Timbre-invariant Audio Features for Style Analysis of Classical Music

Christof Weiß
Fraunhofer IDMT, Ilmenau
wes@idmt.fraunhofer.de

Matthias Mauch
Queen Mary University, London
m.mauch@qmul.ac.uk

Simon Dixon
Queen Mary University, London
s.e.dixon@qmul.ac.uk

ABSTRACT

We propose a novel set of chroma-based audio features inspired by pitch class set theory and show their utility for style analysis of classical music by using them to classify recordings into historical periods. Musicologists have long studied how composers' styles develop and influence each other, but usually based on manual analyses of the score or, more recently, automatic analyses on symbolic data, both largely independent from timbre. Here, we investigate whether such musical style analyses can be realised using audio features. Based on chroma, our features describe the use of intervals and triads on multiple time scales. To test the efficacy of this approach we use a 1600 track balanced corpus that covers the Baroque, Classical, Romantic and Modern eras, and calculate features based on four different chroma extractors and several parameter configurations. Using Linear Discriminant Analysis, our features allow for a visual separation of the four eras that is invariant to timbre. Classification using Support Vector Machines shows that a high era classification accuracy can be achieved despite strong timbral variation (piano vs. orchestra) within eras. Under the optimal parameter configuration, the classifier achieves accuracies of 82.5%.

1. INTRODUCTION

The analysis of musical style is a major task in musicology. For the investigation of Western classical music, the most important research topics are the life and works of the composers, as well as their relationships and mutual influences. Finding similarities and trends among composers living at the same time leads to a categorization into historical periods comprising musical works composed under similar artistic premises [1]. In Music Information Retrieval (MIR), the classification of music data into genres is a widely explored task [2]. Some work has been done to obtain a finer resolution of subgenres for Jazz, Pop, and Rock [3] as well as for classifying music into global cultural areas [4]. For such tasks, features describing the timbral properties of the music such as instrumentation, playing and singing style, have been applied successfully since short fragments of music have been shown sufficient to capture the typical sound of a genre.

In contrast, the subdivision of the genre "Classical" has been addressed sparsely for audio data. However, passionate classical music listeners are usually able to identify the historical period or the composer of a work after a few seconds. Since this holds independently of the instrumentation or genre, there must be internal structures in the music that make a Mozart piece sound like Mozart, be it a piano sonata, a string quartet, or a symphony. We show that such structures can be found in the dimensions of tonality, harmony, and melody. Obviously, timbral features will not be able to describe such properties. Therefore, we present a set of timbre-invariant features and evaluate them on a subgenre classification task for classical music audio collections.

Musicologists often prefer the detailed view of single composers or even single works to observe very subtle stylistic differences. They find a great individuality in the style of single composers, together with substantial evolutions and breaks within their oeuvre. Nonetheless, one can observe developmental lines in music history, as well as the breaking of such lines. This is why a classification into eras can be helpful as a first step for analysis, which may be followed by a closer look at individual stylistic tendencies [5, 6]. Most commonly, the classical repertoire, which dominates Western concert halls and classic radio programmes, is divided into historical periods ("eras"). This categorization is a simplification but can provide "a reasonably consistent basis for discussion" [1]. On these grounds, we evaluate our features on the rather superficial problem of classifying music into the periods Baroque, Classical, Romantic, and Modern. Treating this task with success is a first step towards more detailed classification scenarios.

The ideal source for studying composer-specific properties is the musical score since it contains that fraction of a musical performance which is created and controlled by the composers themselves. Approaching scores or symbolic data, several studies have been published: McKay and Fujinaga have performed hierarchical classification into root and leaf genres using high-level musical features on MIDI data [7]. As classical subgenres, they have considered the periods Baroque, Romantic, and Modern. In [8], chord profiles have been used for composer style identification. A similar task has been performed in [9] relying on high level interval-based features. Van Kranenburg has evaluated different composer identification tasks on score [10] and MIDI data [11] using interval- and pitch-related features as style markers.

Perttu studied the increasing chromaticism in Western music from the year 1600 to 1900 on score data [12] while

Ventura used symbolic music text representations to identify historical periods from melodic properties [13]. Another melody-based approach has been tested in a study based on the Peachnote Corpus [14] containing statistics of melodic intervals obtained via Optical Character Recognition from open-access graphical scores. On that data, Rodriguez Zivic et al. [15] performed an unsupervised clustering into compositional styles obtaining a division into the eras Baroque, Classical, Romantic, and Modern. Honingh's approach [16, 17] is based on pitch class profiles which are motivated by recent musicological theories and relate to interval categories. The evaluation was performed on several clustering and classification tasks on MIDI representations of individual pieces. De Leon and Iñesta tested a pattern recognition approach for style identification on MIDI data of monophonic melodies [18].

Such high-level representations are not available in many analysis scenarios. For automatic classification tasks on large audio archives as well as for music search and recommendation tasks, algorithms capable of directly handling audio data are necessary. To extract tonal information from audio, chroma features have been used widely. For example, Müller et al. [19] have made use of their capability for audio matching between orchestral and piano versions of the same piece of music. On that account, we build our system on chroma or pitch class features. [1]

The main contributions of this work are the introduction of novel template-based features computed from chroma and the evaluation of their suitability for describing musical style. We test four different chroma feature types as basis features and investigate the time-scale dependence of the features. For evaluation, we present a new large cross-era data set of classical music audio recordings. On this data set, we show different visualizations and perform classification experiments on several 4-class problems. In particular, we examine the timbre invariance of the features. To evaluate this aspect, we investigate piano music as well as orchestral music and compare the results of different configurations of that data.

