TONAL COMPLEXITY FEATURES FOR STYLE CLASSIFICATION OF CLASSICAL MUSIC

Christof Weiss

Fraunhofer Institute for Digital Media Technology
Ilmenau, Germany

Meinard Müller

International Audio Laboratories
Erlangen, Germany

ABSTRACT

We propose a set of novel audio features for classifying the style of classical music. The features rely on statistical measures based on a chroma feature representation of the audio data and describe the tonal complexity of the music, independently from the orchestration or timbre of the music. To analyze this property, we use a dataset containing piano and orchestral music from four general historical periods including Baroque, Classical, Romantic, and Modern. By applying dimensionality reduction techniques, we derive visualizations that demonstrate the discriminative power of the features with regard to the music styles. In classification experiments, we evaluate the features' performance using an SVM classifier. We investigate the influence of artist filtering with respect to the individual composers on the classification performance. In all experiments, we compare the results to the performance of standard features. We show that the introduced features capture meaningful properties of musical style and are robust to timbral variations.

Index Terms— Tonal Features, Musical Style Classification

1. INTRODUCTION

In Music Information Retrieval, the classification of audio data into genres and stylistic categories is a central task [1, 2]. There have been several attempts to obtain a finer class resolution by considering sub-classes of individual genres such as rock [3], electronic [4], or ballroom dance music [5]. Most of these tasks have been addressed using timbral or rhythmic features. In contrast, there are only few methods concerning the subgenre classification for classical music. Some of them use instrument categories as sub-classes [6], others try to identify a small number of composers [7, 8]. Further work has been performed considering only symbolic data [9, 10].

For musicologists and lovers of classical music, a categorization into historical style periods may be helpful. Independent from the timbre and sound of the music, a piano sonata, a string quartet, and a symphony composed in the same musical style are often considered similar. For this kind of similarity, the melodic and harmonic properties of the music are crucial. The use of modulations, individual chord changes, chord types, intervals or the pitch content of the music constitute important style characteristics. Over the history of Western music, certain evolutions can be observed with respect to the underlying pitch content. For early music such as the Gregorian chant and Renaissance vocal music, the pitches of a diatonic scale were sufficient for most of the pieces. In Baroque and Classical music, characteristic modulations to related local keys played an important role to clarify the musical structure. In the Romantic period, an extended use of modulations and more dissonant chords led to an increased complexity in the tonal domain, ending in a fully equal use of all chromatic pitches in dodecaphonic music.

Inspired by these considerations, we propose a novel set of features that correlate to some kind of tonal complexity and test the features' usefulness for style classification. Concepts for describing the tonal complexity of popular music were tested in [11]. The suitability of pitch class set categories for measuring degrees of tonality was evaluated in [12]. Audio-based methods to quantify tonal complexity were presented in [13,14]. To the authors' knowledge, the use of tonal complexity measures for style classification has been addressed sparsely. Some approaches to describe musical style rely on different aspects of tonality such as the usage of particular chords [9], intervals [8], or key-related pitch class occurrences [15] and were tested on MIDI or score data. As the main contribution of this paper, we introduce novel measures for tonal complexity and show their usefulness for style analysis of classical music. To this end, we compare these features against standard features regarding the performance in visualizations and classification experiments. We have discussed the musicological implications of some of these features in [16].

The paper is structured as follows: First, we explain the basic chroma feature types and the smoothing procedure used for the feature calculation. Then, we introduce methods for quantifying tonal complexity. Next, we present the evaluation dataset. Moreover, we show visualizations of the feature space to demonstrate the features' suitability for separating the classes. Finally, we present classification results for different constellations of the features and the data.

2. FEATURES

2.1. Basic Features

To achieve invariance to timbre and orchestration, we build our system on chroma features which have been shown to capture tonal characteristics and to be invariant against timbre variations to a large extent [17, 18]. Let $\mathbf{c} = (c_0, c_1, \ldots, c_{11})^{\mathrm{T}} \in \mathbb{R}^N$ denote a chroma vector of dimension $N := 12$, with $c_n \geq 0$ indicating the energy of the n-th pitch class. The indices $n = 0, 1, 2, \ldots, 11$ correspond to the twelve chroma values C, C$^\sharp$, D, ..., B. Because of the octave invariance, the features show a cyclic nature so that a transposition in pitch leads to circular shift. The temporal resolution is 10 chroma features per second (10 Hz). The features are normalized columnwise with respect to the ℓ^1-norm so that $||\mathbf{c}||_1 = 1$.

