Comparison of Harmonic Mid-level Representations for Genre Recognition

Matthias Gruhne and Christian Dittmar

Fraunhofer IDMT, Metadata,
Ehrenbergstr. 31, 98693 Ilmenau, Germany
`{ghe,dmr}@idmt.fraunhofer.de`
`http://www.idmt.fraunhofer.de`

Abstract. In music information retrieval, acoustic low-level features are well studied and successfully applied in diverse classification tasks. So called mid-level features pose a very useful addition to low-level descriptors because they are specifically designed to bridge the gap between the low-level physical representation of music signals and the high-level semantic and symbolic information. Mid-level features have been proposed for different domains, such as dynamics, harmony or rhythm. Harmonic mid-level features, however, have mostly been used for harmonic analysis, such as key detection or chord boundary detection. They are rarely used as features for music classification. This publication describes different harmonic mid-level features and examines their usefulness for genre recognition. State of the art harmonic mid-level features are evaluated and their original extraction procedures are adapted to yield a satisfactory classification accuracy.

1 Introduction

During the last decade a lot of effort has been spent on the development of suitable features for music classification tasks such as genre or mood estimation. Besides common and well described low-level features (e.g., MFCC [10]) several mid-level representations have been proposed, e.g. in [1]. These descriptors intend to bridge the gap between low-level representations and the semantically meaningful high-level features such as symbolic note transcriptions or statements about music properties (e.g., tempo, key). One of such a mid-level representation is based on chroma features. Chroma features are computed by folding down all audible tones regardless of their octave to a 12-dimensional vector representing their distribution the chromatic scale. Chroma features are well described and often used for tasks like chord recognition and chord boundary detection. Diverse approaches for the computation of chroma features have been proposed, among others by Fujishima [2], Gomez[4], and Harte [5]. A detailed overview about different chroma extraction algorithms and their performance is given in [15]. In this report, the enhanced pitch class profile (EPCP) was reported to perform best with real-world music signals. Therefore, EPCP chroma features were used for all investigations in this paper. In the literature, the original chroma feature vector is often postprocessed in order to derive more suitable features for semantic high-level analysis, such as chord recognition. One could argue that the chroma feature vector already contains semantic knowledge, interpretable by humans. But in order to obtain a high-level transcription of, e.g., the chords in a song, a number of consecutive chroma vectors need to be grouped, analyzed, and postprocessed based on majority decisions. However, a number of algorithms were proposed in literature, which process the chroma vector further and transfer them into a mid-level representation based on music theory. These representations increases the accuracy for the chord classification significantly. The proposed algorithms can be distinguished into two categories: The first category alter the distribution of the chroma vector elements based on human perception (Pitch Distribution Profiles). The second category postprocesses the chroma vector into geometric models (Geometric Pitch Spaces). Both principles are based on cognition, but the additional semantic information supplied to the system differs. The following paragraphs describe the two principles in more detail:

1.1 Pitch Distribution Profiles

Pitch Distribution Profiles (PDP) are widely used for key detection or chord recognition. The research in this area is mostly based on Krumhansls [9], [8] probe tone experiments. In these experiments, human listeners were asked to evaluate how good selected tones ('probe tones') cognitively matched with several major and minor cadences. This evaluation resulted in two key-independent profile vectors for major and minor chords. Temperley [16] identified a number of problems with these profile vectors. The diatonic notes in a chromatic scale should be rated higher than the non-diatonic notes. But in Krumhansls profiles this is not always the case. This often causes problems with 7th chords. Therefore, Temperley revised this profile vector slightly. He performed the tests only on transcribed music, but this procedure resulted in a better classification accuracy by using real music. There were two other approaches for pitch profiles described in literature. One is called flat diatonic profile and concentrates only on the diatonic scale [7] [12]. The triad profile on the other hand focuses only on the triads in the scale [16].

In order to perform a classification, one or more of the proposed profiles are correlated with the actual chroma vector. Applying this procedure for major and minor with all possible root notes yields to 24-dimensional vector. The vector that exhibits the highest correlation is chosen as chord candidate. Often, these chord candidates are postprocessed to obtain better results. Evaluations on real polyphonic music using the different key profiles by Krumhansl, Temperley, etc. for key detection were proposed in [7] and [12].

Pitch Distribution Profiles are generally useful for key detection and chord transcription tasks, because they add semantic knowledge to the system describing the general structure of chords. Their application improves the accuracy of the classification significantly.

