MID-LEVEL CHORD TRANSITION FEATURES FOR MUSICAL STYLE ANALYSIS

Christof Weiß, Fabian Brand, Meinard Müller

International Audio Laboratories Erlangen

ABSTRACT

Chords and their progressions are an important aspect of Western tonal music. Specifically, transitions between subsequent chords within a piece carry style-relevant information. To extract such information from audio recordings, a naive approach first peforms automatic chord estimation for computing chord labels explicitly and then derives transition statistics. Often, this is done with Hidden Markov Models involving the Viterbi decoding algorithm. However, since chords are often ambiguous, deciding on one "optimal" chord sequence can be problematic, which heavily affects the subsequent derivation of transition features. In this paper, we propose novel mid-level features that capture chord transitions in a "soft" way. Our method exploits the Baum–Welch algorithm, which does not involve hard decisions on chord labels. Instead, we obtain probabilistic features that account for ambiguities among chords and chord transitions. In several experiments, we evaluate these features within a style classification scenario discriminating four historical periods of Western classical music. Our soft transition features consistently achieve higher accuracies than comparable hard-decision features, thus demonstrating the descriptive power of the novel features.

***Index Terms*—** chord transitions, musical style, genre classification, computational music analysis

1. INTRODUCTION

The analysis of music recordings with respect to tonal characteristics is a central problem within Music Information Retrieval (MIR) [1]. Typical tasks such as global key detection [2, 3], local key detection [4, 5], or chord recognition [6–8] relate to tonal structures on various temporal scales. Beyond these concrete analysis scenarios, tonal features showed success for classifying music recordings with respect to more abstract categories such as musical styles [9–12]. Concretely, features quantifying the presence of certain chord or interval types [9] or the tonal complexity [10] led to efficient and robust discrimination of the four eras Baroque, Classical, Romantic, and Modern (Figure 1c). Beyond that, features describing transitions between consecutive chords (chord progression bigrams) turned out beneficial for this task [11]. From music theory [13], we know that such chord transitions bear style-relevant information (Section 2), which could be verified in a study based on audio recordings [14].

Using a naive approach, Weiss *et al.* [11, 14] explicitly compute chord labels from audio recordings and derive chord transition features from the resulting label sequence by counting transitions of a particular type. To extract chord labels, they make use of a typical chord recognition pipeline [7] based on a suitable pitch class representation (chromagram) and Hidden Markov Models (HMMs) together with the Viterbi decoding algorithm [15]. This algorithm aims for providing the "single best state sequence" [16] (here: chord label sequence) given a particular chroma sequence. This is a useful approach when the label sequence itself is of interest (e. g. , when providing chord labels for a pop song to an amateur guitar player). However, this "optimal" chord sequence might not be the correct or the *only correct* solution from a musicological perspective. Reasons for this lie in an often restricted chord vocabulary [17], an imperfect chord model [8], and the high level of musical abstraction required for this task [18, 19]. These problems lead to natural ambiguities of estimated chords and their transitions. Such ambiguities cannot be captured by the explicit label sequence produced by a traditional chord recognition system. In Section 2, we discuss this in detail.

To overcome this problem, we aim for mid-level features that describe chord labels and transitions in a "soft" way—providing continuous-valued probabilities rather than deciding on a unique output. This additional information can be useful for music analysis, visualization, and classification purposes where the mid-level features can be interpreted either by a human expert or a machine learning system. In particular, we hypothesize that such features are useful for classifying music recordings according to style categories [11].

As our main contribution, we propose a novel strategy for deriving continuous-valued chord transition features from music recordings without the need to explicitly compute chord labels (Section 3). On the basis of chroma representations, we estimate chord transition probabilities using the Baum–Welch algorithm. In a case study on a choral by J. S. Bach, we show visualizations indicating the potential of these features for music analysis (Section 4). By aggregating local transition probabilities over time, we compute different types of piece-level, transposition-invariant features (Figure 1b). In several experiments (Section 5), we demonstrate the potential of these features for style classification of Western classical music recordings.

