Automatic genre and artist classification based on high-level features by analyzing improvised solo parts

Jakob Abeßer, Christian Dittmar, Holger Großmann {abr,dmr,grn}@idmt.fraunhofer.de

Fraunhofer IDMT, Germany

IDMT Fraunhofer Institut Digitale Medientechnologie

Introduction

Motivation

- Each musician has got an individual conception and understanding of aspects like timing, dynamics and articulation
- Describing the semantics of music performances is suited for high-level music characterization
- Application of performance analysis is expected to improve genre and artist classification results

Goal

- Performance analysis based on high-level features
- Processing of both symbolic (MIDI) and real audio data (MP3)
- Focus on solo parts because they offer the most freedom of individual musical expression to the soloist within a song
- Evaluation of the features' discriminative power within genre and artist classification scenarios

Challenges

- Score parameters are extracted from (polyphonic) audio data via different automatic music transcription methods
- Quality of the calculated high-level features strongly depends on precise and complete transcription results
- Evaluation for genre classification solely based on semantic information derived from solo parts (instead of complete songs)
- Artists playing the same instruments within related musical genres are difficult to distinguish

Related work

High-level features

- Rhythmic high-level features have been derived from
  - rhythmic deviations (swing factor) [1]
  - the percussion-related instrumentation [2]
  - statistical spectrum descriptors related to periodic rhythm patterns [3]
- Melodic high-level features have been derived from
  - basic statistical descriptors [3], [4]
  - complexity-based descriptors [6]

Genre classification

- Commonly based on low- and mid-level features
- McKay [4]:
  - Defined a set of 109 high-level features
  - Investigated 3 root genres each with 3 leaf genres
  - Achieved high classification scores for symbolic audio data

Performance analysis / artist classification

- Analysis of improvisation performed in clinical music therapy [7]
- Artist classification based on the progression of tempo and dynamic within piano performances [8]

Proposed method

General Assumptions

- Simple yet prevalent instrumentation model
  - Melody Instrument (MEL) - Soloist
  - Harmony Instrument (HAR) - Accompaniment
  - Bass Instrument (BAS) - Accompaniment
  - Percussion Instrument (DRU) - Accompaniment
- Excerpts from solo parts of 25s to 40s duration are investigated
- Excluded aspects for feature extraction are
  - timbral characteristics
  - the precise instrumentation (e.g. saxophone, piano)
  - applied expression and playing styles (e.g. vibrato, glissando)

Extraction of score parameters

- Symbolic audio data (MIDI) : *MIDI toolbox for MATLAB [9]*
- Real audio data (MP3) : *Transcription Toolbox [10]*
  - Software toolbox encapsulating four different transcription algorithms for each of the above mentioned instrument groups
  - Automatic extraction of the beat grid enables a projection of all note onsets from their values in milliseconds to multiples of bar lengths

Harmonic analysis

- Simplified harmonic analysis is applied to determine the root notes of all chords played by the harmony instrument via template matching of the most common 2-, 3- and 4-note chords
- All chords are artificially elongated for internal representation to allow an assignment of the harmonic context to each note
- Projection from absolute to functional pitch
  - Corresponds to the interval size (third (3), fifth (5) etc.) between melody note and the corresponding chord root note
  - The type (e.g. major or minor third) is not considered here to provide a scale-independent interval representation

Feature extraction

Melodic features

- Derived from absolute, relative and functional pitch
- Examples:
  - Temporal ratio of polyphonic parts, chromatic note sequences, note sequences with constant pitch, fragments with a constant melodic direction (up- or downwards)
  - Ratio of chord-tones within the melody (in relation to simultaneous sounding chord notes)
  - Statistical descriptors (e.g. zero- and first-order entropy, D'Agostino measure) based on state and transition probabilities of the chromatic pitch distribution, functional intervals (e.g. fifth upwards) as well as functional pitches related to the corresponding chord root note (e.g. third)

