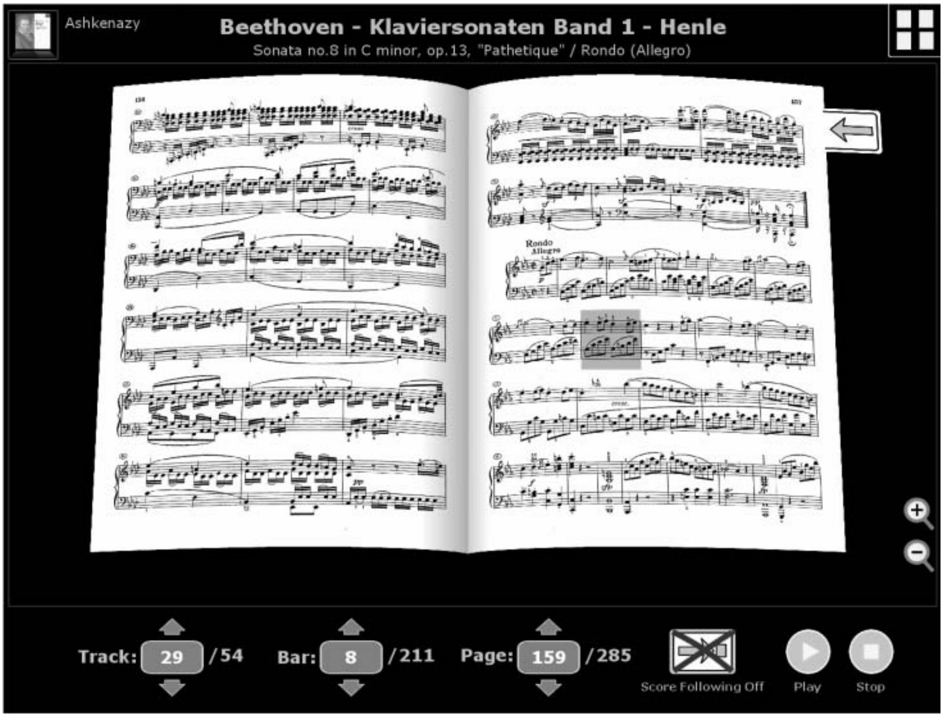


FIGURE 2 Illustration of data types and data transformations relevant for SheetMusic-Audio synchronization.

connection between the audio data and the symbolic data is realised through the same mechanisms. The basic idea is to transform both the symbolic score data as well as the corresponding audio recording into a common mid-level representation, which can then be synchronized as described in the previous section. Note that in both scenarios, one based on OMR and the other based on rendering, one requires an explicit mapping between the musical objects given by the symbolic representation and the 2D coordinates of their depicted counterparts in the image representation. Using this mapping and the synchronization result, a correspondence between spatial regions in the sheet music and temporal regions in the audio recording can be derived. The quality of the resulting synchronization depends on several factors. In particular, differences between the audio and score representation may have a crucial impact on the final synchronization result. Here, such differences may be due to extraction errors in the OMR step. Furthermore, the actual interpretation may deviate from the notated score.

Having computed the links between the images and the audio, one can realise novel ways of experiencing and browsing music. As an example, we present our ScoreViewer interface for presenting sheet music while playing back associated audio recordings, depicted in Figure 3. Here, the main visualization mode is illustrated for two scanned pages of Beethoven's Piano Sonata Op. 13 (Pathétique). When starting audio playback, corresponding measures within the sheet music are synchronously highlighted based on the linking information generated by the SheetMusic-Audio alignment. In Figure 3, a region in the centre of the right page, corresponding to the eighth measure of the 3rd movement (Rondo), is currently highlighted by a surrounding box. When reaching the end of 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 ScoreViewer interface allows for navigating 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



Ashkenazy
Beethoven - Klaviersonaten Band 1 - Henle
Sonata no.8 in C minor, op.13, "Pathetique" / Rondo (Allegro)
Rondo
Allegro
Track: 29 /54
Bar: 8 /211
Page: 159 /285
Score Following Off
Play
Stop

FIGURE 3 The ScoreViewer interface for multimodal music presentation and navigation (from Kurth *et al.* 2008, 334). Synchronously to audio playback, corresponding musical measures within the sheet music are highlighted.