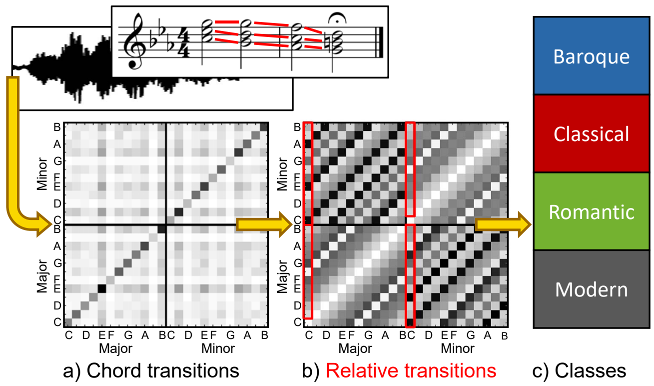

Fig. 1. Extraction, post-processing, and application of mid-level chord transition features (schematic overview).

This work was supported by the German Research Foundation (MU 2686/10-1, MU 2686/11-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS. The authors thank Sebastian Ewert for inspiring and fruitful discussions.

2. MUSIC THEORY BACKGROUND

The notion of chords and their progressions is central for composing, performing, and understanding Western tonal music. We consider a chord as a group of simultaneously sounding notes usually perceived as an entity [20]—the "vertical way" of understanding music [21]. Analyzing musical pieces regarding the underlying chords is of central interest for musicology since the chord types used and their progressions constitute characteristic indicators for musical style [22, 23]. More specifically, transitions between subsequent chords bear style-relevant information [13, 14, 24].

For many purposes, it is sufficient to specify chords at the pitch class level disregarding their concrete realization (octave information / inversions). Assuming enharmonic equivalence, we rely on the twelve-tone equal-tempered scale. Consequently, we define a chord by its root pitch class and its type—a specific set of intervals above the root. Typically, these interval sets comprise successions of major and minor thirds resulting in triads (major, minor, diminished, augmented), seventh chords (dominant 7th, major 7th, ...), or more complex types. This high number of chord types leads to a large vocabulary involving complex relationships among chords. For example, the Cm^7 chord (C-E$^\flat$-G-B$^\flat$) both comprises the Cm triad (C-E$^\flat$-G) and the $\mathrm{E}^\flat\mathrm{M}$ triad (E$^\flat$-G-B$^\flat$). The triads C+ (C-E-G$^\sharp$) and E+ (E-G$^\sharp$-B$^\sharp$) are identical under enharmonic equivalence (B$^\sharp$ = C).

Such relationships become relevant for approaches relying on reduced chord vocabularies—which is often the case in chord recognition systems within MIR [8]. Reducing complex chord types to a vocabulary of the 24 triads (*MajMin*) is problematic and leads to ambiguities [17]. Further ambiguities arise from melodic figuration (non-chord tones) involving subjective decisions [19] (which notes are part of the chord?). Estimating chord labels from audio recordings introduces further uncertainties due to acoustic phenomena (overtones). Thus, the extraction of a "correct" chord label sequence from audio is generally problematic and highly ambiguous.

These ambiguities can heavily affect the estimation of chord transitions. For example, an ambiguous segment with several possible chord labels implies several possible transition to and from this chord. Furthermore, chord segmentation can be problematic. A chord of short duration might be considered or not—leading to two ambiguous transitions. To account for such problems, we want to handle the ambiguities of both chords and chord transitions with a flexible approach, which we present in the following section.

