

---

# 3rd International Workshop on Learning Semantics of Audio Signals, Graz, Austria

---

## Comparison of Harmonic Mid-level Representations for Genre Recognition

Matthias Gruhne, Christian Dittmar

Fraunhofer IDMT, Ilmenau, Germany

Dipl. Ing. (FH) Christian Dittmar

Group Manager Semantic Music Technologies

Graz, 02.12.2009

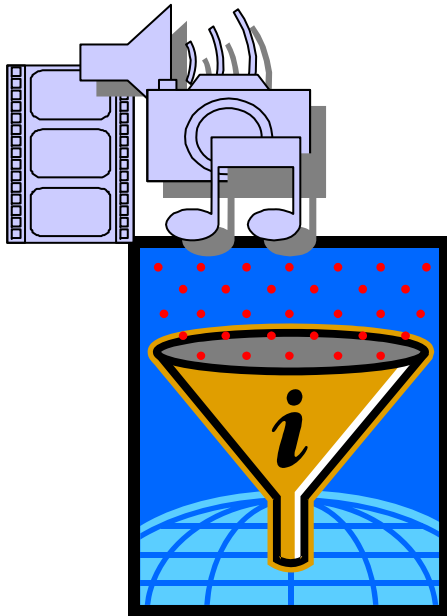
[christian.dittmar@idmt.fraunhofer.de](mailto:christian.dittmar@idmt.fraunhofer.de)

# Presentation outline

---

- Introduction
  - Motivation
  - Goal
- Features
  - Chroma Extraction Method
  - Geometric Pitch Spaces
  - Pitch Distribution Profiles
  - Post-Processing
- Evaluation
  - Test-Set
  - Classification Method
  - Results
- Conclusions

# Introduction: Motivation

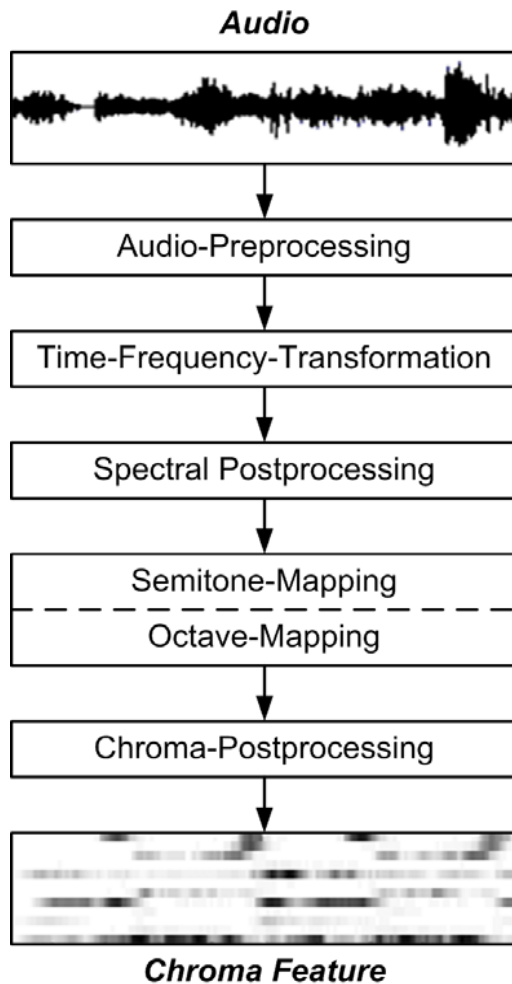


- Drowning in multimedia data → millions of digital music files on download shops, community portals, user's hard-drives
- Conventional metadata (artist, title, album etc.) are of limited use for retrieval & exploration in large music catalogues
- Automatic music annotation:
  - Derive content descriptions for items in large music catalogs
  - Enable specialized search, e.g. music that has the properties "Rock + Aggressive + Minor Tonality"
  - Music recommendation based on Query-by-Example needs justification via human-readable content description