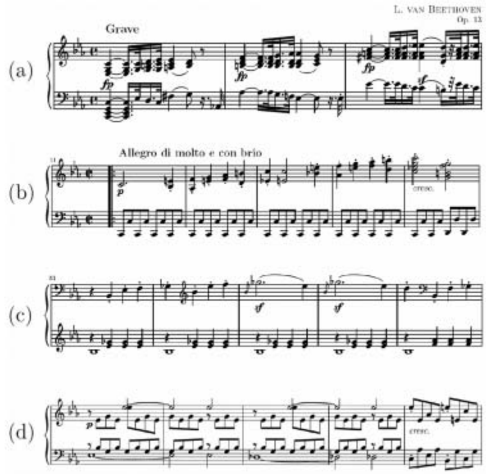
playback. If more than one associated audio recording is available for the currently active musical work, the user may switch between those using an icon list that is shown by clicking on the current icon. The ScoreViewer interface described above is convenient for enjoying a musical work in a multimodal 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 representation yields an intuitive way to search specific parts and to change the playback position in the audio representation.

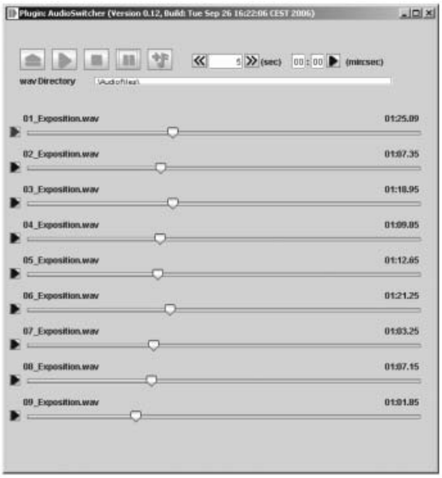
Music education

Computers have become an indispensable tool for storing, processing, and generating music. Even though computer-based methods and interfaces are ubiquitously used for music synthesis, there is still a reluctance in using computers for music education. Research in computer-assisted music education started at the end of the 1960s mainly in the USA and the UK (Brown 1999; 2007; Kiraly 2003, 41; Smith 2009, 59; Stevens 1991, 24).
For example, computers have served as a tool for creative music-making. Furthermore, the method of Computer-Assisted-Instruction (CAI), where the

students are taught a particular skill by a computer, has been applied in areas such as music theory or aural training. Various studies have been conducted to investigate the effect of CAI-based methods within music education (Smith 2009, 59; Kiraly 2003, 41), and the usefulness of such methods seems to be a controversial issue. In particular, musicians and music teachers often seem to be sceptical about the benefits of computer-based methods. Under such circumstances, it remains a challenge to raise the interest of music educators for using, testing, and participating in the development of novel MIR interfaces and for discussing possible application scenarios. With the objective of introducing computer-assisted methods to music education and of testing and evaluating novel user interfaces within a setting of practical relevance, we conducted a first investigation in collaboration with the University of Music Saarbrücken. The experiment was conducted in several steps, which we now describe in more detail.

First, nine piano students from the same piano class (Professor Thomas Duis) were asked to practise the same piece of music. Here, we chose the first movement of Beethoven's Piano Sonata Op. 13 (Pathétique), see Figure 4. This piece is not only a musically outstanding work, which often belongs to the curriculum in a pianist's education, but it is also very rich in contrasts of tempo and dynamics. This offers the pianist a wide range of possibilities in interpreting and shaping the piece by varying certain musical parameters during playing. All of the nine students were recorded playing this piece on the same piano and under the same recording conditions. In each of the nine recording sessions, only the performer, the technical staff, and the scientific investigators were present in the room — the other performing students were




Plugin: AudioSwitcher (Version 0.12, Build: Tue Sep 26 16:22:06 CEST 2006)

« 5 » (sec) 00 : 00 ▶ (micsec)

wav Directory \AudioFiles\

01_Exposition.wav 01:25.09
02_Exposition.wav 01:07.35
03_Exposition.wav 01:18.95
04_Exposition.wav 01:09.85
05_Exposition.wav 01:12.65
06_Exposition.wav 01:21.25
07_Exposition.wav 01:03.25
08_Exposition.wav 01:07.15
09_Exposition.wav 01:01.85

FIGURE 4 First movement of Beethoven's Piano Sonata Op. 13 (Pathétique). The left side shows the beginning of the introduction (measures 1 ff.), of the first theme (measures 11 ff.), of the second theme (measures 51 ff.), and of some connecting passage (measures 89 ff.). The right side shows an instance of the Interpretation Switcher interface for synchronous playback of different audio recordings of the same piece of music. In this example, nine different recordings of the exposition of Beethoven's Pathétique are opened.

not allowed to listen to their fellow students. Being in several education stages, the students played on different performance levels.