2. DATABASE

For our classification task, we built a 1600 track corpus of classical music audio recordings compressed in the MP3 format. The main source for the recordings is a large data set of recordings released by the label NAXOS. We considered music clearly assignable to the four historical periods Baroque, Classical, Romantic, and Modern.

To evaluate the influence of timbre and scoring, we took into account solo piano music as well as orchestral music. For each period, we collected 200 tracks each of piano and orchestra. To avoid the system learning timbral particuliarities, we only selected Baroque piano music performed on the modern grand piano (no harpsichord recordings), and the orchestral data neither includes works featuring voices nor solo concertos. To obtain a subgenre classication rather than capturing individual composer styles, every category contains music from a minimum of five different composers from three different countries.

Since we want to perform a baseline experiment, we did not include composers whose style can be described as lying between two of the periods. [2] To make sure that we do not classify properties other than style-related ones, we tried to include a certain amount of works by the single composers, considering different musical forms (Sonatas, Variations, Suites, Symphonies, Symphonic poems, Overtures, and many more) as well as fast and slow movement types (head movements, minuets, etc.). The keys and modes (major/minor) of the pieces are mixed arbitrarily. The composers and their countries are listed in Table 1.

3. METHOD

We perform a common Machine Learning based classification experiment using a Support Vector Machine (SVM) classifier. First, we obtain the audio signals by decoding the MP3 data. Based on this representation, we calculate four different types of chroma features which have been tested successfully on chord recognition tasks (Section 3.1). To evaluate the influence of time scales and temporal resolution, we compute different smoothed representations of the chroma (Section 3.2). Finally, we calculate a set of interval- and chord-related mid-level features which will be used as input for our classifier (Section 3.3).

3.1 Basis Features

Since early studies have shown the suitability of chroma features for representing tonal characteristics [20, 21], a number of different chroma feature extraction methods have been presented and evaluated. The basic idea of chroma is the mapping of the spectrogram bins into a series of 12-dimensional vectors $\mathbf{c}^i$ representing the energy of the pitch-classes independent of the octave:

$$\left(c_1^i, c_2^i, \ldots, c_{12}^i\right) \;\widehat{=}\; (C, C\sharp, \ldots, B)\,. \qquad (1)$$

c_k^i denotes the k-th element of the i-th chroma vector.

One of the fundamental difficulties of the chroma representation is the handling of the partials: Each note played by an acoustical instrument generates a spectrum showing energy not only at the fundamental frequency but also at the integer multiples of this frequency. While the octave-related harmonics do not cause problems in a chroma representation, harmonics corresponding to other pitches such as the upper fifths may lead to wrong musical interpretations. Several chroma extraction methods try to cope with this issue [22–24]. On this account, we are considering four different chroma computation techniques to test the influence of this processing step:

[1] We know that many harmonic properties cannot be derived this way: In a chroma representation, the separation of the voices is not possible. Therefore, voice leading information is lost. Additionally, characteristics of harmonic intervals depend strongly on the pitch order. For example, a note in perfect fourth distance above the bass note was treated as a dissonance over centuries of Western music whereas the same interval appearing between the upper voices was considered consonant.

[2] For example, no works from Beethoven or Schubert were selected because these composers show influences from both Classical and Romantic styles.

Era	*Scoring*	*Composers*	*Countries*
Baroque	Piano	Bach, J. S.; Couperin, F.; Giustini, L.; Platti, G. B.; Rameau, J.-P.	France, Germany, Italy
	Orchestra	Albinoni, T.; Bach, J. S.; Corelli, A.; Handel, G. F.; Lully, J.-B.; Purcell, H.; Rameau, J.-P.; Vivaldi, A.	England, France, Germany, Italy
Classical	Piano	Cimarosa, D.; Clementi, M.; Dussek, J. L.; Haydn, J.; Mozart, W. A.	Austria, Czechia, England, Italy
	Orchestra	Bach, J. C.; Boccherini, L. R.; Haydn, J. M.; Haydn, J.; Mozart, W. A.; Pleyel, I. J.; Salieri, A.	Austria, England, Germany, Italy
Romantic	Piano	Brahms, J.; Chopin, F.; Faure, G.; Grieg, E.; Liszt, F.; Mendelssohn-Bartholdy, F.; Schumann, C.; Schumann, R.; Tchaikovsky, P. I.	France, Germany, Hungary, Norway, Poland, Russia
	Orchestra	Berlioz, H.; Borodin, A.; Brahms, J.; Bruckner, A.; Dvořak, A.; Grieg, E.; Liszt, F.; Mendelssohn-Bartholdy, F.; Mussorgsky, M.; Rimsky-Korsakov, N.; Saint-Saëns, C.; Schumann, R.; Smetana, B.; Tchaikovsky, P. I.; Verdi, G.; Wagner, R.	Austria, Czechia, France, Germany, Hungary, Italy, Norway, Russia, USA
Modern	Piano	Bartók, B.; Berg, A.; Boulez, P.; Hindemith, P.; Messiaen, O.; Milhaud, D.; Prokofiev, S.; Schoenberg, A.; Shostakovich, D., Stravinsky, I.; Webern, A.	Austria, France, Germany, Hungary, Russia, USA
	Orchestra	Antheil, G.; Bartók, B.; Berg, A.; Britten, B.; Hindemith, P.; Ives, C. E.; Messiaen, O.; Prokofiev, S.; Schoenberg, A.; Shostakovich, D.; Stravinsky, I.; Varese, E.; Webern, A.; Weill, K.	Austria, England, France, Germany, Hungary, Russia, USA

Table 1: Composers contained in the data set, and their countries.

