A MULTI-PERSPECTIVE EVALUATION FRAMEWORK FOR CHORD RECOGNITION

Verena Konz
Saarland University
and MPI Informatik
vkonz@mpi-inf.mpg.de

Meinard Müller
Saarland University
and MPI Informatik
meinard@mpi-inf.mpg.de

Sebastian Ewert
Computer Science III
University of Bonn
ewerts@iai.uni-bonn.de

ABSTRACT

The automated extraction of chord labels from audio recordings constitutes a major task in music information retrieval. To evaluate computer-based chord labeling procedures, one requires ground truth annotations for the underlying audio material. However, the manual generation of such annotations on the basis of audio recordings is tedious and time-consuming. On the other hand, trained musicians can easily derive chord labels from symbolic score data. In this paper, we bridge this gap by describing a procedure that allows for transferring annotations and chord labels from the score domain to the audio domain and vice versa. Using music synchronization techniques, the general idea is to locally warp the annotations of all given data streams onto a common time axis, which then allows for a cross-domain evaluation of the various types of chord labels. As a further contribution of this paper, we extend this principle by introducing a multi-perspective evaluation framework for simultaneously comparing chord recognition results over multiple performances of the same piece of music. The revealed inconsistencies in the results do not only indicate limitations of the employed chord labeling strategies but also deepen the understanding of the underlying music material.

1. INTRODUCTION

In recent years automated chord recognition, which deals with the computer-based harmonic analysis of audio recordings, has been of increasing interest in the field of music information retrieval (MIR), see e. g. [1, 4, 5, 7, 12, 14]. The principle of harmony is a central attribute of Western tonal music, where the succession of chords over time often forms the basis of a piece of music. Such harmonic chord progressions are not only of musical importance, but also constitute a powerful mid-level representation for the underlying musical signal and can be applied for various tasks such as music segmentation, cover song identification, or audio matching [10, 13].

The evaluation of chord labeling procedures itself, which is typically done by comparing the computed chord labels with manually generated ground truth annotations, is far from being an easy task. Firstly, the assignment of chord labels to specific musical sections is often ambiguous due to musical reasons. Secondly, dealing with performances given as audio recording, the ground truth annotations have to be specified in terms of *physical units* such as seconds. Thus, specifying musical segments becomes a cumbersome task, which, in addition, has to be done for each performance separately. On the other hand, musicians trained in harmonics are familiar with assigning chord labels to musical sections. However, the analysis is typically done on the basis of musical scores, where the sections are given in terms of *musical units* such as beats or measures. When dealing with performed audio recordings, such annotations are only of limited use.

As one main contribution of this paper, we introduce an automated procedure for transferring annotations and chord labels from the score domain to the audio domain and vice versa, thus bridging the above mentioned gap between MIR researchers and musicians. Given the score of a piece of music, we assume that musical sections specified in terms of beats or measures are labeled using the conventions introduced by Harte [4]. In case the score is given in some computer-readable format such as MusicXML or LilyPond [6], recent software allows for exporting the score into an uninterpreted MIDI file, where the tempo is set to a known constant value. This allows for directly transferring the score-based ground truth annotations to a MIDI-based ground truth annotation. We then use music synchronization techniques [9] to temporally align the MIDI file to a given audio recording. Finally, the resulting alignment information can be used to temporally warp the audio annotations onto a common musically meaningful time axis, thus allowing a direct comparison to the ground truth annotations.

As a second contribution, we extend this principle by suggesting a novel multi-perspective evaluation framework, where we simultaneously compare chord recognition results over multiple performances of the same piece of music. In this way, consistencies and inconsistencies in the chord recognition results over the various performances are revealed. This not only indicates the capability of the employed chord labeling strategy but also lies the basis for a more detailed analysis of the underlying music material. As a final contribution, we indicate the potential of our framework by giving such detailed harmonic analyses by

means of three representative examples.

The remainder of this paper is organized as follows. First, in Sect. 2 we give an overview about music synchronization. Then, in Sect. 3 we present the multi-perspective evaluation framework. In Sect. 4 we demonstrate our framework giving an in-depth analysis of typical chord recognition errors. Conclusions and prospects on future work are given in Sect. 5.

