A MULTI-PERSPECTIVE USER INTERFACE FOR MUSIC SIGNAL ANALYSIS

Meinard Müller, Verena Konz, Nanzhu Jiang, Zhe Zuo

Saarland University and MPI Informatik
Campus E 1 4, 66123 Saarbrücken, Germany
{meinard, vkonz, njiang, zzuo}@mpi-inf.mpg.de

ABSTRACT

In view of the exploding distribution of digitized audio material, computer-based methods have become indispensable for processing and analyzing the content of music signals. To evaluate analysis results obtained by automated methods, one requires manually generated high-quality labeled data and the feedback by music experts. In this paper, we introduce various novel functionalities for a user interface that opens up new possibilities for viewing, comparing, interacting, and evaluating analysis results within a multi-perspective framework and bridges the gap between signal processing and music sciences. Here, we exploit the fact that a given piece of music may have multiple, closely-related sources of information including different audio recordings and score-like MIDI representations. Our interface then allows a user to interactively generate unifying views of the analysis results across the available music representations. Disclosing musically relevant consistencies and inconsistencies, these views not only afford new evaluation and navigation possibilities but also deepen a user's understanding of the underlying musical material.

1. INTRODUCTION

Significant digitization efforts and internet-based distribution have resulted in huge and unstructured audio collections which comprise music-related documents of various types and formats. In this context, the development of computer-based methods for extracting musically meaningful information from audio material has become a major research strand in the field of *music information retrieval* (MIR). For example, a central MIR task is known as *chord recognition*, where a given audio recording is analyzed with regard to its local harmonic content [14]. Another prominent analysis task is referred to as *music structure analysis* with the goal to divide an audio recording into temporal segments corresponding to musical parts and to group these segments into musically meaningful categories [13]. To evaluate analysis results obtained from automated methods, one requires reliable ground-truth annotations, which often have to be generated by musically trained listeners in a tedious, manual process. Furthermore, when conducting a direct user-centered evaluation, one requires feedback by domain experts such as musicians or music teachers—groups that are often reluctant in using novel computer-assisted methods and user interfaces [10].

In this paper, we introduce a user interface that facilitates novel ways of viewing, comparing, and evaluating analysis results obtained from different methods and computed on the basis of different music representations. Here, we exploit the fact that for a given piece of music one often has multiple, closely-related sources of information, including audio recordings of different performances and score-like representations including MIDI versions. Our interface combines and extends the functionality of known user interfaces for inter- and intra-document navigation [1, 2, 4, 16]. The technical backbone of our interface is the *Interpretation Switcher* [3], which allows a user to select several recordings of the same piece of music and, during playback, to seamlessly switch between these versions (inter-document navigation). We extended this switcher to additionally visualize version-dependent annotations such as chord labels or structure blocks, which can be used for intra-document navigation similar to [4]. As one main contribution, we introduce different modes for adjusting the version-dependent timelines of the music representations. Furthermore, our interface allows for interactively generating multi-perspective views across the different version-dependent analysis results disclosing consistencies and inconsistencies. This allows a user to conveniently locate, playback, and compare musically interesting passages, which not only makes evaluation and annotation easier but also deepens the listener's understanding of the annotations and the underlying audio material. Here, our interface not only allows a technically unexperienced user to interact with the music analysis results and the audio material, but also opens up new possibilities for enriching music education using signal processing techniques.

The remainder of this paper is organized as follows. We start by reviewing the concept of the original Interpretation Switcher while summarizing the underlying music synchronization techniques (Section 2). We then introduce the novel functionality that allows for switching between different timeline modes (Section 3). The usefulness of this functionality is illustrated by means of a case study using Beethovens's Pathétique as example (Section 4). We then discuss the second functionality that

