SAC: G: Using Chord Distance Descriptors to Enhance Music Information Retrieval

Ladislav Maršík
Dept. of Software Engineering,
Faculty of Mathematics and Physics,
Charles University, Malostranské
nám. 25, Prague, Czech Republic
marsik@ksi.mff.cuni.cz

ABSTRACT

Analysis of digital music and its retrieval based on the audio features is one of the popular topics within Music Information Retrieval (MIR). From many applications, however, only a fraction will provide results that are meaningful for musicians [4]. While retrieval algorithms may be precise, it is similarly important to think about the end-users and their understanding of the query and the result. We have developed a music analysis system which is based on music theory and contains visualizations meaningful for those interested in harmony aspects of music. Music is first segmented to chords by known techniques, providing the basis that musicians understand. From there, distances between chords are evaluated by a novel approach. We propose a chord distance descriptor in form of a time series created from chord distances. This descriptor is straightforward for a trained musician, and at the same time can be used for various retrieval tasks. We compare the available chord distance measures and choose the best candidates for time series. We also derive a novel chord distance, based on simplification of the music theory. Finally, we evaluate our selection by analyzing pieces of popular music and achieve a 86% accuracy when using chord approach, while significantly cutting down the execution time when using chord distances for database pruning techniques. The dataset and application are publicly available to achieve replicable research.

Keywords

music information retrieval; chord distance; chord transcription; cover song identification

1. PROBLEM AND MOTIVATION

The gap between music theory and recent MIR applications has been pointed out by multiple researchers, calling for more work on how music theory can help recent retrieval tasks [1][4]. Even if the application provides valid results (e.g. retrieves a correct song), users may have difficulties understanding, why the result was made such. Our work is motivated by this fact and employs music theory in the proposed application. The analysis does not assume to be complete in the context of all music properties, including melody or rhythm. We focus on music harmony, since it is one of the most important aspects to describe the song structure, and can also be used effectively for cover song identification, as shown by Ellis [2].

One of the tasks that can showcase our theoretical approach, is cover song identification (CSI). A *cover song* is an alternative version, performance, or recording of a previously published musical piece. The author of a cover song may choose to differ from the original piece in several musical aspects: instrumentation and arrangement, tonality and harmony, song structure, melody, tempo, lyrics, or language. Thus, CSI task is a very difficult challenge in choosing the best techniques, features and algorithms, and has been a vivid field of research within music information retrieval in the last decade. The state-of-the-art methods are evaluated annually on the benchmarking challenge MIREX (Music Information Retrieval Evaluation Exchange)[1] with up to 8 algorithms posted every year since.

The motivation for our research was ignited by the fact that the best result on the MIREX benchmark was achieved by Serra et al. in 2009 [9] and has not been outperformed since. This implies, that to further improve the results for CSI, the researchers need to change techniques, focus on the partial tasks or use diverse datasets. Chord distance research is a novel way to approach CSI, that focuses on the harmony aspect of the task, and the research cannot be done without a thorough work with datasets, that often involves creating new datasets along the way. In this paper we describe how we have tackled all the aspects of the task: from novel methodology, publicly available application to achieve replicable experiments, all the way to the results and publishing the dataset.

2. BACKGROUND AND RELATED WORK

Extracting high-level features from digital music such as melody, rhythm, tempo, or harmony structure, has been a major task in Music Information Retrieval (MIR), starting with ground-breaking work of Fujishima [8]. In his work, it is the first time we see automatic chord estimation in practice. Since then, it has been a journey of many various techniques [10].

2.1 Chord transcription

To obtain the chord representation, a standard multiple-step process is followed. First, obtaining chroma vectors by a short-time Fourier transformation, where chroma vector is a representation of a short musical moment mapped to 12 tones of the piano keyboard [2]. The next step is chord segmentation and labeling, jointly referred to as chord transcription [5]. Segmentation finds an exact timestamp for every chord change, while labeling names the chord using a chord dictionary.