2. MUSIC SYNCHRONIZATION

For the methods presented in the following sections the concept of music synchronization is of particular importance. In general, the goal of music synchronization is to determine for a given position in one version of a piece of music, the corresponding position within another version. Most synchronization algorithms rely on some variant of dynamic time warping (DTW) and can be summarized as follows. First, two given versions of a piece of music are converted into feature sequences, say $X := (X_1, X_2, \ldots, X_N)$ and $Y := (Y_1, Y_2, \ldots, Y_M)$, respectively. In this context, chroma features have turned out to yield robust alignment results even in the presence of significant artistic variations. In the following we employ CENS (**C**hroma **E**nergy **N**ormalized **S**tatistics) features, a variant of chroma features making use of short-time statistics over energy distributions within the chroma bands, for a detailed description see [9]. Additionally, we consider non-standard tunings similar to Gómez [3]. Then, an $N \times M$ cost matrix C is built up by evaluating a local cost measure c for each pair of features, i. e., $C(n,m) = c(x_n, y_m)$ for $n \in [1:N] := \{1, 2, \ldots, N\}$ and $m \in [1:M]$. Each tuple $p = (n,m)$ is called a *cell* of the matrix. A (global) *alignment path* is a sequence $(p_1, \ldots, p_L)$ of length L with $p_\ell \in [1:N] \times [1:M]$ for $\ell \in [1:L]$ satisfying $p_1 = (1,1)$, $p_L = (N,M)$ and $p_{\ell+1} - p_\ell \in \Sigma$ for $\ell \in [1:L-1]$. Here, $\Sigma = \{(1,0), (0,1), (1,1)\}$ denotes the set of admissible step sizes. The *cost* of a path $(p_1, \ldots, p_L)$ is defined as $\sum_{\ell=1}^{L} C(p_\ell)$. A cost-minimizing alignment path, which constitutes the final synchronization result, can be computed via dynamic programming from C. For a detailed account on DTW and music synchronization we refer to [9].

Based on this general strategy, we employ a synchronization algorithm based on high-resolution audio features as described in [2]. This approach, which combines the high temporal accuracy of onset features with the robustness of chroma features, generally yields robust music alignments of high temporal accuracy.

3. MULTI-PERSPECTIVE VISUALIZATION

A score in a computer readable format such as LilyPond or MusicXML is available for many classical pieces of music [11]. For a trained musician it is much more intuitive to annotate the chords of a piece on the basis of the underlying score than on the basis of an audio recording. However, such an annotation is not directly transferable to an audio recording of the same piece, as both use very different notions of time. Furthermore, this also implies that this annotation cannot be used directly to evaluate the results of an audio-based automatic chord labeling method. In this section, we present a method integrating music synchronization techniques, which allows for a direct comparison of chord labels derived from different versions of a piece of music. This approach has several advantages. Firstly, the manual annotation becomes much more intuitive. Secondly, the position of a chord recognition error in an audio recording can be easily traced back to the corresponding position in the score. This allows for a very efficient in-depth error analysis as we will show in Sect. 4. Thirdly, a single score-based annotation can be transfered to an arbitrary number of audio recordings for the underlying piece.

