ON THE IMPACT OF KEY DETECTION PERFORMANCE FOR IDENTIFYING CLASSICAL MUSIC STYLES

Christof Weiß
Fraunhofer Institute for Digital Media Technology Ilmenau
christof.weiss@idmt.fraunhofer.de

Maximilian Schaab
Fraunhofer Institute for Digital Media Technology Ilmenau

ABSTRACT

We study the automatic identification of Western classical music styles by directly using chroma histograms as classification features. Thereby, we evaluate the benefits of knowing a piece's global key for estimating key-related pitch classes. First, we present four automatic key detection systems. We compare their performance on suitable datasets of classical music and optimize the algorithms' free parameters. Using a second dataset, we evaluate automatic classification into the four style periods Baroque, Classical, Romantic, and Modern. To that end, we calculate global chroma statistics of each audio track. We then split up the tracks according to major and minor keys and circularly shift the chroma histograms with respect to the tonic note. Based on these features, we train two individual classifier models for major and minor keys. We test the efficiency of four chroma extraction algorithms for classification. Furthermore, we evaluate the impact of key detection performance on the classification results. Additionally, we compare the key-related chroma features to other chroma-based features. We obtain improved performance when using an efficient key detection method for shifting the chroma histograms.

1. INTRODUCTION

In the field of Music Information Retrieval (MIR), a considerable amount of research has been performed to classify music audio recordings according to different categories [3,29]. Beyond top-level *genres* such as Rock, Jazz, or Classical, several attempts towards resolving *subgenres* have been made. We dedicate ourselves to the subgenre classification of Western classical music which has been addressed sparsely in previous work.

There are plenty of possibilities to organize classical music archives. Apart from the specific artists—soloists or ensembles—, timbral properties such as the predominant instrument(s) may serve as categories [26]. We think that the rather abstract concept pf *musical style* provides a more appropriate subgenre taxonomy. The specific application of this idea leads to the task of composer identification [4, 11, 15, 22]. Beyond such a detailed taxonomy, we restrict ourselves to more general categories—the *historical periods* Baroque, Classical, Romantic, and Modern. [1] This naturally constitutes a simplification but may provide a convenient starting point for finer analyses [6].

Several researchers have published studies on the basis of symbolic data such as score or MIDI representations [1, 8, 10, 11, 15, 22, 25]. However, we find some benefits when directly dealing with audio recordings. First, the audio incorporates more information than the score by representing the "sounding reality" of the music to a higher degree. [2] Second, audio-based methods enable nice applications for organizing and browsing today's large archives of classical music. Morevoer, such archives provide precious possibilities for data-driven musicological research in a new quantitative dimension.

Studies based on symbolic data often make use of musical properties such as the use of specific intervals [1] or chords [22]. Sometimes, characteristics of polyphony and voice leading are considered as well [1,11]. Other methods rely on more fundamental properties of harmony such as the occurrence of pitch classes [10] and pitch class sets [8]. Usually, researchers statistically analyze these characteristics to obtain classification features. These features are then used as input for machine learning (ML) classifiers.

There are several limitations for harmonic analysis of audio based on state-of-the-art signal processing algorithms. Due to the restricted perfomance of automatic music transcription [3] , we build our method upon chroma features that have been shown to suitably represent the pitch class content of audio [7, 19]. Using chroma features, several musical characteristics such as voice leading properties or interval and chord inversions cannot be resolved. Furthermore, acoustic phenomena such as overtones and timbre show considerable effect on the chroma features. Scholars proposed several attempts to approach these problems by enhancing the robustness of chroma [7, 13, 14, 17].

Researches have proposed several chroma-based fea-

[1] Here, the *Modern* class refers to 20th century art music with some stylistic distance from romantic music.

[2] This observation particularly matters for older music such as the Baroque style, where numerous conventions for practical performance were known by the interpreters without notating them in the scores.

[3] In particular, these algorithms highly depend on the orchestration. On that account, automatic transcription is not reliable when dealing with mixed music for piano, orchestra, and voices.

ture types for classifying musical genres and styles. Tzanetakis uses the predominant pitch class, its relative amplitude, and the size of the predominant interval as features [29]. Others extract chords from audio and classify based on the chord types and progressions [24]. In [32], interval and chord types are estimated from different chroma resolutions. Furthermore, measures for quantifying tonal complexity have been tested as classification features [33].

