AUTOMATIC STYLE CLASSIFICATION OF JAZZ RECORDS WITH RESPECT TO RHYTHM, TEMPO, AND TONALITY

Arndt Eppler[1], Andreas Männchen[1], Jakob Abeßer[1,2], Christof Weiß[1], Klaus Frieler[2]

[1] *Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany*

[2] *Liszt School of Music, Weimar, Germany*

Correspondence should be addressed to: jakob.abesser@idmt.fraunhofer.de

Abstract: **In this paper, we focus on the automatic classification of jazz records. We propose a novel approach where we break down the ambiguous task, which is commonly referred to as genre classification, into three more specific semantic levels.**

First, the rhythm feel (swing, latin, funk, two-beat) characterizes the basic groove organization and most often relates to the rhythm section. Second, the tonality type (functional, blues, bebop harmony) describes the general harmonic organization of the underlying composition, which serves the soloist as a guideline for the improvisation. Third, the tempo class (slow, medium, up) provides a rather broad categorization of the overall tempo of a song. We consider three individual classification tasks with respect to the different descriptors. A set of 229 jazz recordings was selected and labeled by musicology and jazz students as part of the Jazzomat Research Project.

For the three tasks, we conduct several classification experiments in order to investigate the usefulness of certain features and feature groups in general as well as pre-processing methods such as feature selection and feature space transformation. For a baseline classification experiment, we use low-level audio features which are widely used in Music Information Retrieval (MIR) research. As second step, we test several recently proposed mid-level features for modeling rhythmic or harmonic content of audio data. We systematically evaluate different modifications and combinations of these features.

By testing decision trees as classifiers, the obtained decision rules provide an opportunity for a musicological interpretation of the classification process. Beyond that, we also investigate which features were selected for the different tasks by an automatic feature selection algorithm. The results of this interdisciplinary study have potential implications for jazz research as well as for content-based audio analysis tasks such as music similarity search and music recommendation in general.

1. INTRODUCTION

The rise of digital audio and online music streaming services has entailed the use of huge music databases. In order to allow users to browse these databases or make automatic suggestions for similar pieces of music, various descriptors are employed. Among those, the genre and style of a recording are very important.

The manual annotation of genre or style can contain inconsistencies due to the subjective nature of genre definitions [1] and, more importantly, can be very time consuming, given the amount of information needed for large databases. That is why the task of automatic genre and style recognition has become more and more important.

In this work, we propose an approach for the style classification of jazz records. To cover a wide range of information inherent in jazz music we divide the broad task of genre classification into three separate problems on more specific semantic levels: the classification of rhythm feel, tempo, and tonality type. To make a subsequent musicological interpretation of the classification process possible, we aim to model musical phenomena for each of these levels.

This paper is structured as follows: In section 2 we present related publications that deal with genre classification and multilayer taxonomies. We then propose our novel approach to style classification in section 3. We present different evaluation experiments and discuss their results in 4. Finally, we draw conclusions and give an outlook on future work in section 5.

2. RELATED WORK

There is a staggering amount of publications regarding the topic of genre and/or style classification. A good overview is given in [2], where Sturm presents a survey of 467 approaches to music genre recognition.

We would like to highlight some publications that employ multilayer taxonomies in order to facilitate genre classification. Using such taxonomies also allows to gain insights into the relationships between different genres or styles.

The importance of taxonomy generation in the context of musical genre classification is shown in [3]. Here, the authors emphasize the usefulness of hierarchical taxonomies in music databases and the efficiency it allows for regarding classification tasks due to a divide-and-conquer approach in which specially trained classifiers can be used on different levels of the hierarchy. Furthermore, they adress the problem of automatic taxonomy generation and present their own approach that leads to a result that is different from manually generated taxonomies. This indicates that manually generated taxonomies are opimized for human use, while automatically generated taxonomies seem better suited for computational classification tasks.

In [4], the authors develop a three layer taxonomy that first separates music and speech and then differentiates between various genres and subgenres. The classification is based on three different semantic levels expressed in three feature sets: timbral texture features derived from speech and general sound classification as well as newly developed features that describe rhythmic content and pitch content, namely beat histogram and pitch histogram features. The classification tasks included the classification of six jazz styles, where a classification accuracy of $\sim$60% was achieved.