1.2 Geometric Pitch Spaces

Geometric Pitch Spaces (GPS) are a geometric analogy to express musical relations. The relatedness between pitches is geometrically modeled by smaller distances in a one- or more dimensional space if the chords/notes are similar. Depending on the model the pitch spaces are often complex and difficult to present. One of the first described pitch spaces is the circle of fifths [6]. In this circle the progression of keys and musical chords with a smaller distance in the circle of fifth sound harmonic. Therefore, the chords played successively are adjacent, oppositely or close-by in the circle of fifth in the majority of music. This can be shown by the song "Blowin' in the Wind" by Bob Dylan. It contains the subsequent major chords "D", "G", "D", "G", "A". Looking into the circle of fifth shows that these chords are always close to each other. The circle of fifth describes the relatedness between keys in the same tone mode (major, minor). But the connection between different modes is unsatisfactorily demonstrated [6]. Therefore, a number of different alternatives were described in literature. Tzanetakis [18] transformed the chroma vector into a circular pitch space and postprocessed the resulting complex vector to obtain features for music classification.

Harte and Sandler [5] extended this approach by introducing a new circular pitch space model based on thirds for detecting harmonic changes in musical audio. They used three different circular pitch spaces, one based on fifth, one based on major thirds and one based on minor thirds.

Gatzsche introduced another circular pitch space, called 'symmetry model' in [3]. This model extends the proposal by Harte/Sandler by one additional circle and a consolidated triad circle. This procedure enables a more detailed description of several facets of harmonic perception. Gatzsche and Mehnert used this model to perform music analysis tasks, such as chord boundary detection [13] or tonality analysis and synthesis [3].

1.3 Problems and Goal

To date, postprocessing techniques of the basic chroma features have not been analyzed in depth for application in harmonic similarity tasks. It is expected, that they can increase the recognition accuracy for some tasks significantly. Using the raw chroma features for music information retrieval tasks results in the disadvantage, that a comparison between two songs with the same chord progression but with different root notes might result into large distances despite their similar chord progressions. The overall goal is the evaluation of harmonic descriptions as feature for music information retrieval tasks. Therefore, the postprocessing methods for chroma vectors as described above are utilized. PDP and circular pitch space models are clearly key dependent. The goal is therefore to develop and to evaluate further processing steps in order to generate key independent feature vectors for each of the mentioned techniques. The remainder of this paper is organized as follows: Section 2 proposes solutions in order to compute key independent feature vectors. Section 3 describes the evaluation of the proposed methods and section 5 concludes the paper and suggests further work in this area.

2 Adaption to obtain key independent harmonic mid-level features

As already mentioned, harmonic mid-level features are based on chroma vectors. Based on the problems discussed above, harmonic postprocessing methods, such as PDP or circular pitch space models result in a key dependent feature vectors. These vectors may not be directly applicable for many classification tasks, such as genre detection. Therefore, the following modifications are proposed in order to obtain key independent harmonic mid-level feature vectors from pitch distribution profiles and pitch space models:

PDP related features: The pitch distribution profiles compute the most probable chord represented by a chroma vector. A chroma vector containing 12 elements is therefore mapped into one chord label. This label may be described by one or two characters and a Boolean (”C” to ”B” and ”minor/major”). On the other hand, this label can be expressed by a number $e, e \in 0..23$, whereas the first 12 elements of the vector refer to the labels ”C” to ”B” in the major mode and the second 12 elements of the vector refers to the labels ”C” to ”B” in the minor mode. These extracted labels are highly key dependent. Therefore they have to be post-processed when used as classification feature. Based on the chroma feature, the PDP method computes one of such chord labels $e(j)$ per time frame j. In order to examine the key dependency of the feature, several time frames j are observed. A musical cadence might contain the progression ”C major”,”F Major”,”G major” and ”C Major”, e.g. indicated by the progression $e_1(1..4) = \{0, 5, 7, 0\}$. Another cadence, might contain the same progression in another key: ”G major”,”C major”,”D major” and ”G Major”, e.g. indicated by the progression $e_2(1..4) = \{7, 0, 2, 7\}$. Both vectors e_1 and e_2 contain the same progression, but their numbers are different. A comparison of these numbers would result into large distance, despite they are harmonically similar. Computing a simple derivative of both numbers by

$$K(j) = \begin{cases} e(j) - e(j-1), & \text{for } e(j) - e(j-1) \geqq 0 \\ e(j) - e(j-1) + 12, & \text{for } e(j) - e(j-1) < 0 \end{cases} \tag{1}$$