In this paper, we want to test a simpler approach that directly uses *chroma histograms* as input for a classifier (Section 2). For tonal music, the chroma distribution of a piece is mainly influenced by the *musical key*. Usually, the notes of the underlying scale and the most prominent chords obtain high values—such as the tonic note or the dominant note. We therefore test the benefits of knowing the global key for classifying with chroma features. To that end, we first compare four key detection methods (Section 2.2) on suitable datasets of classical music and optimize the algorithms' parameters (Section 3.2). Second, we perform classification experiments on a separate dataset of 1600 classical pieces (Section 3.3). As classification features, we use chroma histograms that are shifted on the basis of different key algorithms or ground truth key annotations. We test the influence of considering the key as well as the effect of training separate models for major and minor keys. Finally, we compare these features' perfomance against other chroma-based features introduced in earlier work [32, 33].

2. PROPOSED METHOD

In Western classical music, tonality and harmony play a central part for establishing musical form, expression, and style. The use of specific pitches, intervals, and chords—as well as their progressions—constitute typical style markers. They hierarchically depend on each other and contribute to the chroma distribution of a piece. Beyond the high importance of the global key, modulations to other keys entail the use of different chords and pitches. That way, sections in foreign keys considerably contribute to the global chroma histogram—depending on their length. Apart from such harmonic characteristics, instrumentation and timbre may affect the shape of the chroma distribution. Let us consider a simple major triad: Depending on the instrumentation, the root, third, or fifth note may be more pronounced leading to different chroma vectors.

Some of these differences may serve to resolve subtler stylistic differences. In Figure 1, we show two chroma histograms of symphony movements by Schumann and Brahms, both in a major key and centered to their respective tonic note (C and F). Though these composers have much in common—a part of their lifetime, the cultural background, and several inspiring persons—the pieces considerably differ in their pitch class histograms. One reason may be the more complex harmony in Brahms' music—the chromatic pitch classes such as F♯, C♯, and A♭ are enhanced compared to Schumann's equivalents. Moreover, Brahms' instrumentation often emphasizes the chords' third notes. This could explain the increased values for D, A, and E—the triad thirds of the main chords B♭M (subdominant), FM (tonic), and CM (dominant), respectively. Another explanation for this observation could be a modulation to the local key A major for a considerable amount of time. Such modulations to third-related *mediant keys* are common in late romantic music.

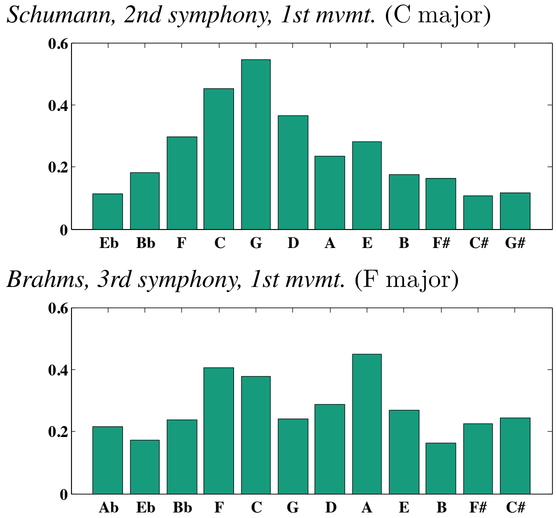

Figure 1. Chroma histograms for Schumann's 2nd symphony, 1st movement in C major (upper plot) and Brahms 3rd symphony, 1st movement in F major (lower plot). The histograms are arranged according to the circle of fifths and centered to the respective tonic note. We normalize the distributions to the ℓ_2 norm in order to ensure comparability.

To describe such characteristics, the *relative* pitch classes are important. Therefore, we need information about the global key. Sometimes, this metadata is provided in musical archives. However, such annotations are often incomplete. For work cycles and multi-movement works, we usually find only one key ("Symphony in F major") which single movements may differ from. For those reasons, we test automatic methods for audio key detection and evaluate the influence of their performance on the overall classification results. We also compare automatic key detection to the use of ground truth key annotations.

Apart from the tonic note, the mode (major / minor) is of high importance, since the harmonic structure of minor pieces fundamentally differs from the one in major. To that end, we split up our data and train a separate model for each mode. Section 2.3 outlines the details of this idea.