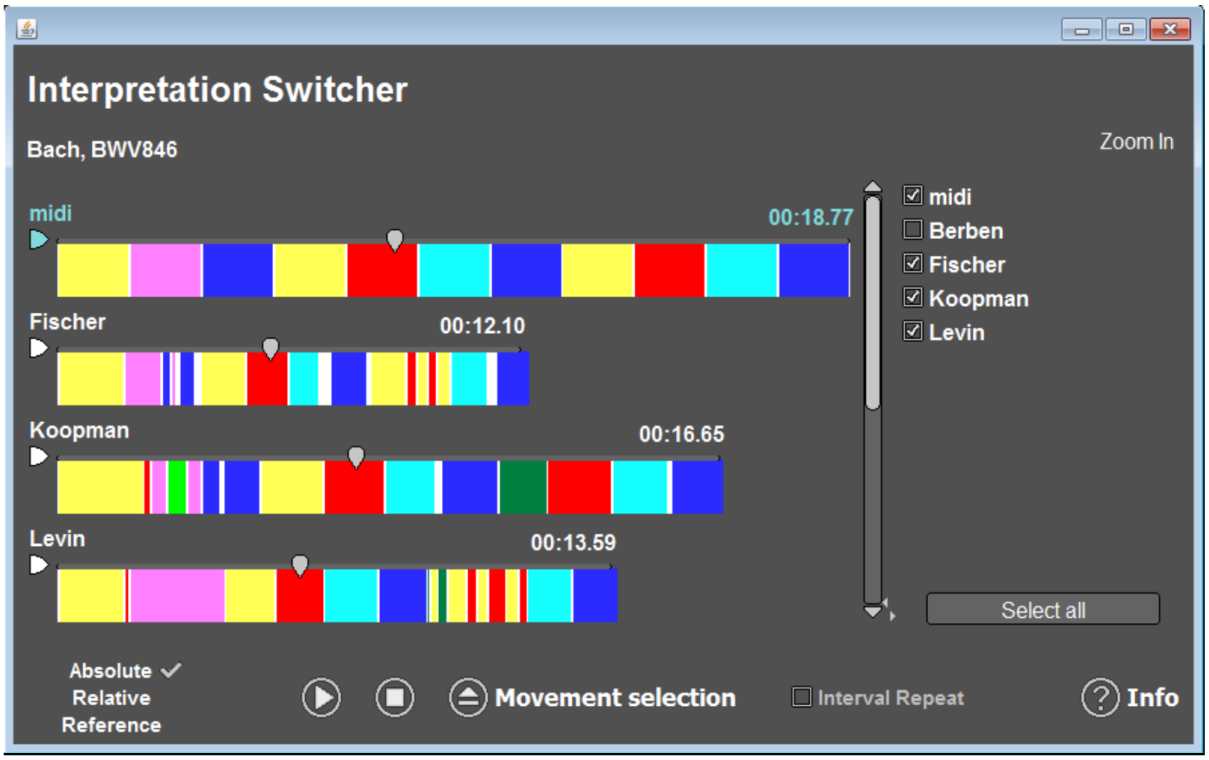

Figure 1: Interpretation Switcher opened with four different versions (MIDI file and three audio recordings) of the first eleven measures of Bach's Prelude in C Major (BWV 846). The annotations correspond to version-dependent chord labels (generated manually for the MIDI version and automatically for the audio versions). In the right part of the interface, the user may select any subset of the available versions (here, four out of five versions are selected).

allows for generating different multi-perspective views (Section 5). Finally, we discuss various applications and indicate future work (Section 6). Further related work is discussed in the respective sections.

2. INTERPRETATION SWITCHER

To make the various music sources (audio recordings, MIDI files) 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 [2, 7, 11, 17]. Here, 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. The technical backbone of our user interface is referred to as *Interpretation Switcher*, which has emerged from the previously developed *SyncPlayer* system [3]. This interface allows a user to select several recordings of the same piece of music, which have previously been synchronized [11]. Each of the recordings is represented by a slider bar indicating the current playback position with respect to the recording's particular timeline, see Fig. 1. The user may listen to a specific recording by activating a slider bar and then, at any time during playback, seamlessly switch to any of the other versions (inter-document navigation).