In real audio recordings, the partials belonging to other pitch classes than the played note have an influence on the chroma distribution. This may lead to a significant difference between the chroma vector and the notated pitch classes at a time. To make chroma features more robust to changes in timbre, several procedures have been proposed. We use four different chroma types for our experiments: The approach presented in [19, 20] uses a multirate pitch filter bank for the chroma extraction. We use this chroma type as our baseline representation (*CP chroma*). Jiang et al. [21] tested several filter-bank-based chroma features on a chord recognition task and found a

significant improvement using logarithmic compression before applying the octave mapping. This *CLP chroma* feature [22] is used with a compression parameter $\eta = 1000$ in our experiments. In [23], a different chord recognition task was tested on several chroma feature types where the Enhanced Pitch Class Profiles (EPCP) by Lee [24] performed best. Here, the partials are considered using a multiplicative method called harmonic product spectrum (HPS). We use this *EPCP chroma* with three iterations of the HPS. Finally, we use the *NNLS chroma* method presented in [25]. Here, a Non-Negative Least Squares (NNLS) algorithm is used to estimate the fundamental frequencies which can be considered as an approximate transcription before the chroma mapping. The approach significantly improved the performance of a chord labeler.

2.2. Scale Dependence

The measurement of tonal complexity crucially depends on the time scale of the observation. On a chromagram with fine resolution, the measures give an estimate of the local complexity of chords and scales. Regarding coarser levels, the complexity of several bars or a whole section is calculated. Using a chroma histogram as input, the complexity value refers to the pitch content of the full movement. To this end, we use several smoothed versions of the chromagram by applying a smoothing procedure to the initial chroma features presented in [17]. For every feature type presented in Section 2.1

$$[\text{Chroma}] \in \{CP, CLP, EPCP, NNLS\}, \qquad (1)$$

we use four time scales with increasing window size: The initial chromagram with a feature rate of 10 Hz $[\text{Chroma}]^{\text{local}}$, a slightly smoothed $[\text{Chroma}]_5^{10}$ and a highly smoothed version $[\text{Chroma}]_{100}^{200}$, as well as a global chroma histogram $[\text{Chroma}]^{\text{global}}$. After smoothing, the features are normalized with respect to the ℓ^1 norm.

2.3. Complexity Features

Motivated by the considerations presented in 1, we want to find a measure, say Γ, that expresses the complexity of the (local) tonal content. To this end, we now propose several statistical measures calculated on a chroma vector. The basic idea of all these features is to calculate a measure for the flatness of a chroma vector. This is motivated by the following considerations. On a fine level, the simplest tonal item may be an isolated musical note represented by a dirac-like ("sparse") pitch class distribution

$$\mathbf{c}^{\text{sparse}} = \left(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\right)^{\text{T}} \qquad (2)$$

to which the lowest complexity value $\Gamma(\mathbf{c}^{\text{sparse}}) = 0$ should be assigned. Furthermore, a sparser chromagram describing, for example, a diatonic scale should obtain a smaller complexity value than an equal ("flat") distribution $\mathbf{c}^{\text{flat}} = \tilde{\mathbf{c}}^{\text{flat}}/||\tilde{\mathbf{c}}^{\text{flat}}||_1$ with

$$\tilde{\mathbf{c}}^{\text{flat}} = \left(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\right)^{\text{T}}. \qquad (3)$$

The latter case where all twelve pitch classes have the same energy could be found in dodecaphonic music and should be rated with the highest complexity value $\Gamma(\mathbf{c}^{\text{flat}}) = 1$. In addition to the aforementioned boundary conditions, our feature values should increase for growing tonal complexity and be scaled to unit range:

$$0 \leq \Gamma \leq 1. \qquad (4)$$

In the following, we present a number of features for capturing such characteristics. Though not every feature fulfills all requirements, it could be able to model an individual aspect of tonal complexity.