The authors of [5] use high-level features as opposed to the usual low-level features in their genre classification task. The features are derived from a symbolic MIDI representation of the music and belong to seven categories: instrumentation, musical texture, rhythm, dynamics, pitch statistics, melody, and chords. The taxonomy consists of three genres with three subgenres each. Each level of the hierarchy has a specially trained classifier ensemble. One classification task in this publication was the classification of three different jazz styles, that resulted in an accuracy of about 85%.

In [6], the authors use a complex four layer taxonomy coupled with a bottom-up classification approach, i. e., classification is performed on the lowest level of the hierarchy first. Four low-level features are employed to describe the signal: spectral roll-off, loudness, bandwidth, and spectral flux. These are grouped together and summarized for 1 second segments. For each segment, all possible genres are compared pairwise and the one with the most wins is considered the correct choice for the segment. The genre that is assigned to the majority of segments is chosen as the genre of the whole signal. This procedure was also applied to jazz recordings and resulted in an accuracy of $\sim$60% for six jazz styles and at least 40 signals for each class.

3. PROPOSED APPROACH

Our approach to automatic style classification of jazz recordings stems from two main ideas: we divide style classification into three separate tasks, i. e., the estimation of rhythm feel, tonality type, and tempo, and use mid-level representations as basis of our feature

extraction.

An overview of our system is shown in Fig. 1. On the rhythm feel and tempo levels, we derive the tempogram and log-lag autocorrelation function (ACF) from the audio signal before calculating further features. The chromagram serves as a basic mid-level representation on the tonality level.

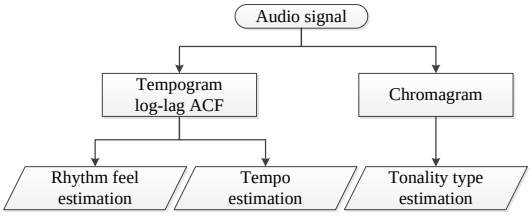

Figure 1: Generalized signal flow in the complete system.

3.1. Tonality

As the first step regarding the tonal domain, we calculate a basic representation of pitch classes commonly referred to as chroma. In this work, two different chroma features have been tested. First, we used Chroma Pitch (CP) features [7] that are calculated from the output of a pitch filter bank with 88 subbands. Second, we tried to decrease the influence of harmonics on the chroma by using Non-negative Least Squares (NNLS) chroma features [8] that are based on an approximate transcription of the tonal audio content.

Furthermore, we tested the application of harmonic-percussive decomposition [9] prior to calculating the CP features, which did not improve the classification accuracy. The CP and NNLS chroma features are calculated with standard parameters and a feature rate of 10 Hz.

Building upon these basic chroma features, we then consider different tempos and chord change frequencies using the approach described in [10]. We perform temporal smoothing and downsampling operations on the chromagram using window length w and downsampling factor d. Together with the local 10 Hz chroma features this results in four different time resolutions: 10 Hz ($w = d = 1$), 1 Hz ($w = d = 10$), 0.1 Hz ($w = d = 100$), and a global chroma vector.

Based on the four chromagram representations we calculate a number of features to describe the tonality type of a jazz recording. They can be separated into three groups which are presented in the following.

Comparing the chromagram to intervals, chords, and scales These features follow the basic idea of comparing individual frames of the chromagram to synthetic chroma vector templates that represent various local musical phenomena such as intervals, chords, and scales. This approach has been used a number of times in the past, e. g., in the fields of chord recognition [11] and key finding [12]. In this work, we use binary vector templates $\mathbf{V} \in \mathbb{R}^{12\times 1}$ that only consist of ones and zeros. Equations 1 through 3 show three examples of chroma templates that we applied.

$$\mathbf{V}_{\text{MinorThirdInterval}} = (1\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0)^{\mathrm{T}} \quad (1)$$

$$\mathbf{V}_{\text{DominantSeventhChord}} = (1\,0\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0)^{\mathrm{T}} \quad (2)$$

$$\mathbf{V}_{\text{DiatonicScale}} = (1\,0\,1\,0\,1\,1\,0\,1\,0\,1\,0\,1)^{\mathrm{T}} \quad (3)$$

For each template, we calculate a multiplicative distance as used in [13] between all 12 transpositions of the template $\mathbf{V}$ and each frame of the chromagram $C \in \mathbb{R}^{12\times N}$. We then take the sum over all transpositions and average over all frames to get the final distance measure F_{local} for one template, as shown in equation 4, where N is the number of chromagram frames.

$$F_{\text{local}} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{m=1}^{12}\prod_{p=m}^{1+(m+11) \bmod 12} C_{p,i}^{V_p}\right) \quad (4)$$

The distances are computed on all four feature rates. This leads to 84 distance measures based on chroma templates.