may result into a similar progression by different keys. Applying this equation to the example above, in $K_1(1..3) = \{0, 5, 2, 5\}$ for the first progression and in $K_2(1..3) = \{7, 5, 2, 5\}$ for the second progression. The only difference is the first element, which needs to be neglected. This procedure enables the comparison of key independent chord progression for chords in the same mode. An additional change of the algorithm must be integrated in the case of successive major and minor chords. If a minor chord is followed by a major chord, their differences might be low, but their harmonic indifference is large (e.g. ”B minor” followed by ”C Major”). Therefore, in case of this occurrence, K(j) is increased by 12:

$$K((e(j) > 11 \ \& \ e(j-1) < 12) \parallel (e(j) < 12 \ \& \ e(j-1) > 11)) = K + 12 \quad (2)$$

This procedure results into key independence for PDP. Since the feature vectors depend on the length of the song, all occurrences $K(j)$ of one song are put into a histogram containing a total size of 24 elements. This histogram is normalized and used as feature vector for music information retrieval tasks.

Features by Harte, Gatzsche, and Tzanetakis: The computation of the features by Harte and Gatzsche and Tzanetakis is conducted in a similar manner. The algorithms of these authors have in common, that they result in vector representations. The length of such vectors represents the strength of the chord characteristic. The angle of the vector represents the direction of the chord center. Both information can be used to classify audio, independently from the original pitch space. The norm of such a vector can be easily used as feature for music information retrieval, since it contains the information independently from the key. The angle of such a vector is clearly chord dependent, because it points into a totally different direction, if the chord changes. This chord dependency leads to a key dependency, because the chord progressions responsible for harmonic similarity are key dependent. A similar idea as already proposed comes into account here. The feature vector in the circular pitch space is time dependent and can be described by $\boldsymbol{v}(j)$, whereas j indicates the number of the frame. The angle $\alpha(j)$ of this vector can be computed by $\alpha(j) = \arctan(v_y(j)/v_x(j))$, whereas x and y indicate the vector cathetuses. The key independence is obtained by differentiating angles of successive frames.

$$\alpha_n(j) = \alpha(j) - \alpha(j-1) \quad (3)$$

The differentiation might result into negative values. Since the angle is represented on a circle, these negative values can be also seen as positive angles, usable as information for a feature vector. The differentiated angle and the norm of the complex vector lead to two dimensional feature vectors. Based on the number of circular pitch spaces considered, three or four of these complex vectors may be appended in order to receive a mid-level feature vector for music information retrieval tasks. For the investigations in this paper, two different mid-level features based on circular pitch spaces were examined: Harte: The circular pitch spaces as described by Harte, circle of fifth, circle of major triads in three modes and circle of minor triads in four modes were postprocessed in order tho receive 8 key independent complex vectors, leading to a mid-level feature vector with 16 dimensions. Gatzsche: The circular pitch spaces as described by Gatzsche, circle of fifth, key related circle of thirds and key related circle of diatonics were postprocessed. This yields to 36 key independent complex vectors, leading to a mid-level feature vector with 72 dimensions. Tzanetakis: Only the circle of fifth is used as mid-level representation, whereby the feature vector consists of 2 dimensions, the norm of the vector and the derivative of the angle.

3 Evaluation

3.1 Test Set

The above described mid-level features were examined for the task of genre classification on polyphonic music by using the mid-level features. Tests could be performed by only classical music with manually annotated chords as ground truth for proving the suitability for harmonic similarity. Such a ground truth does not exist as test set, and creating one takes a lot of effort. Sancho [14] showed that the use of chord progressions is suitable to represent musical genres, because they capture the harmonic rules relevant in each musical period or style. Therefore, a popular music test set containing 775 songs, dividable into 10 genres has been established as ground truth. The 10 genres are: Classical (60), Electronic Music (168), Jazz (112), Pop (74), Rock (91), Folk (53), Urban (52), Speech (31), World Music (120), Others (14), whereby the values in brackets describe the number of songs in this genre. The selection of songs into the specific genre has been agreed upon by 10 expert listeners. The feature vector which serves best in this task are expected to be usable for genre detection in combination with other low- or mid-level features.

