Ewa Łukasik
Institute of Computing Science
Poznan University of Technology

Review of Ph.D. Dissertation authored by
Ladislav Maršik
entitled
Cover Song Identification using Music Harmony Features, Model and Complexity Analysis

1. General Description

Music Information Retrieval (MIR) is the multidisciplinary domain of retrieving information from music. It is rounded in the combined use of theories, concepts and techniques from music, computer science, machine learning, computational intelligence, audio signal processing, psychoacoustics and cognition. This area of research needs automatic generation of features, called descriptors, that capture the sonological or musical characteristics embedded in the audio signals. After more than two decades of the research, MIR has achieved such a level of the development that may be used in the MIR industry (e.g. Playlisting: Pandora, Spotify, YouTube Music, Amazon Prime Music, iHeart, Shazam, Gracenote). However there are still many problems to be solved.

Some exemplary music information facets to be discovered are: rhythm, timbre, tonality, melody or structure. Other topics are more related to the semantic description of music, e.g. the concept of complexity, similarity, music recommendation, genre, mood, social tags or song covers. In this latter topic the seminal work belongs to Serrà et al. (2009) – their results have not been beaten for almost ten years and the candidate took this fact as a point of reference of his dissertation.

Music Information Retrieval requires cooperation between musicologists and computer scientists, or, which is not so often the case of the researcher that has the expertise in both music and computer science domains. An attempt to combine the two fields was made by the candidate in his dissertation, proposing the use of harmony principles to create new music descriptors. The candidate proposed a new harmony model, defined the music complexity and applied them to a specific task, which is Cover Song Identification (CSI). He also used his algorithms in open-source software tools and created a database together with descriptors.

2. Layout and the content

The dissertation consists of the abstract, the foreword, the introduction, six chapters, conclusions, bibliography, lists of figures, tables, abbreviations and a list of Author's publications. There are also five attachments. The structure is deliberate and transparent.

In the foreword the Author describes circumstances and decisions that influenced his research and the final shape of doctoral dissertation.

The introduction is devoted to explaining the background of the thesis: the domain of Music Information Retrieval, especially the importance of exploring music harmony for calculating the similarity of musical pieces and a practical problem of the Cover Song Identification. It also enumerates the main contributions of the dissertation, describes works published by the author and shortly describes the contents of each chapter.

The first chapter introduces the terminology from the domain of musicology and Music Information Retrieval as well as features based on Digital Signal Processing useful for the problem of Cover Song Identification.

The second chapter reviews the works related to the topic of the dissertation - both in the domain of musicology (tonality, harmonicity, chord distance measures) and Music Information Retrieval (music complexity, harmony analysis and Cover Song Identification. It refers to benchmarking results of MIREX task on Audio Cover Song Identification (MIREX is the yearly Music Information Retrieval Evaluation eXchange competition) and analysis of the CSI state-of-the-art results. It also discovers current trends in CSI together with the identification of features used for this task.

The third chapter comprises original contributions to the music harmony description. It presents the new model of music harmony, defines new distances between chords and chroma vectors, a fingerprint of the harmonic content of a song, and the Average Transition Complexity descriptor. Its usage is illustrated by the case study analysis of genres, music formats, popular songs by authors such as Oasis, Queen or Beatles and pieces of classical music

In the subsequent, fourth chapter, the author focuses on the experiments validating the new descriptors (ChordCD, ChromaCD, ATC) to form a useful harmony fingerprint of a song, compares music harmony features, standard and new, Performs a CSI experiment using the best features and scores coming from the previous analysis, with some of his own improvements a search for the best combination of the feature, the technique, and the similarity measure for a well balanced CSI system.

The fifth chapter presents the software used for harmony calculation – harmony-analyser – developed by the author - that is capable of creating, naming, and analyzing chords and visualize chord progressions as well as provide tools for comparison of chord distances

The sixth chapter is devoted to the presentation of KaraMIR project and Kara1k dataset - large, clearly labeled musical dataset.