Features describing the tonal complexity This feature group is used to describe harmonic complexity by considering the dissonance of simultaneous sounds. These features include the flatness and sparseness of chroma vectors as well as the spread of the pitch classes over a perfect fifth axis. Their detailed explanations can be found in [14]. In addition, we use a feature that describes the average chroma vector novelty.

This whole group of tonal complexity features that we denote with $F_{\text{complexity}}$ is computed on all three local feature rates, resulting in 33 features.

Comparing the chromagram to synthetic harmonic progressions Similar to the first group of tonality features, we again employ distance metrics to compare the chromagram of a jazz recording to synthetic templates. Here, however, we use harmonic progressions instead of single vectors. We consider the 12-bar blues as well as I-VI-II-V turnarounds in major and minor. The synthetic progressions are stretched with different factors and shifted over the chromagram. At each position, we calculate the ℓ_1-distance between the chromagram and all 12 transpositions of the template. The resulting features for each template are the absolute smallest distance $F_{\text{prog, min}}$ and the smallest average distance at a certain stretching factor $F_{\text{prog, mean}}$. They are computed on a feature rate of 10 Hz, which leads to eight distance measures between the chromagram and synthetic harmonic progressions. Fig. 2 shows an example of a synthetic blues progression.

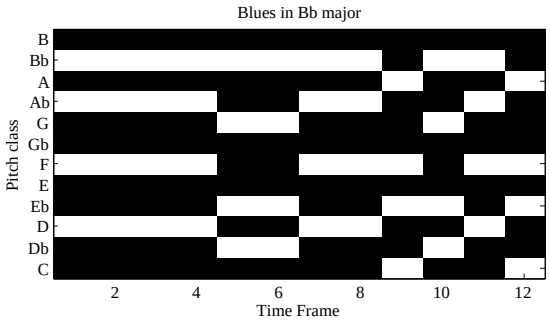

Figure 2: Synthetic chromagram for a blues schema in Bb major.

3.2. Rhythm feel

In contrast to approaches using pattern-based rhythmic genre classification like [15] or [16] we analyze the log-lag ACF and the tempogram feature representations to model the musical phenomena swing ratio and rhythmic complexity.

Modelling swing ratio Swing ratio refers to the subdivision of the beat and therefore relates to the phrasing of the eighth notes in a swing pattern. As explained in [17], the swing ratio is dependent on the tempo and the individual playing style of the drummer. Thus, the swing eighth notes are typically played between straight eighths (1:1), triple (2:1) and dotted (3:1) eighths.

To analyze the rhythmic characteristics of the musical pieces, we aim to define how the swing ratio relates to the output function of the log-lag ACF. To illustrate the effect of different swing ratios, four synthesized ride-patterns with the same tempo are depicted in Fig. 3. As can be seen, the distances d and u between the highest peak M and the adjacent peaks decrease with increasing swing ratio.

It should also be noted, that M can deviate from the actual tempo. Here, the tempo of all ride-patterns is 160 BPM, whereas the highest log-lag peak appears at 80 BPM. This relates to the so-called octave error, which is a common problem in tempo estimation. Since the position of M also differs for varying tempi and pattern types, we introduce the peak ratios δ_{down} and δ_{up} (see eq. 5 and 6).

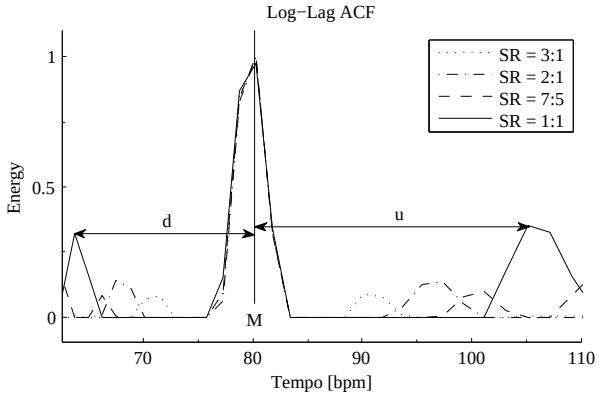

Figure 3: Log-lag ACF representations of various swing ratios (SR) around their main peak.

$$\delta_{\text{down}} = M/d \tag{5}$$

$$\delta_{\text{up}} = u/M \tag{6}$$

In addition to the peak ratios, we compute features that make use of the tempo invariance of the log-lag ACF. By normalizing it to its first peak, we revert the tempo dependant shift. The following mentioned features aim to describe the normalized function in a comprising manner. They consist of the four highest normalized peaks ($L_{\text{norm, P1-P4}}$, $L_{\text{norm, V1-V4}}$), statistical measures ($L_{\text{stat.name}}$) and the sparseness ($L_{\text{Sparseness}}$). This process takes place on a global and frame-wise (0.2 Hz) feature rate.

