MODELLING MUSICAL ATTRIBUTES TO CHARACTERIZE TWO-TRACK RECORDINGS WITH BASS AND DRUMS

Jakob Abeßer
Semantic Music Technologies Group,
Fraunhofer IDMT, Ilmenau, Germany

Olivier Lartillot
Finnish Centre of Excellence in Interdisciplinary
Music Research, University of Jyväskylä, Finland

ABSTRACT

In this publication, we present a method to characterize two-track audio recordings (bass and drum instruments) based on musical attributes. These attributes are modelled using different regression algorithms. All regression models are trained based on score-based audio features computed from given scores and human annotations of the attributes. We compare five regression model configurations that predict values of different attributes. The regression models are trained based on manual annotations from 11 participants for a data-set of 70 double-track recordings. The average estimation errors within a cross-validation scenario are computed as evaluation measure. Models based on Partial Least Squares Regression (PLSR) with preceding Principal Component Analysis (PCA) and on Support Vector Regression (SVR) performed best.

1. INTRODUCTION

A lot of music pieces show stylistic influences from multiple music genres. These influences usually can be linked to the individual instrument tracks of a song. Instead of modelling music pieces as a whole, we believe that it is more meaningful to characterize them on a track-level. In this publication, we investigate double-track recordings including bass and drum instruments. Both instruments are essential parts of the so-called "rhythm section" that establishes the rhythmic and harmonic foundation of a band that performs a piece of music. The bass track and drum track usually follow a repeating, pattern-based structure.

The contribution of this paper is two-fold. First, we present new features for the rhythmic and tonal analysis of instrument tracks. Second, we investigate the applicability of regression models to model semantic attributes of instrument tracks based on human ratings. Since these attributes have a continuous scale, we use regression algorithms rather than classification algorithms to automatically predict their values for a given recording. The attributes introduced in this work (see Sect. 5) allow to describe a piece of music on a more abstract level than features derived from music theory allow. This semantic level opens up a more general perspective to characterize, to compare, and to retrieve music pieces. It is furthermore accessible to a broader selection of users since it does not require detailed musical knowledge.

2. GOALS & CHALLENGES

We aim to develop a regression-based prediction system that automatically characterizes double-track bass and drum recordings in terms of five different tonal and rhythmic attributes. Since the recordings we investigate cover various music styles from different regional backgrounds, we need to identify features that allow a robust semantic description independent of stylistic idiosyncrasies.

3. PREVIOUS WORK

In the last decade, score-based audio features (high-level features) were mainly applied for classification tasks such as genre classification [2, 3, 7]. In contrast to low-level and mid-level audio features such as the spectral flux or the Mel-Frequency Cepstral Coefficients (MFCC), high-level features relate to expressions of music theory to characterize instrument tracks in terms of rhythmic and tonal properties. These features are derived based a score representation of a music piece, which can be generated either by an automatic transcription of real audio files or directly from symbolic formats such as MIDI. In the past, most methods to extract high-level features comprise a statistical analysis of note onsets, pitches, and intervals [3, 8]. In [4], different regression algorithms were compared to predict different emotion ratings based on extracted audio features. Music recordings with guitar, bass guitar, and drums were analyzed as presented in [1] based on rhythmic high-level features. In this publication, three different configurations of regression models were compared to model 8 different musical attributes related to different instruments.

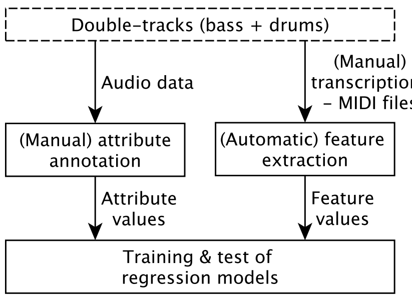

Figure 1. Processing steps including the manual annotation step, feature extraction, and regression analysis.

4. DATASET

In this study, we use a collection of 70 two-track recordings including a drum track and a bass track taken from instructional bass literature [10] as dataset. These tracks cover different Western music styles such as blues, funk, boogie, and modern jazz, Non-western styles from Latin and South America such as Cuban mambo, reggae, and samba as well as some African styles. All audio recordings were performed by professional musicians in a recording studio. The processing steps pursued in this study are depicted in Fig. 1. We used the audio recordings for the manual annotation of the given attributes as explained in Sect. 5. In addition, we extracted a score representation of the bass track based on the related score sheets and manually transcribed the drum track. Both track transcriptions were stored as MIDI files for further analysis. The question of automatic bass and drum transcription is not within the scope of this paper.