In the second step, the nine audio recordings were temporally aligned, and integrated in a user interface referred to as Interpretation Switcher (Fremerey *et al.* 2007, 131; Müller 2007). This interface allows the user to select several previously synchronized recordings of the same piece of music (see Figure 4). Each of the selected recordings is represented by a slider bar indicating the current playback position with respect to the recording's particular timescale. The audio recording that is currently used for audio playback, in the following referred to as reference recording, is represented by a red marker. The slider of the reference recording moves at constant speed while the sliders of the other recordings move according 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 marker located on the left of each slider. 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. One can also jump to any position within any of the recordings by directly selecting the position of the respective slider. This also automatically triggers a switch of the reference to the respective recording. In our experiment, we restricted ourselves basically to the switching functionality. Here, our motivation was to keep the interface as simple and intuitive as possible to avoid any rejections from the user side. After a short explanation of the main switching functionality, none of the music students reported any difficulties in using our interface.

In the next step of our experiment, each student was provided with a computer running the Interpretation Switcher interface and with earphones. Upon using this interface, the music students were asked to analyse the anonymized performances according to a well designed questionnaire. Here, the questions were designed in such a way that the participants naturally started to use the switching functionality of the interface, thus getting familiar with the Interpretation Switcher in a concrete application of musical relevance. The questionnaire consisted of two main parts. In the first part, the students had to listen, to compare, and to rate the nine different interpretations with respect to various performance aspects. Among others, the students were asked to rate the performances (using a rating scale ranging between 1 and 10) within specific passages (the ones indicated in Figure 4) and with respect to the different performance aspects such as dynamics, articulation, and agogics. While doing so, the participants not only had to switch between the different performances but also to find the corresponding entry points of the various passages within the recordings. With this task, the students had to constantly switch between and jump within the audio recordings so that they were forced to use the functionality of the Interpretation Switcher interface extensively. In the second part of the questionnaire, the participants were then asked to give feedback on the usefulness and operability of the Interpretation Switcher itself. In this respect, the performance evaluation in the first part served as a means to confront the students with our novel interface without being aware of the interface evaluation.

Finally, we evaluated the completed questionnaires with respect to the ratings given for the performances as well as for the interface. As it turned out, most of the participants found the handling and functioning of the Interpretation Switcher very intuitive, even music students who had had only little experience with computers. Furthermore, most students found the Interpretation Switcher very useful for tasks such as performance analysis, music comparison, and other analysis tasks. One student remarked that he would have appreciated having had an additional functionality for displaying the musical score during playback, similar to the one described in the previous section. Also user-defined auxiliary markers that can be freely fixed, adjusted, and removed along the various slider control bars should be introduced for additional orientation and navigation purposes. All but one of the participants affirmed that they could imagine using the Interpretation Switcher within their studies or even for private use. In particular, they said that the interface may be useful in the context of special seminars, where the comparison of different performances plays an important role. One student was enthusiastic about the features offered by the Interpretation Switcher. He usually recorded his piano lessons in order to listen to and study his own playing afterwards. In such a context, he would significantly benefit from novel switching and navigation functionalities for comparing and analysing the recorded audio material. Also, the Interpretation Switcher could be very useful for compactly documenting the history of learning progress over a longer period of time. For example, it could synchronously present the various performances of a specific musical passage recorded in different piano lessons over the semester.