Modelling rhythmic complexity In this work, we quantify the complexity of the rhythmic organisation of a musical piece by the number of rhythmic layers and their variation over time. The periodicity-time representation of the tempogram offers the opportunity to measure such phenomena. By summing up the magnitude of the tempogram computed with a long (9s) DFT window (w_1) and selecting the peaks, we locate the dominant rhythmic layers. This way we get a good resolution of periodicity. By changing towards a short (1s) DFT window during the tracking phase, a good time resolution can be achieved (w_2). In this step, we compute the sum of absolute differences (SAD) of the dominant periodicities $i \in [1 : I]$, where $t \in [1 : l]$ denotes the time index. This measure for the rhythmic fluctuation over time is computed on the four feature rates 0.2 Hz, 0.1 Hz, 0.02 Hz, and global. Afterwards, we normalize it with respect to the frame duration l as shown in equation 7. We then use different measures to describe the resulting SAD functions. Beside common vector statistics $S_{\text{stat.name}}$, the mean count of dominant periodicities S_{count} and its variation from frame to frame S_{cvar} as well as the mean $S_{\text{complexity}}$, which refers to the variance of the difference of the SAD vector, serve as descriptors. This results in a 78-dimensional feature vector. In Fig. 4 the tempogram and the corresponding SAD vector are shown.

$$S(i) = \frac{\sum_{i=1}^{I} |w_2(i,t+1) - w_2(i,t)|}{l} \tag{7}$$

3.3. Tempo

Regarding the tempo, we try to classify the pieces into the classes slow, medium, and up. These tempo ranges were defined as shown in Tab. 1.

Although an approach using an estimated BPM value for classification was also followed, in this work we focus on the classification based on machine learning. Both state-of-the-art tempo estimation methods and our own implementation could not reach the performance of the machine-learning-based approach described in the following.

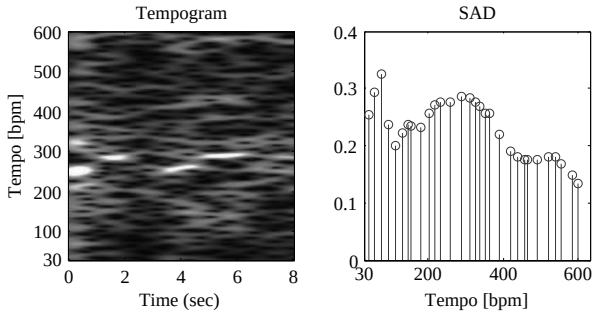

Figure 4: First seconds of Pat Metheny's 'Cabin Fever' as tempogram and corresponding vector with the sum of absolute differences (SAD).

Table 1: Tempo class names and the corresponding tempo ranges.

Tempo class	BPM
slow	< 100
medium	100 - 140
up	$140 >$

The main idea is to extract meaningful information from recently proposed rhythm- and tempo-based feature representations, namely the tempogram [18] and the log-lag ACF [19].

For the tempogram, we used a song-wise feature extraction. To get a time-independent representation, we first transform the three-dimensional tempogram to a two-dimensional median tempogram. That way, common function descriptors can serve as features. This step is shown in Fig. 5. Here, we use the positions of the six highest peaks and their values (T_{P1-P6}, T_{V1-V6}) as well as the flatness (T_{Flatness}), the centroid (T_{Centroid}) and a set of statistical features to describe the median tempogram in detail ($T_{\text{stat.name}}$).

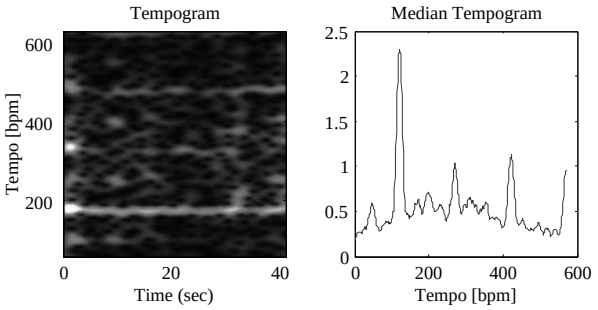

Figure 5: Tempogram and median tempogram representation of Lester Young's 'D.B. Blues'.