(1) Sum of chroma differences: To account for harmonic similarity of pitch classes, we re-sort the chroma values to an ordering of perfect fifth intervals (7 semitones) $\mathbf{c}^{\text{fifth}}$:

$$c_n^{\text{fifth}} = c_{(n \cdot 7 \bmod 12)}. \qquad (5)$$

Then, we compute the absolute differences between all neighboring chroma values:

$$\tilde{\Gamma}_{\text{Diff}}(\mathbf{c}) := \sum_{n=0}^{N-1} |c_{n+1}^{\text{fifth}} - c_n^{\text{fifth}}|. \qquad (6)$$

Since $\tilde{\Gamma}_{\text{Diff}}(\mathbf{c}^{\text{flat}}) = 0$ and $\tilde{\Gamma}_{\text{Diff}}(\mathbf{c}^{\text{sparse}}) = 2$, we rescale this feature, with $\gamma_1 := 2$:

$$\Gamma_{\text{Diff}}(\mathbf{c}) := 1 - \tilde{\Gamma}_{\text{Diff}}(\mathbf{c})/\gamma_1. \qquad (7)$$

(2) Standard deviation of the chroma vector:

$$\tilde{\Gamma}_{\text{Std}}(\mathbf{c}) = \left(\frac{1}{N-1} \sum_{n=0}^{N-1} \left(c_n - \frac{1}{N} \sum_{m=0}^{N-1} c_m\right)^2\right)^{1/2} \qquad (8)$$

The standard deviation reaches its maximum for a sparse distribution $\gamma_2 := \Gamma_{\text{Std}}(\mathbf{c}^{\text{sparse}}) = 1/\sqrt{12} \approx 0.29$ so that we calculate the rescaled feature in the following way:

$$\Gamma_{\text{Std}}(\mathbf{c}) = 1 - \tilde{\Gamma}_{\text{Std}}(\mathbf{c})/\gamma_2. \qquad (9)$$

(3) Negative slope of a linear function: We re-order the chroma vector entries to a descending series

$$\mathbf{c}^{\text{descend}} = \left(c_{\max}, \dots, c_{\min}\right). \qquad (10)$$

To measure the flatness, we apply linear regression assuming c_i^{descend} being dependent on the index i. The slope $\lambda(\mathbf{c}^{\text{descend}})$ of the line that best fits $\mathbf{c}^{\text{descend}}$ in a least squares sense serves as feature value. For a sparse chroma vector, the fitted line has a slope of $\gamma_3 = |\lambda(\mathbf{c}^{\text{sparse}})| \approx |-0.039|$. Hence, we rescale this feature to

$$\Gamma_{\text{Slope}}(\mathbf{c}) = 1 - |\lambda(\mathbf{c}^{\text{descend}})|/\gamma_3. \qquad (11)$$

(4) Shannon entropy of the chroma vector:

$$\Gamma_{\text{Entr}}(\mathbf{c}) = -\left(\sum_{n=0}^{N-1} c_n \cdot \log_2(c_n)\right)/\log_2(N). \qquad (12)$$

(5) Non-Sparseness feature based on the relationship of ℓ^1- and ℓ^2-norm [26], inverted by subtraction from 1:

$$\Gamma_{\text{Sparse}}(\mathbf{c}) = 1 - \left(\sqrt{N} - \frac{||\mathbf{c}||_1}{||\mathbf{c}||_2}\right)/\left(\sqrt{N} - 1\right). \qquad (13)$$

(6) Flatness measure describing the relation between the geometric and the arithmetic mean [27]:

$$\Gamma_{\text{Flat}}(\mathbf{c}) = \left(\prod_{n=0}^{N-1} c_n\right)^{1/N}/\left(\frac{1}{N} \sum_{n=0}^{N-1} c_n\right). \qquad (14)$$

(7) Angular deviation of the fifth-ordered chroma vector: We re-sort the chroma values according to Eq. 5 obtaining a circular distribution of the pitch class energies—similar to the circle of fifths. Therefore, we interpret $\mathbf{c}^{\text{fifth}}$ as angular data and calculate the angular deviation describing the spread of the pitch classes:

$$\Gamma_{\text{Fifth}}(\mathbf{c}) = \left(1 - \left|\frac{1}{N} \sum_{n=0}^{N-1} c_n^{\text{fifth}} \exp\left(\frac{2\pi i n \pi}{12}\right)\right|\right)^{1/2}. \qquad (15)$$

All of the proposed features take values between 0 and 1, where $\Gamma(\mathbf{c}^{\text{flat}}) = 1$ and $\Gamma(\mathbf{c}^{\text{sparse}}) = 0$ is fulfilled. Γ_{Diff} and Γ_{Fifth} respect the ordering of the chroma entries and penalize distant relations

in a perfect fifth sense. The remaining features are invariant against permutation of chroma values. With this set of features, we consider several flatness-related aspects of a chroma vector.

For each frame (chroma vector), we compute the introduced complexity features. Then, we calculate the arithmetic mean and the standard deviation to obtain classification features for the full track, resulting in $2 \times 7 = 14$ features per track. This procedure is conducted for the four chroma representation presented in Section 2.1 and for each temporal resolution explained in Section 2.2, resulting in $14 \times 4 \times 4 = 224$ feature dimensions in total.