In conclusions the Author recapitulates the dissertation and points out possible future directions of the research.

The Bibliography consists of 149 well chosen bibliographic entries reflecting the main achievements in the field of MIR, especially in relation to harmony research and CSI.

The attachments present in a concise form the fundamentals of music theory concerning tones, intervals, scales, chords and harmonic functions, very useful for a general reader.

The dissertation is well illustrated by 47 figures and 15 tables.

The list of Author's publications comprises nine articles, mainly co-authored.

The work is thoroughly edited with a good layout and is written clearly using good English, often in fluent narrative style. Each Chapter (excluding Chapter 6) starts with a brief introduction pointing out the main contributions and finishes with conclusions that facilitate the reading.

It was a real pleasure to read the nice hard-copy version of the dissertation.

3. The main contribution

The dissertation makes several key contributions. In terms of novel research, the first contribution is a model of the tonal harmony useful for Music Information Retrieval. It is

based on the tonic-subdominant-dominant representation of chords (TSD) and is called the TSD distance model. To represent a harmonic complexity of a musical piece the Author has defined new distances between chords and chroma vectors. Of practical character are the descriptors proposed: Chord complexity distance (ChordCD), Chroma complexity difference (ChromaCD) and Average Transition Complexity (ATC). Grounded in MIR they are also understandable by the community of musicians/musicologists and may be further improved and modified by researchers from both communities. The Author has shown the usefulness of new descriptors in certain case studies featuring analysis of popular songs and classical music the results of which have been promising (Chapter 3). This is somehow curious for me, why the experiments on music harmony features comparison from Chapter 4 (Section 4.2) have not been included in Chapter 3, especially that they do not concern CSI, but genre classification.

I would have liked to have seen further in-depth systematic studies on the problem of harmonic complexity to discover more characteristics of music useful for Music Information Retrieval. However the Author has chosen a different, more practical direction of research, concentrating on a challenging problem of the Cover Song Identification and has measured the performance of the proposed harmonic descriptors in this specific task. I would not have treated it as the lost opportunity as the topic is interesting in itself and is a good test bed for new methods.

The next contribution is establishing a new harmony fingerprinting technique using a time series of chord distances and applying it to the Cover Song Identification task. The cover song fingerprints were tested using five versions of Dynamic Time Warping (DTW), standard method for time series comparison. The results of comparison of harmony fingerprints for exemplary datasets have been all worse than for standard raw chroma vectors. The same trend remained in the case of karaoke dataset. This might have been rather disappointing outcome of the research, but, in my opinion, this is a good lesson to be learned for researchers who challenge the signal processing based features with those reflecting human understanding of music. In search of efficient features we have to admit that very often, despite the lack of understandable musical interpretations of DSP-based features, especially when aggregated, they often outperform the common musical features used by musicologists.

On the other hand it would have been interesting to see results of manual verification of the correctness of calculated Chord and Chroma distances on some examples. I agree with the Author, that the poor results do not exclude the new descriptors from the use as features in MIR tasks. He suggests to apply them e.g. in a first layer of database pruning techniques, especially that they are computationally more efficient than the others. Chords are well-suited features for the initial check. In my opinion there is also a room for further studies of computational harmony as I suggested earlier in this review.

The contribution of practical character highlighting the candidate's IT skills is *harmony-analyser.org* project published to the open-source community applying the theoretical findings on harmonic complexity and used for the Author's own experiments. It contains *chord analyzer* package and *chroma analyzer* package and acts as a framework for other packages. It has got attention form the MIR community.

The culmination of the candidate's work on CSI and a substantial contribution to MIR community is the *KaraMIR* project and *Kara1k* dataset, whose descriptions the Author decided to include in this dissertation. Karaoke may be treated as a simplified version of cover song and the dataset was utilized in some of the CSI experiments. and may be used for another task of the singing voice analysis.