In addition to the switching functionality, we have extended the Interpretation Switcher to also indicate available version-dependent annotations below each individual slider bar, where labeled segments are represented by color-coded blocks. Such annotations may encode the chord labels generated manually or obtained by some automated chord recognition procedure [14]. Or, such annotations may correspond to the repetitive structure or the musical form, which may have been extracted from the respective recording using automated structure analysis procedures [13]. Based on these annotations, the Interpretation Switcher also facilitates intra-document navigation, where the user can directly jump to the beginning of any structural element simply by clicking on the corresponding block, see Fig. 1.

3. TIMELINE MODES

We have further extended the functionalities of the Interpretation Switcher by realizing three different modes for representing the timelines of the versions. In the *absolute mode*, each timeline encodes absolute timing, where the length of a particular slider bar is proportional to the duration of the respective version, see Fig. 2 (top). In the *relative mode*, each timeline encodes relative timing, where the length of all slider bars coincide, see Fig. 2 (middle). In other words, in the relative mode all timelines are linearly stretched to yield the same length. The third mode,

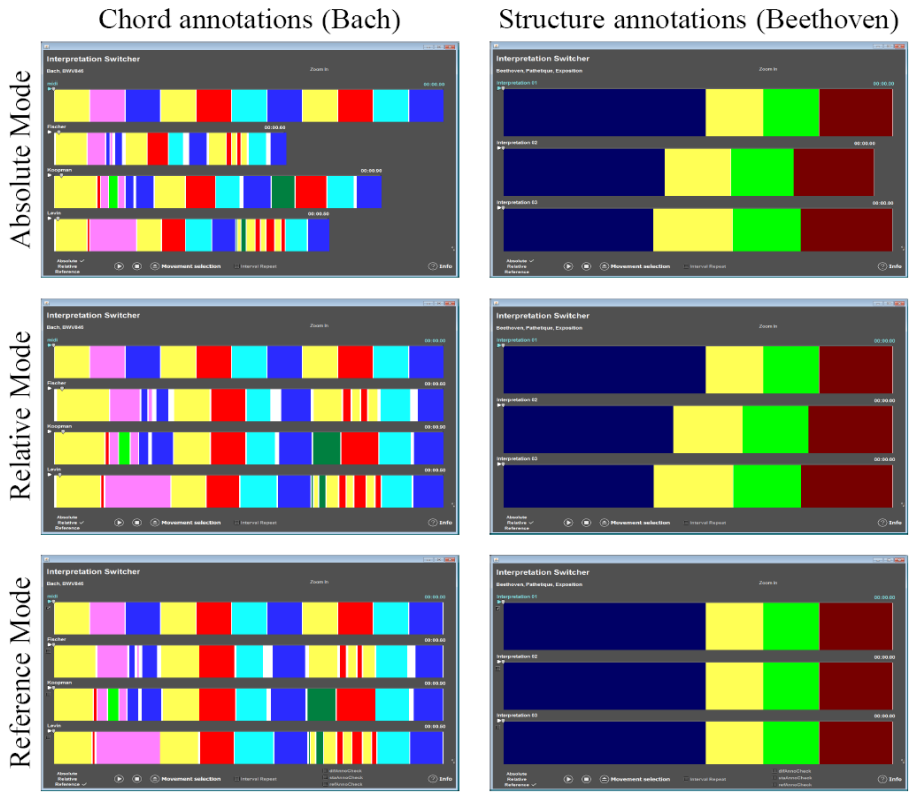

Figure 2: Different timeline modes showing annotations in the absolute mode (**top**), the relative mode (**middle**) and the reference mode (**bottom**) using the first versions as reference, respectively. The left column continues the Bach example from Fig. 1. The right column shows the Interpretation Switcher opened with three different recordings of the exposition of Beethoven's Pathétique Sonata. Here, the annotations correspond to structural information indicating four musically meaningful parts of the exposition.

which is referred to as *reference mode*, is the most interesting one. Here, an arbitrary but fixed version can be selected to act as a reference. Then, all timelines of the other versions are temporally warped to run synchronously to the reference timeline, see Fig. 2 (bottom).