3. EVALUATION

3.1. Dataset

For our evaluation, we use a dataset comprising 1600 tracks of classical music recordings, categorized into four historical periods Baroque, Classical, Romantic, and Modern. From a musicological point of view, this constitutes a rather superficial categorization. However, the obtained simplified classification task can serve as a first step towards a finer style analysis [28]. To investigate the timbre dependence of the features, we collected for each period 200 tracks of orchestra pieces and 200 tracks of piano music. Each of these eight classes contains pieces by a minimum of five composers from at least three different countries. Critical composers who cannot be assigned clearly to one of the periods (such as, for example, Beethoven or Schubert, who can be considered both as late classical or early romantic composers) have been avoided. To preserve the variety of movement types with respect to rhythm and emotion (major/minor keys, slow/fast tempo, duple/triple meter), we included all parts for most of the work cycles. More information about the dataset can be found in [29], where this dataset was used for testing a different class of features.

3.2. Dimensionality Reduction

To analyze the separation capability of the proposed features, we perform a dimensionality reduction technique know as Linear Discriminant Analysis (LDA). This supervised method projects the feature space onto a low-dimensional subspace while separating the classes as much as possible. The procedure can also be used as a preprocessing step for classification experiments. For convenient visualization, we choose two subspace dimensions. First, we perform LDA on all complexity features (Figure 1). With these features, the Classical, Romantic, and Modern styles can be separated well from each other. Between the Modern era and the other periods, the best separation is obtained. This indicates that our features can discriminate between tonal (low complexity) and atonal (high complexity) music. The desired separation of the Romantic style and the Classical style may be the result of the higher tonal complexity of Romantic music compared to Classical music. However, the classes of Classical and Baroque music could not be separated well. Even though one would expect a considerable difference between Barouqe and Classical harmony, these characteristics were not captured by the used features and classification scheme.

To compare with common methods, we also test standard audio features for calculating LDA visualizations of our data. We consider Mel Frequency Cepstral Coefficients (MFCC), Octave Spectral Contrast (OSC), Zero Crossing Rate (ZCR) and Audio Spectral Envelope (ASE), Spectral Flatness Measure (SFM), Spectral Crest Factor (SCF), as well as Spectral Centroid and Modulations thereof (CENT). In addition to these timbre-related features, we calculate
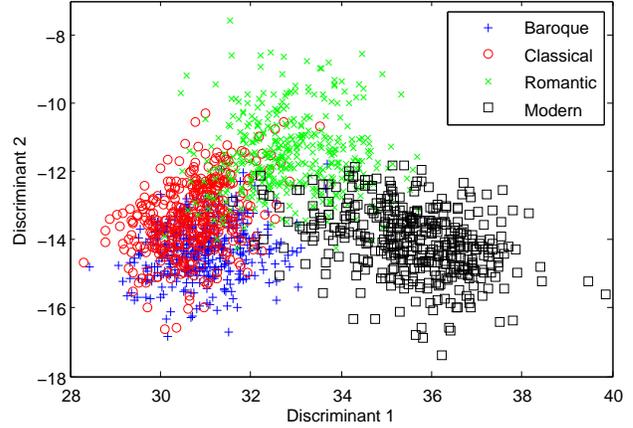
Fig. 1: LDA visualization for the full dataset, using complexity features ($224 \rightarrow 2$).
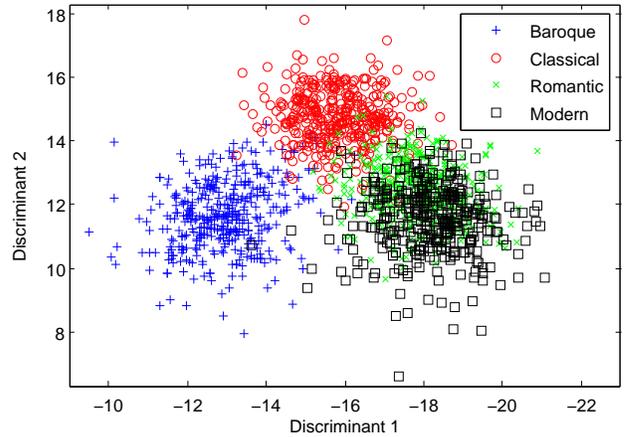
Fig. 2: LDA visualization for the full dataset, using standard features ($238 \rightarrow 2$).
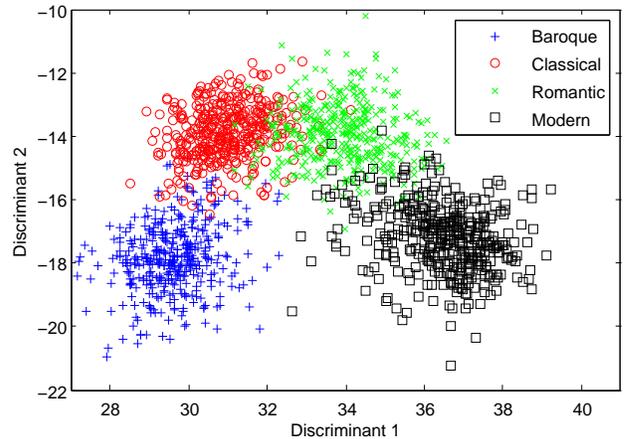
Fig. 3: LDA visualization for the full dataset, using both complexity features and standard features ($462 \rightarrow 2$).

