V. Kůrková et al. (Eds.): ITAT 2014, Part II, pp. 13–17
Institute of Computer Science AS CR, Prague, 2014

Improving Music Classification Using Harmonic Complexity

Ladislav Maršík[1], Jaroslav Pokorný[1], and Martin Ilčík[2]

[1] Dept. of Software Engineering, Faculty of Mathematics and Physics
Charles University, Malostranské nám. 25, Prague, Czech Republic
{marsik, pokorny}@ksi.mff.cuni.cz
[2] The Institute of Computer Graphics and Algorithms,
Vienna University of Technology, Favoritenstraße 9-11, Vienna, Austria
ilcik@cg.tuwien.ac.at

Abstract: Publicly available multimedia systems provide users with plenty of music files offered in different genres. These systems should process the music as fast as possible while satisfying the needs of their users as well. In this context, reliable music classification represents one of the major challenges. Classification systems without deeper knowledge of music structure and composition yield to considerable errors. In some cases, music can not be classified clearly due to an overlap in genres. However, in other cases, we can clarify the classification simply by using the approach of a skilled musician. In this paper, we develop a new approach to automatic music classification inspired by the theory of neural networks, enhanced by deeper knowledge of tonal harmony. Based on a new measure derived from harmonic movements, harmonic complexity, our supporting experiments proved a significant improvement in classification accuracy.

1 Introduction

Recent expansion of music-oriented systems has increased the importance of reliable music classification and tagging. Music recommendation services and radios such as Spotify[1] or Pandora[2] rely on music tagging done manually by human experts or amateur users, with respect to the genre, mood, or other semantic category. This is referred to as *social tagging* [13]. In addition to tagging, majority of the systems use collaborative filtering and recommend the music based on other users' preferences [5]. On the other hand, automatic music classification and recommendation becomes more and more popular, and systems such as Mufin[3] use purely algorithmic content-based approach to find similarities between musical pieces [17].

The major problem in manual tagging is, that it is time-consuming and expensive for large collections of music [13]. Many attempts were introduced to automatize the process. However, to the best of our knowledge, no method has been proven to be particularly suited for this task. For a comprehensive survey, we refer the reader to the book chapter summarizing the state-of-the-art techniques by Bertin-Mahieux et al. [2].

In this paper, we focus on improving the state-of-the-art techniques, by introducing a new musical feature, *harmonic complexity* [11]. Based on tonal harmony and the chord progression in a musical piece, the harmonic complexity can play a role of one of the principal descriptors for music classification.

1.1 Music Descriptors

Feature extraction is the first step in any content-based music discovery algorithm. Various features can be considered to form a „fingerprint" of a musical piece prior to retrieval, including tempo, rhythm structure, instrumentation or volume. Generally, we distinguish between low-level features obtained directly after signal processing, and high-level features, extracted by further analysis of the frequency-domain input. A popular music discovery system Shazam[4] uses low-level features extracted from a short excerpt to retrieve a musical piece from the database [21]. On the contrary, Mufin internet radio uses more than 40 characteristics and high-level features, to automatically recommend music [17]. Various teams around the world work on establishing the set of features for the most effective music retrieval. However, standards have not been established yet and choosing the right features remains not only the academic, but also a commercial challenge.

1.2 Outline

In Section 2, we will look at the related works, from both sides, extracting harmony features from music, and the music classification challenge. In Section 3, we explain the underlying model of harmonic complexity, that was used to aid the classification. Then, in Section 4, we show the results of our supporting experiments. We conclude our paper with a summary and the outline of future work.

2 Related Works

In this section, we look at the works related to ours. First, we show what other harmony features have been proposed

[1] http://www.spotify.com
[2] http://www.pandora.com
[3] http://www.mufin.com/usecase/music-recommendation
[4] http://www.shazam.com

in recent studies. Second, we look at the recent works of music classification.

For introduction to basic theory of music, please refer to Schönberg's Theory of Harmony [16] or more recent works on musicology such as Krumhansl [7]. Please note that the term key has a different meaning in music than in computer science.