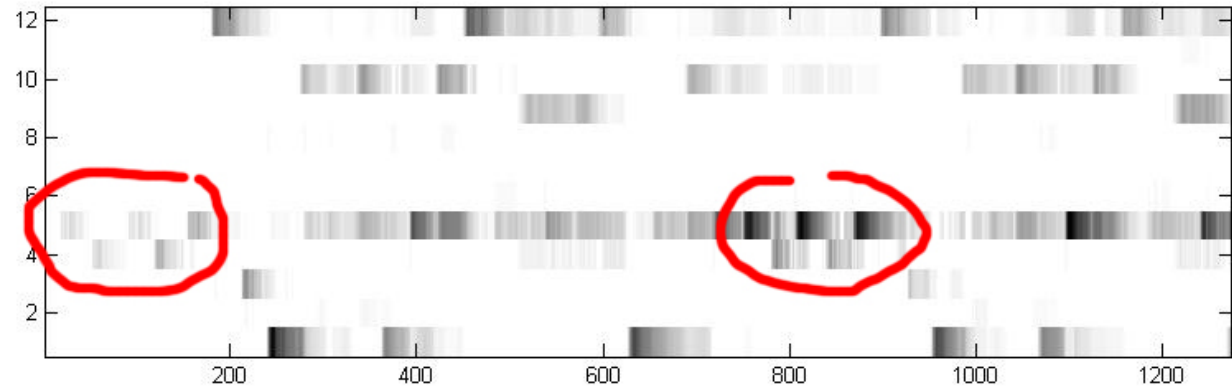
## Introduction: Goal

- Tonality: In MIR often described with chroma-vectors and related mid-level features
- Although genres are seldomly defined by tonality, a certain discriminatory power is inherent to tonal mid-level features
- Problem: Chroma-Vectors & derived mid-level features are not key independent
- Goal: Development and evaluation of post-processing steps to achieve key independence

# Features: Chroma Extraction Method



- EPCP: Enhanced Pitch Class Profiles (Lee 2006)
- Based on Harmonic Product Spectrum (Schroeder 1968)
  - Decimations by power of two ( $q=2^m$   $m=0\dots2$ )