One feature of our timeline adjustment functionality is that the annotations indicated below the slider bars are also adjusted according to the respective mode. Thus, the different timeline modes allow for generating different views on these annotations. For example, using the reference mode, all annotations are temporally warped onto a common timeline, which then facilitates a direct comparison of the annotations across the versions. This is a very useful feature, in particular when the reference corresponds to ground-truth annotations. Furthermore, when the reference corresponds to an uninterpreted MIDI version representing a musical score, the reference mode allows for presenting all version-dependent annotations with respect to a musically meaningful timeline, where time is given in measures and bars rather than seconds.

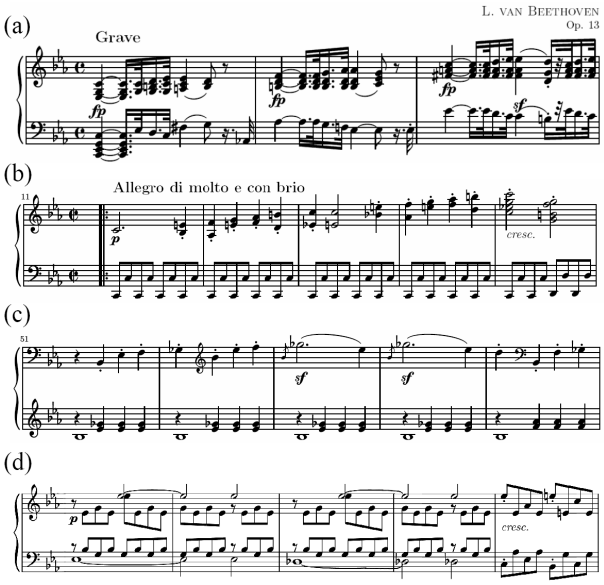

Figure 3: First movement of Beethoven's Pathétique Sonata Op. 13 (score obtained from [12]). **(a)** Beginning of the introduction (Part A, mm. 1 ff.) **(b)** First theme (Part B, mm. 11 ff.) **(c)** Second theme (Part C, mm. 51 ff.) **(d)** Part D (mm. 89 ff.)

4. CASE STUDY

In the following case study, we exemplarily discuss the effect of the different timeline modes by means of the Beethoven example in Fig. 2, right column. Here, the Interpretation Switcher is opened with three different recordings of the exposition of Beethoven's Pathétique Sonata Op. 13 for which structure annotations are indicated. For each recording the respective structure an-

notation consists of four blocks (A (blue), B (yellow), C (green) and D (red)), which correspond to musically meaningful parts of the exposition. The Pathétique is a musicologically outstanding work, for which numerous detailed descriptions and scientific literature exist [15]. Furthermore, being a very famous work it belongs to the standard repertoire of many pianists resulting in numerous audio recordings for this piece. The sonata is characterized by its richness in contrast concerning tempo as well as dynamics.

To better understand the structure annotations shown in Fig. 2, right column, we now describe the Pathétique's exposition in more detail. The first block (A) of the structure annotation of each recording corresponds to the introduction of the exposition. Beginning with the slow introductory theme marked *Grave* (measures (abbreviated mm.) 1-10, see Fig. 3a)), the work starts very dramatically. The introduction is characterized by its contrasts in dynamics—fortissimo passages are followed by subito piano and vice versa. This contrast in dynamics is underlined by contrasts in rhythm, articulation, and mood. Ending with the chromatic run, the introduction leads in the first theme (mm. 11 ff., see Fig. 3b) of the sonata, corresponding to the second block (B) of the structure annotation. The first theme is characterized by the tremolo in octavos in the left hand giving it a dramatic touch. In contrast, the second theme (mm. 51 ff., see Fig. 3c), which corresponds to the third block (C) of the structure annotation, sounds more playful. It is based on the call and response principle and is characterized by a play with articulation. The last block (D) of the structure annotation refers to the fourth part of the exposition, introduced by a third theme in E flat major (mm. 89 ff., see Fig. 3d).