Similar to the rhythm level, the log-lag ACF is processed on a global and frame-wise (0.2 Hz) basis. Besides using the highest peaks (L_{P1-P6}, L_{V1-V6}), the description of their differences in position is done with statistical measures ($L_{\text{diff, stat.name}}$). We also used the count of these peaks and the sparseness ($T_{\text{Sparseness}}$) as additional information. Some features from the rhythmic level processing step were also included. Especially the newly introduced peak ratio from section 3.2 proved to be beneficial for the tempo classification.

4. Evaluation

4.1. Dataset

The dataset used in this work consists of 229 jazz recordings of different styles. The recordings have been annotated manually with respect to rhythm feel, tonality, and tempo by musicology and jazz students at Liszt School of Music Weimar as part of the Jazzomat Research Project. Tab. 2 shows the item distribution across the three semantic levels.

Table 2: Item distribution in our dataset.

Level	Class	#Items
Rhythm feel	Swing	47
	Latin	32
	Funk	42
	Two-Beat	43
Tonality	Functional	41
	Blues	28
	Bebop Harmony	35
Tempo	Slow	29
	Medium	34
	Up	35

4.2. Evaluation methods

A number of different machine learning techniques were used in this evaluation. As a first step, all features described in section 3 were normalized to zero mean and unit standard deviation.

For dimensionality reduction prior to classification, we used the Inertia Ratio Maximization with Feature Space Projection (IRMFSP) feature selection algorithm as described in [20]. The features selected via the IRMFSP also serve as a starting point for musicological interpretation.

Furthermore, we used three different classifiers. We employed multiclass Support Vector Machines (SVMs), Gaussian Mixture Models (GMMs), and decision tree classifiers. The classification performance with each classifier was tested with leave-one-out cross-validations.

As can be seen in Tab. 2, the distribution of items between the classes of each semantic level is imbalanced. To counteract this fact, we randomly duplicated training data for each fold of the leave-one-out cross-validations in such a way that all classes had the same amount of items. Since this step introduced random deviations into the evaluation, we used the median of 20 leave-one-out cross-validations for each experiment.

4.3. Tonality

The performance of the tonality type estimation was tested with a variety of different configurations. Here, we only present the most important results. Tab. 3 shows the median classification accuracies for the three different tonality feature groups and for all tonality features, whereby $\tilde{A}$ denotes the median accuracy of 20 leave-one-out cross-validations using an SVM classifier and selecting a maximum of ten features.

Table 3: Classification results for the tonality type estimation using an SVM classifier.

Features	$\tilde{A}$	
	CP	NNLS
Local templates only	.48	.49
Complexity features only	.46	.51
Harmonic progressions only	.40	.44
All features	.48	.49

As can be seen, the usage of NNLS chroma features led to a slightly better classification performance overall compared to CP features. The exclusive usage of the complexity features led to the best median classification accuracy of 51%, while using the distances between the chromagram and synthetic harmonic progressions resulted in a median accuracy of only 44%. Nevertheless, the harmonic progression features allowed for the best recognition of blues tonality out of all feature configurations, which is shown in an exemplary confusion matrix in Tab. 4. This suggests that the 12-bar blues schema was often identifiable using the NNLS chroma features.

By way of comparison, we conducted a baseline experiment using standard audio features related to log-lag autocorrelation, CENT and EPCP chroma features, log-loudness, octave-based spectral contrast, spectral flatness and crest factor, and the zero crossing rate.

Table 4: Confusion matrix for the tonality level using harmonic progression features only.

		Class (classified)		
		Functional	Blues	Bebop harm.
Class (correct)	Functional	.22	.29	.49
	Blues	.14	.72	.14
	Bebop harm.	.34	.17	.49

With a GMM classifier, a mean accuracy of 62% was achieved.

One possible explanation for this substantially better result is the lack of an artist or album filter in our evaluation. With such a filter in place, the classifier is never trained with data from the same artists or albums that are found in the test set [21]. Since we did not make use of an artist or album filter, some of the features used in the baseline experiment may describe characteristic sonic properties of a recording setup or instrumentation that was used on multiple recordings in the same class.

In this work, we are especially interested in the results of the feature selection and their musicological interpretation. Tab. 5 shows the five features that were picked most frequently by the IRMFSP feature selection algorithm while using NNLS chroma features and all tonality feature groups.