3. TECHNICAL APPROACH

Hidden Markov Models (HMMs) have shown great success for modelling real-world sequence data. Besides the popular application for speech recognition [16], they have been extensively applied to automatic chord recognition [7, 8]. HMMs comprise a "hidden" probabilistic model determined by the Markov property (the current state only depends on the previous one), which generates visible "observations." In this paper, we deal with continuous HMMs where the emission probabilities are modelled by continuous-valued functions. A specific HMM Θ is defined by a set of components (see Table 1) comprising the I states α_i, the time-invariant transition probabilities a_{ij}, the initial state probabilities c_i, and the continuous-valued emission probability functions $b_i : \mathbb{R}^{12} \to \mathbb{R}^+$ with $b_i(\boldsymbol{o})$ indicating the probability of observing $\boldsymbol{o}$ while being in state α_i where $i, j \in [1 : I]$. Since only the observation sequence $O := (\boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_N)$ is directly accessible, there are three classical algorithmic problems concerning the application of HMMs [1, 16]:

$\mathcal{A}$	Set of states $\{\alpha_1, \ldots, \alpha_I\}$
A	State transition probabilities a_{ij} for $i, j \in [1 : I]$
C	Initial state probabilities c_i for $i \in [1 : I]$
B	Emission probabilities $b_i(\boldsymbol{o})$ for observation $\boldsymbol{o}$, with $i \in [1 : I]$

Table 1. Components specifying a continuous-valued HMM Θ.

Forward variable $v_n\,(i)$	
Initialization	$v_1\,(i) = c_i b_i(\boldsymbol{o}_1)$
Iteration	$v_{n+1}\,(i) = \left[\sum_{j=1}^{I} v_n\,(j)\, a_{ji}\right] b_i(\boldsymbol{o}_{n+1})$
Backward variable $w_n\,(i)$	
Initialization	$w_N\,(i) = 1$
Iteration	$w_n\,(i) = \sum_{j=1}^{I} a_{ij} b_j(\boldsymbol{o}_{n+1}) w_{n+1}\,(j)$

Table 2. Forward–backward algorithm for recursively computing the variables $v_n\,(i)$ and $w_n\,(i)$, with $i \in [1 : I]$ and $n \in [1 : N-1]$.

- The *evaluation problem* indicates how to compute the overall probability $P[O|\Theta]$ of an observation sequence O given an HMM Θ. We can efficiently solve this problem using the forward part of the forward–backward algorithm [25, 26].
- The *uncovering problem* consists in finding the hidden state sequence $S^* = (s_1^*, \ldots, s_N^*)$ with $s_n^* \in \mathcal{A}$, which maximizes $P[O, S|\Theta]$. This problem becomes relevant for the chord recognition scenario. We can efficiently compute S^* using the Viterbi algorithm [15] obtaining a chord label sequence.
- The *estimation problem* deals with reestimating the model parameters (A, C, B) by maximizing the probability $P[O|\Theta]$ given an observation sequence O. An efficient method to solve this is the Baum–Welch algorithm [26, 27]—a variant of the EM algorithm [28] that iteratively updates the parameters until convergence using the forward–backward algorithm.

Applying HMMs to audio chord recognition, one usually considers the chord labels as the hidden states s_n and a suitable feature representation computed from the music recording as the observation sequence $O = (\boldsymbol{o}_1, \ldots, \boldsymbol{o}_N)$. Typically, chroma features $\boldsymbol{o} \in \mathbb{R}^{12}$ are used for this purpose since they capture relevant tonal information in a suitable way [1]. In our experiments, we extract NNLS chroma features [29] with a resolution of 10 Hz using the *Chordino* Vamp plugin. For simplicity, we compute the emission probabilities via template matching between the chroma vectors $\boldsymbol{o}_n$ and binary chord templates $\mathbf{t}_i$ (using the cosine similarity):

$$b_i(\boldsymbol{o}_n) = \frac{\mathbf{t}_i \cdot \boldsymbol{o}_n}{||\mathbf{t}_i||_2 ||\boldsymbol{o}_n||_2}. \quad (1)$$

with $i \in [1 : I]$, $n \in [1 : N]$, and $|| \cdot ||_2$ denoting the ℓ_2 norm.