Now, we again consider the three recordings of the Pathétique Sonata. The three different timeline modes of the Interpretation Switcher allow for generating different views on the structure annotations. Firstly, using the absolute mode (see Fig. 2, right column, top), where each timeline encodes absolute timing, enables to visually compare the absolute durations of the three recordings in an intuitive way. For example, one directly observes that the lengths of the first and the third slider bar roughly agree with each other, whereas the second slider bar is noticeably shorter. In other words, Pianist 1 and Pianist 3 choose a slower overall tempo in their performances of the exposition (resulting in a total duration of 224 seconds), whereas Pianist 2 plays the exposition much faster (resulting in a total duration of only 213 seconds), see also Table 1.

Secondly, the relative mode (see Fig. 2, right column, middle) allows for visually comparing the relative durations of the particular structure blocks with respect to the total durations of the recordings. In this way, performance characteristics concerning the tempo shaping in the four parts can be investigated easily. For example, one can notice that Pianist 1 plays the introduction of the exposition (Part A) rather slowly compared to the two other pianists (covering 52.2% of the duration of his/her whole performance). On the contrary, Pianist 3 plays the introduction much faster so that its duration amounts to only 38.4% of the total duration. However, one observes that Pianist 1 plays all the three subsequent parts (B, C, D) faster than the two other pianists, see Table 1. Indeed, Pianist 1 plays the introduction in a slow and expressive way but changes to a faster tempo level at the actual beginning of the exposition (Part B).

		Exp.	A	B	C	D
Pianist 1	Time[sec]	224	117	33	32	42
	Rel. time[%]	100	52.2	14.7	14.3	18.8
Pianist 2	Time[sec]	213	93	38	36	46
	Rel. time[%]	100	43.7	17.8	16.9	21.6
Pianist 3	Time[sec]	224	86	46	39	53
	Rel. time[%]	100	38.4	20.5	17.4	23.7

Table 1: Absolute and relative time durations for the structural parts of the Pathétique's exposition. The table shows for each performance the absolute durations (in seconds) and the relative durations (in %) of the considered structural parts (A, B, C, D) with respect to the total duration of the respective performance.

Thirdly, in the reference mode (see Fig. 2, right column, bottom) all timelines are temporally warped to run synchronously to the reference timeline, where every recording can be selected to act as a reference. In this example, the first recording serves as the reference. The reference mode allows now for a direct comparison of the annotations across the recordings. One directly notices that the annotations of the three recordings agree with each other. Here, the underlying reason is that the structure annotations are consistent across the recordings and perfectly reflect the musical structure of the exposition. Actually, in this example, the annotations were generated manually. However, the situation changes when annotations are computed by automated procedures for each of the versions independently. Then, one typically encounters analysis errors and inconsistencies, which become apparent in the Bach example (Fig. 2, left column). This example will be described in more detail in the subsequent section.

5. MULTI-PERSPECTIVE VIEWS

As a further contribution, we have realized a functionality that facilitates the generation of multi-perspective views across different version-dependent analysis results. We discuss this functionality by means of our Bach example, where we consider a score-like uninterpreted MIDI file (with manually generated ground-truth chord labels) and three audio recordings (with automatically extracted chord labels). Applying the interface's zooming functionality, Fig. 4b shows the version-dependent chord labels of the Bach example (mm. 7-9) in the reference mode, which enables for a simultaneous comparison of the various chord labels over multiple versions of the same piece of music. The various colors correspond to the different labels. In our example, the first slider bar corresponds to the MIDI version which, in this example, is used as the