3.2 Computed Features

From the described music collection the chroma features were extracted as follows: In order to be sample rate independent, all songs were re-sampled to 44100 Hz. A time-frequency transformation was applied on the windowed frames of each songs with a frame size of 2048 and the overlap of 512 samples. This procedure allows a fair frequency resolution, suitable to capture the notes suitably. A higher time resolution was obtained by an analysis window overlap of 75%. Based on the windowed frames, EPCP chroma features were computed. The resulting chroma features consisted of a vector with 12 bins with a relative strength of each tone in common western representation. Based on these chroma vectors, five harmonic mid-level feature vectors were computed:

- The original chroma feature (named Chroma) has been used in order to compare these results with the results obtained by the other postprocessing algorithms. A chord in music usually is audible for a few seconds. In order to save computation time, the original chroma feature has been grouped to a size of two seconds. In order to compute the feature, 200 successive chroma frames were averaged to one chroma feature. This results into one 12-dimensional vector approximately 2 seconds of music.
- The second examined feature is the chroma length (named Length) as described in chapter 2. The chroma length has been computed by these methods and grouped to a time of 2 seconds for the reasons explained above. This resulted in a one dimensional feature vector every 2 seconds.
- The PDP related features (named PDP) were computed as described in chapter 2. This feature resulted in one 12 dimensional histogram per song.

- The features by Harte (named Triad) were computed based on the underlying chroma features. The computation of the features is described in chapter 2. The size of this harmonic mid-level feature vector consists of 16 dimensions. Since the original chroma features contained a large time resolution, 200 successive features were grouped to obtain one mid-level chroma feature.
- The features by Gatzsche (named Symm) were extracted as described in chapter 2. The feature size consists of 72 dimensions and also a grouping was applied on order to receive one mid-level chroma feature vector every 2 seconds.
- The features by Tzanetakis (named Tzan) were extracted as described in [17] with the exception, that the chroma computation as described in chapter 1 was used. The circle of fifths and the weighting has been applied on the chroma vector and the result was postprocessed as described in Tzanetakis' publication.

3.3 Test Methodology

In order to prepare the test, all previously described extraction algorithms were applied on each song of the test set. The test evaluated the general suitability of each of the proposed harmonic mid-level feature vectors for genre detection. In order to enable practically relevant classification results, the above introduced test set was randomly divided into training- and test set, whereby 70% of the songs were used for training and 30% of the songs utilized as test set. As dimension reduction method, a Linear Discriminant Analysis had been applied. The number of reduced feature dimensions has been reduced to the number of classes minus one. If the feature dimension size was smaller or equal the class size, no LDA has been applied. For the classification, a Gaussian Mixture Model classifier with three mixtures has been trained and classified. Often in music information retrieval, the accuracy of the results is measured by the percentage of the songs correctly classified. This procedure leads to insignificant results in the case of unequal distribution of classes. Therefore, the evaluation in this paper was based on pairwise precision and recall as normally applied on segmentation [11]. The recall indicates the probability that a relevant song is found amongst all songs and the precision indicates the probability that a found song is relevant. The f-measure indicates the overall result by considering precision and recall. In case of the pair-wise precision and recall as performed in this paper, the appearance of pairs of similar classes between the successive result and reference list is evaluated. The comparison of these pairs yields to false positives, false negatives and true positives, which indicates overall results independently from the number of items in each class. The measures precision, recall and f-measure were computed for each test run and for each examined feature vector. In order to rely not only on one compilation of training- and test set, the random separation between test- and training set has been repeated nine times. This enabled a computation of minimum, mean, and maximum. Two test scenarios were run for this publication:

- Test of harmonic mid-level features: This test scenario shows the genre classification accuracy based on the harmonic mid-level features. Therefore, all five post-processed key independent mid-level representations were evaluated in order to examine, which feature vector is most suitable for music information retrieval tasks.
- Test of described post-processing method: The post-processing methods described in this paper enable a key-independence. This allows to find similar chord progressions, even if the key is different. This test scenario evaluates the difference in classification accuracy between the harmonic mid-level representations by Harte and Gatzsche as described in their publications and the post processed mid-level representations as described in this paper.

4 Results and Discussion

Table 1 shows the results for all harmonic mid-level representations with the music collection of 10 genres. Using such a large test set enables an evaluation under real conditions, since for real-life genre detection, more than 9 genres should be distinguished. The three vertical boxes show precision, recall and f-measure. Each of the boxes shows the results of minimum, maximum and average of the 10 classifications. In order to obtain a quick overview about the results, the mean f-measure as well as its standard deviations are depicted in figure 1.