- *CP chroma*: Müller and Ewert [19, 25] presented a chroma extraction method using a multirate pitch filter bank [26]. We use the basic *Chroma Pitch* (CP) as baseline representation. The code was published in the Chroma Toolbox package [26].

- *CLP chroma*: Jiang et al. [27] tested several filter-bank-based chroma features on a chord recognition task. They found a significant improvement when using logarithmic compression before applying the octave mapping. We test the *Chroma Logarithmic Pitch* (CLP) with compression parameter $\eta = 1000$ performing best in this evaluation.

- *EPCP chroma*: A different chord labeler was tested on a number of chroma feature types in [28]. The *Enhanced Pitch Class Profiles* (EPCP) by Lee [23] came out best in this study. They used an iterative approach called harmonic product spectrum (HPS). We use three HPS iterations in our work.

- *NNLS chroma*: In [24], an approximate transcription method using a *Non-Negative Least Squares* (NNLS) algorithm was presented for chroma extraction. The features were used as input to a high-level model for chord transcription which was tested on the MIREX Chord Detection task with good results. The code was published as "Vamp" plugin. [3]

We computed all chroma feature representations with an initial feature rate of 10 Hz using a step size of 4410 at an audio sample rate of 44100 Hz. The features are normalized to the Euclidean norm (ℓ^2 norm) to eliminate the influence of dynamics.

3.2 Multi-Scale Feature Smoothing

Tonal characteristics of music can be regarded at various time scales. On a rough scale, local keys and modulations play an important role. Regarding a finer level, chords and chord changes provide more detailed information. Finally, considering the melodic and voice leading properties will give an insight into the relationship of the pitches to the underlying chords. These layers of tonal characteristic are crucial for musical style recognition: Analyzing a piece of dodecaphonic music, we will find a complex tonality making use of most of the chromatic pitches on a fine scale as well as on a global scale. A Romantic symphony may look similarly complex globally due to numerous modulations while being built on simple harmony on a fine level.

Therefore, we have to consider different temporal resolutions for the computation of our classification features. To do this, we start with the 10 Hz chroma features introduced in Section 3.1 and apply a feature smoothing with different resolutions. We use the approach introduced in [21] for the CENS features with smoothing window length w and downsampling factor d given as numbers of frames. The smoothing procedure is part of the MATLAB Chroma Toolbox [26]. After the smoothing, the feature frames are normalized by the ℓ^2 norm again. Together with the local 10 Hz features and global chroma statistics, we have seven different temporal resolutions (Table 2).

temporal resolution \ *feature type*				
	$\mathrm{CP}^{\mathrm{global}}$	$\mathrm{CLP}^{\mathrm{global}}$	$\mathrm{EPCP}^{\mathrm{global}}$	$\mathrm{NNLS}^{\mathrm{global}}$
	CP^{200}_{100}	CLP^{200}_{100}	$\mathrm{EPCP}^{200}_{100}$	$\mathrm{NNLS}^{200}_{100}$
	CP^{100}_{20}	CLP^{100}_{20}	EPCP^{100}_{20}	NNLS^{100}_{20}
	CP^{20}_{10}	CLP^{20}_{10}	EPCP^{20}_{10}	NNLS^{20}_{10}
	CP^{10}_{5}	CLP^{10}_{5}	EPCP^{10}_{5}	NNLS^{10}_{5}
	CP^{4}_{2}	CLP^{4}_{2}	EPCP^{4}_{2}	NNLS^{4}_{2}
	$\mathrm{CP}^{\mathrm{local}}$	$\mathrm{CLP}^{\mathrm{local}}$	$\mathrm{EPCP}^{\mathrm{local}}$	$\mathrm{NNLS}^{\mathrm{local}}$

Table 2: Feature type $[\,\mathrm{Chroma}\,]^{w}_{d}$ for different time scales specified by the smoothing parameters w and d.

3.3 Classification Features

Relying on the chroma feature types listed in Table 2, we then compute semantic mid-level features describing the

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

Cat.	*Interval*	*Dist. semitones*
PC1	minor second / major seventh	1 / 11
PC2	major second / minor seventh	2 / 10
PC3	minor third / major sixth	3 / 9
PC4	major third / minor sixth	4 / 8
PC5	perfect fourth / perfect fifth	5 / 7
PC6	tritone / diminished fifth	6

Table 3: Pitch Class Set Categories PCa, their characteristic intervals, and the interval distance in semitones.

tonal content of the audio data at several time scales. Since we do not want our features to depend on the global or local key, these features have to be invariant under cyclic shifts of the chroma vector. Motivated by music theory, we start with simple binary templates modeling the interval and chord content of the music. Inspired by the Pitch Class Set Theory, Honigh and Bod [16, 17] performed classification and tonal analysis experiments on MIDI data which showed that pitch class sets can be valuable style markers. A pitch class (PC) set is characterized by its predominant interval class (Table 3). From these classes, so-called prototypes with different numbers of notes can be built. The occurences of these categories are used as classification features.