Rhythmic features

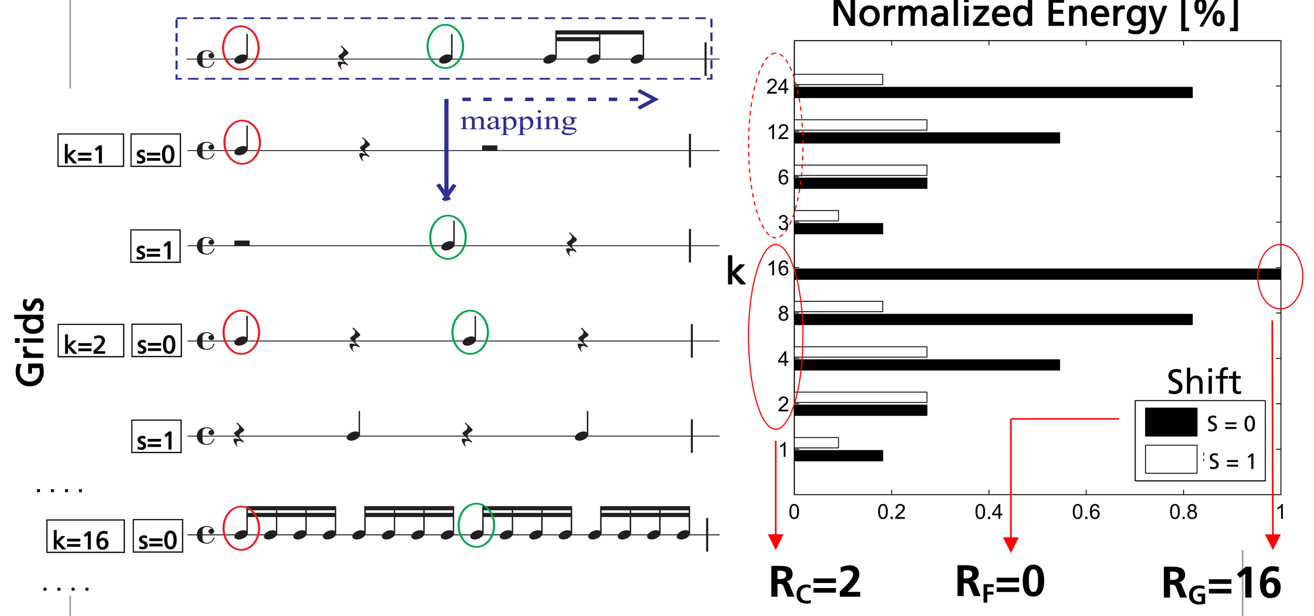
- Examples:
  - Measure of perceived rhythmical precision as an inverse measure of the quantization costs of all onsets to different grids (4th, 8th, 16th, 32th note grid)
  - Measure of swing factor (similar approach as shown in [1])

- **Rhythmical Structure Profile (RSP)**
  - Rhythmical representation that is independent from tempo and bar measure
  - It is based on partitioning each bar length into **k** equidistant grid points (corresponding binary and ternary values like 2-3 etc.) and a subsequent note mapping onto these grids