2.1 Chroma Features

In audio signal processing, chroma features have been shown to suitably represent tonal characteristics [7, 19]. For a chromagram, the spectrogram bins are mapped into a series of 12-dimensional chroma vectors $\mathbf{c} = (c_0, c_1, \ldots, c_{11})^{\mathrm{T}} \in \mathbb{R}^{12}$. These vectors represent the energy of the pitch classes that are independent from the octave. To reduce the influence of overtones and timbral characteristics, several chroma extraction methods have been proposed. We consider six different approaches:

(i) **CP**. This algorithm [21] is based on a multirate filter bank and published in the Chroma Toolbox [18]. We use the *Chroma Pitch* as our baseline feature.

(ii) **CLP**. Jiang et al. found improvement of chord recognition when using logarithmic compression before octave summarization. We use the *Chroma Logarithmic Pitch* with compression parameter $\eta = 1000$ which performed best in [9].

(iii) **CRP**. Müller and Ewert proposed a method to eliminate timbral information using the Discrete Cosine Transform—*Chroma-DCT-Reduced Log Pitch* [17].

(iv) **HPCP**. These *Harmonic Pitch Class Profiles* consider the overtones for the chroma computation [7].

(v) **EPCP**. In [27], *Enhanced Pitch Class Profiles* [13] performed best in a chord matching experiment. This algorithm makes use of an iterative procedure (harmonic product spectrum). We use three iterations.

(vi) **NNLS**. Mauch introduced an approximate transcription step based on a *Non-Negative Least Squares* algorithm [14]. The resulting chroma features led to a considerable boost of chord recognition performance. The code is published as a “Vamp” plugin. [4]

We compute the initial chroma features with a resolution of 10 Hz. In order to eliminate the influence of dynamics, we normalize to the ℓ^1 norm so that

$$||\mathbf{c}||_1 = \textstyle\sum_{n=0}^{N-1} |c_n| = 1. \qquad (1)$$

2.2 Key Detection Algorithms

For automatic key detection, we compare four approaches that have been tested successfully on classical music data.

(1) **Template matching**. For this standard method, the distance between a chroma histogram and a key profile is computed for each of the 24 keys. The profile minimizing the distance gives the global key [28].

(2) **Profile learning**. Van de Par et al. improved the profile matching algorithm by using a learning procedure for the key profiles [30]. Furthermore, they emphasize the beginning and ending section of the pieces. We extend this idea by separately weighting beginning and ending section. Therefore, we introduce new parameters β and γ to emphasize the beginning and ending, respectively—along with the parameter α from [30].

(3) **Symmetry model**. Another class of key finding algorithms makes use of geometrical pitch models [2, 5]. We use the symmetry model by Gatzsche and Mehnert that was evaluated for key detection in [16].

(4) **Final chord**. The algorithm proposed in [31] considers the final chord to estimate the tonic note of the global key—combined with a profile matching for estimating the mode. This algorithm was tested on three datasets of classical music.

[4] http://isophonics.net/nnls-chroma

2.3 Classification Features

The basic idea of this paper is to directly use chroma histograms for classification of music styles. We therefore sum up the M chroma vectors $\mathbf{c}^1, \ldots, \mathbf{c}^M$ of a piece in order to obtain a ℓ_1 normalized chroma histogram $\mathbf{h}$:

$$\widehat{\mathbf{h}} = \textstyle\sum_{i=1}^{M} \mathbf{c}^i \;, \quad \mathbf{h} = \widehat{\mathbf{h}} \;/\; ||\widehat{\mathbf{h}}||_1 \qquad (2)$$

In order to compare the impact of the chroma computation method, we use four different chroma algorithms from the ones presented in Section 2.1: CP, CLP, EPCP, and NNLS.

As the main contribution of our work, we want to evaluate the relevance of key information for classification. To this end, we test different combinations of key estimation and classification algorithms. Using 3-fold cross-validation, we randomly split our dataset into a training fold (2/3) and a test fold (1/3). For the training stage, the ground truth key annotations are used to split up the data into pieces in major and minor modes. With the same key information, we circularly rotate the chroma histograms so that the tonic note is on the first position:

$$h_k^{\text{rotated}} = h_{(k-k^*) \bmod 12}, \qquad (3)$$

with $k \in [0:11]$ and k^* denoting the chroma index of the tonic note ($k^* = 0$ for C, etc.). For testing, we use one of the four automatic key detection algorithms presented in Section 2.2. With this key information, we split up the test data according to the mode and again rotate each chroma histogram with respect to the tonic note. The full processing chain of our approach is shown in Figure 2.