5. ANNOTATION PROCESS

For the annotations, we recruited 11 participants of different levels of musical education (most of them being semi-professional musicians). The participants were asked to annotate each audio track according to the attributes *harmonic clarity* (**HClar**), *harmonic predictability* (**HPred**), *rhythmic clarity* (**RClar**), *rhythmic coherence* (**RCoh**), and *dancability* (**Dan**) using a 7-point numeric scale between 1 (very low) and 7 (very high) with 4 being the neutral value. All attributes were introduced to the participants based on explanatory questions as shown in Tab. 1. The Annotation Tool previously presented in [11] was used for the subjects to manually assign attribute values for all recordings within the dataset. The participants were allowed to skip single annotations if they were unsure of their annotations for those particular tracks.

6. FEATURE EXTRACTION

We used the MIDI toolbox [5] to extract the basic score parameters absolute pitch $\theta_{P,A}$ of all notes of the bass track as well as onset φ_O (in fractions of bar lengths) and duration φ_D (in fractions of bar lengths) of all notes of both tracks from the MIDI files. Based on these note parameters, we compute high-level features related to rhythmic and tonal properties of both tracks as explained in the following sections. Both φ_O and φ_D provide a tempo-independent rhythmic representation of the bass line. In addition to the bass track (**BA**), we split the drum track (**DR**) into the three instrument sub-tracks bass-drum (**BD**), snare drum & rim-shot (**SD**), and hi-hat & cymbals (**HH**). In this section, we first explain pre-processing steps and then illustrate the extracted rhythmic and tonal features. For each feature, the corresponding instrument tracks are given in brackets.

6.1 Rhythmic features

6.1.1 Pre-processing

Metric level
In order to emphasize notes that occur on strong metric positions, we compute the metric level l_i of each note within the metric hierarchy of the corresponding bar. All examples within the dataset are in a $\frac{4}{4}$ time signature, thus we define the quarter notes as the *beat-level*. If the note onset corresponds to a beat position within the beat-level, we obtain $l_i = 1$, if it is not on the beat level but still on the first sub-beat level (eight-notes), we obtain $l_i = 2$, and so forth. For simplification, we assign both triplets as well as duplets to the same rhythmic level.

Similarity matrix (based on Levenshtein distance)
For each instrument track and each bar, we extract sequences made of the corresponding notes. Each note is represented by its modified note onset $\hat{\varphi_O} = \varphi_O \bmod 1$. This representation neglects the associated bar number of a note and only takes its relative position within its bar into account. We compute a rhythmic similarity measure $s_{m,n}$ between bar m and bar n based on the Levenshtein distance measure $d_{m,n}$ as $s_{m,n} = 1 - d_{m,n}/d_{max}$. For each pair of bars, the scaling factor d_{max} corresponds to the length of the longer note sequence. See [1] for further details.

6.1.2 Features

Average metric level (**BA**, **DR**)
In order to characterize the rhythmic complexity of an instrument track, we compute the average metric level l_i over all notes of this track as feature.

Tempo (**All**)
We use the tempo in BPM derived from the function *get-tempo* from the MIDI toolbox.

Attribute	Related instrument track(s)	Explanatory questions
HClar	**BA**	How clear could you imagine the harmonic content / harmonic progression of the music by just listening to the bass line?
HPred	**BA**	When listening to the excerpt for the first time, did you find the harmonic progression implied by the bass line predictable, or was it on the contrary surprising and unexpected?
RClar	**BA& DR**	How clear could you perceive the rhythmic structure (beat positions) by listening to the bass and the drums?
RCoh	**BA& DR**	Did the two instruments contribute to a coherent rhythmic structure, or did they contradict?
Dan	**BA& DR**	While listening to the music, could you imagine that it is easy to dance to it or not?

Table 1. Attributes used for manual annotations.

Note density (**BA**, **DR**)
We compute the number of notes N_m per bar. Then we take the mean and standard deviation of all values of N_m as features.

Rhythmic similarity within instrument tracks (**BA**, **DR**)
We compute mean and standard deviation over all similarity values $s_{m,n}$ with $m \neq n$ to measure the average similarity between all bars of an instrument track as well as its variance.