In conclusion, this first experiment indicated that computer-based interfaces may constitute valuable tools for supporting the learning process. Most participants affirmed the usefulness of our interface for comparing and analysing performances or simply for music listening and enjoyment. By testing and evaluating our interface within a concrete application of practical relevance, we not only acquainted a new group of prospective users with MIR methods but also obtained valuable feedback from music experts. As a side-effect of our experiment, we also obtained audio material that could be used freely for research purposes without any copyright restrictions. Having various interpretations of the same piece of music makes this material interesting for tasks such as automated performance analysis (Sapp 2007, 497; Widmer *et al.* 2003, 111). Also, the material has been annotated by means of the ratings obtained from the performance evaluation. Finally, using a Yamaha Disklavier, we also obtained MIDI data along with the audio recordings, which is valuable ground truth material for transcription (Klapuri and Davy 2006) or synchronization tasks (Müller 2007). The experiment presented here only constitutes the beginning of a planned collaboration with music educators and students who are usually not aware of the developments in music information retrieval. For the future, we plan to conduct similar experiments on a larger scale. One further idea is to participate regularly in the lessons of piano students to record their playing. We then plan to process (segment, classify, synchronize) the audio material automatically and to integrate it suitably in our Interpretation Switcher to document and to analyse the students' learning process. Furthermore, we plan to integrate and test novel navigation and analysis functionalities in the user interface for supplying appropriate feedback mechanisms which can be used either in the music lessons directly or in subsequent self-studies. Such tools should help the students not only to question and to analyse their own playing style, but also to directly compare it with the playing style of different pianists.

Tempo estimation

In the previous two application scenarios, we have shown how music synchronization techniques are of fundamental importance for implementing interfaces that allow users to browse, compare, or simply enjoy music in its various manifestations. We now indicate how synchronization results can also be used for automatically extracting temporal performance attributes from expressive music recordings. The motivation for such a task is that musicians give a piece of music their personal touch by continuously varying tempo, dynamics, and articulation. Instead of playing mechanically they speed up at some places and slow down at others in order to shape a piece of music. Similarly, they continuously change the sound intensity and stress certain notes. Such performance issues are of fundamental importance for the understanding and perception of music. The automated analysis of different interpretations, also referred to as *performance analysis*, has become an active field of research (Langner and Goebl 2003, 69; Sapp 2007, 497; Widmer *et al.* 2003, 111). Here, one goal is to find commonalities between different interpretations, which allow for the derivation of general performance rules. A kind of orthogonal goal is to capture what is characteristic for the style of a particular interpreter. Before one can analyse a specific performance, one requires the information about when and how the notes of the underlying piece of music are actually played. Therefore, as the first step in performance analysis, one has to annotate the performance by means of suitable attributes that make explicit the exact timing and intensity of the various note events. The extraction of such performance attributes constitutes a challenging problem, in particular, in the case of audio recordings.

Many researchers manually annotate the audio material by marking salient data points in the audio stream. However, being very labour-intensive, such a manual process is prohibitive in view of large audio collections. Another way to generate accurate annotations is to use a computer-monitored *player piano*. Equipped with optical sensors and electromechanical devices, such pianos allow for recording the key movements along with the acoustic audio data, from which one directly obtains the desired note onset information. The advantage of this approach is that it produces precise annotations, where the symbolic note onsets perfectly align with the physical onset times. The obvious disadvantage is that special-purpose hardware is needed during the recording of the piece. In particular, conventional audio material taken from CD recordings cannot be annotated in this way. Therefore, the most preferable method is to automatically extract the necessary performance aspects directly from a given audio recording. Here, automated approaches such as *beat tracking* (Dixon 2007, 39) and *onset detection* (Bello *et al.* 2005,

1035) are used to estimate the precise timings of note events within the recording. Even though great research efforts have been directed towards such tasks, the results are still unsatisfactory, in particular, for music with weak onsets and strongly varying beat patterns. In practice, semi-automatic approaches are often used, where one first roughly computes beat timings using beat tracking software, which are then adjusted manually to yield precise beat onsets.

Now, instead of trying to derive the tempo information only on the basis of a given music recording, one can exploit the fact that, for many pieces, there exists a kind of 'neutral' representation, which can be used as a reference. Such a reference representation may be given in the form of a musical score or MIDI file, where the notes are played with a known constant tempo (measured in beats per minute or BPM) in a purely mechanical way. Using music synchronization techniques, one can temporally align the MIDI note events with their corresponding physical occurrences in the music recording. From the synchronization result, it is possible to derive a *tempo curve* that reveals the relative tempo differences between the actual performance and the neutral MIDI reference representation. Assuming that the time signature of the piece is known, one can recover measure and beat positions from MIDI time positions. This information suffices to convert the relative values given by the tempo curve into musically meaningful absolute values. As a result, one obtains a tempo curve that describes for each musical position (given in beats and measures) the absolute tempo of the performance (given in BPM).