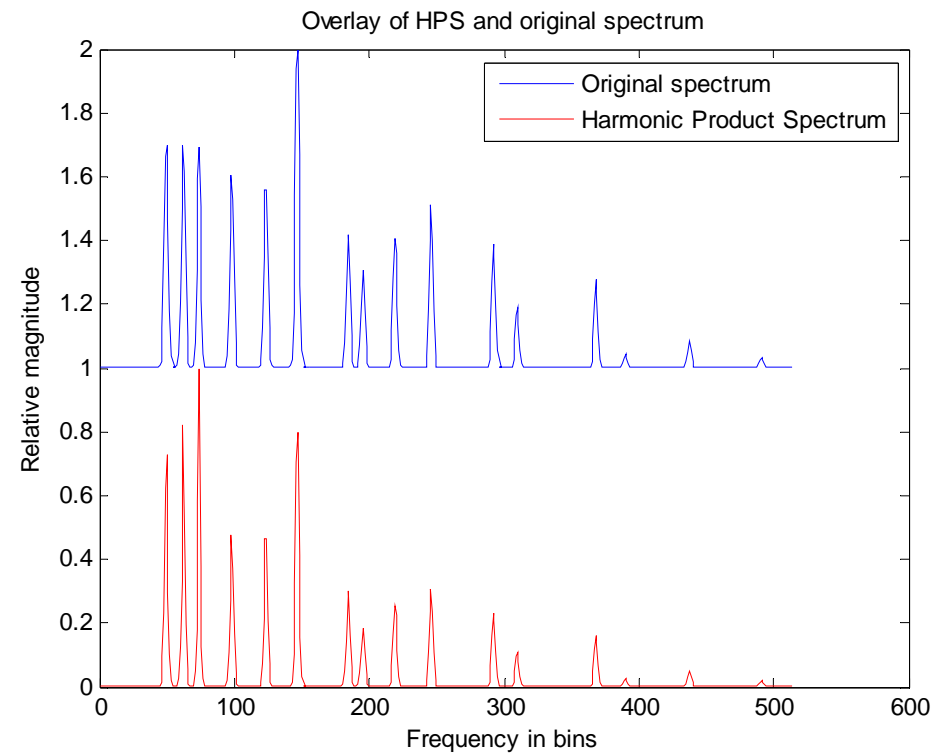
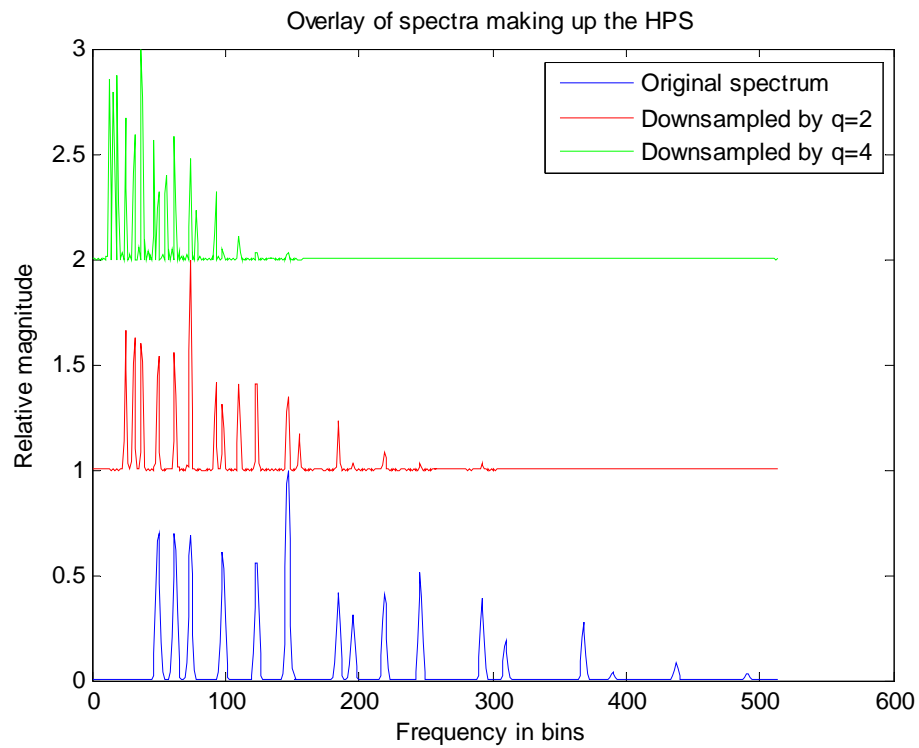
- Quiz: Which piece do we have here ???



# Features: Chroma Extraction Method



Artificial chord signal



# Features: Pitch Distribution Profiles (PDP)

- Chord templates originally derived from probe tones (Krumhansl 1984)
- Major and minor profiles for all possible root notes yields 24-dimensional vector
- Profiles yielding maximum correlation with chroma vector indicate local chord candidates
- Post-processing strategies needed to achieve stable results