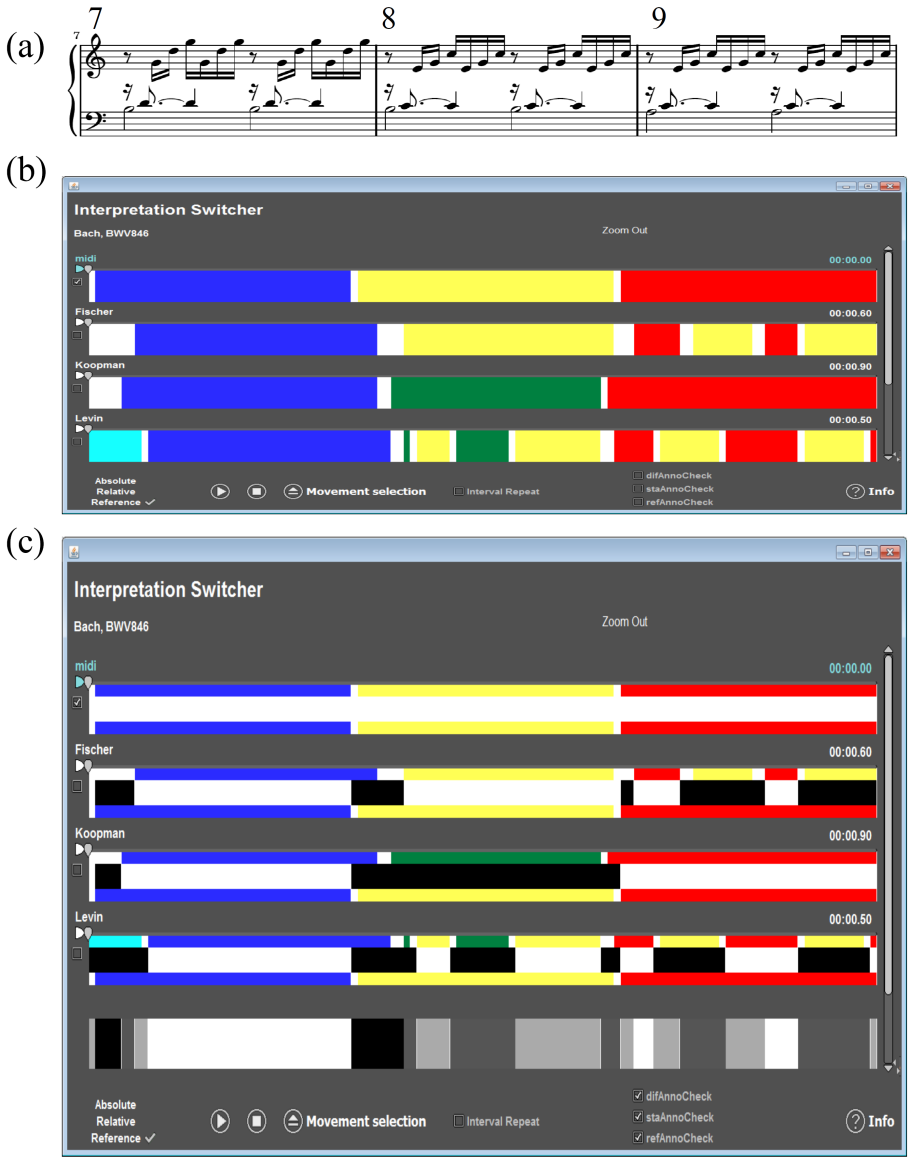

Figure 4: Various multi-perspective views for the Bach example zooming into measures 7 to 9 of the first eleven measures as shown in Fig. 2. **(a)**: Score of measures 7 to 9. **(b)**: Interpretation Switcher in the reference mode (using the first version as reference). **(c)**: Multi-perspective view showing copies of the reference annotations below each of the version-dependent annotations and the pairwise consistencies (white) and inconsistencies (black). The bottom visualizes the degree of consistency (gray values) across all versions.

reference. (However, note that *any* version may be selected to serve as the reference.) Here, mm. 7 is labeled as G major (blue), mm. 8 as C major (yellow), and mm. 9 as A minor (red). The white color indicates unannotated passages. In the reference mode, the interface allows for placing a copy of the reference annotations below each of the version-dependent annotations. Furthermore, the pairwise consistencies (indicated by white) and inconsistencies (indicated by black) across annotations can be visualized.