In the following, we assume that an audio recording and a score in computer readable format are given for a piece of music. Additionally, chord labels manually annotated by a trained musician on the basis of a score are given as well as labels automatically derived from the audio recording via some computer-based method. In a first step, we export the score to a MIDI representation. This can be done automatically using existing software. Beat and measure positions are preserved during the export, such that the score-based annotations are still valid for the MIDI file. In a next step, we derive CENS features from the MIDI as well as from the audio as mentioned in Sect. 2, say $X := (X_1, X_2, \ldots, X_N)$ and $Y := (Y_1, Y_2, \ldots, Y_M)$, respectively. Since each CENS feature corresponds to a time frame, we can also create two binary chord vector sequences, $A := (A_1, \ldots, A_N)$ and $B := (B_1, \ldots, B_M)$, which encode the given chord labels in a framewise fashion. Here, $A_n, B_m \in \{0,1\}^d$ for $n \in [1:N]$ and $m \in [1:M]$. The constant d equates the number of considered chords. A value of one in a vector component encodes the chord prevalent in the corresponding time frame. As we consider in the following only the 24 major and minor chords ($d = 24$), we have to map the given chord labels in a meaningful way to one of these. To this end, we employ the interval comparison of the dyad, which was used for MIREX 2009 [8] and takes into account only the first two intervals of each chord. Thus, augmented and diminished chords are mapped to major and minor respectively, as well as any other label having a major or minor third as its first interval. Using the first four measures of Chopin's Mazurka Op. 68 No. 3 as an example, we illustrate the sequences A for the score and B for the audio in Fig.1(b) and 1(c), respectively. Note that in Fig.1(b) the time is expressed in terms of measures, while in Fig.1(c) the time is given in seconds. This different notion of time prevents a comparison of A and B at this point.

The next step consists of synchronizing the two CENS features sequences X and Y as mentioned in Sect. 2. The resulting alignment path $p = (p_1, \ldots, p_L)$ encodes temporal correspondences between elements of X and Y. Following the same time frame division, the alignment path also encodes correspondences between the sequences A and B. Using this linking information, we locally stretch and contract the audio chord vector sequence B according to the warping information supplied by p. Here, we have to consider two cases. In the first case, p contains a subse-