	Precision			Recall			F-Measure		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Chroma	19.25	21.33	22.95	17.02	19.07	20.18	19.14	20.10	20.93
Length	16.35	18.26	20.27	17.05	18.36	20.72	16.69	18.30	20.49
PDP	15.60	21.36	28.00	17.17	21.83	27.36	16.35	21.57	27.67
Triad	28.00	30.59	36.34	33.04	35.46	42.67	30.31	32.84	39.25
Symm	25.60	30.01	33.07	28.72	34.17	38.79	27.07	31.94	35.70
Tzan	17.44	20.81	23.17	21.12	23.31	25.13	19.11	21.97	23.91

Table 1. This table shows minimum, maximum and mean precision, recall and f-measure for genre recognition in percent based on different harmonic mid-level features for the musical test set. The numbers are given in percent.

The recognition rates (mean f-measures) are between 16% and 32%. From a first view, these numbers seem quite low. But a random classification of 10 genres in the test collection has an average probability of 12.9%. This shows that the actual numbers are significantly higher than random guessing. For the further evaluation, the average f-measure is used to compare the usefulness of the harmonic mid-level features. The Chromalength features obtained worst results with approximately 18%. The Chromagram feature performed slightly better with approximately 20%. The harmonic mid-level features by Tzanetakis and the PDP performed almost similar with approximately 22%. The best results were obtained by using Hartes and Gatzsches features with approx. 32%. The significance results are 5%, 7%, 8%, 8%, 18%, and 18% for the features Chromlength, Chromagram, PDP, Tzanetakis, Gatzsche, and Harte, respectively. Inspection of these results shows that all features perform better than random guessing. The pitch space models, which used the circle of fifths and other circles improved the classification rates significantly. It shows that the described postprocessing steps generally perform better than the raw chroma features. It shows also, that the proper usage of circular pitch spaces improves the genre recognition in comparison to PDPs.

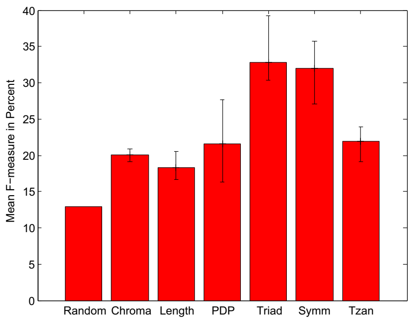

Fig. 1. This figure depicts the results for the genre detection based on different harmonic mid-level features. The genre detection was performed on a test/training set of 10 genres. The values are given in percent.

The second test examined the improvements when using the postprocessing method which computes key independent feature vectors. Table 2 shows the results in percent, whereas the label Symm represents the results of the system by Gatzsche and Triad represents the results by Harte/Sandler. The addition "Post" indicates, that the proposed postprocessing techniques were applied. Both original feature vectors by Gatzsche and Harte performed quite similar with an average f-measure of about 22%. In case, the postprocessing technique was applied, the recognition accuracy increased to about 31%. The results by Gatzsche and Harte might be similar because both authors used similar circles, whereby Gatzsche applied additionally the circle of diatonics. But it seems that this circle has no influence on the results.

In order to obtain a quick overview about the examination, the same results are also depicted in figure 2. The results show, that the postprocessing technique which enables a key independence improves the classification accuracy significantly.

	Precision			Recall			F-Measure		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Symm	18.87	21.91	24.85	20.43	23.26	25.66	19.62	22.56	24.95
Symm Post	25.60	30.01	33.07	28.72	34.17	38.79	27.07	31.94	35.70
Triad	18.75	22.27	28.15	20.77	23.67	27.26	20.38	22.91	26.77
Triad Post	28.00	30.59	36.34	33.04	35.46	42.67	30.31	32.84	39.25

Table 2. This table shows minimum, maximum and mean precision, recall and f-measure for genre recognition in percent based on the basis circular pitch space models by Harte (Triad) and Gatzsche (Symm). Additionally, the results for the classification with the same test data but the proposed changes are given (Symm Post, Triad Post). The numbers are given in percent.