Such a multi-perspective view is shown in Fig. 4c. Here, a tripartite panel is associated to each version showing the original version-dependent annotations (top), the pairwise consistency information (middle), and the reference annotations (bottom). For example, this view immediately reveals that there are inconsistent annotations in mm. 8, which is labeled as C major (yellow) in the first version (reference) and labeled as E minor (green) in the third version. Actually, this misclassification has musical reasons: in mm. 8 a C major seventh chord is played, which is simplified to C major in the manual annotation. However, due to the added seventh (B) all the tones for E minor (E,G,B) are also present leading to the misclassification E minor.

Additionally, the interface can also provide statistics that indicate the degree of consistency with respect to the reference across all available versions. These statistics are visualized as an additional gray-scaled panel as shown at the bottom of Fig. 4c. Here, the degree of consistency is reflected by the luminance of the grayscale. In particular, a white entry at a given reference time position indicates that all chord labels agree with the reference label

across all versions, whereas a black entry indicates that all non-reference chord labels differ from the reference label. This visualization points the user to problematic passages, which were labeled inconsistently. These inconsistencies may be due to weaknesses of the used labeling procedure (analysis errors), to synchronization inaccuracies, or to musical ambiguities in the piece of music (ill-posed problem, inadequate model assumptions). In our Bach example, the multi-perspective view reveals that for mm. 7 the chord labels (blue) agree across all versions (except for some smaller inconsistencies at the left boundary that may stem from synchronization inaccuracies). On the contrary, for mm. 8-9, the multi-perspective view indicates several inconsistencies. Hence, these two measures seam to be problematic passages in the piece of music. Actually, looking at the score one finds out that seventh chords are present in both measures (C major seventh in mm. 8, A minor seventh in mm. 9), which produces a certain chord ambiguity resulting in misclassifications.

Our interface offers various ways, a user can interactively modify the views including zooming and selection options. In particular, *every* version can be selected to serve as a reference, where the view immediately adjusts upon selection. Note that in this case not only the timeline is changed, but also the copied reference annotations are replaced and the statistics are recomputed. As another feature, for a given set of versions, one can select an arbitrary subset to be considered in the multi-perspective view. For example, in Fig. 1, four of the five available versions are selected.

6. APPLICATIONS AND CONCLUSIONS

In this section, we indicate various application scenarios for our advanced Interpretation Switcher Interface. First of all, as indicated in the previous section, our user interface may serve as a valuable tool for the evaluation of automated music analysis and labeling procedures. Using the reference mode, a multi-perspective view can be generated that yields a synchronized and compact overview of version-dependent analysis results across multiple music representations of a given piece of music. Here, annotation consistencies and inconsistencies can be visualized in a pairwise mode, where each version is compared with the reference separately, as well as in a comprehensive mode comprising all versions. Here, inconsistencies typically point to misclassifications that may be due to analysis errors of automated methods or to intrinsic musical ambiguities. On top of the visual feedback, our interface allows for immediate playback of any position within any version simply by clicking on a color-coded block. Such a block visually represents either a labeled segment or a derived segment that indicates consistency information. This allows a user to easily identify interesting musical passages by means of the visual cues and then to playback the corresponding underlying acoustic material. Having such audio-visual navigation and feedback functionalities, a researcher is greatly supported in performing an in-depth error analysis while deepening his or her understanding of the underlying musical material.

At this point, we want to emphasize that our multi-view evaluation interface may yield interesting information even in the case that no ground-truth annotations are available. For example, in chord recognition, most research is evaluated on the basis of a corpus of Beatles songs, for which high-quality manual chord transcripts have been prepared [6]. However, such special-purpose manual annotations are rarely available. Therefore, one may exploit the fact that one often has large quantities of different versions (e. g., various performances) of a given piece of music, which present opportunities for generating substitutes for manual ground-truth using music synchronization techniques [7, 11, 17]. First multi-perspective approaches to automatically evaluate algorithms have been applied to chord recognition [9] and to beat tracking [5]. In this context, our user interface supports such approaches by supplying immediate visual and acoustic feedback.