2.2 Chord distances

The concept of chord distance has been studied in the literature previously, but remains ambiguous, as there are multiple

[1] http://www.music-ir.org/mirex/wiki/MIREX_HOME

definitions to choose from Rocher et al. [7]. As a result, similarity of chord sequences or their distances is an unexposed area, with only a few recent studies, such as the one by De Haas et al. [11]. The authors have chosen to compare the whole chord series to a common chord representing the key of the piece, achieving satisfactory results for Jazz pieces. In this work it is also stressed, that there is surprisingly little research conducted chord series comparison, given the popularity of chord annotations.

We propose chord distances to fulfil the goal of analyzing and retrieving music. In MIR, chords are usually represented either as 12-dimensional vectors similar to chroma vectors, or as a string of tones (e.g. *CEG*). If we want to obtain distance between 2 chords, we can therefore use vector or string distances. However, other chord distances based on music theory are being proposed, that make the decision of choosing one distance challenging. Musicology standard is to use "Circle of Fifths" [3] and more profound models from which the most acknowledged is Tonal Pitch Space (TPS) from Lerdahl [3]. In addition to the knowledge of the chord, many methods need to be context-aware - in particular, having the knowledge of the music key. The whole variety of possible chord distances was summarized in the work of Rocher et al. [6].

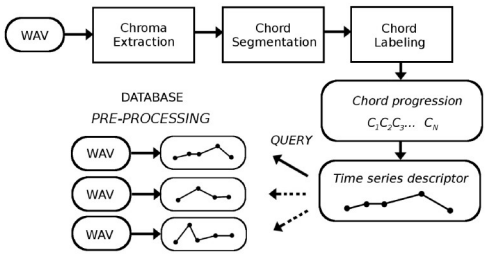

Figure 1: System outline for descriptor creation and retrieval query

3. APPROACH AND UNIQUENESS

Our approach starts with standard music processing tasks – obtaining chroma vectors and segmentation, as described in Section 2.1. We use methods proposed by Mauch and Dixon [5] for these tasks, allowing the use of other methods at the same time, as the analysis is independent from lower-level processing. Our approach differentiates from all known methods in the further steps: labeling (chord representation), key finding, chord distance and visualization, while common retrieval techniques can be used to evaluate the innovative approach. Figure 1 shows the outline of the whole system.

3.1 Chord representation

While most of the MIR works are using only a subset of all possible chords (chord dictionary), we think that the added value of dissonances can enhance the task. Therefore, we prefer to leave all sounding tones in the chord. As a result, even chords that are difficult to describe by music theory, containing dissonances, can appear in our analysis. There are advantages and disadvantages to this approach. Dissonant tones can be treated as invalid information towards retrieval, such as noise. However, non-chord tones can also play an important role in the character of the musical piece. An example can be a singer singing the voice above the chord accompaniment, causing a temporary dissonance.

3.2 Chord distance time series descriptor

After the chords are represented, we continue by evaluating transitions between every two subsequent chords and plotting them on the color temperature graph, such as on Figure 3. Using chord distances for obtaining time series is an innovative approach on its own – to the best of our knowledge, there are very few studies focusing on this aspect [1][6], despite the developed musicology background. Instead of using classic methods for chord distances such as TPS [3], we proceed by employing our new Chord Complexity Distance (ChordCD), based on a novel model, which is an alternative to TPS. Our innovation lies in taking every chord as a sentential form of a grammar-like system, and evaluating chord distance as the number of steps of derivation of one sentential form to another (see Figure 2). This way we can work with dissonant tones. Also, we can abstract from a separate key-finding method, as ChordCD model finds key and chord simultaneously, similar to the work of Rocher et al [7]. For a trained musicians, the proposed rules are understandable, since they are based on tonal harmony [3]. Once we extract time series descriptor, we can use this descriptor as a basis for a specific music retrieval task. We look for musical pieces with a similar series, as is depicted on the last steps on Figure 1. The descriptor is simple (most of the songs having less than 100 transitions), yet describing the whole chord progression, independent from the music key. As such it can be understood as a fingerprint of the chord progression and a way how to speed up music search.