Since in the chroma vector the octave information is missing, we cannot discriminate between the intervals and their complements. Thus, the six interval categories related to PC1 ... PC6 are the only information left. On every chroma vector $\mathbf{c}^i$ (see Equation 1), we compute a score for the joint appearance of two chroma values related by the respective interval class by multiplying their values. For example, for the feature F_5 related to the perfect fourth/fifth (PC5), we multiply the C chroma with the F chroma (distance of 5 semitones):

$$F^i_{5,1} = c^i_1 \cdot c^i_{1+5} \tag{2}$$

We are interested only in the type of the interval, and not in the specific pitches. Therefore, we want to equally weight all keys and chords and sum over all cyclic shifts:

$$F^i_5 = \sum_{m=1}^{12} F^i_{5,m} = \sum_{m=1}^{12} c^i_m \cdot c^i_{1+(m+5-1) \bmod 12} \tag{3}$$

Finally, we sum over all chroma frames i and divide by the total number of frames N to obtain the average likelihood of this interval on the given time resolution:

$$F_5 = \frac{1}{N} \sum_{i=1}^{N} F^i_5 \tag{4}$$

We can generalize this expression using binary templates $\mathbf{T}^{(a)}$ of exponents for the different interval classes PCa:

$$F_{(a)} = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{m=1}^{12} \prod_{p=m}^{1+(m+11) \bmod 12} \left(c^i_p\right)^{T^{(a)}_p} \right) \tag{5}$$

with the interval templates

$$\begin{aligned}
\mathbf{T}^{(1)} &= (1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0) \\
\mathbf{T}^{(2)} &= (1\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0) \\
\mathbf{T}^{(3)} &= (1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0) \\
\mathbf{T}^{(4)} &= (1\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0) \\
\mathbf{T}^{(5)} &= (1\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0) \\
\mathbf{T}^{(6)} &= (1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0).
\end{aligned} \tag{6}$$

This procedure can easily be extended to sets with three or more notes. As the most basic harmonic vocabulary of Western tonality, we considered the triad types Major, Minor, Diminished, and Augmented:

$$\begin{aligned}
\mathbf{T}^{(7)} &= (1\,0\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0) \\
\mathbf{T}^{(8)} &= (1\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0\,0) \\
\mathbf{T}^{(9)} &= (1\,0\,0\,1\,0\,0\,1\,0\,0\,0\,0\,0) \\
\mathbf{T}^{(10)} &= (1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\,0).
\end{aligned} \tag{7}$$

Note that also the triad inversions are considered by this approach. All of the template-based features $F_1 \ldots F_{10}$ are calculated for every chroma feature type of Table 2 resulting in $10 \times 7 \times 4 = 280$ different features per track.

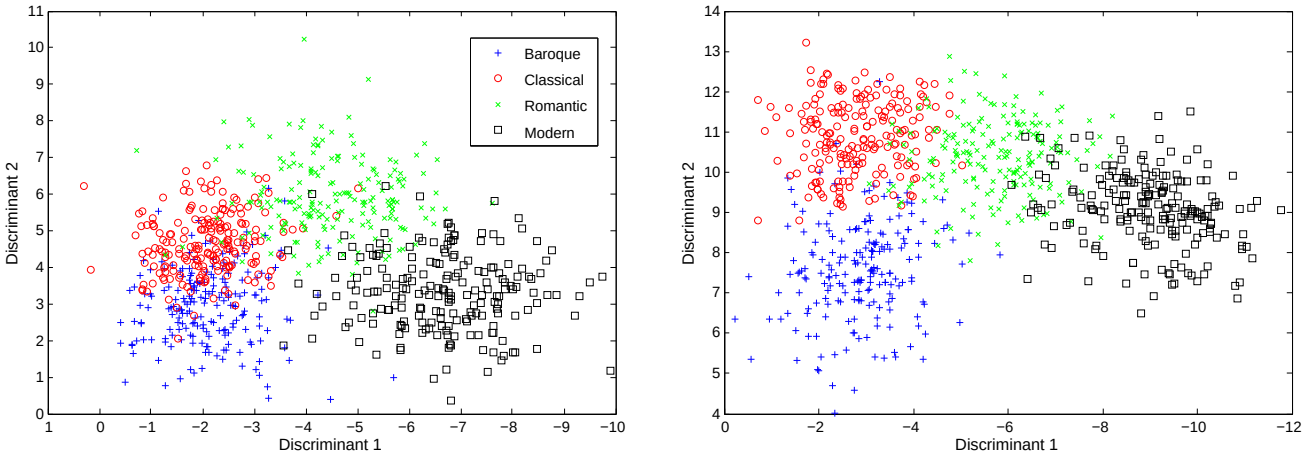
4. RESULTS

4.1 Visualization

To visualize the discriminative power of the proposed features, we apply a dimensional reduction technique known as Fisher transformation or Linear Discrimant Analysis (LDA). This supervised decomposition reduces the dimensions of the feature space in such a way that the classes Baroque, Classical, Romantic, and Modern are optimally separated [29]. The procedure has been used for a similar task in [11]. The results for the full data set are shown in Figure 1, and the visualizations of the piano and orchestra data can be seen individually in Figure 2. A rough separation for the full data seems to be possible with this type of feature; the scenarios considering piano or orchestral music only show slightly better separation of classes. The clustering procedure groups the classes in accordance with their historical ordering. To a great extent, overlapping regions only occur between neighbouring periods.