As another major benefit, our Interpretation Switcher alleviates interdisciplinary research by bridging the gap between music information retrieval (MIR) and music sciences. Usually, MIR methods are evaluated by MIR researchers in their own lab environment, and music experts are rarely incorporated in the evaluation process. Here, one reason is the lack of communication between MIR researchers, who often do not have an adequate musical background, and music experts, who are often reluctant in using novel computer-assisted methods. Our interface allows even a technically unexperienced user to perform an error analysis of automatically generated annotations. Being pointed to problematic passages by the interface, a music expert can employ his or her musical knowledge and trained ear for an in-depth audio-visual analysis of specific passages. This process can be supported by an MIR researcher who provides the knowledge about the details of the employed annotation methods. In this way, our interface opens the way for an interdisciplinary collaboration, which, on the one hand, supports the MIR researcher in improving the employed methods using the valuable feedback from the music expert, and, on the other hand, familiarizes the music expert with novel computer-assisted methods and interfaces.

For the future, we plan to apply our advanced Interpretation Switcher to support interdisciplinary research going far beyond evaluation. In the context of musicology, one project consists in determining tonal centers (i. e., passages dominated by a certain key) within a large musical work or even entire music corpora. Here, first experiments show that our multi-perspective audio-visual navigation functionalities considerably alleviates the work of musicologists. As a second interdisciplinary project, we have started to introduce computer-based methods into the context of music education [8]. Here, our user interface may help to conduct more user-centered analyses of MIR

methods within natural, music-oriented settings [10].

Acknowledgement. The authors are supported by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University.

7. REFERENCES

[1] D. Damm, C. Fremerey, F. Kurth, M. Müller, and M. Clausen, "Multimodal presentation and browsing of music," in *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI)*, Chania, Crete, Greece, Oct. 2008, pp. 205–208.

[2] S. Dixon and G. Widmer, "Match: A music alignment tool chest," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, GB, 2005.

[3] C. Fremerey, F. Kurth, M. Müller, and M. Clausen, "A demonstration of the SyncPlayer system," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, Sep. 2007, pp. 131–132.

[4] M. Goto, "A chorus section detection method for musical audio signals and its application to a music listening station," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1783–1794, 2006.

[5] P. Grosche, M. Müller, and C. S. Sapp, "What makes beat tracking difficult? A case study on Chopin Mazurkas," in *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, Utrecht, Netherlands, 2010, pp. 649–654.

[6] C. Harte, M. Sandler, S. Abdallah, and E. Gómez, "Symbolic representation of musical chords: A proposed syntax for text annotations," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, GB, 2005.

[7] N. Hu, R. Dannenberg, and G. 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, October 2003.

[8] V. Konz and M. Müller, "Introducing the Interpretation Switcher interface to music education," in *Proceedings of the 2nd International Conference on Computer Supported Education (CSEDU)*, Valencia, Spain, 2010, pp. 135–140.

[9] V. Konz, M. Müller, and S. Ewert, "A multi-perspective evaluation framework for chord recognition," in *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, Utrecht, Netherlands, 2010, pp. 9–14.

[10] M. Lesaffre, M. Leman, B. De Baets, H. De Meyer, L. De Voogdt, and J.-P. Martens, "How potential users of music search and retrieval systems describe the semantic quality of music," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 5, pp. 1–13, 2008.

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

[12] Mutopia Project, http://www.mutopiaproject.org, Retrieved 12.05.2009.

[13] J. Paulus, M. Müller, and A. Klapuri, "Audio-based music structure analysis," in *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, Utrecht, Netherlands, 2010, pp. 625–636.

[14] A. Sheh and D. P. W. Ellis, "Chord segmentation and recognition using EM-trained hidden Markov models," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Baltimore, USA, 2003.

[15] E. R. Sisman, "Pathos and the Pathétique: Rhetorical stance in Beethoven's C-minor Sonata, Op.13," *Beethoven Forum*, vol. 3, 1994.

[16] Sonic Visualiser, http://www.sonicvisualiser.org/, Retrieved 12.05.2009.

[17] R. J. Turetsky and D. P. Ellis, "Force-aligning MIDI syntheses for polyphonic music transcription generation," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Baltimore, USA, 2003.