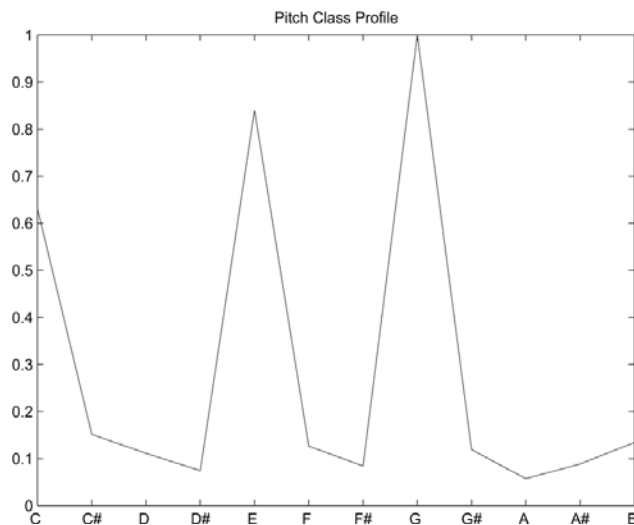
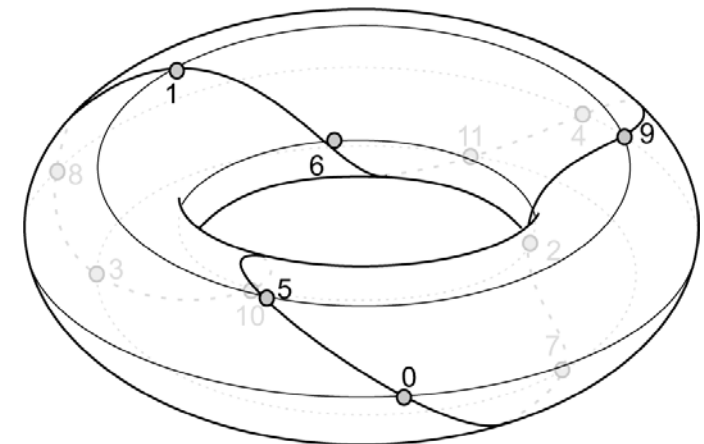
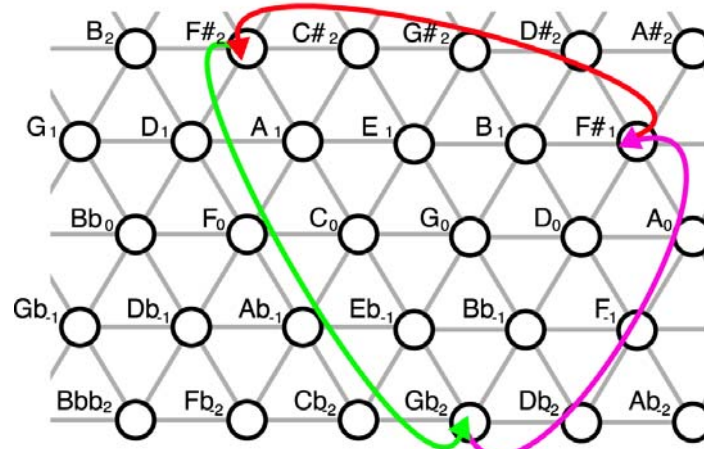


Figure: Kyogu Lee, *Automatic Chord Recognition Using Enhanced Pitch Class Profile*, in *Proceedings of International Computer Music Conference, 2006*

# Features: Geometric Pitch Spaces (Triad)

- Geometric analogy to express musical relations
- Harmonic Network (Tonnetz) as presented by Harte in 2006
- Enharmonic and octave equivalence leads to projection of chroma vectors onto 3 circularities:
  - Circle of fifths
  - Circle of major thirds
  - Circle of minor thirds



Figures: C. Harte, M. Sandler, and M. Gasser, *Detecting harmonic change in musical audio*, in Proceedings of ACM Multimedia 2006



# Features: Proposed Post-Processing

- Main idea: use derivative between adjacent frames to achieve key independence
- Applied to chord candidates resulting from PDP
  - 12 major chords coded as [0 ... 11]
  - 12 minor chords coded as [12 ... 23]
  - Offset of 12 is added in case of change from minor to major and vice-a-versa
- Applied to angles resulting from Geometric Pitch Spaces
  - Vector angle points to tonal center per pitch space → only relative change is preserved
  - Vector length is key independent by default

# Evaluation: Test-Set

Table 2. Root and leaf-genres contained in the test set.

Root-genres	Corresponding leaf-genres
Classical	Choir, Ensemble, Guitar, Piano, Art Songs, Opera, Popular Classical Music
Electronic	Big Beat, Hi-NRG Disco, Euro Dance, House, Techno, Lounge, Dark Drum&Bass, Bright Drum&Bass, Experimental, New Age, Trip Hop
Jazz	Dixieland, Swing, Blues, Entertainer, Bebop, Cool Jazz, Blue Note, Bossa Nova
Pop	Dance Pop, Country Music, Ballads, Italo Pop, 70ties Disco
Rock	Hard Rock, Nu Rock, Rock&Roll, Beatmusic, Rock Ballads, Psychedelic, Metal, Punk/Hardcore, Ska
Ger. Pop	German Ballads, German Beatmusic, German Dance, German Disco, German Latin, German Traditional
Urban	Soul/RnB, Gospel, Reggae, Rap, Female Funk, Modern Funk
Speech	Speech with Sounds, Speech without sounds
World	Asian, Flamenco, Indian, Irish Ballads, Nubenebra
Misc.	Children's Songs