a normalized and a logarithmic version of the specific loudness, after grouping again into sub-bands [30]. For each audio track, we

calculate mean and standard deviation of the frame-wise features resulting in 238 features per track. When performing LDA using these features, we observe a different distribution of the data (Figure 2). In particular, Baroque and Classical music are separated well here. This may be the result of a considerable change between these periods regarding the sound of the music. Indications for such a change may be, for example, the disappearance of the figured bass (basso continuo) in orchestral music—which is usually played with the involvement of a harpsichord—or a different use of musical registers. Romantic music, however, cannot really be identified with standard audio features but is mixed up with Classical and even more with Modern music. A possible reason for this may be the rather continuous evolution of instrumentation from the Classical period on. For example, the scoring of an orchestra was extended step by step from a small Classical orchestra (Haydn) to a huge Romantic orchestra (Bruckner) which many modern composers have changed only slightly (Shostakovich).

Because of the complementary behaviour of complexity features and standard features, the separation capability may benefit from a combination of the two feature types. Figure 3 confirms this assumption. Using both feature sets, Baroque and Classical music can be discriminated very well, thanks to the standard features. The good separation between Romantic and Modern may originate mostly from the use of the complexity features. Discrimination of Classical and Romantic music also benefits from the joint usage of the features, but is still more difficult than separating the other classes. This is in accordance with musicological expectations since the stylistic change from the Classical to the Romantic period is the least distinctive one between all neighboring periods.

3.3. Classification Experiments

Finally, we want to evaluate the performance of the proposed features in style classification experiments. Therefore, we use a standard Support Vector Machine (SVM) classifier with a Radial Basis kernel function (RBF), using the implementation published in [31]. As an alternative, a Gaussian Mixture Model (GMM) classifier was tested leading to similar results. For evaluation, we conduct a three-fold cross validation (CV). Since overtraining constitutes a problem for a higher number of features (the so-called "curse of dimensionality"), we apply LDA to the initial feature space of the training data. For our four-class problem, we use three feature dimensions as input for the SVM classifier. To optimize the RBF kernel parameters of the SVM, we perform a grid search on the training folds.

To investigate the timbre dependence of the approach, we consider different subsets of the data: *Full Data* contains all 1600 tracks; for the *Piano* and *Orchestra* subsets, only one type of instrumentation is considered (800 tracks). As mentioned in Section 3.1, the classes usually contain more than one track from an album. Thus, we have to take care of the "album" or "artist effect": If both training and test folds contain items from the same CD recording, the system can adapt to technical artifacts or the specific sound of a recording rather than learning musically meaningful properties [32,33]. Additionally, we want to avoid substantial influence of a specific composer style on the classification but capture the overall style characteristics of a period. Motivated by these considerations, we apply a "composer filter" which forces a composer's works to be in the same fold, thus avoiding the album effect and a "composer effect" at the same time. [1]

The classification results are shown in Table 1. Let us start with the results when using a three-fold CV without composer filter. All

---

[1] The dataset does not contain works by different composers that are on one album.