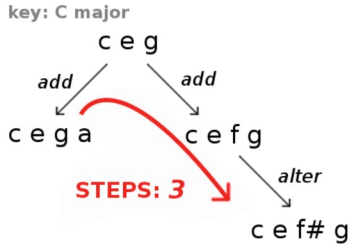

Figure 2: ChordCD model for distances based on adding/removing tones from the same key, and altering the tones outside the key. The key is evaluated for each chord tuple.

3.3 Visualization and important music parts

One of the main advantages of the time series descriptor introduced in previous section is the ease of visualization of a musical piece. As seen on Figure 3, both color temperature or line graph can be employed to show the data. This visualization can be easily used while playing the music in music player. When user retrieves a similar musical piece based on this descriptor, the visualization provides an understanding of why the piece was selected. Another advantage is, that user sees important chord sequences in the visualization easily, as the contrasting areas in the color temperature graph, or as the peaks in the line graph. These parts usually relate to interesting harmony movements, as are often in the bridge or before the start of the chorus.

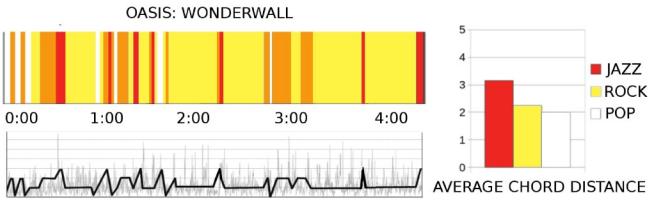

Figure 3: Visualization of the song Wonderwall by Oasis. On the top left, the chord progression is shown as a color temperature graph, on the bottom left as a line graph. On the right there are average chord complexity distances for 3 genres: Jazz, Rock and Pop made on a corpus of 120 songs.

4. RESULTS AND CONTRIBUTIONS

4.1 harmony-analyser.org project

We have developed a system[2] capable of automatic analysis described in this abstract. The application provides visualizations which are easy to understand for musicians. In the *harmony-analyser.org* project, we provide GUI tools published as executable JAR archives, to allow for a custom harmony analysis of WAV or MIDI input. The tools itself are using the JHarmonyAnalyser Java library, which we describe in details in the more technical report [17]. The screenshot of the application is displayed on the Figure 4. Showcasing multiple tools, one of the use-cases is to use the MIDI keyboard plugged in via the USB port, or use a text input field, to specify two chords, to obtain the chord distance feature.

Another feature of the application is the analysis and visualization of a musical piece. On a sample analysis on Figure 3, we can notice chord distance peaks around 0:40, 1:30 and 2:30, which correspond to the A5 chord followed by B7sus4, as well as a dissonance in the end caused by a guitar ornament. Preliminary results also show different average chord distances for different genres and promising results for music classification task (Figure 3 on the right). Such system can popularize MIR research among musicians and can be a useful plugin for today's music players. The system is distributed under GNU Public License, it fills in the gap in MIR studies and encourages further research on chord distances.

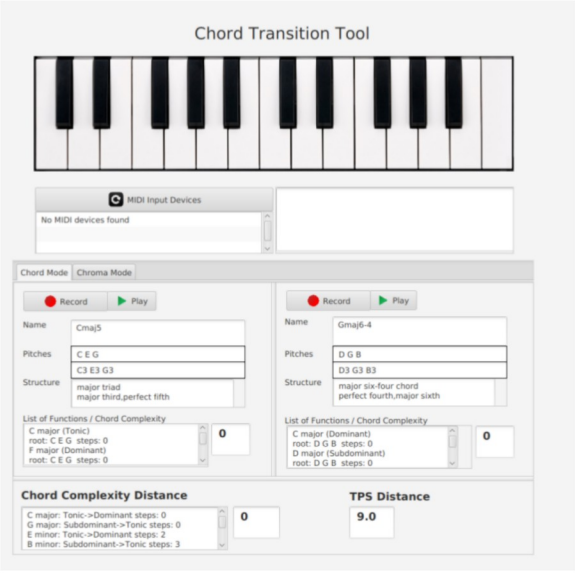

Figure 4: Chord Transition Tool from harmony-analyser.org; capturing the MIDI input and outputting the chord labels, functions and the chord distances. C major and G major chords are analysed.