- Manually annotated genre set provided by customer
- 775 songs in total, from 10 genres and 60 sub-genres
- Similarities between sub-genres evaluated in listening-tests

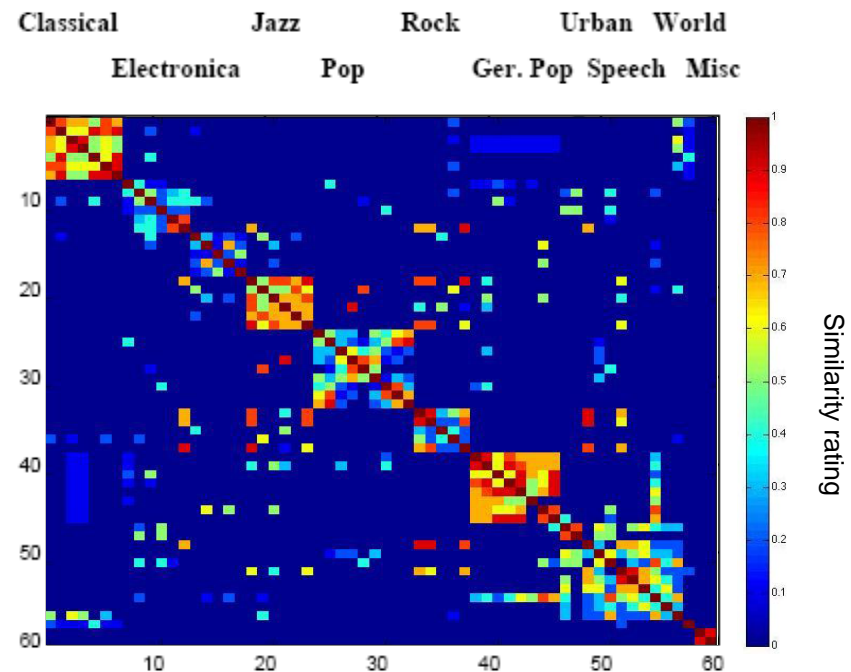
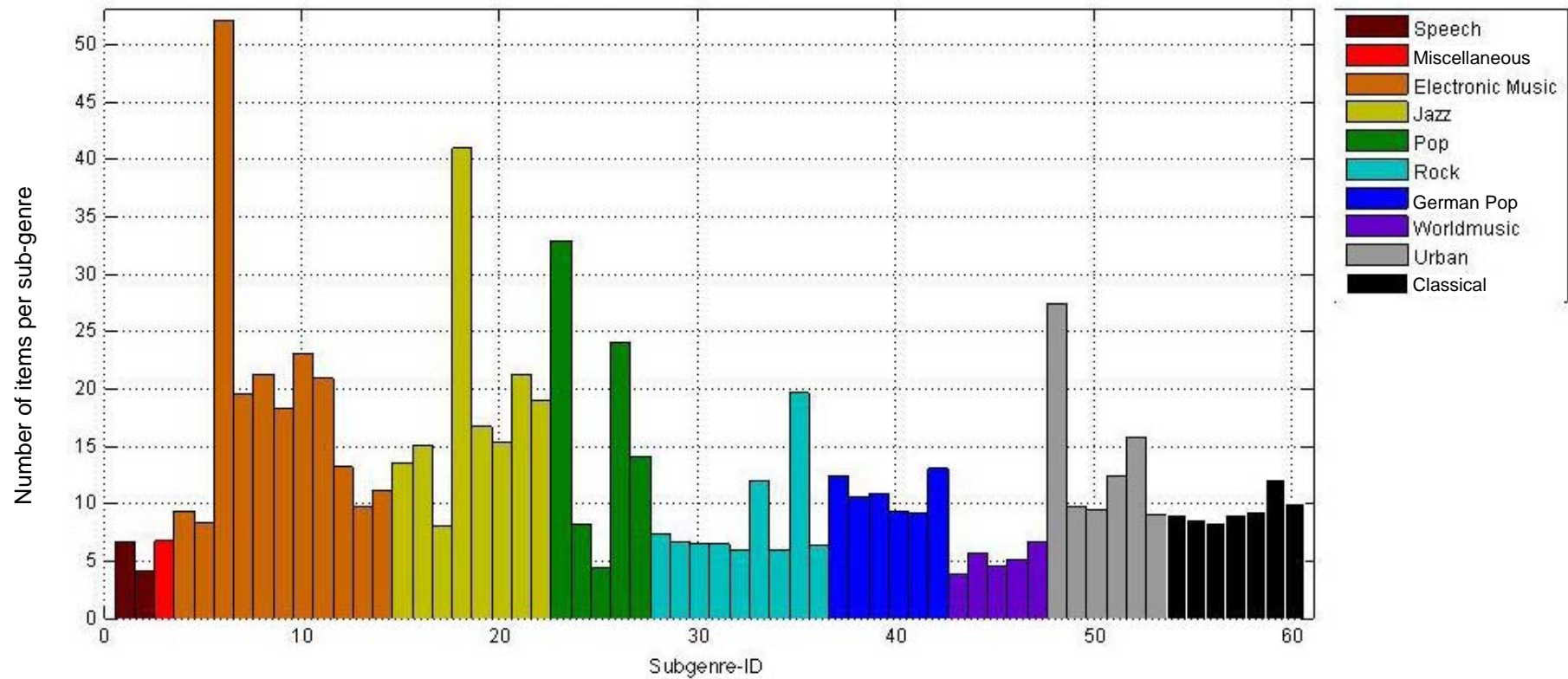