As an example, we consider the Schubert song *Der Lindenbaum* (D. 911 No. 5). For this piece, we computed tempo curves for thirteen different interpretations. The first seven measures (piano introduction) as well as the corresponding parts of the tempo curves are shown in Figure 5. As indicated by the curves, all interpretations exhibit an accelerando in the first few measures followed by a ritardando towards the end of the introduction. Interestingly, some of the pianists start with the ritardando in measure 4, whereas most of the other pianists play a less pronounced ritardando in measure 6. This example indicates that the automatically extracted tempo curves are accurate enough to reveal interesting performance characteristics.

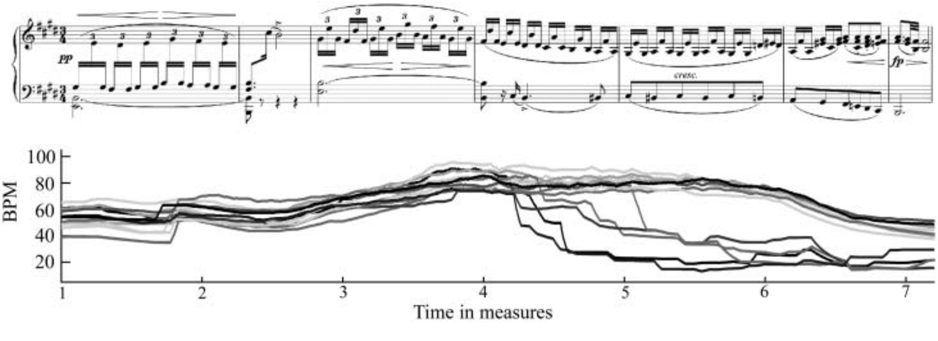


FIGURE 5 Tempo curves of 13 different performances shown for the first seven measures of the Schubert song *Der Lindenbaum*. The tempo is given in beats per minute (BPM).

We close this section with a more philosophical and critical discussion of the limitations of automated tempo extraction approaches. Generally speaking, the feeling of pulse and rhythm is one of the central components of music closely relating to what one commonly refers to as tempo. In order to define some notion of tempo, one requires a proper reference to measure against. Western music, for example, is often structured in terms of measures and beats, which allows for organizing and sectioning musical events over time. Based on a fixed time signature, one can then define the tempo as the number of beats per minute. Obviously, this definition requires a regular and steady musical beat or pulse over a certain period in time. Also, the very process of measurement is not as well defined as one may think. Which musical entities (e.g., note onsets) characterize a pulse? How precisely can these entities be measured before getting drowned in noise? How many pulses or beats are needed to obtain a meaningful tempo estimation? With these questions, we want to indicate that the notion of tempo is far from being well defined. Furthermore, due to discretization and synchronization errors, one needs numerically robust procedures to extract the tempo information by using average values over suitable time windows. Here, the window size constitutes a delicate trade-off between susceptibility to synchronization errors and sensibility towards timing nuances of the performance. In practice, it becomes a difficult problem to determine whether a given change in the tempo curve is caused by synchronization errors or whether it is the result of an actual tempo change in the performance. As shown by our experiments on harmony-based Western music, our approach allows for capturing at least the overall tempo flow and for certain classes of music, even finer expressive tempo nuances.

Conclusions

The task of music synchronization is of fundamental importance for bridging the gap between various music representations by automatically finding semantically meaningful correspondences between various instances of the same musical work. In the synchronization context, we have discussed three case studies that indicate how alignment techniques in combination with novel user interfaces can greatly enhance the human involvement with music.