Rhythmic similarity between instrument tracks (**BA-BD**, **BA-SD**, **BA-HH**, **BD-SD**, **BD-HH**, and **HH-SD**)
Similar to the previously explained feature, we compute the bar-wise similarity between the bass and drum track pairs **BA- BD**, **BA- SD**, **BA-HH**, and the drum track pairs **BD-SD**, **BD-HH**, and **HH-SD**. For instance, this allows to detect whether the bass and the bass-drum track play rhythmically in unison or not. The participants agreed that this particular configuration contributes to the perception of a high rhythmic coherence between the bass and the drum instrument.

Divergence from a (Western) prototype rhythm (**DR**)
In accordance to the statements of various participants, we identified a *prototypic drum rhythm* [1] as illustrated in Fig. 2 that was said to serve as a rhythmic orientation for locating the beat positions in an unknown bass and drum groove. Therefore, we assume that the similarity between a given drum track and this prototype rhythm can be interpreted as a measure that is proportional to the perceived rhythmic clarity. For each of the drum instrument tracks **BD**, **SD**, and **HH**, we compute the similarity based on the Levenshtein distance as explained above between the real drum track and the corresponding track in the prototype rhythm. Finally, we average the similarity over all three instruments to derive an overall similarity measure for the complete drum track. This measure is averaged over all bars and taken as feature.

[1] This rhythm can be found in different Western music genres. Considering that most of the participants said to have only minor listening experience with Latin American and African rhythms, we only take this rhythm as a basis of comparison even though a couple of Latin American bass and drum grooves are present in the database.

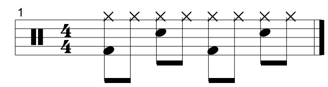

Figure 2. (Western) prototype drum-rhythm. The three drum classes introduced in Sec. 6 are represented by the lowest note (bass drum - **BD**), the middle note (snare drum - **SD**), and the cross-note (hi-hat - **HH**).

6.2 Tonal features

6.2.1 Pre-processing

Chromatic pitch representation
The chromatic pitch class $\theta_{P,C}$ represents all absolute pitch values mapped to one octave as $\theta_{P,C} = \theta_{P,A}$ mod 12 with $\theta_{P,C} \in [0, 11]$. The note name C corresponds to $\theta_{P,C} = 0$.

Diatonic interval representation
θ_I denotes the intervals between adjacent notes in semitones. After all intervals are mapped to a maximum absolute value of 12, we derive a diatonic interval representation $\theta_{I,D}$ that corresponds to the musical interval labels unison ($\theta_{I,D} = 1$), second ($\theta_{I,D} = 2$), and so forth up to seventh ($\theta_{I,D} = 7$). The octave ($\theta_I = 12$ or $\theta_I = -12$) is considered as a unison ($\theta_I = 0$) here according to the modulo-12 operation. For reasons of simplifications, we convert all descending intervals $\theta_{I,D} < 0$ into their complementary intervals, i.e., a descending second ($\theta_{I,D} = -2$) to an ascending seventh ($\theta_{I,D} = 7$) etc.

Bass note detection
We aim to detect the dominant *bass note* in each bar. Since no other instrument track is available for harmonic analysis, we use this bass note as harmonic reference for the compu-

tation of different tonal features. First, we retrieve all chromatic pitch classes $\boldsymbol{\hat{\theta}_{P,C}}$ apparent in a bar of the bass line. Then, we compute a *chromatic presence* value α, which accumulates the duration values $\varphi_{D,i}$ of all notes associated to the same chromatic pitch class $\hat{\theta}_{P,C,k}$ within this bar:

$$\alpha(\hat{\theta}_{P,C,k}) = \sum_{\forall i \leftrightarrow \theta_{P,C,i}=\hat{\theta}_{P,C,k}} \frac{1}{l_i}\,\varphi_{D,i} \qquad (1)$$

Each note is weighted according to its metrical level by the weighting factor $1/l_i$ (see Sect. 6.1.1). This is because notes on strong metric positions are assumed to be more likely perceived as bass notes than notes on weak metric positions. Finally, we obtain the chromatic pitch class of the bass note $\theta_{P,C,B}$ in this bar by maximizing α over all apparent chromatic pitch classes as

$$\theta_{P,C,B} = \hat{\theta}_{P,C,k^*} \leftrightarrow k^* = \arg\max_k \ \alpha(\hat{\theta}_{P,C,k}). \qquad (2)$$

6.2.2 Features

Percentage of bass note changes (BA)
Since we assume that the bass note acts as an indicator for the predominant harmony in a bar, we compute the number of bass note changes in a bass line divided by its length in bars as feature.