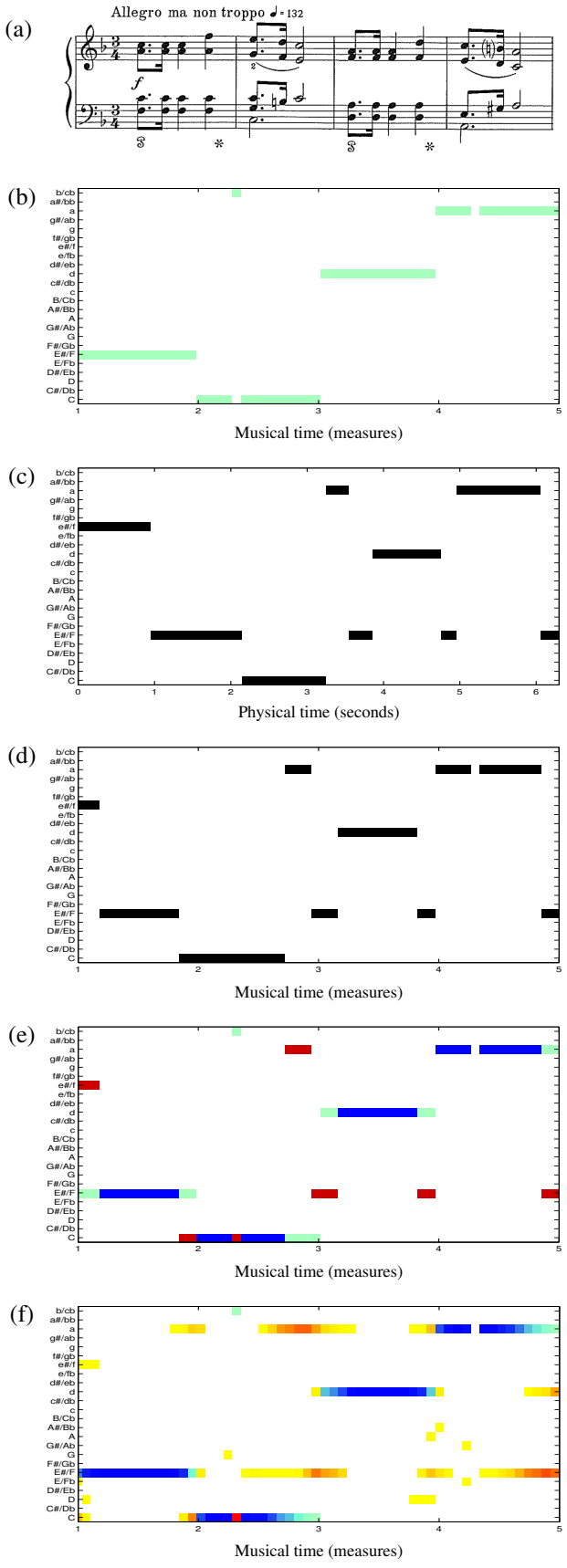

Figure 1. Various chord annotations visualized for the Chopin Mazurka Op. 68 No. 3 (F major), mm. 1-4. **(a)** Score. **(b)** Score-based ground truth chord labels. **(c)** Computed audio chord labels (physical time axis). **(d)** Warped audio chord labels (musical time axis). **(e)** Overlayed score and audio chord labels. **(f)** Multi-perspective overlay of score and audio chord labels.

quence of the form

$$(n, m), (n + 1, m), \ldots, (n + \ell - 1, m)$$

for some $\ell \in \mathbb{N}$, i. e., the ℓ score-related vectors $A_n, \ldots, A_{n+\ell-1}$ are aligned to the single audio-related vector B_m. In this case, we duplicate the vector B_m by taking ℓ copies of it. In the second case, p contains a subsequence of the form

$$(n, m), (n, m + 1), \ldots, (n, m + \ell - 1)$$

for some $\ell \in \mathbb{N}$, i. e., the score-related vector A_n is aligned to ℓ audio-related vectors $B_m, \ldots, B_{m+\ell-1}$. In this case, we replace the ℓ vectors by the vector $B_{m+\lfloor \ell/2 \rfloor}$. The resulting warped version of B is denoted by $\bar{B}$. Note that the length of $\bar{B}$ equals the length N of A, see Fig. 1(d). For the visualization we set all vectors in $\bar{B}$ to 0, where no groundtruth chord label is available, as for example in the middle of measure (abbreviated mm.) 4, see Fig. 1(d).

Overall, we have now converted the physical time axis of the audio chord vector sequence B to the musically meaningful measure axis, as used for A. Finally, we can visualize the differences between the score-based and the audio-based chord labels by overlaying A and $\bar{B}$, see Fig. 1(e). Here, the vertical axis represents the 24 major/minor chords, starting with the 12 major chords and continuing with the 12 minor chords. Blue entries now indicate areas, where the ground truth labels and the audio chord labels coincide. On the contrary, green and red encode the differences between the chord labels. Here, green entries correspond to the ground truth chord labels derived from the score, whereas red entries correspond to the audio chord labels. For example, at the beginning of mm. 2 the score as well as the audio chord labels indicate a C major chord. On the contrary, at the end of mm. 2 there is a C major chord specified in the score, while the chord labels derived from the audio incorrectly specify an A minor chord. Using the measure-based time information, we can look directly at the corresponding position in the score and analyze the underlying reason for this error. We will demonstrate this principle extensively in Sect. 4, where we present an in-depth analysis of typical errors produced by automatic chord labeling methods.

Next, we extend the just developed concept by introducing a multi-perspective visualization, see Fig. 1(f). Here, we make use of the fact that for classical pieces usually many different interpretations and recordings are available. Visualizing the chord recognition results simultaneously for multiple audio recordings of the same piece, we can analyze the consistency of errors across these recordings. On the one hand, if an error is not consistent, then this might indicate a chord ambiguity at the corresponding position. On the other hand, a consistent error might point to an intrinsic weakness of the automatic chord labeler, or an error in the manual annotations. This way, errors might be automatically classified before they are manually inspected.

In Fig. 1(f) the multi-perspective visualization for the first four measures of the Chopin Mazurka is represented. Here, we warped the automatically generated chord labels for 51 different audio recordings onto the musical time axis