A synchronous audio-visual presentation of scanned sheet music and audio recordings offers a novel way for experiencing and exploring music. While listening to an interpretation of a piece of music, the user can visually track the corresponding notes or measures within the sheet music representation. Furthermore, spatio-temporal links between sheet music and audio data facilitate multimodal music interaction, where a user can exploit the benefits of each of the modalities to access music data in a more convenient way. For example, the user may revert to the visual domain by marking certain measures in the sheet music, which can then be used to search in the acoustic domain by identifying all corresponding passages within available audio recordings, where these measures are actually performed. For the future, we plan to extend our synchronization framework to account for various formats comprising text, audio, image, and video. Here, our vision is to supply the user with multimodal interfaces that allow him or her to interact, explore, compare, and relate any type of music-related document. For example, listening to an opera, the interface presents the music, the staging, as well as the lyrics along with structural information that relates the various scenes of the opera and facilitates convenient navigation. Based on the presented data, the user may trigger a search using any combination of audio, video, or text elements to retrieve information from the worldwide web or from digital music libraries on the composer, the musicians, the musical work, or a specific performance.

In our second case study, we described an experiment conducted at the University of Music Saarbrücken with the main objective of introducing MIR user interfaces with novel switching and navigation functionalities to music teachers and students. Even though this group tends to be sceptical about using computer-based methods in music education, most participants affirmed the usefulness of such interfaces for comparing and analysing performances or simply for music listening and enjoyment. In co-operation with music educators and students, we plan to extend the navigation and analysis functionalities by supplying appropriate feedback mechanisms, which can be used either in the music lessons directly or in subsequent self-studies. Such tools should help the students not only to question and analyse their own playing style, but also to compare it directly with the playing style of different musicians. In introducing novel functionalities, one main challenge will be to keep the operability of the interface as simple and intuitive as possible to avoid overstrains and rejections from the users' side.

Finally, we indicated how automated synchronization methods can be used to extract tempo curves from expressive music recordings by comparing the performances with neutral reference representations. For the future, we plan to extract other performance aspects on dynamics and articulation using alignment information as well. In this context, we have seen that it is of crucial importance to improve further the temporal accuracy of synchronization strategies in order to recover even the fine agogic timing nuances of a performance, which often make the difference between an average and a master performance. It is our vision to develop and supply automated methods and interfaces that support musicologists, musicians, and music lovers in their effort to gain a deeper understanding of music. Here, music synchronization techniques constitute a powerful tool for generating musical meaning by identifying, relating, and aggregating the various music sources given for a music work.

Bibliography

Arifi, Vlora, Michael Clausen, Frank Kurth, and Meinard Müller. 2004. Synchronization of music data in score-, MIDI- and PCM-format. *Computing in Musicology* 13: 9–33.

Bello, Juan Pablo, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. 2005. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing* 13(5): 1035–47.

Brown, Andrew. 1999. Music, media and making: humanising digital media in music education. *International Journal of Music Education* 33(1): 10–7.

Brown, Andrew. 2007. *Computers in Music Education*. London: Routledge.

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

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

Dannenberg, Roger, and Ning Hu. 2003. Polyphonic audio matching for score following and intelligent audio editors. *Proceedings of the ICMC, San Francisco, USA*, 27–34.

Dixon, Simon. 2007. Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research* 36: 39–50.

Dannenberg, Roger, and Christopher Raphael. 2006. Music score alignment and computer accompaniment. *Communications of the ACM, Special Issue* 49(8): 39–43.

Dixon, Simon, and Gerhard Widmer. 2005. MATCH: A music alignment tool chest. *Proceedings of the ISMIR, London, GB*, 492–7.

Fujihara, Hiromasa, Masataka Goto, Jun Ogata, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno. 2006. Automatic synchronization between lyrics and music CD recordings based on Viterbi alignment of segregated vocal signals. *Proceedings of the Eighth IEEE International Symposium on Multimedia*, 257–64.

Fremerey, Christian, Frank Kurth, Meinard Müller, and Michael Clausen. 2007. A demonstration of the SyncPlayer system. *Proceedings of the ISMIR, Vienna, Austria*, 131–2.

Fremerey, Christian, Meinard Müller, Frank Kurth, and Michael Clausen. 2008. Automatic mapping of scanned sheet music to audio recordings. *Proceedings of the ISMIR, Philadelphia, USA*, 413–8.

Klapuri, Anssi, and Manuel Davy. 2006. *Signal processing methods for music transcription*. Springer.

Kurth, Frank, David Damm, Christian Fremerey, Meinard Müller, and Michael Clausen. 2008. A framework for managing multimodal digitised music collections. *Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries (ECDL)*, 334–45.