Features	Dim.	Full Data	Piano	Orchestra
Classification accuracy without composer filter				
Complexity	224	.86	.87	.86
Standard	238	.87	.89	.86
Combined	462	.92	.86	.81
Classification accuracy with composer filter				
Complexity	224	.69	.64	.75
Standard	238	.54	.29	.71
Combined	462	.67	.46	.67

Table 1: Results for different constellations of classification features. The upper block shows the results for the combination of all features without using composer filtering. In the last block, these experiments are repeated using the composer filter. "Dim." denotes the number of feature dimensions before performing LDA.

feature constellations yield high results. For example, the classification with complexity features achieves an accuracy of .87 for piano music and .86 for orchestral music as well as .86 for the full dataset. In comparison to the complexity features, standard features perform similar on this task, with a slightly better result of .89 for piano music. Combining the two feature types, the classification accuracy for the full dataset improves to .92. This is in accordance with the observation of Section 3.2, where the separation of classes for the LDA visualizations could be improved as well. In contrast, the accuracies for the isolated datasets are worse for the combined features.

Now, let us consider the classification with composer filter. In general, the results get much worse. Looking at the complexity features, the accuracy for the full data decreases to .69. In general, standard features are more affected by the composer filter than complexity features. This may originate from the album effect or composer-specific characteristics. We find an extreme case for piano data: Using standard features, the classification fails (with .29 slightly above chance level) whereas complexity features (.64) still yield some meaningful result. This may indicate that our complexity features capture some "musical" information that is not related to timbre but to tonal aspects. Moreover, complexity features perform even better than all features combined when using a composer filter.

We have conducted first experiments to study the influence of parameters such as the chroma feature type, the temporal resolution of the chroma, or the individual computation method for the complexity features. The results of such studies will be part of future work.

4. CONCLUSIONS

We have proposed a novel set of timbre-invariant audio features for the classification of classical music style. The features are based on chroma representations and relate to the tonal complexity of the music. Experiments on piano and orchestra recordings point out that our proposed tonal complexity features can be used in combination with standard audio features to better separate the historical periods. We have shown this by visualizing supervised reductions of the feature space. In classification experiments, the proposed complexity features exhibit a high robustness to composer- and artist-specific properties. Thus, the features may describe some musically meaningful information which standard features do not capture.

5. REFERENCES

[1] 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.

[2] George Tzanetakis and Perry Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[3] Valeri Tsatsishvili, *Automatic Subgenre Classification of Heavy Metal Music*, Master's thesis, University of Jyväskylä, Jyväskylä, Finland, 2011.

[4] Daniel Gärtner, Christoph Zipperle, and Christian Dittmar, "Classification of electronic club-music," in *Proceedings of the DAGA 2010: 36. Jahrestagung für Akustik*, 2010.

[5] Simon Dixon, Elias Pampalk, and Gerhard Widmer, "Classification of dance music by periodicity patterns," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*, 2003.

[6] 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*, pp. 444–458. Springer, Berlin and Heidelberg, 2006.

[7] Zhen Hu, Kun Fu, and Changshui Zhang, "Audio classical composer identification by deep neural network," *Computing Research Repository*, 2013.

[8] Peter von Kranenburg and Eric Baker, "Musical style recognition - a quantitative approach," in *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, 2004.

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

[10] 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.

[11] Mitchell Parry, "Musical complexity and top 40 chart performance," 2004.

[12] 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)*, 2010.

[13] Matthias Mauch and Mark Levy, "Structural change on multiple time scales as a correlate of musical complexity," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, 2011.

[14] Sebastian Streich, *Music Complexity: A Multi-Faceted Description of Audio Content*, Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[15] Francis J. Kiernan, "Score-based style recognition using artificial neural networks," in *Proceedings of the 1st International Symposium on Music Information Retrieval (ISMIR)*, 2000.

[16] Christof Weiss and Meinard Müller, "Quantifying and visualizing tonal complexity," in *Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM)*, 2014.

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

[18] Emilia Gómez, *Tonal Description of Music Audio Signals*, Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2006.

[19] Meinard Müller, *Information Retrieval for Music and Motion*, Springer, Berlin and Heidelberg, 2007.

[20] 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)*, 2011.

[21] 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*, 2011.

[22] Meinard Müller and Sebastian Ewert, "Towards timbre-invariant audio features for harmony-based music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 649–662, 2010.

[23] Michael Stein, Benjamin 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.

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

[25] 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)*, 2010.

[26] Patrick O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[27] Geoffroy Peeters, "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," Technical report, 2004.

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

[29] 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.

[30] Hugo Fastl and Eberhard Zwicker, *Psychoacoustics: Facts and models*, Springer, Berlin and Heidelberg, 1990.

[31] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[32] Elias Pampalk, Arthur Flexer, and Gerhard Widmer, "Improvements of audio-based music similarity and genre classificaton," in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, 2005.

[33] Arthur Flexer, "A closer look on artist filters for musical genre classification," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007.