using the steps described above. By overlaying the resulting chord vector sequences $\bar{B}$ for all pieces, we get a visualization similar to the previous one in Fig. 1(e), so that the visualization for one audio recording can be seen as a special case of the multi-perspective visualization. In the multi-perspective visualization, we distinguish two cases using two different color scales: one color scale ranges from dark blue to bright green, and the other color scale ranges from dark red to yellow. The first color scale from blue to green serves two purposes. Firstly, it encodes the score-based ground truth chord labels. Secondly, it shows the degree of consistency between the automatically generated audio labels and the score labels. For example, the dark blue entries at the beginning of mm. 2 show, that a C major chord is specified in the score labels, and most audio-based labels coincide with the score label here. At the end of mm. 2 the bright green shows that the score specifies a C major, but most audio-based results differ here from the score label. Analogously, the second color scale from dark red to yellow also fulfills two purposes. Firstly, it encodes the audio-based chord labels that differ from the score labels. Secondly, it shows how consistent an error actually is. For example, at the beginning of mm. 2 there are no red or yellow entries, since the score labels and the audio labels coincide here. However, at the end of mm. 2, most audio-based chord labels differ from the score labels. Here most chord labels either specify an F major or an A minor chord.

4. EVALUATION

None of the currently available automatic chord labeling approaches works flawlessly. Errors can either be caused by the inherent ambiguity in chord labeling, or by a weakness special to the employed chord labeler. An in-depth analysis allowing for a distinction between these error sources is a very hard and time-consuming task. In this section, we show how this process can be supported and accelerated using the evaluation and visualization framework presented in Sect. 3. To this end, we manually created score-based chord annotations for several pieces of music. Furthermore, we implemented a very simple baseline chord labeler to study very common sources of error in chord labeling.

4.1 Annotations

For the following evaluation, a trained musician (Verena Konz) manually annotated the chords for three pieces of Western classical music. Firstly, Mazurka in F major Op. 68 No. 3 by Chopin. Secondly, Prelude in C major BWV 846 by Bach. Thirdly, the first movement of Beethoven's Fifth Symphony, Op. 67. Using the underlying score, the annotations were created on the beat-level, and in the case of the Bach Prelude on the measure-level. The format and naming conventions used for the annotation were proposed by Harte [4]. The annotator paid much attention to capture even slight differences between adjacent chords. Hence, the bass tone as well as missing or added tones in chords are marked explicitly using the corresponding shorthands.

4.2 Baseline-method for chord recognition

A baseline chord labeler can be implemented using only a few simple operations. Given an audio recording, we first extract CENS features (see Sect. 2) resulting in a feature sequence $Y := (Y_1, Y_2, \ldots, Y_M)$. We derive ten features per second, with each feature considering roughly 1100 ms of the original audio signal. Non-standard tunings are considered as described in Sect. 2. Then, we define a total of 24 chord templates, 12 templates for the major chords and 12 for the minor chords. The considered templates are 12-dimensional vectors, in which the respective three tones of the corresponding major(minor) chord (the root note, the major(minor) third and the fifth) are set to 1 and the rest to 0. Thus, we obtain e. g. for C major the template

$$(1,0,0,0,1,0,0,1,0,0,0,0)$$

and for C minor the template

$$(1,0,0,1,0,0,0,1,0,0,0,0).$$

Let T in the following denote the set of all 24 chord templates. In a next step, we choose a distance function d, which measures the distance of the i-th feature vector Y_i to a template $t \in T$.

$$d : [0,1]^{12} \times [0,1]^{12} \mapsto [0,1]$$
$$d(t, Y_i) = 1 - \frac{\langle t, Y_i \rangle}{\|t\| \cdot \|Y_i\|},$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\|\cdot\|$ the Euclidean norm. By minimizing over $t \in T$ we can find the best matching chord template t^* for the i-th feature vector.

$$t^* = \underset{t \in T}{\operatorname{argmin}} \quad d(t, y_i)$$

The chord label associated with t^* constitutes the final result for the i-th frame.

4.3 Experiments

We start our evaluation by looking again at our running example, Chopin Mazurka Op. 68 No. 3. Our proposed visualization method clearly reveals various chord recognition errors, see Fig. 1(e). Making use of the musical time axis, these errors can now easily be traced back to the corresponding position in the score and analyzed further. For example, at the beginning of the piece, the score-based ground truth annotation corresponds to F major, whereas the computed audio-based annotation corresponds to F minor. A mix-up of major and minor often appears in the chord recognition task. The next misclassification occurs at the end of mm. 1, where the ground truth still corresponds to F major, but the computed annotation specifies a C major, which is actually the subsequent chord in the ground truth. This may be a boundary problem or an error in the synchronization.