2.1 Harmony Features

Harmony is an important component in music, and studies have shown its importance for music retrieval [1]. We believe that a new feature summarizing the harmonic movements in the musical piece can be a great contribution to the recently used sets of music descriptors.

There have been several approaches to utilize harmony in music retrieval tasks. Still, the standards for evaluating harmonic movements have not been set and the approaches vary from work to work. We show the commonly used methods in a short survey.

Key modulations. A piece of music or its part can belong to a certain class, called the *key*. Knowing that some parts of music really belong to a certain key can be useful for music classification, since some genres are known to modulate from key to key more often and others do not (i.e. 20th century music contrary to medieval music). Krumhansl [6] was the first to use a probabilistic model based on music cognition, to discover what key is the sequence of music played in. Other authors such as Sapp [15] or De Haas et al. [4] were building up on this knowledge and used the clarity of a certain key in certain parts of music to achieve their results. In general, we can conclude that using measures related to the key and its clarity is one way of obtaining a harmony-based music descriptor.

Harmony models. One of the more difficult ways to evaluate the harmonic movements in the piece is to build a model of harmony and use that model to obtain the needed measures. De Haas et al. [4] have built a tree structure of tonal harmony rules and then used their model to derive a music sequence similar to the one from the musical piece being analyzed. Then, the total number of errors in the derivation of a chord progression was used to improve the state-of-the-art chord labeling techniques. We find the method of De Haas et al. as a good recipe that shows us how to use music harmony effectively to achieve other results.

Works on musicology. It is important to notice, that works on musicology provide us with comprehensive mathematical theories, that evaluate the distances between keys, chords, and other entities clearly (cf. Lerdahl [8]). However, these formalisms are still not implemented in the state-of-the-art music services, probably due to the complex and high-level approach they imply.

2.2 Music Classification

In order to show that our new feature will be efficient for music classification, we must first observe what are the common techniques used. There have been a great amount of proposals, since music classification is one of the recent challenges. The most common are using hidden Markov models [18], self-organizing maps [14], k-nearest neighbor [20], support vector machines [22], or neural networks [19] of different kinds. The state-of-the-art techniques include compressive sampling [3], or low-rank semantic mapping [13]. Generally, we may conclude that there are two major approaches: classifying music based on short excerpts or based on aggregating the values extracting from the whole piece. Music classification is an interesting challenge for the recent research, which can be documented by multiple surveys summarizing the state-of-the-art techniques, such as the one from Bertin-Mahieux [2].

3 Harmonic Complexity

We propose a new feature for music classification, called *harmonic complexity*. Our motivation is that the harmonic movements are important to characterize the music, nevertheless, similar measure has not been introduced yet. Only the main definitions and formulas are summarized in this chapter. For the comprehensive theory and guides on how to calculate the harmonic complexity for a musical piece we refer the reader to our earlier work [11], and for the main definitions of tonal harmony we suggest the introduction by Krumhansl [7].

3.1 Basic Definitions

First, we assume that the musical piece is segmented into a finite number of harmonies, called *chords*. We consider every chord in its *sentential form*. Sentential form is the chronological sequence of *tones* of which the chord consists. Only one occurrence of the same tone is considered and we limit the alphabet to the 12 tones of the *octave*.

Example 1. *An example chord* C major *in the sentential form: ceg.*

Further, we define two basic rules that can be applied to the sentential form (see [11] for definitions). These rules were designed in the accordance with the known tonal harmony rules. The *add* rule adds a tone from the same key to the sentential form. The *alter* rule moves one tone outside the key, altering it by a *semitone* up or down. By the two rules, we can derive more complex chords from the basic chord.