4.2 Classification

To measure the features' performance for the 4-class era classification problem, we conduct experiments using a standard Support Vector Machine (SVM) algorithm implemented in the LIBSVM library [30]. We are making use of a Radial Basis Kernel Function (RBF kernel) with standard parameters as suggested in [30] and perform a 10-fold cross validation (CV) to study the individual features' influence on the classification performance. All classification experiments are conducted for five configurations of the data performing classification on (1) the *Full* data set, (2) the *Piano* data only, and (3) the *Orchestra* data only,



(a) *Piano* data set. (b) *Orchestra* data set.

Figure 2: LDA visualization of the data subsets.

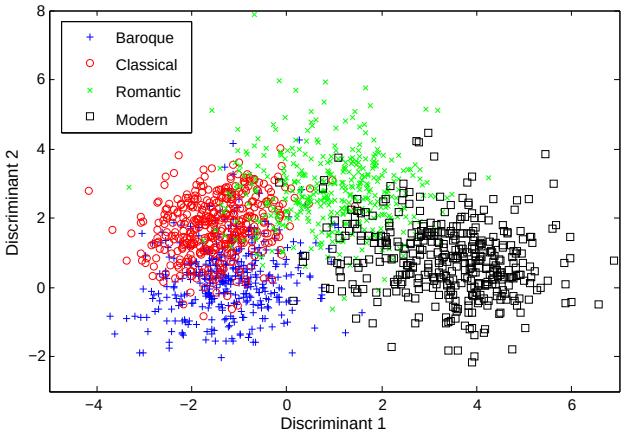


Figure 1: LDA visualization for the full data set.

	Full	*Piano*	*Orch*	*P/O*	*O/P*
CP	63.1 %	61.0 %	65.4 %	46.4 %	48.8 %
CLP	63.1 %	60.6 %	67.6 %	48.9 %	37.5 %
EPCP	64.7 %	62.3 %	69.3 %	52.0 %	41.9 %
NNLS	**79.8 %**	**79.5 %**	**84.9 %**	**65.6 %**	**50.5 %**
all	81.9 %	81.8 %	86.5 %	64.5 %	55.6 %

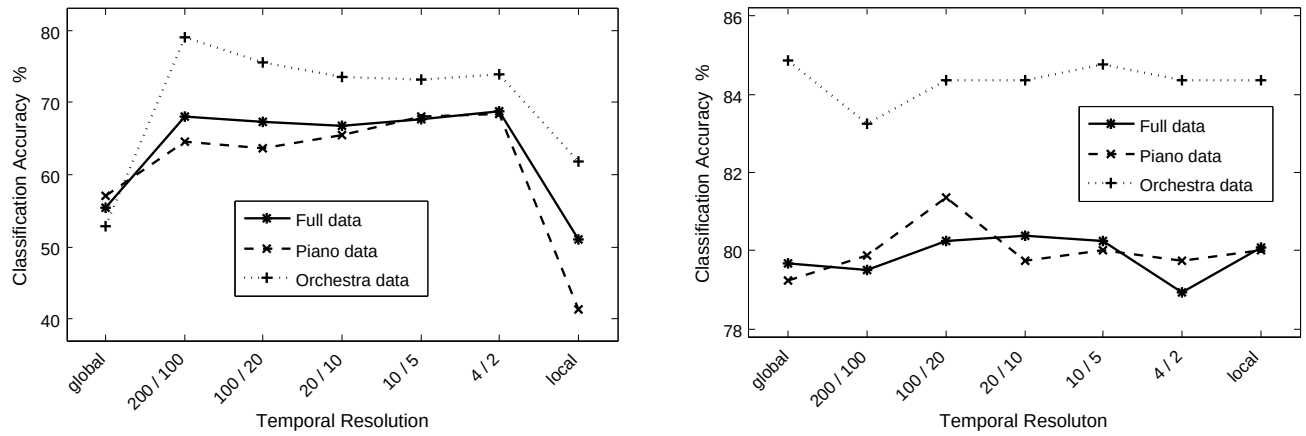
Table 4: SVM classification accuracy for the different types of basis chroma features in a 10-fold (*Full*, *Piano*, *Orch*) and 2-fold (*P/O*, *O/P*) cross validation.

as well as a 2-fold CV (4) training on the piano and evaluating on the orchestra data *P/O* and (5) vice versa *O/P*. The latter two configurations serve to test our hypothesis of invariance against orchestration and timbre.