Diatonic intervals related to the bass note (BA)
In each bar, we compute the interval between the chromatic pitch class of all bass notes and the chromatic pitch class of the estimated bass-note $\theta_{P,C,B}$. Then, we derive the diatonic representation $\theta_{I,D}$ of this interval in the same way as previously explained in Sect. 6.2.1. If the bass note relates to the root note of the current chord and the bass line plays mainly thirds and fifths ($\theta_{I,D} = 3$, $\theta_{I,D} = 5$), we expect the harmonic predictability to be high since the bass uses main chord tones. Therefore, we compute $n(\theta_{I,D} = 3) + n(\theta_{I,D} = 5))/\sum n(\theta_{I,D})$ as feature with $n(\theta_{I,D})$ indicating the number of notes with the given diatonic interval value. If only a small number of different diatonic intervals are present, we assume the harmonic complexity of a bass line to be low. Therefore, we compute the zero-order entropy over the probability values $p(\theta_{I,D}) = n(\theta_{I,D})/\sum n(\theta_{I,D})$ as second feature:

$$H_0 = -\sum p(\theta_{I,D})\log_2\left[p(\theta_{I,D})\right] \qquad (3)$$

Tonal similarity between subsequences (BA)
To measure the tonal complexity of a bass line, we investigate, whether it is repeated after a certain number of bars. Therefore, we subdivide the bass line into adjacent sub-sequences of a length of $L = 1$, $L = 2$, and $L = 4$ bars. Each sub-sequence is represented by the absolute pitch values of the included notes. Again, we compute a similarity measure based on the Levenshtein distance as described in the previous section. Bass lines are often repeated after a few bars but played in a transposed form, i.e., translated in pitch by a constant term. To cope with that, we subtract the lowest pitch value from all absolute pitch values in each sub-sequence that is to be compared. Finally, we average the similarity values between all adjacent pairs of sub-sequences (e.g. for $L = 1$, we compare bar 1 with bar 2, bar 2 with bar 3, and so on) and derive one feature value for each sub-sequence length.

7. EVALUATION

7.1 Regression analysis

We compare 5 different configurations of regression models based on Robust Regression (RR), Partial Least-Squares Regression (PLSR), and Support Vector Regression (SVR). The RR uses an iteratively algorithm to assign a weight to each data point within the training data. This way, outliers have a smaller influence on the regression model. A different approach is followed by PLSR. A smaller number of less correlated predictor variables is derived from a linear combination of the initial feature dimensions. For the PLSR, we investigate the influence of a preceding feature selection via Principal Component Analysis (PCA). Therefore, we select all feature dimensions with eigenvalues $\lambda > 1$ during the PCA. We then determine the optimal model order for the PLSR models by minimizing the Akaike information criterion (AIC). For the SVR, we compare ν-SVR and ϵ-SVR as provided by the LibSVM toolbox [6]. We used the RBF kernel with parameter γ and cost factor C for both configurations. Based on a three-stage grid search, we determine the optimal parameters $\{C, \gamma, \nu\}$ for the ν-SVR and $\{C, \gamma, \epsilon\}$ for the ϵ-SVR [2] by minimizing the mean squared error (MSE) value. For more details on the regression methods, see for instance [6] and [9].

For each attribute, we select the features that are used for the model training as illustrated in the third column of Tab. 3. Leave-one-out cross-validation is used to evaluate each configuration-attribute pair and to avoid model overfitting, i.e., a different sample is used within each fold for testing and the remaining 69 samples are used for training of the regression models. Within each fold, all vectors of the training set are normalized to zero mean and unit variance. Then, the feature vector of the test set is normalized using the mean and standard deviation vectors derived from the training set. The MSE is computed between the predicted values and the ground-truth values of the test set and averaged over 70 folds. For each configuration-attribute pair,

[2] Search area for ν is $0.01 : .05 : .5$ and for ϵ is $(0.1 : 0.1 : 1) \cdot 10^{-3}$. The parameters C and γ are selected via grid-search as proposed in [6].

we store the test set ground truth values as well as the corresponding model predictions over all folds in two vectors. Then, we compute the sample correlation between both vectors. The correlation is considered as significant if $p < .05$ holds true for the corresponding p-value.