Example 2. *An example derivation from ceg to* $cef\sharp g$*:*

$$ceg \xrightarrow[add(f)]{} cefg \xrightarrow[alter(f,\sharp)]{} cef\sharp g$$

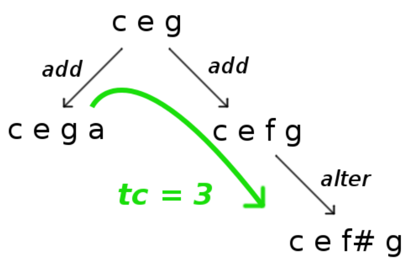

Figure 1: The transition complexity of the chords *cega* and *cef♯g*

Given that the derivation always starts in the basic harmony function, such as *ceg*, we define our first complexity: the *chord complexity*.

Definition 1. *A* chord complexity $c(h)$ *of chord* h *is the minimal length of derivation of the chord's sentential form in the given key*

Example 3. *The chord complexity for the chord* $h = cef\sharp g$ *is:* $c(h) = 2$.

We can now proceed with an informal definition of the *transition complexity* of two chords (the full definition is out of the scope of this paper and can be found in our previous work [11]). The reader can note that the transition complexity is a variant of Levenshtein's edit distance for strings [9].

Definition 2. *A* transition complexity $tc(h_1, h_2)$ *of the chords* h_1 *and* h_2 *is the minimal number of steps needed to derive harmony* h_2 *from* h_1*, using rules* add, alter *and their inverse rules (see Figure 1).*

3.2 Average Transition Complexity

We can now define an aggregation measure for the whole musical piece, that we can use as a feature for music classification. It is a simple mean of all transition complexities between successive chords in the piece.

Definition 3. *An* average transition complexity (ATC) *for a musical piece M, a sequence of chords* $\{C_i\}_{i \leq l}$ *of the length l and a sequence of its transition complexities* $\{t_i\}_{i \leq l-1}$ *is defined as follows:*

$$ATC(M) = \frac{\sum_{i=0}^{l-1} t_i}{l-1}$$

4 Supporting Experiments

We have conducted a series of experiments proving that harmonic complexity is a useful feature for automated music classification. As a classification method we have decided to use machine learning and the theory of artificial neural networks. The motivation for this method was, that we can validate our results on the same data set easily by re-training the network multiple times. We were interested in comparing the performance of a neural network in two scenarios:

1. Each musical piece is represented by a basic set of features
2. Each musical piece is represented by the same set of features with an added ATC value (see Definition 3)

4.1 Basic Set of Features

We have picked features commonly used for music classification tasks. The basic set of features needs to be strong enough for a satisfying classification, so we may observe the added value of the harmonic complexity.

1. **Mel Frequency Cepstral Coefficients (MFCC) mean value and covariation matrix.** MFCC values are commonly used for music classification and speech recognition [10].
2. **Root Mean Square (RMS) Amplitude.** RMS Amplitude represents the volume of the piece.
3. **Tempo in Beats per Minute (BPM).** The values were extracted by GPL-Licensed Queen Mary Vamp plugins set[5].
4. **Chord transition probabilities.** The matrix of probabilities of transitions between the pairs of chords. The chords in the matrix were simplified and represented only by the root tone of the chord.
5. **Number of similarity segments.** Number of segments that are similar in the piece, derived from self-similarity matrix, extracted by Queen Mary Vamp plugins set.
6. **Other harmonic measures.** We completed the feature vector by meta features describing the harmonic content: number of distinct keys, number of modulations, and number of distinct chord roots, acquired by Queen Mary Vamp plugins set.

4.2 ATC Feature Extraction

The whole process of obtaining the ATC values starts with the low-level feature extraction. For that purpose, another Vamp plugins, NNLS Chroma and Chordino[6] in version 0.2.1, were used. Developed by Mauch [12], using the method of non-negative least squares, the NNLS Chroma plugin provided us with chromas, i.e. the presence of frequencies mapped into semitone-spaced 12-dimensional vectors. Chordino plugin was used to locate the points of

[5] http://www.vamp-plugins.org
[6] http://isophonics.net/nnls-chroma/

time when the harmony changes significantly, segmenting the piece into small parts. We then chose the 4 tones with the highest presence in the segment to represent the chord. Finally, a model of harmonic complexity was used to obtain the ATC values.

