



**TECHNISCHE UNIVERSITÄT ILMENAU**

Fakultät für Informatik und Automatisierung

# Diplomarbeit

Automatisierte Charakterisierung von Soloparts in Musikstücken

Eingereicht von:	Jakob Abeßer
Geboren am:	03.05.1983 in Jena
Matrikelnummer:	35081
Studiengang:	Ingenieurinformatik
Studienrichtung:	Telekommunikations- und Messtechnik
Anfertigung bei:	Fraunhofer Institut für Digitale Medientechnologie (IDMT) Ehrenbergstraße 31, 98693 Ilmenau
Verantwortlicher Professor:	Prof. Dr.-Ing. Karlheinz Brandenburg (Institut für Medientechnik, Fakultät für Elektrotechnik und Informationstechnik)
Wissenschaftlicher Betreuer:	Dipl.-Ing. Holger Großmann (Fraunhofer IDMT)
Wissenschaftlicher Betreuer:	Dipl.-Ing. (FH) Christian Dittmar (Fraunhofer IDMT)
Inventarisierungsnummer:	2008-08-15/095/II02/2181
Datum der Ausgabe:	15.02.2008
Datum der Abgabe:	15.08.2008



## Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Ilmenau, 13.08.2008

Jakob Abeßer

## Zusammenfassung

Die vorliegende Diplomarbeit beschreibt die Extraktion von High-Level-Merkmalen aus transkribierten Soloparts in Musikstücken. Die aus den Partiturnformationen gewonnenen Merkmale charakterisieren die Spuren von Solo- und Begleitinstrumenten sowohl rhythmisch, melodisch, harmonisch als auch strukturell. Zusätzlich wurden Ansätze zur Beschreibung der Interaktion zwischen den beteiligten Musikern implementiert.

Die Anwendbarkeit der entwickelten High-Level-Merkmale wurde im Kontext der Genre- und Musikerklassifikation anhand der vorliegenden Soloparts evaluiert. Hierzu wurden jeweils zwei Test-Datensätze zusammengestellt, die sowohl Audiodateien im symbolischen (MIDI) als auch im subsymbolischen Format (MP3) enthalten. Für die Genreklassifikation wurde zusätzlich ein Hörtest mit 25 Probanden durchgeführt, dessen Ergebnisse als Vergleichsdaten zu Verfügung stehen. Neben einem SVM-Klassifikator mit vorangegangener Diskriminanzanalyse (LDA) wurde ein neuer Ansatz für einen Klassifikator entwickelt, der mit Hilfe des in dieser Arbeit vorgestellten „rhythmischen Strukturprofils“ in Kombination mit einem *Nearest-Neighbor*-Algorithmus versucht, den durch die Begleitinstrumente vorgegebenen rhythmischen Kontext abzubilden.

Über eine Ensembleentscheidung der instrumentenbezogenen Einzelklassifikatoren wurde für den LDA-SVM-Klassifikator eine mittlere Erkennungsquote von 83,9 % für MIDI-Dateien und 63,4 % für Audiodateien in einer Genretaxonomie bestehend aus sechs Genres erreicht. Da weder klang- noch instrumentierungsbezogene Merkmale berücksichtigt wurden, motivieren die erreichten Ergebnisse weiterführende Forschung in diesem Aufgabenfeld. Die mittleren Klassifikationsquoten für die beiden jeweils aus 4 Künstlern bestehenden Musikertaxonomien betragen 58,7 % und 56 %. Der zweite beschriebene Klassifikator erreichte bei der Genreklassifikation eine Erkennungsquote von 73,2 %.

Nach einer Auswertung der erreichten Ergebnisse werden in dieser Arbeit abschließend einige Verbesserungsvorschläge und Anwendungsszenarien skizziert.

## Abstract

This thesis introduces a set of high-level features to describe instrumental solo-parts. The set contains 119 features related to melodic, rhythmic, harmonic and structural properties of each instrument track. A simple yet common instrumentation model has been applied to describe both the soloing and the accompanying instruments as well as rhythmical and melodic interaction between them. To evaluate the features discriminative power related to different musical styles, an evaluation for both genre and artist classification has been performed each with two different test sets consisting of symbolic and real audio data.

Two different classifier approaches have been utilized, one commonly used SVM classifier with preliminary discriminant analyses (LDA) and one new approach based on the „rhythmical structure profile“ which allows a representation of the rhythmical context provided by the accompanying instruments during the solo part. Using ensemble decisions based on single instrument-related classifiers, the two classifier approaches achieved best mean classification results of 83.9 % and 73.2 % for genre-classification within a 6-class-taxonomy. The first classifier furthermore achieved a classification score of 58.7 % and 56 % for artist-classification within two different 4-class-taxonomies.

Considering that timbre- and instrumentation-related features have not been taken into account within this thesis, the results are encouraging for further research within this topic. Finally, suggestions for improvement as well as some application scenarios are outlined.

## Danksagungen

Für die große Unterstützung im Laufe meines Studiums und meiner Diplomarbeit danke ich besonders meinen Eltern, Monika und Harald Abeßer, meinen Brüdern Michel und Wolfram Abeßer, meiner Freundin Franziska Albert sowie all meinen Freunden.

Weiterhin möchte ich mich bei meinen Betreuern Christian Dittmar und Holger Großmann bedanken, die immer ein offenes Ohr für Fragen hatten und mir mit vielen konstruktiven Ideen und Vorschlägen weiterhalfen.