To compare against existing methods, we use other types of chroma-based classification features. In [32], a set of *template-based features* for estimating the occurence of interval and chord types has been proposed. To this end, chroma features are smoothed to different temporal resolutions followed by a multiplication of chroma values according to interval and chord templates. Another group—*tonal complexity features*—makes use of statistical measures on the chroma distribution in order to estimate the tonal complexity of the music on different time scales [33].

3. EVALUATION

In order to estimate the classification performance on unseen data, we apply a two-step evaluation strategy. First, we test the key detection performance of the four methods presented in Section 2.2 and optimize the algorithms' free parameters (Section 3.2). Second, we perform classification experiments on a different dataset using a Random Forest classifier with chroma histograms as input features. We train separate models for major and minor pieces, respectively. For estimating the importance of the algorithm's elements, we conduct several baseline experiments.

3.1 Datasets

In our studies, we make use of different datasets. To evaluate key detection performance and optimize param-

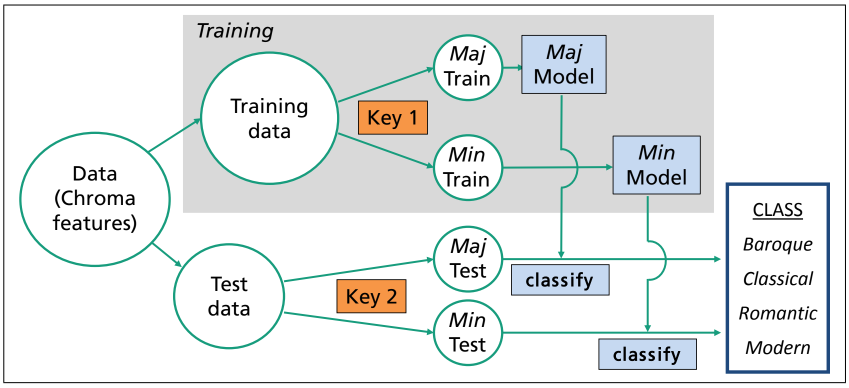

Figure 2. Flow diagram for the classification procedure. For performing cross-validation, the data is split up into training and test set. Each set is sorted with respect to the mode by using different key algorithms or ground truth key annotations (key 1 / key 2), respectively. The trained models for major and minor keys are then used to classify the respective test data.

eters, we use three datasets of classical music recordings with corresponding key annotations. This data has been used for evaluating key detection in published work [23, 30, 31]. The first set (*Symph*) comprises classical and romantic symphonies—each with all movements—from 11 composers containing **115 tracks** in total. The second one—a selection from *Saarland Music Data Western Music* (*SMD*) [20]—includes music for solo instruments, orchestra, and chamber music. The key annotations for **126 selected tracks** that show clear tonality are available on the corresponding website. [5] Third, we recompiled a dataset of piano music recordings (*Pno*) used for key detection in [23, 30]. This data comprises **237 piano pieces** (Bach, Brahms, Chopin and Shostakovich). We consider these datasets as training data for the key detection step, thus justifying the overfitting procedure for the parameters.

For the classification experiments, we make use of another dataset (*Cross-Era*) containing **1600 audio recordings** of classical music as used in [32, 33]. The data is balanced with respect to the historical periods (each 400 tracks for the Baroque, Classical, Romantic, and Modern period) and instrumentation (200 piano pieces and 200 orchestral pieces per class). We collected expert annotations for the key of 1200 tracks. The modern class has not been considered due to a high amount of atonal pieces. For atonal pieces, we assume little influence of key detection on classification with chroma histograms. [6] The data is not balanced with respect to the key or the mode (major/minor). We show the key distribution in Figure 3.

3.2 Key Detection Experiments

For estimating the optimal parameters, we run each algorithm with different parameter settings in a stepwise fashion. To that end, we optimize each parameter by maximizing the weighted total performance Λ_t

$$\Lambda_t = (115\,\Lambda_{\mathrm{Symph}} + 126\,\Lambda_{\mathrm{SMD}} + 237\,\Lambda_{\mathrm{Pno}})\,/478 \quad (4)$$

[5] http://www.mpi-inf.mpg.de/resources/SMD

[6] For example, a dodecaphonic piece of music shows nearly equal pitch class distribution. Thus, its chroma distribution is practically invariant to cyclic shifts.