4.3 Data Set and Choosing the Genres

We selected 5 genres from the modern music: **Electronic, Jazz, Metal, Rock and Pop**. It is obvious, that these genres are difficult to recognize and that they may overlap. As a data set we have used 100 songs (20 for each genre), featured in the 2013 and 2014 music charts[7].

4.4 Network Parameters and Validation

We used 70 songs represented as normalized feature vectors for the training of neural network. 15 songs were used for validation of the training, and 10 songs for testing. The network contained 110 input neurons for basic feature set and 111 input neurons for the feature set with ATC. In the hidden layer, there were 40 hidden neurons to achieve multi-layered learning. Finally, 5 output neurons were used to classify the music – each representing one genre. The neuron with the highest activation determined the genre.

To validate our results, we have used cross-validation with 10 rounds, in each round new sets of songs were used for training, training validation and testing.

4.5 Results

The resulting overall confusion matrices for all tests are shown in the Figure 2. If we consider the classification as a retrieval system, we can observe the *precision* and *recall* values. On the example of Jazz music, precision is the number of real Jazz pieces classified as Jazz, divided by all pieces classified as Jazz. Recall is the number of real Jazz pieces classified as Jazz, divided by all Jazz pieces.

We can see that the use of harmonic complexity yields to up to 4% improvement in the overall precision (74% against 70%). While observing the successive rounds of cross validation, we noticed that, when ATC was used, the overall precision fell below 70% only in one round, as opposed to 4 rounds when the basic features were used.

After a deeper exploration of the results, we found an interesting paradox, concerning the use of ATC. Even though the overall precision was higher with ATC, its classification was less able to distinguish between Electronic and Jazz music. Knowing that Jazz music usually achieves the highest complexity values, we attribute this to the fact, that Electronic music contains segments in which the harmony is not clear (noises, raising or lowering pitches), and therefore obtains higher values too.

[7] http:/www.billboard.com

1: Basic features

REAL GENRE

CLASSIFIED GENRE	E	J	M	R	P	PRECISION
Electronic	**15**	1	1	4	2	0.65
Jazz	0	**15**	0	2	2	0.79
Metal	1	0	**13**	0	0	0.93
Rock	2	3	0	**12**	1	0.67
Pop	2	1	6	2	**15**	0.58
RECALL	0.75	0.75	0.65	0.60	0.75	OVERALL: 0.70

2: with ATC

REAL GENRE

CLASSIFIED GENRE	E	J	M	R	P	PRECISION
Electronic	**13**	1	0	1	4	0.68
Jazz	2	**17**	0	0	0	0.89
Metal	1	0	**17**	0	1	0.89
Rock	2	2	0	**16**	4	0.67
Pop	2	0	3	3	**11**	0.59
RECALL	0.65	0.85	0.85	0.80	0.55	OVERALL: 0.74

Figure 2: Results of classification: 1. Scenario with basic set of features, 2. Scenario with added ATC

We were mostly interested in finding out, which genres are better classified when ATC is used. Besides Metal music, all four other genres were better classified with ATC. The best result, up to 10% improvement, we observed in Jazz music, which yields to the expectations, since Jazz music usually contains complex harmonic movements.

5 Conclusion and Future Work

We have proposed a new feature, harmonic complexity, as a useful descriptor for music classification task. First we have provided a short survey on what are the other similar harmonic features available, but we have concluded that none of them is standardized or widely used. We have then provided the basic definitions for harmonic complexity and ATC as a proposed value for music classification. In the supporting experiments we have shown, that adding the ATC values indeed enhances the music classification using the neural network method. For the future work, experiments on the larger data sets need to be performed. A comparative study of all possible harmonic features would be also needed to establish the harmonic complexity.

Acknowledgments. The study was supported by the Charles University in Prague, project GA UK No. 708314 and by project SVV-2014-260100.