Figure 5. Structure of genre relation Matrix R.

# Evaluation: Measures

- Inhomogeneous distribution of items per sub-genre
- Utilization of pair-wise precision, recall and f-measure



# Evaluation: Classification Method

- 9x repeated random split of ground truth data into 70% train items and 30% test items → mean f-measure & minimum and maximum
- Utilization of Linear Discriminant Analysis → dimensionality reduction + enhanced class separability
- Gaussian Mixture Models with 3 Gaussians and diagonal covariance matrix

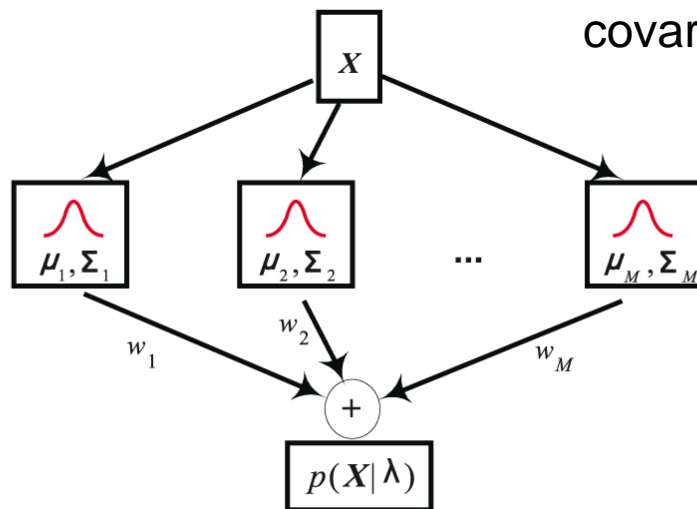
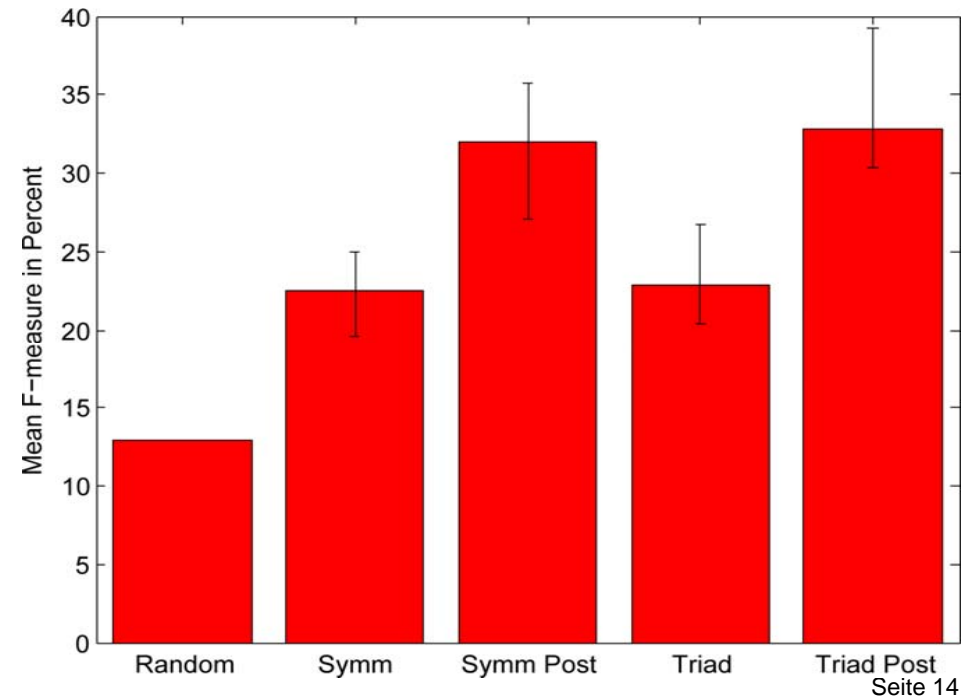
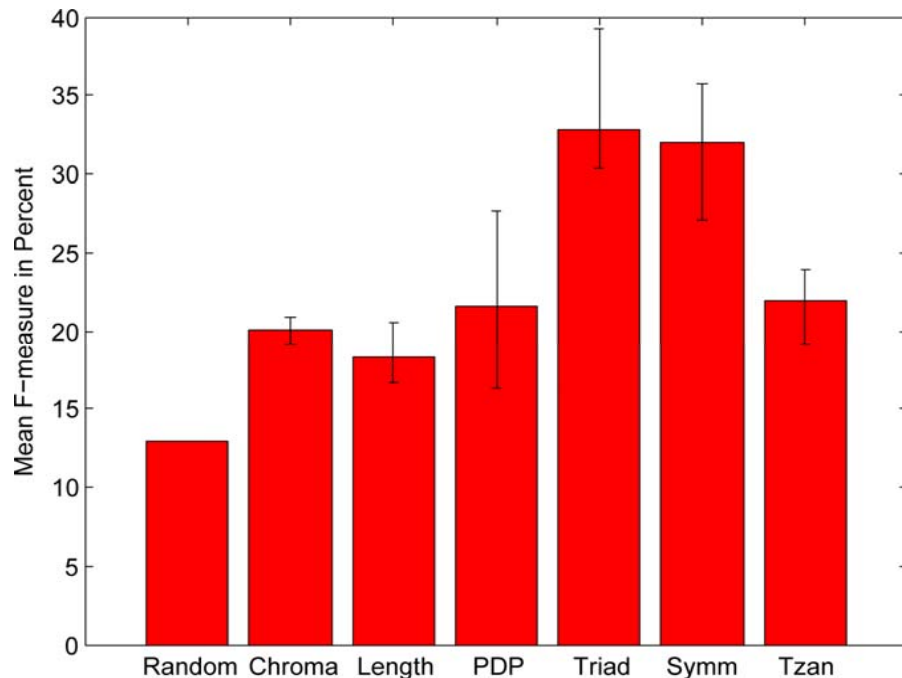


Fig. 4. Schematic representation of Gaussian Mixture Model

# Evaluation: Results

- Random guessing achieves around 13% f-measure
- PDP perform only slightly better than raw chroma
- Geometric pitch space models perform best
- Key-invariance through post-processing is crucial



Seite 14

christian.dittmar@idmt.fraunhofer.de

# Conclusions

---

- Geometric pitch space based harmonic mid-level features show promising results in genre classification task
- Musical knowledge is inherent in mid-level features and post-processing
- Usefulness in combination with timbral and rhythmical features yet to be assessed
- Publicly available genre data must be used in the future (e.g., ISMIR-2004 set)

---



Thank you for your attention!