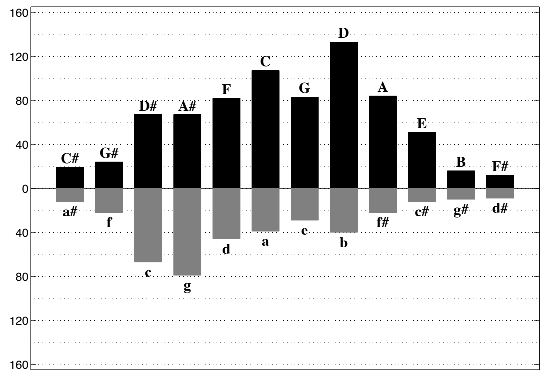

Figure 3. Key distribution (annotations) of the periods Baroque, Classical, and Romantic (1200 pieces) in the dataset *Cross-Era*. Major keys are shown in black and upward direction, minor keys are in grey downwards. The tonic notes are arranged according to the circle of fifths.

and fix the remaining parameters to default or best fit values. For the basic chroma features, we test the six types presented in Section 2.1. We obtain the following results for the different algorithms:

(1) **Template matching**. We test three pairs (maj / min) of profiles proposed by Krumhansl [12], Temperley [28], and Gomez [7]. In our study, the latter ones performed best. Though these profiles have been developed in combination with HPCP features, NNLS features outperformed these features (84.7 %) followed by CLP.

(2) **Profile learning**. For the profile training, we performed a cross-validation with 98 % training data, 2 % test data, and 5000 repetions following [30]. We found best performance for CLP chroma features (92.3 %)—closely followed by NNLS—together with parameters $\alpha = 2$, $\beta = 1$, and $\gamma = 0.25$. We could not reach the result presented in [30] (98 % on the *Pno* dataset). As a reason for this, we assume that the specific chroma features presented in that work (including a masking model) provide additional benefits.

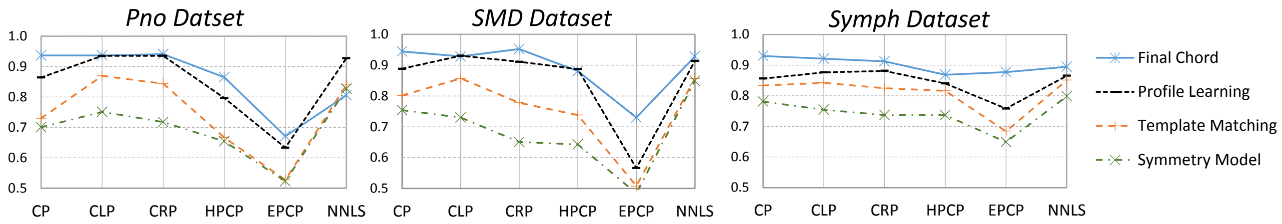

Figure 4. Evaluation of different key detection algorithms. Here, we show the individual key recognition accuracies for the three datasets of classical music. As the basic feature, we compare six types of chroma features.

(3) **Symmetry model**. This algorithm worked best in conjunction with NNLS chroma. The optimal pitch set energy threshold was found at $f_{\mathrm{TR}} = 0.12$. The angular vector value came out best at $w_{\mathrm{sym}} = 0.53$ leading to a total performance of 82.6 %.

(4) **Final chord**. The final chord algorithm obtained optimal results on the basis of CP chroma features. For the parameters, $N = 19$ final frames, a root-scale weight exponent of $s = 0.9$, an energy threshold of $f_e = 0.19\,\%$, and the weight exponents template $\mathbf{m}^{(2)}$ have come out best (93.7 % accuracy).

The overall results for the key detection evaluation are shown in Figure 4 for the individual datasets. All algorithms considerably depend on the chroma extraction method—especially when the data includes piano music (*Pno*, *SMD*). NNLS features often obtained the best results and seem to be the most stable basis for key detection methods. EPCP features are not a good choice for this purpose. The profile learning and the final chord algorithms performed similarly. Hereby, the first one is rather data-dependent whereas the final chord algorithms requires a fine parameter tuning. In the following, we use the final chord algorithm that showed a slightly better total rate (93.7 %) compared to the profile training method (92.3 %).

Finally, we test the four key detection methods on a subset of the *Cross-Era* dataset (Section 3.1) using 1200 tracks with key annotations. For each method, we use the feature and parameter setting performing best in the previous experiments. [7] We obtain a performance of **83.9 %** for the template matching algorithm (1), **87.1 %** for the profile learning (2), **80.4 %** for the symmetry model (3), and **85.4 %** for the final chord based method (4). Compared to the optimization datasets, the performance is worse and the differences between the methods are smaller. Profile learning and final chord stay with best results. However, the learning strategy (2) seems to be more robust than the parameter-dependent final chord algorithm (4).