[2] http://harmony-analyser.org

4.2 Cover song identification experiments

Our chord and chord distance approach was tested on various musical datasets. We chose to use the publicly-available datasets covers80 [12] and SecondHandSongs [13], as well as our own Kara1k dataset [14]. Notably, the Kara1k dataset was created as a by-product of our Chord distance research. As we describe in our Kara1k publication, the publicly-available datasets to-date have not triggered new results in several years, or have not provided sufficient features. We tried to tackle this by making our set of features for karaoke songs public as Kara1k dataset [14].

We define our CSI task as: "For a given query song, an algorithm has to find the corresponding cover version out of N songs.", where N is the size of the song database. To do so, we took features discussed is Section 3.2, for each query song (chord distance features, ChordCD). Along with our distance we also extracted other MIR features, such as raw chroma vectors and chroma complexity differences (ChromaCD)[3] [15], to compare chord distances with more complex features. Although the distances cannot compete with "raw" data in terms of precision, we were interested how will our "fingerprints" be suited for the CSI task in the broad sense.

4.2.1 Method and evaluation metrics

We compare the query song features with the features from the database of cover songs using DTW method [15], and we built a similarity matrix containing each song combination. Each matrix is thus composed of N query songs and N cover songs. Cover songs are then sorted by their similarities to the query song. The first song in this list is then considered to be the detected cover.

We use three metrics to compare the features. The first metric is the Mean arithmetic of Average Precision (MAP) that is commonly used by MIREX benchmark. The MAP takes into account the whole sorted list of cover songs and assigns a weight to the ground-truth song according to its corresponding rank. The MAP can range between 0.0 and 1.0 and the higher the MAP the better the algorithm. The second metric is the Mean Average Rank (MAR) that averages the rank of the ground-truth song for all the queries. The MAR can range from 1.0 to the number of songs in the musical dataset and the lower the MAR the better the algorithm. The third metric is the percentage of correctly detected cover songs out of the N queries.

4.2.2 Datasets

The covers80 dataset consists of 160 songs organized as 80 musical works, each in two versions. The first versions of the songs are used as queries for the search in the whole set of the second versions of the songs. SecondHandSongs is a set of

[3] ChromaCD is in its definition similar to chord distances, only the distance is extracted from subsequent chromas. We reference the reader to [15] for the thorough explanation.

18,196 tracks with 5,854 cover song clusters (average cluster size is 3.11). For our comparison, we have taken a chunk of 999 songs (295 clusters) from SecondHandSongs train set - the first 999 songs listed in the official dataset information file. Our Kara1k[4] dataset references a list of 1,000 cover songs provided by karoke application Karafun, and a list of 1,000 original songs corresponding to the cover songs. The cover songs contain the singing voice and the instrumental tracks and were recorded in a studio with professional musicians following the original song scores.

4.2.3 Results and discussion

Score	Raw chroma vectors	ChromaCD	ChordCD
score (4.1)	**0.482**	0.094	0.142
score (4.2)	0.103	0.070	0.071
score (4.3)	0.417	0.174	0.156
score (4.4)	0.454	0.061	0.114
score (4.5)	0.082	0.041	0.034

Table 1. MAP results for each DTW score and feature for covers80 dataset.

Score	Raw chroma vectors	ChromaCD	ChordCD
score (4.1)	**0.107**	0.031	0.019
score (4.2)	0.021	0.014	0.014
score (4.3)	0.029	0.035	0.021
score (4.4)	0.043	0.015	0.012
score (4.5)	0.008	0.008	0.009

Table 2. MAP results for each DTW score and feature for SecondHandSongs dataset.

We can see the results for covers80 and SecondHandSogns datasets in the Tables 1 and 2. Scores (4.1)-(4.5) used are the most common DTW scores, from which we wanted to determine the best score for our future experiments (which is Score (4.1) as discussed in details in [15]).