The dissertation has also introduced fundamentals of Music Information Retrieval and harmony analysis research to any reader not accustomed with these interdisciplinary topics. It has been done very well so I would like to mention another contribution - providing the comprehendible didactic text.

The quality and contribution of this dissertation is strong. More in-depth studies of harmonic similarity could be explored in future work. Overall, the candidate is able to pose research questions and to search for their solutions. The dissertation makes a solid research contribution. Each of the contributing studies draws interesting and useful conclusions that advance the state of knowledge in this area of research and practice.

4. Critical remarks

Even though the thesis was written in a correct way, there are some minor errors to correct. These shortcomings do not substantially affect the quality of the work and do not disqualify it in any way. However I would like to point out certain issues for further discussion.

1. The candidate has formalized the theoretical considerations by introducing definitions. However some of them are not consistent enough and do not have the character of definition, e.g. the definition of the mel-frequency cepstral coefficients (MFCC) – page 33/34 - or the definition of the spiral distance of the two chords – page 44. It would have been advisable to provide the bibliographic reference of the definition. Also it would have been helpful, if the definitions have been numbered.

2. Dynamic Time Warping (DTW) is a method still under scientific investigation in search of the time and precision optimality. The Author has used five different algorithms for DTW calculation, however the level of detail in their presentation is too low as for the methods for providing crucial results. DTW is a dynamic programming problem attributed to Sakoe and Chiba (1998) who introduced it for the problem of isolated word recognition. However the reference has been given to M. Müller's book (2007).

3. The sixth chapter related to KaraMIR project has different structure and contains elements not related directly to the topic of the dissertation (singing voice analysis). It disturbs the consistency of the whole work. The chapter presents the contents of the article published in International Journal of Semantic Computing (Maršik et al., 2018), so maybe it would have been better to provide the paper as an attachment to the dissertation.

4. Section 4.2.1 – the Author uses the term "music classification" instead of "music genre classification" that he really means. As a survey publication in this domain I would recommend, instead of Bertin-Mahieux et al. [2010], a broader one:

Sturm B.L. (2014) A Survey of Evaluation in Music Genre Recognition. In: Nürnberger A., Stober S., Larsen B., Detyniecki M. (eds) Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation. AMR 2012. Lecture Notes in Computer Science, vol 8382. Springer, Cham

5. The author's publications show multi-author works related to the thesis. In all of them the candidate is the first author. However, I would value a simple statement about his role in their preparation or at least his percentage involvement in those publications.

There are some singular typos that do not disturb fluent reading of the text.

5. Knowledge of the Candidate

The candidate exhibits extensive knowledge of the subject areas. For example, the background section provides a very detailed and systematic study of prior research in the area of modeling harmony CSI. It also provides a very extensive discussion of existing features, which creates a useful resource for others who may wish to explore this space. Each part of the problem is formalized or defined. In addition, the candidate demonstrates solid knowledge of research methodology. His research works have been many times recognized and awarded on national and international fora. The number and quality of publications is satisfactory and the participation in international research groups and candidate's research enthusiasm promise well for his future research career.

6. Conclusions

In conclusion, I find this body of work to be well conducted and well presented. The thesis represents a solid contribution to the field of Music Information Retrieval and to the domain of Computer Science, and I would certainly consider it to be of sufficiently high quality for a PhD dissertation. Taking into consideration my comments above my evaluation of the dissertation is following:

- The dissertation contains original and valuable scientific results. The rank of the results is highlighted by the fact that some of them were published in journals.
- The dissertation definitely presents an original solution to a scientific problem.
- The candidate has general theoretical knowledge and understanding of the discipline of Computer Science, and particularly the area of the Music Information Retrieval.
- The dissertation definitely supports the claim that the candidate is able to conduce scientific work.

In my opinion, the dissertation of Ladislav Maršik undoubtedly fulfills all the conditions for gaining the Ph.D. degree in Computer Science; therefore it is recommended.

In Poznań, September 2019

Ewa Łukasik