Due to reasons discussed in Section 2, we are interested in a "soft" estimation of chord labels and chord transitions while taking advantage of the probabilistic nature of HMMs, which allows for context-sensitive smoothing. To this end, we exploit some internal variables of the Baum–Welch algorithm corresponding to probabilities for local states and transitions. Following the definitions in [1, 16], we denote the probability of being in state α_i at time n and in state α_j at time $n + 1$ as

$$\xi_n\,(i, j) = P[s_n = \alpha_i, s_{n+1} = \alpha_j | O, \Theta] \quad (2)$$

with $n \in [1 : N]$. This variable precisely provides the information we are interested in. To efficiently compute $\xi_n\,(i, j)$, we use the forward–backward algorithm, which computes the probability of observing the partial sequence $(\boldsymbol{o}_1, \ldots, \boldsymbol{o}_n)$ while ending in state α_i at time n (forward variable):

$$v_n\,(i) = P[\boldsymbol{o}_1 \boldsymbol{o}_2 \ldots \boldsymbol{o}_n, s_n = \alpha_i | \Theta] \tag{3}$$

and the probability of observing the partial sequence $(\boldsymbol{o}_{n+1}, \ldots, \boldsymbol{o}_N)$ while starting in state α_i at time n (backward variable):

$$w_n\,(i) = P[\boldsymbol{o}_{n+1} \boldsymbol{o}_{n+2} \ldots \boldsymbol{o}_N | s_n = \alpha_i, \Theta]. \tag{4}$$

We compute these variables recursively using the expressions in Table 2. On this basis, we estimate local transition probabilities via

$$\xi_n\,(i,j) = \frac{v_n\,(i)\, a_{ij} b_j(\boldsymbol{o}_{n+1}) w_{n+1}\,(j)}{P[O|\Theta]} \tag{5}$$

with $n \in [1 : N-1]$. $P[O|\Theta]$ is a suitable normalization factor, which we compute using the forward variable:

$$P[O|\Theta] = \textstyle\sum_{i=1}^{I} v_N\,(i)\,. \tag{6}$$

We further obtain the probability for being in state α_i at time n

$$\gamma_n\,(i) = \textstyle\sum_{j=1}^{I} \xi_n\,(i,j) = \dfrac{v_n\,(i)\, w_n\,(i)\,,}{P[O|\Theta]} \tag{7}$$

which constitutes a "soft" estimation of chord labels.

4. FEATURE VISUALIZATION

We now illustrate the potential of the introduced mid-level features and compare those to a Viterbi-based approach by analyzing a choral by J. S. Bach as a case study (Figure 2). For simplicity, we consider a reduced chord vocabulary comprising only the major and minor triads ($I = 24$). We equally distribute the initial state probabilities $c_i = 1/I$ and compute the emission probabilities with Equation 1.

For HMM-based chord recognition, the choice of the transition matrix A plays a crucial role [8]. Since the chord change rate is typically lower than the feature resolution, the self-transitions a_{ii} assume higher values than the off-diagonal elements of A. Cho and Bello [8] showed that the main improvement of HMM-based chord estimation over frame-wise strategies is due to context-sensitive smoothing, which is determined by the relative size of the diagonal entries in A. "Musical knowledge" encoded in the off-diagonal transitions only has a minor effect. Thus, we pursue a similar approach as in [8] defining a uniform, diagonal-enhanced transition matrix A^{u} with elements

$$a_{ij}^{\mathrm{u}} = \begin{cases} \rho & \text{for } i = j, \\ \dfrac{1-\rho}{I-1} & \text{for } i \neq j. \end{cases} \tag{8}$$

With this definiton, the self-transition probability only depends on ρ and not on the number of chords in total while the transition probabilities sum up to one for each row. For the Viterbi algorithm, we directly use the resulting transition matrix A^{u}. For the Baum–Welch algorithm, we initially set $A = A^{\mathrm{u}}$ (prior probabilities) and compute the variables $\xi_n\,(i,j)$ and $\gamma_n\,(i)$ as introduced in Section 3.