First, we test the influence of the basis feature type and perform a classification using all templates and time scales (70 features) for each of the chroma types. The results are shown in Table 4. Compared to the simplest chroma approach (CP) resulting in 63.1% accuracy on the full data set, the enhancement of weaker components via a logarithmic compression (CLP) does not improve the classification performance (63.1%) except for a little increase on the orchestral data. This is in contrast to the results of [27] where this procedure improved the performance of a chord labeler. The consideration of the harmonics leads to a weak improvement in the case of the EPCP features (64.7%), whereas the NNLS features show a better performance of 79.8% accuracy reaching almost the result of all chroma feature types combined (81.9%). The reasons for this substantial difference have to be examined in detail in the future. Due to this result, we choose the NNLS chroma as basis feature in the following. Interestingly, the algorithm performs better on the orchestral data compared to the piano data for all feature types. This may be a hint to the fact that composers showed a higher degree of individuality when writing piano music. Another explanation could be the existence of remaining timbral information or peculiarities of the instrumentation in the chroma, which are used by the classifier to determine the era.

To understand the influence of the different time scales, we performed two studies for each of the seven temporal resolutions in Table 2, once (a) using only the respective temporal resolution (10 features) and once (b) leaving out the respective time scale (60 features). The results shown in Figure 3 confirm our assumption that for a powerful classification more than one time scale is needed. Only relying on the global scale leads to bad results since a 12-dimensional global chroma statistics cannot be representative for the tonal characteristics of the music. Nonetheless, also the local and fine scales alone are not sufficient for a good classification either. Leaving out one of the medium resolutions only slightly affects the performance. Thus, we confine ourselves to use only four different time scales for our final experiments, while keeping the variety of different resolutions including global and local scale: $\text{NNLS}^{\text{global}}$, NNLS^{200}_{100}, NNLS^{10}_{5}, and $\text{NNLS}^{\text{local}}$.

On these four temporal resolutions of the NNLS chroma, we test the performance's dependence on the type of the templates. To do this, we first use the two-part interval templates only ($6 \times 4 = 24$ features, see Equation 6) compared to using the three-part triad templates ($4 \times 4 = 16$



(a) Single time scale

(b) Single time scale removed

Figure 3: Classification accuracy for different temporal resolutions.

	Full	*Piano*	*Orch*	*P/O*	*O/P*
Intervals	**68.7 %**	**65.4 %**	**75.6 %**	60.0 %	41.5 %
Triads	64.3 %	60.6 %	75.3 %	**60.1 %**	**46.3 %**
all	75.1 %	71.1 %	80.9 %	69.6 %	45.1 %

Table 5: SVM classification accuracy (10-fold CV) for the different template types on 4 selected temporal resolutions.

	Full	*Piano*	*Orch*	*P/O*	*O/P*
5-dim	78.1 %	77.9 %	84.4 %	17.6 %	25.0 %
4-dim	77.5 %	78.5 %	84.1 %	17.1 %	25.0 %
3-dim	**78.5 %**	**78.5 %**	**84.5 %**	15.9 %	25.0 %
2-dim	68.3 %	68.0 %	82.5 %	7.6 %	25.0 %
1-dim	60.9 %	54.6 %	67.5 %	25.0 %	25.0 %
no red.	75.1 %	71.1 %	80.9 %	**69.6 %**	**45.1 %**

Table 6: SVM classification accuracy (10-fold CV) including a reduction to a different number of dimensions.

	Full	*Piano*	*Orch*	*P/O*	*O/P*
Fold 1	83.4 %	83.0 %	87.8 %		55.8 %
Fold 2	81.6 %	82.5 %	86.3 %	58.0 %	
Comb.	**82.5 %**	**82.8 %**	**87.0 %**	**56.9 %**	

Table 7: SVM classification accuracy of the grid search. For the last two columns, the folds 1 and 2 are identical to the piano and orchestra part of the data, respectively.

features, Equation 7). The results are listed in Table 5. The interval templates are performing slightly better than the triads. However, considering all template types leads to the best results. This may be seen as a motivation to test advanced templates modeling more complex chords. Interestingly, the triad templates show a higher capability to generalize in the cross-instrumentation test.

Keeping this 40-dimensional feature space (4 temporal resolutions $\times$ 10 templates), we finally test if the dimensional reduction technique used for visualization in Section 4.1 improves the classification. To do this, we calculate the decomposition matrix on the training folds and multiply the feature vectors of the test data to this matrix before applying the SVM classifier. Table 6 shows the results for different numbers of dimensions remaining. As we expect for a 4-class problem, the ideal number of feature dimensions after the LDA is 3. In total, classification performance only slightly improves compared to the usage of the full feature space. The cross-instrumentation task fails completely when using LDA. The reason for that may be the preference of different features for the two data sets. The most important dimensions for separating piano music seem to be different from those separating orchestral eras. Nevertheless, there must be features capable of separating both of them well, otherwise classification of the *Full* set would lead to worse results.

For the previous experiments, the parameters c and γ in the RBF Kernel of the SVM classifier have been fixed to standard values. To examine the final classification performance of our system, we conduct a three-stage grid search on these two parameters to optimize the classifier (search area as suggested in [30]). To this, the data set is split into training and test set with equal numbers of classes in each fold (Stratified Cross Validation). On the training fold, the best parameters are selected in another 5-fold cross validation. We measure the classifier's performance with these parameters on the test set and repeat the procedure commuting training and test set. The final results are shown in Table 7, and the confusion matrices for the three sets are displayed in Table 8. The averaged confusion matrix for the cross instrumentation experiment (train/test with either piano or orchestra data) is shown in Table 9. Applying a grid search improves performance from 75.1% to 82.5%.