Needless to say, raw chroma vectors have outperformed all other features. We attribute this to the fact that 12-dimensional chroma vectors contain 12-time more information than the one-dimensional ChromaCD or ChordCD. However, the execution time was significantly better for ChromaCD, and even better so for ChordCD. To obtain similarity matrices was: 56s for raw chroma vectors, 51s for ChromaCD and 25ms for ChordCD time series for covers80 dataset. On SecondHandSongs dataset the execution time was 550s for raw chroma vectors and 100s for ChordCD time series. It means that for a new ChordCD feature, that is 5-to-1000 time faster (depending on the dataset), and 12-times smaller in size, we can still achieve results that can detect the basic cover songs (MAP score 0.156 detected at least 10 cover songs from covers80 dataset and the rest was ranked in the upper-half of the songs). This is an interesting result, in the light of *database pruning* techniques described by Osmalskyj et al. [16]. It was proposed only recently, that the easier "fingerprints" should be used as the pruning technique on the first level of analysis. These fingerprints can select song candidates for the second level, where more complex feature is used (e.g. raw chroma feature).
To further elaborate the contribution of chords and chord distances, we chose to do another set of experiments, for karaoke songs on our own Kara1k dataset. The motivation for our own dataset was: better sets of features, as well as the use of karaoke songs, which are easier to be paired than cover songs because of the same music background, and as such the karaoke song identification sub-task can help narrow the research for the CSI task [14].

[4] http://yannbayle.fr/karamir

	MAP	MAR	% detected
Chroma	**0.899**	48.112	**88.9%**
MFCC	0.878	**25.161**	86.3%
Chord	0.865	43.943	84.4%
ChromaCD	0.442	70.994	36.4%
TPS	0.257	141.941	19.3%
Key	0.203	109.954	11.0%

Table 3. Results for DTW technique on Kara1k dataset

Table 3 compares the MAP, the MAR and the percentage of correctly detected Kara1k songs for multiple features. As a baseline we chose chroma and MFCC features, and we experimented with chord, TPS chord distance and key fingerprints (more information in [14]). Again, the best MAP and accuracy is achieved by the chroma features. However, the MFCC and chord features detect a similar amount of karaoke songs as the chroma features, while, again, chord feature is a significantly smaller feature than the chroma feature. Chord feature represent a reduced information from float vectors to binary vectors. We attribute this result to the fact, that the karaoke songs are typically produced in the same key as the original songs. This emphasizes the use of chords, in place of raw features, when applicable, as the accuracy is similar.

The TPS feature [14] is the representant of the chord distance feature in this experiment (a different model used than in ChordCD, as we describe in Section 2.2). The comparison of ChordCD and TPS models is an ongoing research, but early results show slight improvement for TPS [15], while arguably ChordCD can be easier to understand for musicians [4]. While the TPS accuracy is lower than the more complex features, we again emphasize the 20% detected covers and the possible use for database pruning techniques. Indeed, the TPS feature is the least complex feature out of all in Table 3, as for every frame of the song, only a scalar is given, instead of 12-dimensional vectors for all other features.

5. CONCLUSION AND FUTURE WORK

We have demonstrated the importance of chords and chord distance descriptors for music information retrieval, which is believed to be an underestimated part of the recent research. By comparing the recent features and applying it to analysis, we can see chord distance descriptors as a vital measure for all relevant retrieval tasks. From the experiments we can conclude that:

- Fingerprints motivated by chord distances are fast to obtain and use, small in size and detect a reasonable amount of cover songs, to be used as the first layer for database pruning techniques
- Chord features are smaller in size than chroma or MFCC features, yet the retrieval accuracy is similar, for tasks such as karaoke song identification

Along with developing the harmony-analyser.org application and providing a Kara1k dataset with valid chord features, the research can move forward in the years to come.