3.3 Classification Experiments

By using the method and parameters performing best in Section 3.2, we now test the influence of key detection on automatic style classification based on the *Cross-Era* dataset. We use a Random Forest (RF) classifier. In order to avoid problems due to the *curse of dimensionality*,

[7] For the profile learning approach, the profiles are also trained on the previously used datasets *Symph*, *SMD*, and *Pno*.

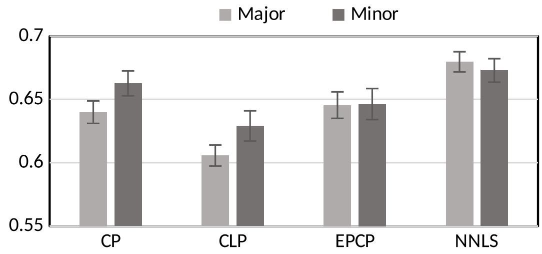

Figure 5. Classification accuracies for different types of chroma features for classification, four classes, *key 1* = *key 2* = final chord. Bars and error marks indicate mean and standard deviation over 100 initializations of the cross-validation. Here, we do not use LDA (only twelve-dimensional features).

we transform the feature space using Linear Discriminant Analysis (LDA) with three output dimensions. For evaluation, we conduct a 3-fold cross-validation. We use the chroma histograms over the full piece as classification features. As our basic idea, we rotate the chroma histograms to the tonic note (Section 2.3). In the ideal setting, we use the ground truth key annotations for the training data (*key 1*). For the test data (*key 2*), we use the automatically detected key from the final chord algorithm (see Section 3.2).

Major and minor keys exhibit very different tonal structures resulting in distinct typical chroma distributions. The mode-related properties in the chroma distribution may heavily overlay the more subtle differences originating from style. We therefore split up the data into major and minor pieces by using key annotations (training set, *key 1*) or automatic key detection (test set, *key 2*), respectively. On the resulting training data sets, we train separate classification models for major and minor keys. The test data is then classified into style periods using the appropriate classifier model. This procedure is visualized in Figure 2. We then repeat the classification by using the next fold as test data. The whole cross-validation is performed 100 times with new random initialization of the folds.

First, we test the influence of the specific chroma feature implementation on the classification performance. In this experiment, we use the automatic key (final chord algorithm) for both training and test. The results are shown in Figure 5. Classification performance considerably depends on the chroma type. Here, logarithmic compression (CLP)—enhancing weak components—does not improve classification performance. CP and EPCP features perform

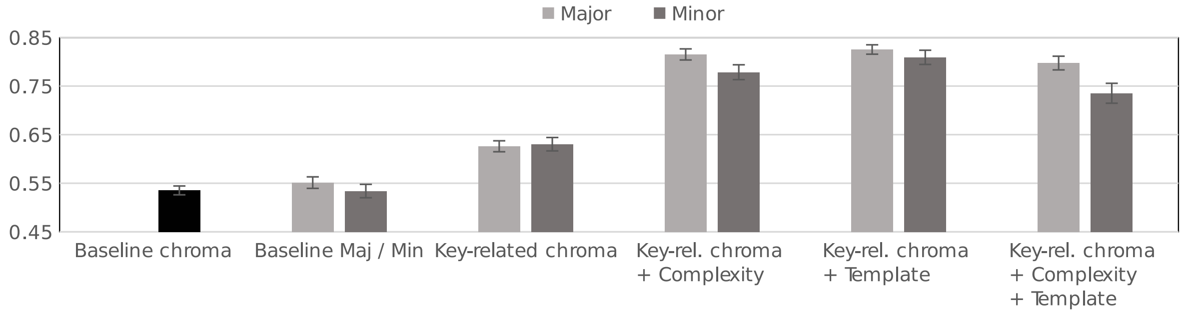

Figure 7. Classification accuracies for different combinations of chroma-based features, four classes, NNLS features. The varying dimensionality of the feature collections is reduced to three dimensions by using LDA.

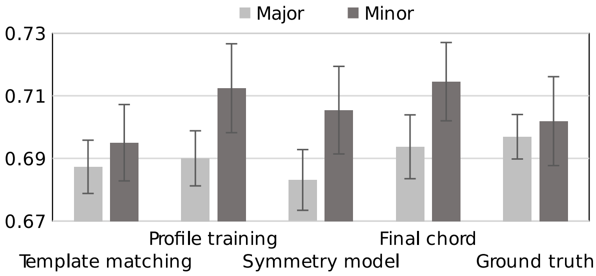

Figure 6. Classification accuracies based on different key detection methods (for *key 1* and *key 2*), three classes, NNLS features. Here, we do not use LDA transformation.