In the middle of mm. 2, we note that the ground truth chord is B minor, whereas the computed chord is C major. Having a look at the score, one can see that the chord in question is actually a B diminished chord. Due to the reduction of the manual annotation to major/minor chords, this chord is mapped to a B minor chord in the ground truth. Causing a misclassification here, this is often a problem in the major/minor evaluation based on the comparison of the dyad.

The next misclassifications are due to the musical ambiguity of chords. At the end of mm. 2 we observe in the score a C major chord, where the fifth is missing. Comparing on the dyad level, this chord is mapped to a C major chord in the ground truth. However, all the notes of the chord (C,E) are also part of an A minor chord, which is actually computed at this position. A similar problem occurs at the beginning and at the end of mm. 3, where the ground truth annotation corresponds to D minor, whereas the computed annotation corresponds to F major. The same phenomenon appears a last time at the end of mm. 4, where F major is recognized instead of A minor. This phenomenon is caused by ambiguities inherent to the chord labeling task and constitutes a very common problem. The chords in classical music rarely are pure major or minor chords, because tones are often missing or added. Hence, the recognition as well as the manual annotation process become a hard task.

Next, we illustrate what kind of additional information our multi-perspective visualization can provide compared to the just discussed visualization that only makes use of a single audio recording. Here, we consider again the first four measures of the Chopin Mazurka. Instead of using only one audio recording we overlay the chord recognition results for 51 different audio recordings in our multi-perspective visualization, see Fig. 1(f). Looking for consistencies and inconsistencies, it is possible to classify and investigate single errors even further. For example, the misclassified F minor chord in the beginning of mm. 1 (see Fig. 1(e)) seems to be an exception for the specific recording. This can be clearly seen from the multi-perspective visualization where only for a few of the 51 audio recordings F minor is computed instead of F major. Also, the misclassification at the end of mm. 4 (F major instead of A minor) is not consistent across all considered audio recordings. On the contrary, some of the misclassifications which we observed in the case of one audio recording (Fig. 1(e)), are consistently misclassified for most of the other audio recordings. For example, the diminished chord in the middle of mm. 2, the chord ambiguity problem occuring at the end of mm. 2 (A minor instead of C major), the beginning of mm. 3 (F major instead of D minor) and the end of mm. 4 (F major instead of A minor). Overall, the multi-perspective chord recognition allows for a classification of recognition errors into those specific to a recording and those independent of a recording.

As a further example we now consider the famous Bach Prelude in C major, BWV 846. The multi-perspective visualization for 5 different audio recordings for mm. 19-24 (see Fig. 2) again reflects the chord recognition problems

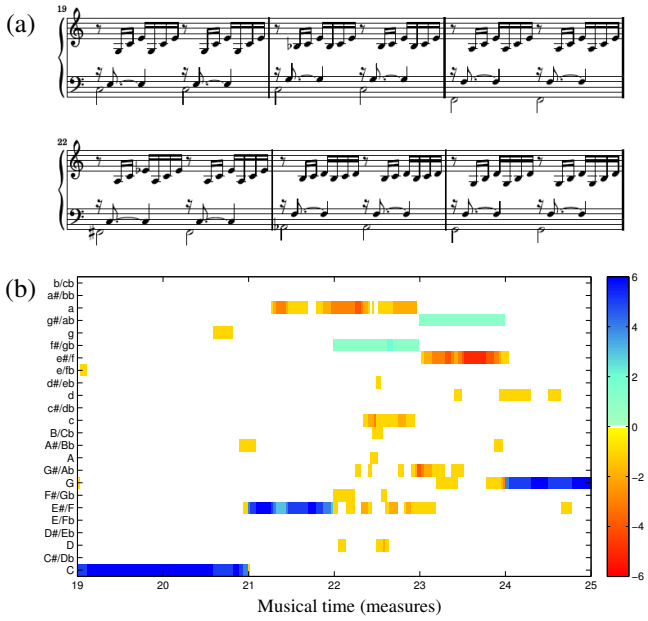

Figure 2. Bach BWV 846, mm. 19-24. **(a)** Score, **(b)** Multi-perspective overlay of score and audio chord labels.