	HClar	HPred	RClar	RCoh	Dan
HClar	/	0.82*	0.48*	0.5*	0.24
HPred	0.82*	/	0.5*	0.58*	-
RClar	0.48*	0.5*	/	0.77*	0.38
RCoh	0.5*	0.58*	0.77*	/	0.31
Dan	0.24	-	0.38	0.31	/

Table 2. Correlation coefficients r between human annotations of different attributes. Only significant correlations ($p < .05$ or $p < .001^*$) are shown.

7.2 Results

Correlation between attributes
As illustrated in Tab. 2, the annotations show that many of the attributes are significantly correlated, especially the two tonal attributes **HClar** and **HPred** ($r = .82$) and the two rhythmic attributes **RClar** and **RCoh** ($r = .77$). The dancability of a bass and drum groove seems to be mainly influenced by its rhythmic attributes ($r_{\text{Dan,RClar}} = .38$, $r_{\text{Dan,RCoh}} = .31$).

Regression experiment
The results of the regression experiments outlined in Sect. 7.1 are illustrated in Tab. 3. As depicted in the upper part of the table, the SVR models lead to the smallest MSE values for all 5 attributes where the PLSR models performed only slightly worse. The models for **HClar** and **RClar** show the smallest prediction errors, while harmonic predictability show the highest errors. The RR performed worse for all attributes.

The sample correlation coefficients and the corresponding p-values are given in the lower part of the table. In contrast to the MSE values, highest (significant) correlation coefficients can be observed for the attributes **HPred** with $r = .59$ and **Dan** with $r = .46$. All significant correlations can be observed for models based on PLSR with preceding PCA or based on ϵ-SVR. No model show significant correlation for the attribute **HClar**.

Comments of participants
We identified a couple of problems during additional interviews with the participants after the annotation step. Two participants generally had difficulties to distinguish between clarity and predictability. The attributes **HClar** and **HPred** were said be the most complicated ones to annotate since the majority of the participants were not used to listen just to the bass and the drum instrument without any accompanying harmony instrument. Since the attribute **HPred** achieved the highest estimation errors as shown above, we assume that further score-based features need to be extracted from the harmony track of a given music recording in order to model this attribute.

8. CONCLUSION

In this paper, we compared five different regression algorithms for the estimation of values related to five different tonal and rhythmic attributes to characterize two-track recordings of bass and drums. Score-based features were extracted and used as predictor variables and manual user annotations of 70 audio excerpts were used as response variables to train and evaluate the regression models. For all five attributes, the PLSR+PCA model and the SVR models performed best (and comparably well) in terms of estimations errors. Significant correlations between annotated and estimated attribute values were only observed for four of the attributes and in particular for PLSR+PCA models and the ϵ-SVR models. Since the highest (significant) correlation coefficient is $r = .59$, we assume that further important aspects of the musical performance are not well captured by the applied features so far.

In general, we believe that the presented approach can be generalized to multi-track recordings including other instruments. However, we think that human attribute ratings should be based on listening to the isolated tracks instead of listening to the mixture signal. One issue of future work is to investigate how strong the perception of these attributes differs when human annotators listen to mixture of multiple instruments instead.

9. ACKNOWLEDGEMENTS

The authors would like to thank all participants who took part in the annotation process. The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project *Songs2See* [3], enabling transnational cooperation between Thuringian companies and their partners from other European regions. Furthermore, this work has been partly supported by the German research project *SyncGlobal* [4] funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S11007D).

10. REFERENCES

[1] J. Abeßer, O. Lartillot, C. Dittmar, T. Eerola, and G. Schuller. Modeling musical attributes to characterize