Table 1. Classification results for different key method combinations, three classes, NNLS features.

Key 1	*Key 2*	*Major*	*Minor*
Ground truth	Final chord	70.1 %	66.7 %
Final chord	Final chord	69.4 %	71.5 %

similar whereas NNLS features outperform the others by several percentage points. We therefore use NNLS chroma features for the remaining experiments.

Next, we evaluate the dependence of the key-related chroma features on the performance of the automatic key detection. To this end, we once use each of the four methods from Section 3.2 both for training and test data. Since we have no ground truth key annotations for the modern era, we just perform classification of the remaining three classes (1200 pieces). Classification results are similar with all key methods (Figure 6). For profile learning and final chord key detection, the results partly outperform the classification based on ground truth key annotations. We conclude that some of the errors in key detection may have beneficial effects on classification performance. Comparing the classification results with the key detection perfomance on *Cross-Era*, we find similar behaviour. Thus, a good key detection leads to better classification, sometimes outperforming the use of ground truth key annotations. When using ground truth key for training (*key 1*) and an automatic method for testing (*key 2*), performance values change but do not generally increase (Table 3.3).

In the last study, we compare different types of classification features (Figure 7). For the *baseline chroma* experiment, we do not use any key information but use the original (absolute) NNLS histograms as classification features—without Major / Minor discrimination (one model for all). *Baseline Maj / Min* makes use of ground truth key annotations for mode selection. This does not lead to increased classification results. For the *key-related chroma* method, we use NNLS rotated with respect to the final chord key, for training and test. [8] The use of key detection boosts classification results by almost 10 %. Next, we combine the key-related chroma histograms with other chroma-based features such as tonal *complexity* or *template*-based features (Section 2.3) leading to improvements of almost 20 %. Combining all three types of features does not further increase classification accuracies.

When comparing our results with the outcome of [32, 33], we do not obtain a general performance boost through adding key-related chroma features. Both complexity [33] and template features [32] alone performed similar in the respective experiments—compared to combining them with our features. However, we already obtain remarkable results with key-related features only. These features can be computed with a high computational effiency. [9] As the main difference, complexity and template features capture local properties whereas global chroma histograms do not.

4. CONCLUSION

We evaluated four automatic key detection methods and optimized their parameters using three datasets of classical music. On a separate dataset, we performed style classification experiments using key-related chroma histograms as classification features. With such features, the use of an efficient key detection algorithm improves classification accuracy. Thus, automatic key detection constitutes a useful step for such music classification systems. However, involving local chroma-based features leads to a better performance than only using global chroma histograms.

Acknowledgments: C. W. has been supported by the Foundation of German Business (Stiftung der Deutschen Wirtschaft). He thanks Daniel Gärtner for fruitful discussions and Judith Wolff for contributing to key annotations.

[8] The difference between this result and the NNLS performance in Figure 5 is due to using LDA transformation here.

[9] Since we only use global chroma, a very coarse time resolution for the time-frequency transform could be applied.

5. REFERENCES

[1] Eric Backer and Peter van Kranenburg. On Musical Stylometry: A Pattern Recognition Approach. *Pattern Recognition Letters*, 26(3):299–309, 2005.

[2] Ching-Hua Chuan and Elaine Chew. Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2005.

[3] Roger B. Dannenberg, Belinda Thom, and David Watson. A Machine Learning Approach to Musical Style Recognition. In *Proceedings of the International Computer Music Conference (ICMC)*, 1997.

[4] Ofer Dor and Yoram Reich. An Evaluation of Musical Score Characteristics for Automatic Classification of Composers. *Computer Music Journal*, 35(3):86–97, 2011.

[5] Gabriel Gatzsche, Markus Mehnert, David Gatzsche, and Karlheinz Brandenburg. A Symmetry Based Approach for Musical Tonality Analysis. In *Proceedings of the 8th International Society for Music Information Retrieval Conference (ISMIR)*, pages 207–210, 2007.

[6] Iriving Godt. Style Periods of Music History Considered Analytically. *College Music Symposium*, 24, 1984.

[7] Emilia Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, Barcelona, 2006.

[8] Aline Honingh and Rens Bod. Pitch Class Set Categories as Analysis Tools for Degrees of Tonality. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pages 459–464, 2010.