In Figure 2, we show the results of the two strategies with an empirically chosen parameter value of $\rho = 0.4$. Figure 2a shows the output of the Viterbi algorithm compared to an expert annotation. The high number of true positive estimates (TP) shows that this algorithm performs well and obtains sharp edges between the chords. However, the algorithm sometimes prefers to stay in a chord and, thus, misses some of the annotated chord labels (e. g., in measure 3, around frame 100). From the computed chord labels (hard decision), we derive chord transition features by looking at the changes of the "uncovered" states—indicated by the red lines in Figure 2b.

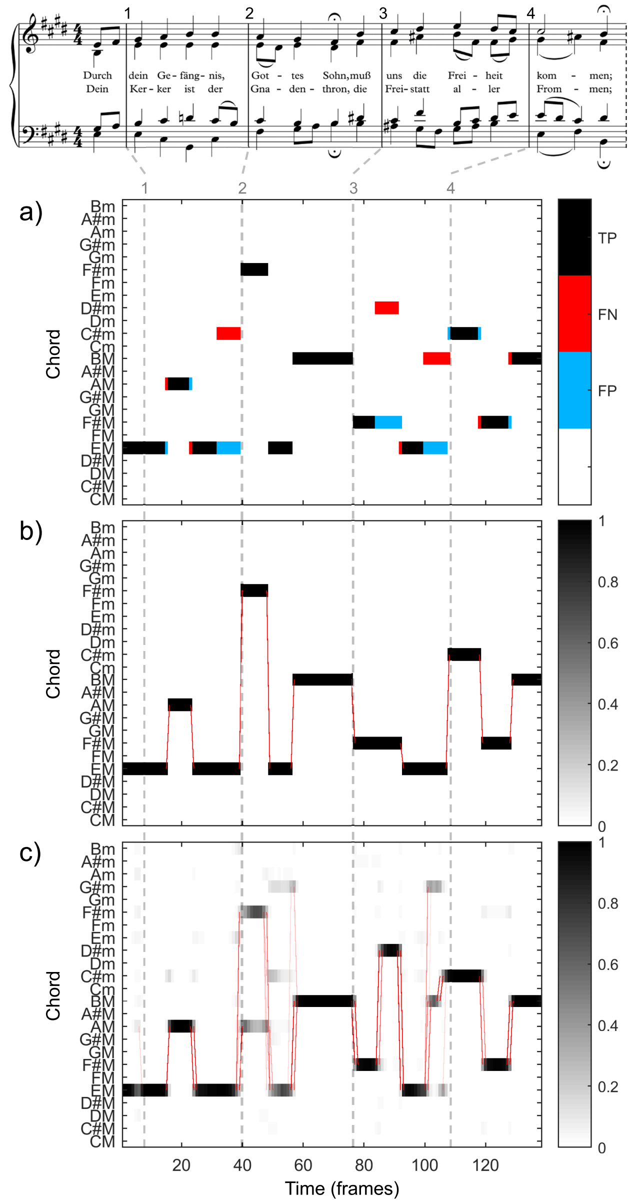

Fig. 2. J. S. Bach, Choral No. 22 from *Johannespassion* BWV 244. (a) Viterbi output compared to reference annotation. (b) Viterbi output with hard transitions (red lines). (c) Baum–Welch output with soft transitions (transparent red lines). We ignore self-transitions.

In contrast to the Viterbi algorithm, our Baum–Welch-based strategy (Figure 2c) produces a soft output for the chord labels (grayscale bars), derived from $\gamma_n\,(i)$.[1] Overall, the soft chord estimates look similar to the Viterbi output. However, some passages show a different behavior. In measure 3 (around frame 90), we see a high probability for the $\mathrm{D}^\sharp\mathrm{m}$ triad, which the Viterbi algorithm smoothed out. A similar effect occurs for the chord transitions,

[1] We visually enhance the chord probabilities using the soft-max function.