Bibliography

[1] Absolu, B., Li, T., Ogihara, M.: Analysis of Chord Progression Data. In: *Advances in Music Information Retrieval*, *Studies in Computational Intelligence*, vol. 274. Springer (2010)

[2] Bertin-Mahieux, T., Eck, D., Mandel, M.I.: Automatic Tagging of Audio: The State-of-the-Art. In: *Machine Audition: Principles, Algorithms and Systems*. IGI Publishing (2010)

[3] Chang, K.K., Jang, J.S.R., Iliopoulos, C.S.: Music Genre Classification via Compressive Sampling. In: *Proceedings of the 11th International Society for Music Information Retrieval Conference*. ISMIR 2010 (2010)

[4] De Haas, W.B., Magalhães, J.P., Wiering, F.: Improving Audio Chord Transcription by Exploiting Harmonic and Metric Knowledge. In: *Proceedings of the 13th International Society for Music Information Retrieval Conference*. ISMIR 2012 (2012)

[5] Konstas, I., Stathopoulos, V., Jose, J.M.: On Social Networks and Collaborative Recommendation. In: *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '09, ACM (2009)

[6] Krumhansl, C.L.: Cognitive Foundations of Musical Pitch. Oxford University Press, New York (1990)

[7] Krumhansl, C.L.: The Geometry of Musical Structure: A Brief Introduction and History. Computers in Entertainment 3/4 (2005)

[8] Lerdahl, F.: Tonal Pitch Space. Oxford University Press, Oxford (2001)

[9] Levenshtein, V.I.: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics-Doklady 10/8 (1966)

[10] Logan, B.: Mel Frequency Cepstral Coefficients for Music Modeling. In: *Proceedings of the 1st International Society for Music Information Retrieval Conference*. ISMIR 2000 (2000)

[11] Marsik, L., Pokorny, J., Ilcik, M.: Towards a harmonic complexity of musical pieces. In: *Proceedings of the 14th Annual International Workshop on Databases, Texts, Specifications and Objects (DATESO '14). CEUR Workshop Proceedings*, vol. 1139. CEUR-WS.org (2014)

[12] Mauch, M., Levy, M.: Structural Change on Multiple Time Scales as a Correlate of Musical Complexity. In: *Proceedings of the 12th International Society for Music Information Retrieval Conference*. ISMIR 2011 (2011)

[13] Panagakis, Y., Kotropoulos, C.: Music Classification by Low-Rank Semantic Mappings. EURASIP Journal on Audio, Speech and Music Processing 2013 (2013)

[14] Rauber, A., Pampalk, E., Merkl, D.: Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structuring of Music by Musical Styles. In: *Proceedings of the 3rd International Society for Music Information Retrieval Conference*. ISMIR 2002 (2002)

[15] Sapp, C.S.: Visual Hierarchical Key Analysis. Computers in Entertainment 3/4 (2005)

[16] Schönberg, A.: Theory of Harmony. University of California Press, Los Angeles (1922)

[17] Schönfuss, D.: Content-Based Music Discovery. In: *Exploring Music Contents*, *Lecture Notes in Computer Science*, vol. 6684. Springer (2011)

[18] Shao, X., Xu, C., Kankanhalli, M.: Unsupervised Classification of Music Genre Using Hidden Markov Model. In: *Proceedings of 2004 IEEE International Conference on Multimedia and Expo*. ICME '04 (2004)

[19] Soltau, H., Schultz, T., Westphal, M., Waibel, A.: Recognition of Music Types. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. ICASSP '98 (1998)

[20] Tzanetakis, G., Cook, P.: Musical Genre Classification of Audio Signals. IEEE Transactions on Speech and Audio Processing 10 (2002)

[21] Wang, A.L.: An Industrial-Strength Audio Search Algorithm. In: *Proceedings of the 4th International Society for Music Information Retrieval Conference*. ISMIR 2003 (2003)

[22] Xu, C., Maddage, N.C., Shao, X., Cao, F., Tian, Q.: Musical Genre Classification Using Support Vector Machines. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. ICASSP '03 (2003)