Table 5: Results for the IRMFSP feature selection on the tonality level using NNLS chroma features.

Rank	Feature
1	$F_{\text{local, 10 Hz}}$(minor second interval)
2	$F_{\text{local, 1 Hz}}$(whole tone scale)
3	$F_{\text{complexity, Flatness, 10 Hz}}$
4	$F_{\text{local, 0.1 Hz}}$(whole tone scale)
5	$F_{\text{prog, mean}}$(minor blues schema)

Interestingly, features from all three tonality feature groups appear in this list. The three groups seem to complement each other. Moreover, the importance of $F_{\text{local, 10 Hz}}$(minor second interval) and $F_{\text{complexity, Flatness, 10 Hz}}$ suggest that the chromaticity of a musical piece is a strong decision factor in this classification task. The fact that $F_{\text{prog, mean}}$(minor blues schema) was selected as often is surprising considering the prevalence of major blues records.

To gain further musicological insights we performed classification with a decision tree classifier. This allowed us to see which features were chosen for the decision nodes and where each numerical decision boundary was located. The classification with a decision tree led to an accuracy of 44%. A decision tree for the classification of the tonality type based on NNLS chroma features is shown in Fig. 6.

Again, features from all three feature groups appear in the tree. The importance of features that describe the chromaticity is confirmed due to the inclusion of F_{local}(chromatic scale) on two different time resolutions as well as $F_{\text{complexity, Flatness, 10 Hz}}$. It should be noted here, that the strong relevance of chromaticity might result from the fact that soloists on jazz recordings are often comparatively loud in the mix of instruments. Since jazz soloists utilize chromatic notes and passing notes frequently, this assumption seems justified.

All in all, the results suggest that the tonality features used in this work are mostly unsuitable to describe the differences between the three jazz tonality types. This could be a result of the features being too focused on local tonality rather than the greater harmonic structure of a musical piece. Additionally, the three classes chosen for this task may be too broad and seem to have some overlap.

4.4. Rhythm feel

On the rhythm feel level, the best classification performance was achieved selecting a maximum of 15 features, and employing an SVM classifier. The median classification accuracy for this task reached 57%. The performance for each class is shown as a

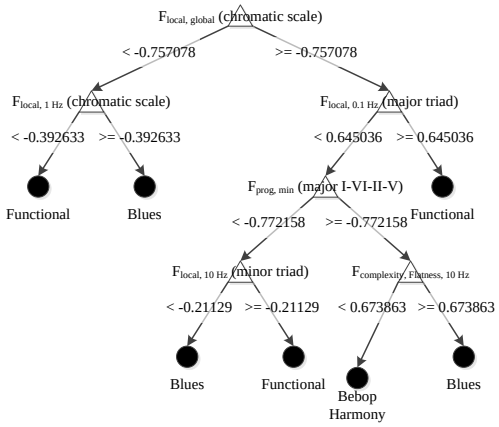

Figure 6: Decision tree for tonality type classification.

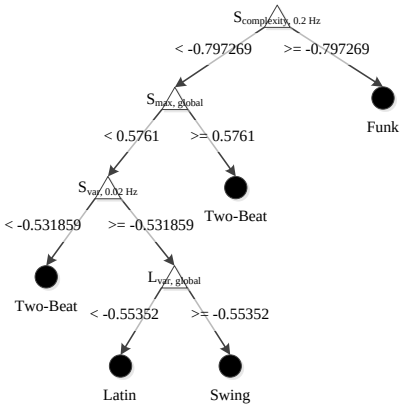

Figure 7: Decision tree for rhythm feel classification.

confusion matrix in Tab. 6. While the distinction between funk and two-beat seems adequate, it was more problematic to distinguish latin from swing pieces.

To gain further insight into the classification process, the most frequently selected features from the automatic feature selection are listed in Tab. 7. It shows that the rhythmic complexity features play an important role in this classification task. For example, the feature listed first, $S_{\text{max, global}}$, refers to the maximum fluctuation over time in one rhythmic layer and the feature listed second, $S_{\text{count, 0.05 Hz}}$, refers to the overall count of these rhythmic layers.

To examine the relation between the features we used and the actual rhythm feel classes, a decision tree is depicted in Fig. 7. The classification with a decision tree yielded an accuracy of 46%. As can be seen, the classes funk and two-beat, which achieved better accuracies in the confusion matrix, are discerned in the first few nodes of the decision tree. Especially $S_{\text{complexity, 0.2 Hz}}$ seems to reflect the descision threshold between funk and the other classes. As a measure for the parallel mixture of strong and weak fluctuations of single rhythmic layers, it supports the definition of the funk class that includes simple funk and rock grooves in combination with complex 16th note phrasings.