Data Subset	*Full*		*Pno*		*Orch*	
Baseline (136)	74.7		73.3		80.1	
Vocabulary | algorithm	*Vit.*	*B–W*	*Vit.*	*B–W*	*Vit.*	*B–W*
MajMin (46)	66.9	70.9	58.0	66.8	72.7	76.1
Triads (132)	66.9	75.0	56.9	68.8	71.6	78.5
Sevenths (464)	66.0	76.6	59.0	70.2	71.4	79.7
Baseline+*Triads* (268)	73.5	78.2	67.4	73.2	77.2	83.2

Table 3. Classification accuracies in % using baseline features and transition features based on Viterbi (*Vit.*) and Baum–Welch (*B–W*) algorithm. In parentheses, we indicate the feature dimensionality.

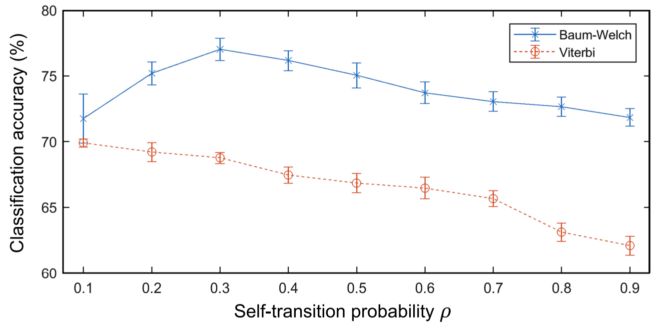

Fig. 3. Results for the *Full* subset using the *Triads* vocabulary.

which we derived from $\xi_n(i,j)$—shown as transparent red lines in Figure 2c.[2] We observe relevant chord transitions from and to $\mathrm{D}^\sharp\mathrm{m}$, respectively. Together with the self-transition probabilities (not shown here), this nicely shows how our algorithm accounts for different possibilities. The "soft" label estimates better represent the ambiguities of this passage. We observe a similar phenomenon at the beginning of measure 2 (frame 40). Here, the musical score indicates an $\mathrm{F}^\sharp\mathrm{m}^7$ chord (F♯-A-C♯-E). The reference annotation denotes this chord as $\mathrm{F}^\sharp\mathrm{m}$ triad but it also embraces the AM triad. Our soft features reflect this ambiguity by showing approximately equal probability for both triads as well as relevant transitions to approach and leave both chords. In general, our mid-level features indicate several possible transitions as soon as there is more than one possible chord label. This means that ambiguous situations lead to characteristic structures without deciding on explicit transitions, which always needs to ignore other options. Due to these properties, our visualization strategy might be useful for music analysis.

5. STYLE CLASSIFICATION EXPERIMENTS

We now test our chord transition features for style classification of Western classical music recordings. To this end, we make use of the *Cross-Era* dataset [14], which contains 1600 tracks—each 200 movements of piano (subset *Pno*) and orchestra music (subset *Orch*) for the four historical periods Baroque, Classical, Romantic, and Modern. The dataset served for evaluation in [9–11].[3] For our experiments, we use a basic machine learning pipeline consisting of feature extraction, dimensionality reduction (PCA followed by Linear Discriminant Analysis with three output dimensions), and classification (GMM classifier with only one Gaussian). We perform 3-fold cross-validation and apply composer filtering (forbid composer splitting across folds) to avoid overfitting due to the album effect [31]. To suppress the influence of the concrete cross-validation split, we repeat the experiment ten times with randomly initialized splits and report mean accuracies. For details, we refer to [11, Chapter 8].

Based on NNLS chroma features (compare Section 3), we compute both Viterbi- and Baum–Welch-based transition features. In both cases, we use a uniform transition matrix A^{u} with $\rho = 0.5$ (we later discuss this choice). We average over the local chord transition features of all N frames in order to obtain piece-wise classification features. To achieve transposition-invariant features that do not depend on the musical key, we shift all possible transitions to a common root note, thus capturing only *relative* transitions (red frames in Figure 1b). We ignore self-transition probabilities, which capture the frequency of chord changes (harmonic rhythm) rather than their type.[4] For musical reasons, we consider different sizes of the chord vocabulary, which also affects the number of transitions T. Because of enharmonic equivalence, some chord types show a symmetric behavior—such as the augmented triad, which only exists in four transpositions—leading to a reduced number of transitions T. For the vocabulary *MajMin* ($I = 24$), we end up with $T = 46$ possible transitions (23 transitions for starting from each of the two chord types). We further consider the vocabularies *Triads* including diminished and augmented triads ($I = 40$, $T = 132$) and *Sevenths* comprising seven types of seventh chords ($I = 75$, $T = 464$).