4.3 Discussion and Outlook

The presented results show that our chroma-based features are able to discriminate classical music styles. The hypothesis of timbre invariance can be verified since the classfication on the full data set leads to similar results as the individual piano or orchestra1 classification. Inspection of the confusion matrices suggests that the best recognition rates can be found for the Modern style. This is no suprise because our "Modern" data contains mostly atonal

	Bar	*Class*	*Rom*	*Mod*
Baroque	**.85**	.10	.05	.01
Classical	.11	**.81**	.09	.00
Romantic	.06	.07	**.76**	.11
Modern	.02	.01	.09	**.89**

(a) *Full* data set.

	Bar	*Class*	*Rom*	*Mod*
Baroque	**.82**	.12	.06	.01
Classical	.09	**.84**	.08	.00
Romantic	.03	.13	**.80**	.06
Modern	.03	.00	.12	**.86**

(b) *Piano* data set.

	Bar	*Class*	*Rom*	*Mod*
Baroque	**.83**	.11	.05	.01
Classical	.09	**.84**	.07	.00
Romantic	.06	.03	**.89**	.03
Modern	.01	.00	.07	**.93**

(c) *Orchestra* data set.

Table 8: Confusion matrices of the grid search classification on the different data sets.

	Bar	*Class*	*Rom*	*Mod*
Baroque	**.32**	.22	.25	.22
Classical	.19	**.49**	.29	.03
Romantic	.11	.04	**.58**	.28
Modern	.03	.00	.09	**.89**

Table 9: Averaged confusion matrix of the grid search classification in the cross instrumentation test (*P/O* and *O/P*).

music and music with a very advanced tonality so that the harmonic material does not consist of triads and common chords anymore. The worst rates are found for the Romantic period. This can have a couple of reasons: Firstly, the transition from the Classical to the Romantic style happened gradually so that these styles may be more similar than other neighbouring eras [1]. On the other hand, late Romantic composers used historical citations and elements from older styles—including also the Baroque style—as an artistic means. Lastly, late Romantic music anticipates the movement towards complex tonality in the 20th century.

In all experiments, the orchestral data can be classified better than the piano or the combined data. We suggest two explanations for this: firstly, the style characteristics could be more pronounced for orchestral music. This could arise from the fact that orchestral music is dedicated to a larger audience and thus may be less complex than piano music. Secondly, our features could still contain timbral information which may be more useful when classifiying on a purely orchestral data set.

There are several open questions that should be addressed in further work. We aim to look into the data in more detail as well as develop more elaborate features to further improve classification performance. To underline the suitability of timbre-invariant features for the analysis of musical styles, the method should be tested against a classification approach using standard features such as Mel Frequency Cepstral Coefficients. The templates used in this work describe interval and basic triad types. Since more complex chords such as seventh or ninth chords can represent style characteristics, templates with more non-zero entries should be included. Furthermore, templates modeling voice leading phenomena such as suspended chords or characteristic dissonances should be tested.

Concerning the data, experiments with finer "stylistic resolution" such as the classification of sub-eras (Early Romanticism, Late Romanticism, etc.) would be interesting contributions. This includes the composer identification task, or even beyond that: Can we see that the early Beethoven sonatas are closer to the Classical area than the later sonatas, which are "more Romantic"?

5. CONCLUSIONS

In this work, we proposed chroma-based features which quantify the occurrence of interval and triad types at different temporal resolutions. Our approach links to more recent ideas in musicology such as the Pitch Class Set Theory. As basis features, we tested four chroma extraction methods (three of them are public code). After a multi-scale feature smoothing, we obtained seven different temporal resolutions. Based on these features, we computed ten classification features making use of a template-matching strategy for intervals and triads.

To test the hypotheses of stylistic differences and timbre invariance, we compiled a 1600 track data set containing piano and orchestral music from composers who can be assigned clearly to one of the four historical periods Baroque, Classical, Romantic, and Modern. Using Linear Discrimant Analysis, we showed the features' capability for separating these classes and for producing nice visualizations. We performed several classification experiments using a Support Vector Machine classifier. In these studies, we evaluated each of the different feature extraction steps. As basis feature, the Nonnegative Least Squares chroma worked best with our features (79.8%), reaching almost the result of using all basis features combined (81.9%). We showed that for a proper classification, more than one time scale is needed and finally considered four temporal resolutions. The test of different templates resulted in a better performance when using interval features (68.7%) rather than triad templates (64.3%), but best performance was obtained with all templates together (75.1%). Combining the most successful features, we performed a grid search to optimize the classifier (82.5%). The results on the orchestra data outperformed the full results by up to 5 percentage points, the piano results were similar or worse than the full data classification. Separating training and test fold between piano and orchestra yielded worse accuracies but still above chance level.

These results indicate that classical music style can be analyzed directly from audio recordings. Apart from the difficulties of the categorization into four eras, the features are able to describe the main stylistic differences of these classes while showing a high degree of timbre invariance. In further studies, we will test the method on tasks with

finer resolution such as sub-era and composer classification. Together with the proposed features, modeling more complex harmonic properties such as tonal complexity and chord sequences will allow us to gain insights into further aspects of musical style and influences between composers.