[3] http://www.songs2see.net
[4] http://www.syncglobal.de

Attribute	(short)	Features	RR	PLRS (+ PCA)	PLRS	ν-SVR	ϵ-SVR
Average MSE							
Harmonic clarity	(**HClar**)	Tonal	0.38	0.32	0.33	**0.29**	0.33
Harmonic predictability	(**HPred**)	Tonal	0.59	0.5	0.59	**0.49**	0.51
Rhythmic clarity	(**RClar**)	Rhythmic	0.43	0.27	0.3	0.26	**0.26**
Rhythmic coherence	(**RCoh**)	Rhythmic	0.56	0.4	0.43	**0.35**	0.36
Dancability	(**Dan**)	Rhy. & Ton.	0.96	0.47	0.53	**0.39**	0.43
Average model order - number of features							
Harmonic clarity	(**HClar**)	Tonal	11 - 11	1.03 - 11	1.37 - 11	11 - 11	11 - 11
Harmonic predictability	(**HPred**)	Tonal	11 - 11	1.37 - 11	2.99 - 11	11 - 11	11 - 11
Rhythmic clarity	(**RClar**)	Rhythmic	17 - 17	1.3 - 17	2.34 - 17	17 - 17	17 - 17
Rhythmic coherence	(**RCoh**)	Rhythmic	17 - 17	1.53 - 17	2.86 - 17	17 - 17	17 - 17
Dancability	(**Dan**)	Rhy. & Ton.	28 - 28	1.27 - 28	1.06 - 28	28 - 28	28 - 28
Sample correlation (absolute value) between ground truth & prediction - p-values							
Harmonic clarity	(**HClar**)	Tonal	0.04 - 0.76	0.05 - 0.69	0.06 - 0.62	0.2 - 0.1	0.17 - 0.17
Harmonic predictability	(**HPred**)	Tonal	0.14 - 0.26	**0.59 - 0**	0.2 - 0.1	0.09 - 0.46	**0.34 - 0**
Rhythmic clarity	(**RClar**)	Rhythmic	0.02 - 0.86	0.22 - 0.07	0.03 - 0.83	0.07 - 0.58	**0.34 - 0**
Rhythmic coherence	(**RCoh**)	Rhythmic	0.1 - 0.41	**0.32 - 0.01**	0.01 - 0.93	0.05 - 0.66	0.03 - 0.82
Dancability	(**Dan**)	Rhy. & Ton.	0.12 - 0.32	**0.24 - 0.04**	0.23 - 0.05	0.08 - 0.5	**0.46 - 0**

Table 3. Results of the regression analysis for 5 attributes as introduced in Sect. 5 based on 5 different configurations as explained in Sect. 7.1. The mean squared error (MSE) and the model order of the regression models were averaged of 70 folds of a leave-one-out cross-validation. The number of input features are given for each attribute. In the lower part of the table, the sample correlation value and p-value between the test set ground truth values and the predicted attribute values (collected over all folds) are shown. Significant correlations ($p < .05$) are denoted in bold print.

ensemble recordings using rhythmic audio features. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011.

[2] J. Abeßer, H. Lukashevich, P. Bräuer, and G. Schuller. Bass playing style detection based on high-level features and pattern similarity. In *Proc. of the Int. Society of Music Information Retrieval (ISMIR), Utrecht, Netherlands*, 2010.

[3] P. J. Ponce de Léon and J. M. Iñesta. Pattern recognition approach for music style identification using shallow statistical descriptors. *IEEE Transactions on System, Man and Cybernetics - Part C : Applications and Reviews*, 37(2):248–257, March 2007.

[4] T. Eerola, O. Lartillot, and P. Toiviainen. Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan*, 2009.

[5] Tuomas Eerola and Petri Toiviainen. *MIDI Toolbox: MATLAB Tools for Music Research*. University of Jyväskylä, Jyväskylä, Finland, 2004.

[6] Chih-wei Hsu, Chih-chung Chang, and Chih-jen Lin. A Practical Guide to Support Vector Classification. 1(1):1–16, 2010.

[7] C. McKay and I. Fujinaga. jSymbolic: A feature extractor for MIDI files. In *Int. Computer Music Conf. (ICMC), New Orleans, USA*, pages 302–305, 2006.

[8] C. Pérez-Sancho, P. J. Ponce de León, and J. M. Iñesta. A comparison of statistical approaches to symbolic genre recognition. In *Proceedings of the Int. Computer Music Conf. (ICMC), New Orleans, USA*, pages 545–550, 2006.

[9] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett. New support vector algorithms. 12:1207–1245, 2000.

[10] P. Westwood. *Bass Bible*. AMA, 1997.

[11] P. Woitek, P. Bräuer, and H. Großmann. A novel tool for capturing conceptualized audio annotations. In *Proceedings of the Audio Mostly Conf., Piteå, Sweden*, 2010.