[9] Nanzhu Jiang, Peter Grosche, Verena Konz, and Meinard Müller. Analyzing Chroma Feature Types for Automated Chord Recognition. In *Proceedings of the 42nd AES International Conference on Semantic Audio*, pages 285–294, 2011.

[10] Francis J. Kiernan. Score-based Style Recognition Using Artificial Neural Networks. In *Proceedings of the 1st International Symposium on Music Information Retrieval (ISMIR)*, 2000.

[11] Peter van Kranenburg. Composer Attribution by Quantifying Compositional Strategies. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, 2006.

[12] Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford Psychology Series. Oxford University Press, 1990.

[13] Kyogu Lee. Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile. In *Proceedings of the International Computer Music Conference (ICMC)*, 2006.

[14] Matthias Mauch and Simon Dixon. Approximate Note Transcription for the Improved Identification of Difficult Chords. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, pages 135–140, 2010.

[15] Lesley Mearns and Simon Dixon. Characterisation of Composer Style Using High Level Musical Features. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, 2010.

[16] Markus Mehnert, Gabriel Gatzsche, and Daniel Arndt. Symmetry Model Based Key Finding. In *Proceedings of the 126th AES Convention*, 2009.

[17] Meinard Müller and Sebastian Ewert. Towards Timbre-Invariant Audio Features for Harmony-Based Music. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):649–662, 2010.

[18] Meinard Müller and Sebastian Ewert. Chroma Toolbox: MATLAB Implementations for Extracting Variants of Chroma-Based Audio Features. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, pages 215–220, 2011.

[19] Meinard Müller, Frank Kurth, and Michael Clausen. Chroma-Based Statistical Audio Features for Audio Matching. In *Proceedings Workshop on Applications of Signal Processing (WASPAA)*, pages 275–278, 2005.

[20] Meinard Müller, Verena Konz, Wolfgang Bogler, and Vlora Arifi-Müller. Saarland Music Data. In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, 2011.

[21] Meinard Müller, Frank Kurth, and Michael Clausen. Audio Matching via Chroma-Based Statistical Features. In *Proceedings of the 6th International Society for Music Information Retrieval Conference (ISMIR)*, pages 288–295, 2005.

[22] Mitsunori Ogihara and Tao Li. N-Gram Chord Profiles for Composer Style Identification. In *Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR)*, 2008.

[23] Steffen Pauws. Musical key extraction from audio. In *Proceedings of the 5th International Society for Music Information Retrieval Conference (ISMIR)*, 2004.

[24] Carlos Pérez-Sancho, D. Rizo, José Manuel Iñesta, Pedro José Ponce de León, S. Kersten, and Rafael Ramirez. Genre Classification of Music by Tonal Harmony. *Intelligent Data Analysis*, 14(5):533–545, 2010.

[25] Pedro José Ponce de León and José Manuel Iñesta. A Pattern Recognition Approach For Music Style Identification Using Shallow Statistical Descriptors. *IEEE Transactions on System, Man and Cybernetics - Part C : Applications and Reviews*, 37(2):248–257, 2007.

[26] Christian Simmermacher, Da Deng, and Stephen Cranefield. Feature Analysis and Classification of Classical Musical Instruments: An Empirical Study. In *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, volume 4065 of *Lecture Notes in Computer Science*, pages 444–458. Springer, Berlin and Heidelberg, 2006.

[27] Michael Stein, B. M. Schubert, Matthias Gruhne, Gabriel Gatzsche, and Markus Mehnert. Evaluation and Comparison of Audio Chroma Feature Extraction Methods. In *Proceedings of the 126th AES Convention*, 2009.

[28] David Temperley. *The Cognition of Basic Musical Structures*. MIT Press, 2001.

[29] George Tzanetakis and Perry Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[30] Steven van de Par, Martin F. McKinney, and André Redert. Musical Key Extraction From Audio Using Profile Training. In *Proceedings of the 7th International Society for Music Information Retrieval Conference (ISMIR)*, 2006.

[31] Christof Weiß. Global Key Extraction from Classical Music Audio Recordings Based on the Final Chord. In *Proceedings of the 10th Sound and Music Computing Conference (SMC)*, 2013.

[32] Christof Weiß, Matthias Mauch, and Simon Dixon. Timbre-Invariant Audio Features for Style Analysis of Classical Music. In *Proceedings of the Joint Conference 40th ICMC and 11th SMC*, 2014.

[33] Christof Weiß and Meinard Müller. Tonal Complexity Features for Style Classification of Classical Music. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.