Table 3 shows the results of our classification experiments. As a baseline, we use chroma-based features (interval / chord types, tonal complexity) evaluated in [9–11], which results in an accuracy of 74.7 % on the *Full* dataset. Using the *MajMin* vocabulary, the performance of our transition features is below this baseline. Hereby, Baum–Welch-based features (70.9 %) perform clearly better than Viterbi-based features (66.9 %). This observation holds for the vocabularies *Triads* and *Sevenths*. On the *Full* dataset, *Sevenths* based on the Baum–Welch algorithm (76.6 %) outperforms our baseline. Combining baseline and chord transition features, the accuracies do not exceed the baseline when using Viterbi-based features. With Baum–Welch-based features, in contrast, we observe a clear improvement for the subsets *Full* (78.2 %) and *Orch* (83.2 %). This is a remarkable result. Indeed, the mid-level chord transition features seem to add relevant information to the generally well-performing tonal features tested in [11]. This is not the case for Viterbi-based transition features, which cannot account for local ambiguities.

Finally, we investigate the role of the self-transition probability ρ. Figure 3 shows the classification accuracies for both feature types as a function of this parameter (using the *Triads* vocabulary). We see that the performance highly depends on ρ for both algorithms, each with a specific global maximum. Our choice of $\rho = 0.5$ seems to be in a stable region though not optimal for this dataset. Remarkably, the proposed features outperform the Viterbi-based features for *any* value of ρ—a strong hint to the benefit of our soft-transition features.

In summary, our results confirm that mid-level features—capturing chord transitions in a soft way—have a higher descriptive power than features based on hard-decision chord estimation. Given these encouraging results, we assume that the proposed features might be beneficial for further applications. This includes the analysis, visualization, and classification of Western classical music in other scenarios. Furthermore, there is a high potential in applying such methods to other musical genres such as pop, rock, or jazz.

[2] For visualization, we ignore the self-transitions and apply a rescaling procedure for enhancing musically relevant transitions described in [30].

[3] The chromagrams used as basis for our experiments are available at `https://www.audiolabs-erlangen.de/resources/MIR/cross-era`

[4] In general, harmonic rhythm is an interesting aspect for style analysis. In this study, however, we want to restrict ourselves to chord transition types.

6. REFERENCES

[1] Meinard Müller, *Fundamentals of Music Processing*, Springer Verlag, 2015.

[2] Geoffroy Peeters, "Musical key estimation of audio signals based on hidden markov modeling of chroma vector," in *Proc. of the Int. Conference on Digital Audio Effects (DAFx)*, Montreal, Kanada, 2006.

[3] Björn Schuller and Benedikt Gollan, "Music theoretic and perception-based features for audio key determination," *Journal of New Music Research*, vol. 41, no. 2, pp. 175–193, 2012.

[4] Wei Chai and Barry Vercoe, "Detection of key change in classical piano music," in *Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR)*, London, UK, 2005, pp. 468–474.

[5] Hélène Papadopoulos and Geoffroy Peeters, "Local key estimation from an audio signal relying on harmonic and metrical structures," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1297–1312, 2012.

[6] Takuya Fujishima, "Realtime chord recognition of musical sound: A system using common lisp music," in *Proc. of the Int. Computer Music Conference (ICMC)*, Beijing, China, 1999, pp. 464–467.

[7] Alexander Sheh and Daniel P. W. Ellis, "Chord segmentation and recognition using EM-trained hidden Markov models," in *Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR)*, Baltimore, Maryland, USA, 2003, pp. 185–191.