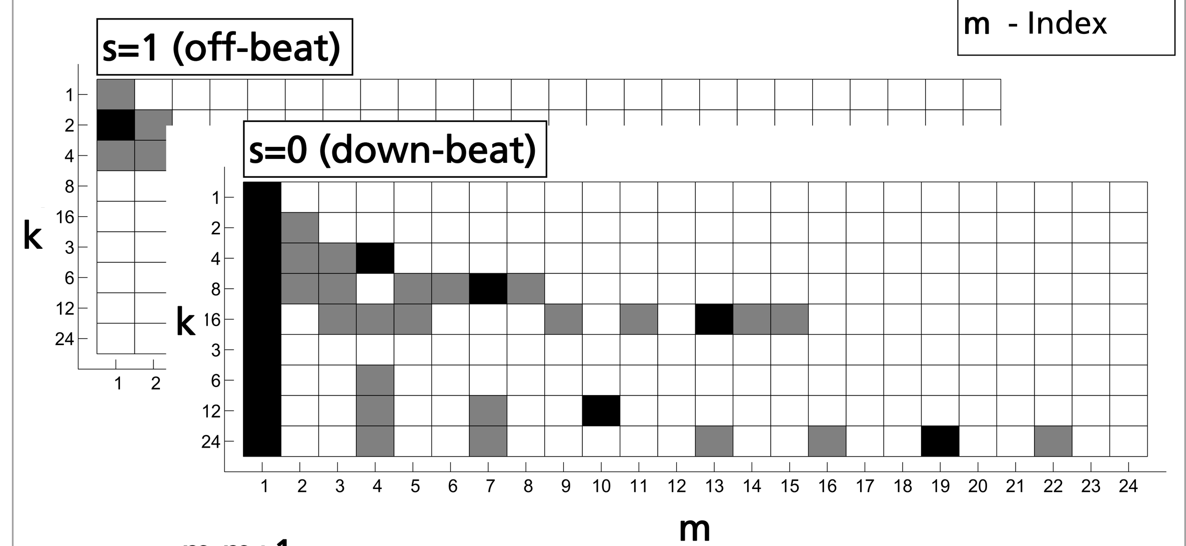
[9] Eerola, T., Toiviainen, P. *„MIDI toolbox: Matlab tools for music research"*. In www.jyu.fi/musica/miditoolbox (last call: 09.10.2008), Jyväskylä, Finland, 2004, University of Jyväskylä

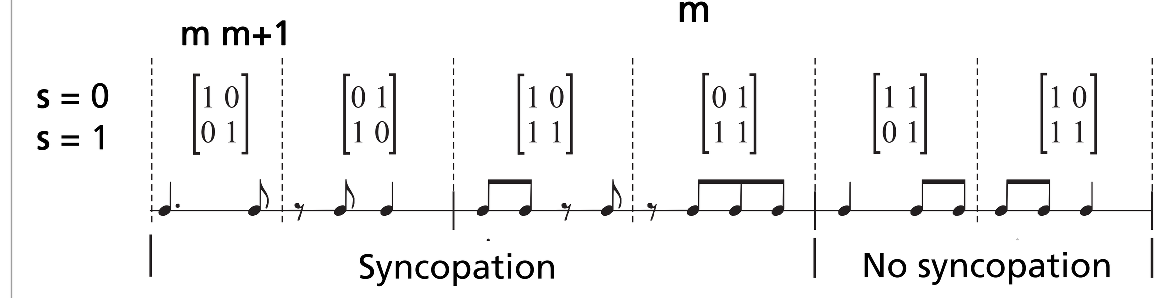
[10] Dittmar, C., Dressler, K., Rosenbauer, K. *„A toolbox for automatic transcription of polyphonic music"*. In Proc. of the Audio Mostly, 2007

- Derived Features
  - Rhythmical grid (containing the majority of all onsets) : $\mathbf{R_G}$
  - Rhythmical feeling (down- or off-beat feeling) : $\mathbf{R_F}$
  - Rhythmical characteristic (binary or ternary) : $\mathbf{R_C}$



- Measure of syncopation within different grids





Structure-related features

- Retrieval of rhythmic and melodic repetitions within instrumental tracks
- Application of a simple pattern search algorithm (*Correlative matrix approach, [5]*) for character strings derived from
  - Absolute pitch
  - Quantized onset & duration
- All repeating patterns are characterized by their parameters length (**l**), incidence rate (**f**) and mean distance (**d**)
- A pattern relevance measure (**r**) calculated from normalized parameter values is suppose to be a measure for the pattern's recall value to the listener

$$\mathbf{r_{Pat} = l_{Pat,Norm} + f_{Pat,Norm} + (1\text{-}d_{Pat,norm})^2}$$

- Features are derived from basic statistical descriptors characterizing the distribution of the 4 pattern parameters (overall 63 features)