For the classification of the two-beat class, $S_{\text{max, global}}$ seems to be most descriptive. But since this feature relates to the overall fluctuation of the rhythmic bands in a piece, a higher value for the two-beat class is rather surprising. This suggests that underlying properties like , e. g., recording quality influence the feature strength for this classification task.

Table 6: Confusion matrix for the rhythm feel level.

		Class (classified)			
		Swing	Latin	Funk	Two-Beat
Class (correct)	Swing	.50	.27	.09	.14
	Latin	.40	.40	.14	.06
	Funk	.18	.16	.58	.08
	Two-Beat	.21	.07	.05	.67

Table 7: Results for the IRMFSP feature selection on the rhythm feel level.

Rank	Feature
1	$S_{\text{max, global}}$
2	$S_{\text{count, 0.05 Hz}}$
3	$S_{\text{complexity, 0.2 Hz}}$
4	$S_{\text{min, 0.2 Hz}}$
5	$S_{\text{median, global}}$

For the sake of comparison, we conducted a baseline experiment using the same features as in the baseline experiment in 4.3. The achieved accuracy of 75% outperformed the approach presented in this work. To interpret these results one has to factor in the feature types that were used. We focused on a more musically relevant and interpretable set of features. Although the common audio features of the baseline experiment achieve a better classification performance, a musicological interpretation is rather difficult.

Another aspect to keep in mind is the rather broad categorization into four classes on the rhythm feel level. A lot of drummers, especially in modern days, incorporate various playing styles into their performances, which results in fluent transitions between, e. g., swing, latin and funk.

An important issue of the categorization used here is the prevalence of different rhythm feel classes during certain eras. Swing, bebop and two-beat are mainly found on older records, whereas funk is linked to later decades. Over this timespan, the advances made in recording technology shaped the overall sound. These timbre-related differences could have been picked up by the features of the baseline experiment.

4.5. Tempo

For the tempo estimation, we conducted a series of classification experiments with different configurations. Again, the best results were achieved with an SVM classifier and the leave-one-out method for cross-validation. This configuration led to a median classification accuracy of 68%. With a decision tree classifier, an accuracy of 56% was achieved.

To draw musically relevant conclusions, we examine the four features that were picked most often by the automatic feature selection. They are listed in Tab. 8. There is again some variation between the features selected by the IRMFSP and the decision tree in Fig. 8. However, in both experiments the lower peak ratio $L_{\delta\text{down, 0.2 Hz}}$ is used as a way to discriminate between the classes slow and up. One explanation is the tempo dependency of the swing ratio, which results in tendencies towards smaller swing ratios in higher tempo ranges. But since the data set comprises not only swing pieces, a rather risky interpretation could be that the microtiming gets equalized in faster tempi.

The flatness of the globally computed median tempogram is also selected in both classification experiments. A higher flatness value can relate to rhythmic changes, tempo changes, and a higher number of rhythmic layers. These phenomena seem to be suitable to distinguish between the medium and up tempo classes.

Both the lower peak ratio and the flatness of the globally computed median tempogram allow for a broad categorization between slow and up, whereas the absolute peak positions of the median tempogram are used to classify the medium class. Nevertheless, there are still some ambiguities concerning this task.

Again, a baseline experiment with common audio features was conducted. The achieved accuracy of 68% shows the comparability of the proposed method.

Table 8: Results for the IRMFSP feature selection on the tempo level.

Rank	Feature
1	$L_{\delta\text{down, 0.2 Hz}}$
2	$T_{\text{Flatness, global}}$
3	$T_{\text{P1, global}}$
4	$L_{\text{diff, median, 0.2 Hz}}$

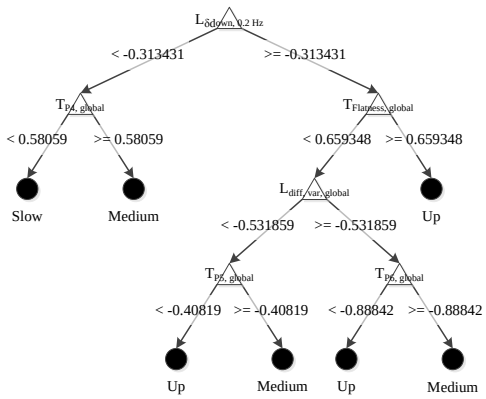

Figure 8: Decision tree for tempo classification.