related to diminished chords. At the beginning of the excerpt (mm. 19-21) and at the end (mm. 24) the chord recognition result for all audio recordings consistently agrees more or less with the ground truth. However, one can observe two passages with green entries in mm. 22-23. Looking at the corresponding position in the score, we find two diminished seventh chords, in mm. 22 an *F#:dim7* and in mm. 23 an *Ab:dim7*. Due to the reduction to major/minor chords these two chords are mapped to F# minor and Ab minor in the ground truth annotation, respectively, see Fig. 2. However, in most audio recordings an A minor chord is detected instead of *F#:dim7*, having two tones (A and C) in common. And instead of the *Ab:dim7* chord an F minor chord is found, for which even all three tones are present (F, Ab and C) due to the additional passing note C in the *Ab:dim7*. While the seventh chord in mm. 20 is recognized well for all recordings, we see that in mm. 21 the F major seventh chord was mistaken for an A minor chord, again due to chord ambiguity reasons.

As a last example we now consider the first movement of Beethoven's Fifth Symphony in 37 different audio recordings. Actually, this piece of music is much more complicated in terms of harmonic aspects than the previously considered Chopin and Bach examples. In the Beethoven example, we can often find the musical principles of suspension, passing notes or "unisono" passages. Here, the automatic chord recognition as well as the manual annotation are challenging and ambiguous tasks. One example for the use of nonharmonic tones in chords can be found in mm. 470-474, visualized in Fig. 3. Looking at the score, we observe in the left hand a D major chord with a missing fifth (mm. 470-473), but in the right hand a G is added in octaves to this D major chord. Being the fourth of D, the G can be seen as a nonharmonic tone in D major. This causes a chord misclassification for about 15 recordings, where G major or alternatively G minor is computed.

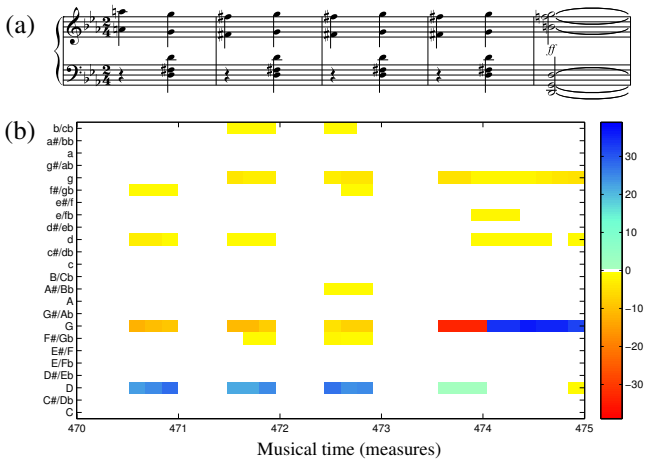

Figure 3. Beethoven's Fifth, mm. 470-474. **(a)** Score, **(b)** Multi-perspective overlay of score and audio chord labels.

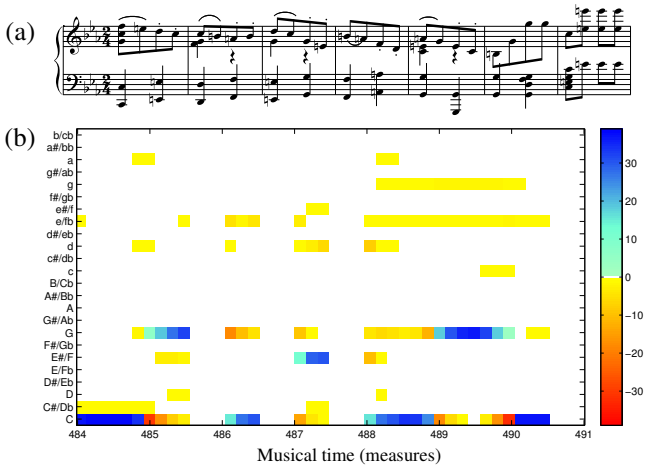

Figure 4. Beethoven's Fifth mm. 484-490. **(a)** Score, **(b)** Multi-perspective overlay of score and audio chord labels.