Interaction-related features

- Are derived from course of the euclidean distance between barwise RSP's (rhythmical similarity between the tracks)
- Are derived from course of the bar-wise calculated chord-tone-ratio (harmony-relatedness of the solo melody)

Evaluation & Results

Genre classification

- 6 genres (Swing, Latin, Funk, Blues, Pop-Rock, Metal-Hardrock)
- Two test-sets (MIDI: 6x50 solos, MP3: 6x40 solos), a maximum of 148 high-level features derived from each instrumental track
- Two classifier approaches
  - SVM classifier (RBF kernel function) with preliminary LDA
  - Nearest-Neighbor classifier based on Euclidean distance between instrument-related RSPs (rhythmical similarity)
- Use of single (instrument-related) and ensemble classifiers (ENS)
- Additional listening test with 25 test persons and 3 scenario

Classifier	LDA-SVM	LDA-SVM	RSP	Human
Input	MIDI	MP3	MIDI	Resynth. MIDI
MEL	63.8	44.4	-	37.6
HAR	57.3	45.1	63.7	-
MEL+HAR	71.7	-	-	58.8
BAS	70.1	51.8	66.3	-
DRU	62.2	35.9	61.0	-
ENS (ALL)	84.0	63.4	73.2	63.1

***Table 1.** Genre classification results in %*

Artist classification

- Only melody-related features have been applied
- Evaluation with LDA-SVM classifierr
- Two test-sets each with 4x30 = 120 solos
  - E-Guitar (E. Clapton, R. Gallagher, J. Hendrix, S. R. Vaughan)
  - Saxophone (J. Coltrane, D. Gordon, C. Parker, J. Redman)

Classifier	LDA-SVM
Input	MP3
Electric guitar	58.8
Saxophone	56.0

***Table 2.** Artist classification results in %*

Outlook

- Classifiers based on high-level features outperform human test person in genre classification solely based on solo parts
- Considering that timbre- and instrumentation-related features have not yet been taken into account, the results are encouraging for further research
- Additional use of low- and mid-level features to describe individual playing styles, timbral characteristics as well as the instrumentation are expected to improve artist classification results
- Transcription-based semantic audio features are furthermore expected to have huge potential for other classification tasks

Selected References

[1] Gouyon, F., Fabig, L., Bonada, J. *„Rhythmic expressiveness transformations of audio recordings - swing modifications"*, in Proc. of the DAFx, 2003

[2] Herrera, P., Sandvold, V., Gouyon, F. *„Percussion-related semantic descriptors of music audio files"*, in Proc. of the AES Conf., 2004

[3] Lidy, T., Rauber, A., Pertusa, A., Iñesta, J. M. *„Improving genre classification by combination of audio and symbolic descriptors using a transcription system"*, in Proc. of the ISMIR, 2007

[4] McKay, C., Fujinaga, I. *„Automatic genre classification using large high-level musical feature sets"*, in Proc. of the ISMIR, 2004

[5] Hsu, J.-L., Liu, C.-C., Chen, A. L. P. *„Discovering nontrivial repeating patterns in music data"*, in IEEE Transactions on Multimedia, vol. 3, pp 311 – 324, Sept. 2001

[6] Madsen, S. T., Widmer, G. *„A complexity-based approach to melody track identification in midi files"*, in Proc. of the MUSIC-AI, 2007

[7] Erkkilä, J., Lartillot, O., Luck, G., Riikkilä, K., Toiviainen, P. *„Intelligent music systems in music therapy"*, in Music Therapy Today, vol. 5, 2004

[8] Saunders, C., Hardoon, D. R., Shawe-Taylor, J., Widmer, G. *„Using string kernels to identify famous performers from ther playing style"*, in Proc. of the ECML, 2004