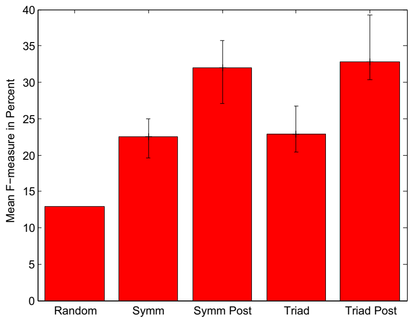

Fig. 2. This figure depicts the results for the genre detection based on two different harmonic mid-level features, the symmetry model and the triad model. It shows the influence of the postprocessing. The genre detection was performed on a test/training set of 10 genres. The values are given in percent.

5 Conclusions and Future Work

This paper evaluated the usefulness of different harmonic mid-level features for genre detection. Two of the examined approaches used the original chroma features or a slight modification for classification. One approach was based on post-processed PDPs and the two other approaches were computed from the model of circular pitch spaces. They showed a clear ranking for the recognition accuracy from the chroma length feature with the lowest performance to the circular pitch space model by Harte and Gatzsche with the highest performance. The results by considering the significance increased by a factor of more than two from 7% to 18%, when applying mid-level postprocessing steps to the original chroma features. The second test showed the influence of the key independent postprocessing technique on the results. The results improved to 23% to 33% when applying this technique. For a real genre detection task, these numbers are of course unsatisfactory. Therefore, other mid- and low-level features must be considered in addition to improve the results.

6 Acknowledgements

This work has been partly supported by the PHAROS Integrated Project (IST-2005-2.6.3), funded under the EC IST 6th Framework Program. Additionally, this project has been funded by the MetaMoses project (nr. 183217) from the Norwegian research council.

References

1. C. Dittmar, C. Bastuck, and M. Gruhne, *Novel mid-level audio features for music similarity*, in Proceedings of the International Conference on Music Communication Science (ICoMCS), 2007.
2. T. Fujishima, *Realtime chord recognition of musical sound: a system using common lisp music*, in Proceedings of the International Computer Music Conference (ICMC), 1999, pp. 464–467.
3. G. Gatzsche, M. Mehnert, D. Arndt, and K. Brandenburg, *Circular pitch space based musical tonality analysis*, in Proceedings of the 124th Audio Engineering Society (AES) Convention, 2008.
4. E. Gomez, *Tonal Description of Music Audio Signals*, PhD thesis, Universitat Pompeu Fabra, Barcelona, 2006.
5. C. Harte, M. Sandler, and M. Gasser, *Detecting harmonic change in musical audio*, in Proceedings of ACM Multimedia 2006 (ACMMM'06), 2006.
6. J. D. Heinichen, *Der Generalbass in der Komposition, 2nd Edition*, G. Olms, 1994.
7. O. Izmirli, *Template based key finding from audio*, in Proceedings of the International Computer Music Conference (ICMC), 2005.
8. C. Krumhansl, *Cognitive foundations of musical pitch*, Oxford University Press, 1999.
9. C. L. Krumhansl and E. J. Kessler, *Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys*, Psychological Review **89** (1982), pp. 334–368.
10. B. Logan, *Mel frequency cepstral coefficients for music modeling*, in Proceedings of 1st International Symposium on Music Information Retrieval (ISMIR), 2000.
11. H. Lukashevich, *Towards quantitative measures of evaluating song segmentation*, in Proceedings of 9th International Symposium on Music Information Retrieval (ISMIR), 2008.
12. S. Madsen and G. Widmer, *Key-finding with interval profiles*, in Proceedings of the International Computer Music Conference (ICMC), 2007.
13. M. Mehnert, G. Gatzsche, D. Arndt, and K. Brandenburg, *Circular pitch space based harmonic change detection*, in Proceedings of the 124th Audio Engineering Convention (AES), 2008.
14. C. Perez-Sancho, D. Rizo, S. Kersten, and R. Ramirez, *Genre classification of music by tonal harmony*, in International Workshop on Machine Learning and Music, International Conference on Machine Learning, 2008.

15. M. Stein, B. Schubert, M. Gruhne, G. Gatzsche, and M. Mehnert, *Evaluation and comparison of audio chroma feature extraction methods*, in Proceedings of the Audio Engineering Conference (AES), 2009.
16. D. Temperley, *The Cognition of Basic Musical Structures*, MIT Press, 2001.
17. G. Tzanetakis and P. Cook, *Musical genre classification of audio signals*, IEEE transactions on Speech and Audio Processing **10**, no. 5 (2002), pp. 293–302.
18. G. Tzanetakis, G. Essl, and P. Cook, *Automatic musical genre classification of audio signals*, in Proceedings of the 2nd International Conference on Music Information Retrieval (ISMIR), 2001.