Kurth, Frank, Meinard Müller, Christian Fremerey, Yoon-Ha Chang, and Michael Clausen. 2007. Automated synchronization of scanned sheet music with audio recordings. *Proceedings of the ISMIR, Vienna, Austria*, 261–6.

Kiraly, Zsuzsanna. 2003. Solfeggio 1: a vertical ear training instruction assisted by the computer. *International Journal of Music Education* 40(1): 41–58.

Langner, Jörg, and Werner Goebl. 2003. Visualizing expressive performance in tempo-loudness space. *Computer Music Journal* 27(4): 69–83.

Müller, Meinard, Frank Kurth, David Damm, Christian Fremerey, and Michael Clausen. 2007. Lyrics-based audio retrieval and multimodal navigation in music collections. *Proceedings of the 11th European Conference on Digital Libraries (ECDL)*, 112–23.

Müller, Meinard, Frank Kurth, and Tido Röder. 2004. Towards an efficient algorithm for automatic score-to-audio synchronization. *Proceedings of the ISMIR, Barcelona, Spain*, 365–72.

Müller, Meinard, Henning Mattes, and Frank Kurth. 2006. An efficient multiscale approach to audio synchronization. *Proceedings of the ISMIR, Victoria, Canada*, 192–7.

Müller, Meinard. 2007. *Information retrieval for music and motion*. Springer.

Raphael, Christopher. 2004. A hybrid graphical model for aligning polyphonic audio with musical scores. *Proceedings of the ISMIR, Barcelona, Spain*, 387–94.

Sapp, Craig Stuart. 2007. Comparative analysis of multiple musical performances. *Proceedings of the ISMIR, Vienna, Austria*, 497–500.

Smith, Kenneth H. 2009. The effect of computer-assisted instruction and field independence on the development of rhythm sight-reading skills of middle school instrumental students. *International Journal of Music Education* 27(1): 59–68.

Soulez, Ferréol, Xavier Rodet, and Diemo Schwarz. 2003. Improving polyphonic and poly-instrumental music to score alignment. *Proceedings of the ISMIR, Baltimore, USA*, 143–148.

Stevens, Robin S. 1991. The best of both worlds: An eclectic approach to the use of computer technology in music education. *International Journal of Music Education* 17(1): 24–36.

Turetsky, Robert J., and Daniel P. W. Ellis. 2003. Ground-truth transcriptions of real music from force-aligned MIDI syntheses. *Proceedings of the ISMIR, Baltimore, USA*, 135–141.

Widmer, Gerhard, Simon Dixon, Werner Goebl, Elias Pampalk, and Asmir Tobudic. 2003. In search of the Horowitz factor. *AI Magazine* 24(3): 111–30.

Wang, Ye, Min-Yen Kan, Tin Lay Nwe, Arun Shenoy, and Jun Yin. LyricAlly: automatic synchronization of acoustic musical signals and textual lyrics. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on multimedia*, 212–219. New York, NY, USA: ACM Press.

Notes on contributors

Meinard Müller is a member of the Saarland University and the Max-Planck Institut für Informatik where he leads the research group Multimedia Information Retrieval and Music Processing within the Cluster of Excellence on Multimodal Computing and Interaction.
Correspondence to: meinard@mpi-inf.mpg.de

Michael Clausen has been Professor at the Department of Computer Science at Bonn University since 1989, where he leads the research group Multimedia Signal Processing. Among his special research interests are music information retrieval and digital music libraries.

Verena Konz studied music (Staatsexamen) at the Hochschule für Musik Köln and mathematics (Diplom) at Cologne University, Germany. Since May 2008, she has been a PhD student within the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University working in the field of Music Information Retrieval.

Sebastian Ewert finished his Master degree (Diplom) in computer science with a topic on music synchronization and currently pursues his doctoral degree at Bonn University under the supervision of Meinard Müller. His research interests cover music information retrieval, source separation, and applications of machine learning techniques to automated music processing.

Christian Fremerey is a PhD student in Professor Clausen's research group. His research interests concern the automatic linking and synchronization of sheet music and audio data, the implementation and integration of Music Information Retrieval tools into digital library workflows, as well as the development of intuitive and application-oriented user interfaces.