[8] Taemin Cho and Juan Pablo Bello, "On the relative importance of individual components of chord recognition systems," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 477–492, 2014.

[9] Christof Weiß, Matthias Mauch, and Simon Dixon, "Timbre-invariant audio features for style analysis of classical music," in *Proc. of the Joint Conference 40th Int. Computer Music Conference (ICMC) and 11th Sound and Music Computing Conference (SMC)*, Athens, Greece, 2014, pp. 1461–1468.

[10] Christof Weiß and Meinard Müller, "Tonal complexity features for style classification of classical music," in *Proc. of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, 2015, pp. 688–692.

[11] Christof Weiß, *Computational Methods for Tonality-Based Style Analysis of Classical Music Audio Recordings*, Ph.D. thesis, Ilmenau University of Technology, Ilmenau, Germany, 2017.

[12] Gissel Velarde, Carlos Cancino Chacón, David Meredith, Tillman Weyde, and Maarten Grachten, "Convolution-based classification of audio and symbolic representations of music," *Journal of New Music Research*, vol. 47, no. 3, pp. 191–205, 2018.

[13] Zsolt Gárdonyi and Hubert Nordhoff, *Harmonik*, Möseler, Wolfenbüttel, Germany, 1990.

[14] Christof Weiß, Matthias Mauch, Simon Dixon, and Meinard Müller, "Investigating style evolution of Western classical music: A computational approach," *Musicae Scientiae*, 2018.

[15] Andrew J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[16] Lawrence R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[17] Brian McFee and Juan Pablo Bello, "Structured training for large-vocabulary chord recognition," in *Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, 2017, pp. 188–194.

[18] Hendrik Vincent Koops, W. Bas de Haas, Dimitrios Bountouridis, and Anja Volk, "Integration and quality assessment of heterogeneous chord sequences using data fusion," in *Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR)*, New York City, USA, 2016, pp. 178–184.

[19] Yizhao Ni, Matt McVicar, Raúl Santos-Rodríguez, and Tijl De Bie, "Understanding effects of subjectivity in measuring chord estimation accuracy," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2607–2615, 2013.

[20] Joel Lester, "Rameau and eighteenth-century harmonic theory," in *The Cambridge History of Western Music Theory*, Thomas Christensen, Ed., The Cambridge History of Music, pp. 753–777. Cambridge University Press, Cambridge, UK, 2002.

[21] Arnold Schönberg, *Theory of Harmony: Translated by Roy E. Carter*, University of California Press, Berkeley, California, USA, 1983.

[22] Jan LaRue, *Guidelines for Style Analysis*, Harmonie Park Press, Michigan, USA, 1992.

[23] Victor Belaiev, "The signs of style in music," *Musical Quarterly*, vol. 16, no. 3, pp. 366–377, 1930.

[24] Lajos Bárdos, *Modális Harmóniák (Modal Harmonies)*, Ed. Zenemukiadó, Budapest, Hungary, 1961.

[25] Leonard E. Baum and John A. Egon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bulletin of the American Metereological Society*, vol. 73, pp. 211–227, 1967.

[26] Leonard E. Baum and George Sell, "Growth transformations for functions on manifolds," *Pacific Journal of Mathematics*, vol. 27, no. 2, pp. 211–227, 1968.

[27] Leonard E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," in *Inequalities III: Proceedings of the 3rd Symposium on Inequalities*, Oved Shisha, Ed., 1972, pp. 1–8.

[28] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

[29] Matthias Mauch and Simon Dixon, "Approximate note transcription for the improved identification of difficult chords," in *Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands, 2010, pp. 135–140.

[30] Fabian Brand, "Chord transition features for style classification of music recordings," Master Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2018.

[31] Arthur Flexer, "A closer look on artist filters for musical genre classification," in *Proc. of the Int. Society for Music Information Retrieval Conference (ISMIR)*, Vienna, Austria, 2007, pp. 341–344.