Acknowledgments

C. W. thanks Jakob Abeßer, Daniel Gärtner, Alexander Loos, Christian Dittmar, and the PhD students at C4DM for fruitful discussions and help with specific tasks. Furthermore, he gives thanks to QMUL for the organization and to the Foundation of German Business (Stiftung der Deutschen Wirtschaft) for the funding of an extended research stay in London.

6. REFERENCES

[1] I. Godt, "Style periods of music history considered analytically," *College Music Symposium*, vol. 24, 1984.

[2] R. B. Dannenberg, B. Thom, and D. Watson, "A machine learning approach to musical style recognition," in *Proceedings of the International Computer Music Conference (ICMC)*, 1997.

[3] V. Tsatsishvili, "Automatic subgenre classification of heavy metal music," Master's thesis, University of Jyväskylä, Jyväskylä, 2011.

[4] A. Kruspe, H. Lukashevich, J. Abeßer, H. Großmann, and C. Dittmar, "Automatic classification of musical pieces into global cultural areas," in *Proceedings of the 42nd AES International Conference on Semantic Audio*, 2011.

[5] J. Webster, "The eighteenth century as a music-historical period?" *Eighteenth Century Music*, vol. 1, no. 01, pp. 47–60, 2004.

[6] P. L. Frank, "Historical or stylistic periods?" *Journal of Aesthetics and Art Criticism*, vol. 13, no. 4, pp. 451–457, 1955.

[7] C. McKay and I. Fujinaga, "Automatic genre classification using large high-level musical feature sets," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, 2004, pp. 525–530.

[8] M. Ogihara and T. Li, "N-gram chord profiles for composer style identification," in *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, 2008.

[9] L. Mearns and S. Dixon, "Characterisation of composer style using high level musical features," in *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010.

[10] P. v. Kranenburg and E. Backer, "Musical style recognition - a quantitative approach," in *Proceedings of the Conference on Interdisciplinary Musicology (CIM 2004)*, 2004.

[11] P. v. Kranenburg, "Composer attribution by quantifying compositional strategies," in *Proceedings of the 7th Internations Conference on Music Information Retrieval (ISMIR)*, 2006.

[12] D. Perttu, "A quantitative study of chromaticism: Changes observed in historical eras and individual composers," *Empirical Musicology Review*, vol. 2, no. 2, pp. 47–54, 2007.

[13] M. Ventura, "Detection of historical period in symbolic music text," *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 4, no. 1, pp. 32–36, 2014.

[14] V. Viro, "Peachnote: music score search and analusis platform," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, 2011.

[15] P. H. Rodriguez Zivic, F. Shifres, and G. A. Cecchi, "Perceptual basis of evolving western musical styles," *Proceedings of the National Academy of Sciences*, vol. 110, no. 24, pp. 10 034–10 038, 2013.

[16] A. Honingh and R. Bod, "Pitch class set categories as analysis tools for degrees of tonality," in *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010, pp. 459–464.

[17] ——, "Clustering and classification of music by interval categories," in *Proceedings of the Third International Conference on Mathematics and Computation in Music*, ser. MCM'11. Springer-Verlag, 2011, pp. 346–349.

[18] P. P. J. d. Léon and J. M. Iñesta, "Pattern recognition approach for music style identification using shallow statistical descriptors," *IEEE Transactions on System, Man and Cybernetics - Part C : Applications and Reviews*, vol. 37, no. 2, pp. 248–257, 2007.

[19] M. Müller, F. Kurth, and M. Clausen, "Audio matching via chroma-based statistical features," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, 2005, pp. 288–295.

[20] M. A. Bartsch and G. H. Wakefield, "To catch a chorus: Using chroma-based representations for audio thumbnailing," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 15–18.

[21] M. Müller, Frank Kurth, and Michael Clausen, "Chroma-based statistical audio features for audio matching," in *Proceedings Workshop on Applications of Signal Processing (WASPAA)*, 2005, pp. 275–278.

[22] E. Gómez, "Tonal description of music audio signals," PhD thesis, Universitat Pompeu Fabra, Barcelona and Spain, 2006.

[23] K. Lee, "Automatic chord recognition from audio using enhanced pitch class profile," in *Proceedings of the International Computer Music Conference (ICMC)*, 2006.

[24] M. Mauch and S. Dixon, "Approximate note transcription for the improved identification of difficult chords," in *Proceedings of the 11th International Society for Music Information Retrieval Conference*, 2010.

[25] M. Müller, *Information Retrieval for Music and Motion*, 1st ed. Berlin and Heidelberg: Springer Verlag, 2007.

[26] M. Müller and S. 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)*, 2011.

[27] N. Jiang, P. Grosche, V. Konz, and M. Müller, "Analyzing chroma feature types for automated chord recognition," in *Proceedings of the 42nd AES International Conference on Semantic Audio*, 2011.

[28] M. Stein, B. M. Schubert, M. Gruhne, G. Gatzsche, and M. Mehnert, "Evaluation and comparison of audio chroma feature extraction methods," in *Proceedings of the 126th AES Convention*, 2009.

[29] A. Webb, *Statistical Pattern Recognition*, 2nd ed. John Wiley and Sons Ltd, 2002.

[30] C. C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001.