In the future work, more robust analysis on bigger data sets should show the relevance of this descriptor. We are also motivated by the idea of combining audio fingerprinting approaches with CSI approaches, in order to develop fast and

precise retrieval algorithm for cover songs that are resembling the originals. This idea should later be tested not only for karaoke cover songs, but also for other specific cover versions, e.g. live performances from the original artist, or, multiple recordings of the same classical music piece. We will therefore guide our future work toward breaking down the difficult CSI task to meaningful subtasks, with the aim of developing new-generation music discovery applications.

6. ACKNOWLEDGMENTS

This study was supported by the Charles University in Prague, project GA UK No. 1580317, project SVV 260451, and project PRVOUK.

7. REFERENCES

[1] De Haas, W. B., Veltkamp, R. and Wiering, F. Tonal Pitch Step Distance: A Similarity Measure for Chord Progressions. *ISMIR* 2008

[2] Ellis, D. P. W. and Poliner, G. E. Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking. *ICASSP* 2007

[3] Lerdahl, F. *Tonal Pitch Space.* Oxford University Press, Oxford, 2001

[4] Lewis, R. J., Fields, B. and Crawford, T. Addressing the Music Information Needs of Musicologists. *ISMIR* 2015

[5] Mauch, M. and Dixon, S. Approximate Note Transcription for the Improved Identification of Difficult Chords. *ISMIR* 2010

[6] Rocher , T., Robine , M., Hanna , P., and Desainte-Catherine , M. A Survey of Chord Distances With Comparison For Chord Analysis. *ICMC* 2010

[7] Rocher T., Robine M., Hanna P., and Oudre, L. Concurrent Estimation of Chords and Keys from Audio. *ISMIR* 2010

[8] Fujishima, T.: Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music. In: Proceedings of the International Computer Music Conference. ICMC 1999 (1999)

[9] J. Serrà, X. Serra, and R. G. Andrzejak, "Cross Recurrence Quantification for Cover Song Identification," New J. Physics, vol. 11, no. 9, p.093017, 2009.

[10] McVicar, M., Santos-Rodriguez, R., Ni, Y., Bie, T.D.: Automatic Chord Estimation from Audio: A Review of the State of the Art. IEEE/ACM Trans. Audio, Speech & Language Processing 22(2), 556–575 (2014)

[11] De Haas, B., Veltkamp, R., Wiering, F.: Tonal Pitch Step Distance: A Similarity Measure for Chord Progressions. In: Proceedings of the 9th International Conference on Music Information Retrieval (2008)

[12] Ellis, D.P.W., Cotton, C.V.: The 2007 LabROSA Cover Song Detection System. In: Music Information Retrieval Evaluation eXchange. MIREX 2007 (2007)

[13] Bertin-Mahieux, T., Ellis, D.P., Whitman, B., Lamere, P.: The Million Song Dataset. In: Proceedings of the 12th International Society for Music Information Retrieval Conference. ISMIR 2011 (2011)

[14] Bayle, Y., Maršík, L., Rusek M., Robine, M., Hanna, P., K. Slaninová, J. Martinovič, and J. Pokorný: Kara1k: a karaoke dataset for cover song identification and singing voice analysis, in: IEEE International Symposium on Multimedia. ISM 2017 (2017)

[15] L. Maršík, M. Rusek, K. Slaninová, J. Martinovič, and J. Pokorný, "Evaluation of Chord and Chroma Features and Dynamic Time Warping Scores on Cover Song Identification Task," in Proc. 16th Int. Conf. Comp. Inform. Systems and Industr. Manag. App. Bialystok, Poland: Springer, 2017, pp. 205–217.

[16] J. Osmalskyj, S. Piérard, M. Van Droogenbroeck, and J.-J. Embrechts,"Efficient Database Pruning for Large-Scale Cover Song Recognition," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Vancouver, BC, Canada, 2013, pp. 714–718.

[17] L. Maršík, "harmony-analyser.org - Java Library and Tools for Chordal Analysis," in Proceedings of 2016 Joint WOCMAT-IRCAM Forum Conference, Taoyuan City, Taiwan, 2016, pp. 38–43.