5. Conclusions & Outlook

In this work, we implemented a system for the automatic style classification of jazz recordings with respect to rhythm feel, tempo, and tonality type.

On the tempo level, we achieved good classification results compared to our baseline experiment. The rhythm feel estimation was more problematic, especially the classification of latin pieces. Our baseline experiment that used standard audio features showed far better results. The same is true for the tonality classification. However, it should be noted that the usage of the third tonality feature group allowed for good classification results of blues recordings.

One of the biggest problems in our evaluation was the relatively small size of the database concerning the individual classification tasks. It led to an imbalanced item distribution across all classes and entailed the use of multiple recordings from the same artist and/or album. A bigger database of at least 100 items per class is advisable for future studies in order to obtain more representative results.

One focus of future experiments may be the usage of a different taxonomy. Our results suggest that different classes might be better suited for this task. Therefore, the use of unsupervised learning algorithms can lead to new insights regarding this musicological question.

References

[1] M. Sordo, O. Celma, M. Blech, and E. Guaus: *The quest for musical genres: Do the experts and the wisdom of crowds agree?* In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, pages 255–260. Philadelphia, USA, 2008.

[2] B. L. Sturm: *A survey of evaluation in music genre recognition*. In *International Workshop on Adaptive Multimedia Retrieval*, 2012.

[3] M. Ogihara and T. Li: *Music Genre Classification with Taxonomy*. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 197–200, 2005.

[4] G. Tzanetakis and P. Cook: *Musical Genre Classification of Audio Signals*. In *IEEE Transactions on Speech and Audio Processing*, volume 10(5):293–302, 2002.

[5] 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)*, pages 525–530. Barcelona, Spain, 2004.

[6] J. G. A. Barbedo and A. Lopes: *Automatic Genre Classification of Musical Signals*. In *EURASIP Journal on Advances in Signal Processing*, volume 1:157–168, 2007.

[7] 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)*, pages 215–220. Miami, USA, 2011.

[8] 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 (ISMIR)*, pages 135–140. Utrecht, Netherlands, 2010.

[9] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama: *Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram*. In *Proceedings of the 16th European Signal Processing Conference*. Lausanne, Schweiz, 2008.

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

[11] L. Oudre, C. Févotte, and Y. Grenier: *Probabilistic Template-Based Chord Recognition*. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 19(8):2249–2259, 2011.

[12] Ö. Izmirli: *An Algorithm For Audio Key Finding*. In *Proceedings of the 1st Annual Music Information Retrieval Evaluation eXchange (MIREX '05)*. London, UK, 2005.

[13] C. Weiß, M. Mauch, and S. Dixon: *Timbre-invariant Audio Features for Style Analysis of Classical Music*. In *Proceedings of the joint conference ICMC/SMC*. Athen, Griechenland, 2014.

[14] C. Weiß and M. Müller: *Quantifying and Visualizing Tonal Complexity*. In *Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM 2014)*. Berlin, Deutschland, 2014.

[15] M. Leimeister, D. Gärtner, and C. Dittmar: *Rhythmic Classification of Electronic Dance Music*. In *Proceedings of the AES 53rd International Conference on Semantic Audio*, pages 71–79. London, UK, 2014.

[16] S. Dixon, F. Gouyon, and G. Widmer: *Towards characterisation of music via rhythmic patterns*. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, pages 509–516. Barcelona, Spain, 2004.

[17] H. Honing and W. B. de Haas: *Swing Once More: Relating Timing and Tempo in Expert Jazz Drumming*. In *Music Perception*, volume 25(5):471–476, 2008.

[18] P. Grosche and M. Müller: *Tempogram Toolbox: MATLAB Implementations For Tempo And Pulse Analysis Of Music Recordings*. In *12th International Society for Music Information Retrieval Conference (ISMIR)*. Miami, USA, 2011.

[19] M. Gruhne, C. Dittmar, and D. Gärtner: *Improving rhythmic similarity computation by Beat Histogram Transformations*. pages 177–182. Utrecht, Netherlands, 2009.

[20] G. Peeters and X. Rodet: *Hierarchical Gaussian Tree with Inertia Ratio Maximization for the Classification of Large Musical Instruments Databases*. In *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx)*. London, UK, 2003.

[21] E. Pampalk, A. Flexer, and G. Widmer: *Improvements of Audio-Based Music Similarity and Genre Classification*. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 628–633. London, UK, 2005.