On the contrary, the G seventh chord in mm. 474 is recognized very well for all recordings. Note that the first beats of the measures 470-474 are not manually annotated, since the octaves do not represent meaningful chords.

Another example of a musical pattern that is found to be extremely problematic in the chord recognition task, is the principle of suspension. We illustrate the problems related to this musical characteristic using another excerpt (mm. 484-490) of Beethoven's Fifth, see Fig. 4. In each of the measures 484-488, one can find a suspension on the first eighth, which resolves into a major chord on the second eighth. This musical characteristic can easily be spotted in the multi-perspective visualization. Here, we see that at the beginning of each measure the number of audio recordings for which the computed annotation agrees with the ground truth is very low and gets higher afterwards. In mm. 490 finally the first complete pure major chord is reached. Note that the second beats of mm. 485-487 consist of passing notes to the next suspension. Hence, a meaningful chord cannot be assigned resulting in several beats missing a ground truth annotation.

5. CONCLUSIONS

In this paper, we have introduced a multi-perspective evaluation framework that allows for comparing chord label annotations across different domains (e. g., symbolic, MIDI, audio) and across different performances. This bridges the gap between MIR researchers, who often work on audio recordings, and musicologists, who are used to work with score data. In the future, we plan to apply our framework for a cross-modal evaluation of several computer-based chord labeling procedures, some of which working in the symbolic domain and others working in the audio domain. Furthermore, in a collaboration with musicologists, we are investigating how recurrent tonal centers of a certain key can be determined automatically within large musical works. Here, again, our multi-perspective visualization based on a musically meaningful time axis has turned out to be a valuable analysis tool.

Acknowledgement. The first two authors are supported by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University. The third author is funded by the German Research Foundation (DFG CL 64/6-1).

6. REFERENCES

[1] J. P. Bello and J. Pickens. A robust mid-level representation for harmonic content in music signals. In *Proc. ISMIR, London, UK*, 2005.

[2] S. Ewert, M. Müller, and P. Grosche. High resolution audio synchronization using chroma onset features. In *Proc. of IEEE ICASSP*, Taipei, Taiwan, 2009.

[3] E. Gómez. Tonal description of polyphonic audio for music content processing. *INFORMS Journal on Computing*, 18(3):294–304, 2006.

[4] C. Harte, M. Sandler, S. Abdallah, and E. Gómez. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proc. ISMIR, London, UK*, 2005.

[5] K. Lee and M. Slaney. A unified system for chord transcription and key extraction using hidden Markov models. In *Proc. ISMIR, Vienna, Austria*, 2007.

[6] LilyPond. `http://www.lilypond.org`.

[7] M. Mauch, D. Müllensiefen, S. Dixon, and G. Wiggins. Can statistical language models be used for the analysis of harmonic progressions? In *Proceedings of the 10th International Conference on Music Perception and Cognition, Sapporo, Japan*, 2008.

[8] MIREX 2009. Audio Chord Detection Subtask. `http://www.music-ir.org/mirex/2009/index.php/Audio_Chord_Detection`.

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

[10] M. Müller, F. Kurth, and M. Clausen. Audio matching via chroma-based statistical features. In *Proc. ISMIR, London, GB*, 2005.

[11] Mutopia Project. `http://www.mutopiaproject.org`.

[12] J. T. Reed, Y. Ueda, S. Siniscalchi, Y. Uchiyama, S. Sagayama, and C.-H. Lee. Minimum classification error training to improve isolated chord recognition. In *Proc. ISMIR, Kobe, Japan*, 2009.

[13] J. Serrà, E. Gómez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. *IEEE Transactions on Audio, Speech & Language Processing*, 16(6):1138–1151, 2008.

[14] A. Sheh and D. P. W. Ellis. Chord segmentation and recognition using EM-trained hidden Markov models. In *Proc. ISMIR, Baltimore, Maryland, USA*, 2003.