Ich danke Alexander Loos für regelmäßige und fruchtbare Diskussionen im Laufe der Diplomarbeit sowie allen Mitarbeitern und Studenten in der Abteilung Metadaten des Fraunhofer IDMT für die angenehme Arbeitsatmosphäre und die Möglichkeit, stets einen Ansprechpartner für offene Fragen und Probleme zu finden.

Abschließend danke ich der Musik für die Kraft und Inspiration, die sie mir im Laufe meines Studiums gegeben hat.

*„Die Musik drückt das aus, was nicht gesagt werden kann und worüber zu schweigen unmöglich ist.“*

(Victor Hugo, 1802-1885)

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
1.1	Motivation . . . . .	5
1.2	Aufgabenstellung . . . . .	6
1.3	Aufbau . . . . .	7
<b>2</b>	<b>Grundlagen</b>	<b>9</b>
2.1	Musiktheorie . . . . .	9
2.1.1	Melodik und Harmonik . . . . .	9
2.1.2	Rhythmik . . . . .	14
2.1.3	Musikalische Genres . . . . .	17
2.1.4	Instrumentierung . . . . .	17
2.2	Mathematik . . . . .	18
2.2.1	Stochastik . . . . .	18
2.2.2	Statistik . . . . .	19
2.2.3	Informationstheorie . . . . .	20
2.3	Statistische Mustererkennung . . . . .	21
2.3.1	Merkmal und Merkmalsextraktion . . . . .	21
2.3.2	Dimensionsreduktion des Merkmalsraumes . . . . .	22
2.3.3	Klassifikation . . . . .	27
2.4	Audiotranskription . . . . .	29
<b>3</b>	<b>Stand der Technik</b>	<b>31</b>
3.1	<i>High-level</i> -Merkmale . . . . .	31
3.2	Improvisation und Interaktion . . . . .	38
3.3	Genreklassifikation . . . . .	40
3.4	Musikerklassifikation . . . . .	42
<b>4</b>	<b>Konzeption und Implementierung</b>	<b>45</b>
4.1	Vorverarbeitung . . . . .	45
4.1.1	Konvertierung und Transkription . . . . .	45
4.1.2	Zeitliche Projektion . . . . .	45
4.1.3	Zeitliche Quantisierung . . . . .	48
4.1.4	Harmonische Analyse . . . . .	48
4.2	Merkmalsextraktion . . . . .	50

4.2.1	Allgemeine Merkmale . . . . .	51
4.2.2	Melodische und harmonische Ebene . . . . .	51
4.2.3	Rhythmische Ebene . . . . .	55
4.2.4	Strukturelle Ebene . . . . .	63
4.2.5	Interaktionsebene . . . . .	66
4.3	Implementierung . . . . .	67
<b>5</b>	<b>Evaluation</b>	<b>69</b>
5.1	Genreerkennung . . . . .	69
5.1.1	Aufgabenstellung . . . . .	69
5.1.2	Klassifikatoren . . . . .	69
5.1.3	Hörtest . . . . .	79
5.1.4	Ergebnisse . . . . .	80
5.2	Musikererkennung . . . . .	86
5.2.1	Aufgabenstellung . . . . .	86
5.2.2	Ergebnisse . . . . .	86
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>89</b>
6.1	Fazit . . . . .	89
6.2	Ausblick . . . . .	90
6.3	Einsatzszenarien . . . . .	91
<b>A</b>	<b>Algorithmen</b>	<b>93</b>
A.1	Entscheidungsalgorithmus des Ensemble-Klassifikators . . . . .	93
A.2	Correlative-Matrix-Algorithmus . . . . .	94
<b>B</b>	<b>Testsets für Genre- und Musikerklassifikation</b>	<b>97</b>
<b>C</b>	<b>High-Level-Merkmale</b>	<b>107</b>
<b>D</b>	<b>Akronyme und Formelzeichen</b>	<b>113</b>
<b>E</b>	<b>Thesen</b>	<b>117</b>



# Abbildungsverzeichnis

2.1	Die chromatische Tonleiter . . . . .	10
2.2	Töne der C-Dur-Tonleiter . . . . .	12
2.3	Harmonisierung der C-Dur-Tonleiter in Vierklänge . . . . .	12
2.4	Verschiedene Taktarten - betonte Zählzeiten . . . . .	14
2.5	Notenlängen . . . . .	15
2.6	Polyrhythmik : „2 gegen 3“ sowie „4 gegen 3“ . . . . .	16
2.7	Synkopen . . . . .	16
2.8	Swing . . . . .	17
2.9	Typischer Ablauf der Mustererkennung . . . . .	21
2.10	Zentroiden für ein 2-Klassen-Problem . . . . .	25
2.11	2-Klassen-Beispiel mit Trennebene . . . . .	27
4.1	Allgemeiner Ablauf der Analyse . . . . .	46
4.2	Zeitliche Projektion (Beispiel) . . . . .	47
4.3	Harmonische Analyse (Beispiel) . . . . .	49
4.4	Berechnung des Swing-Faktors im Achtelnotenraster ( $Q = 8$ ) . . . . .	56
4.5	Rhythmisches Strukturprofil - Projektion (Beispiel) . . . . .	57
4.6	Energieprofil $\mathbf{\Lambda}_{R1}$ (Beispiel) . . . . .	58
4.7	Energieprofil $\mathbf{\Lambda}_{R2}$ (Beispiel) . . . . .	59
4.8	Synkopierungen und korrespondierende Matrizen $\mathbf{S}_i$ (Beispiele) . . . . .	61
4.9	Musterverteilung der absoluten Tonhöhe (”Desafinado“ - Latin) . . . . .	63
4.10	Musterverteilung der absoluten Tonhöhe (”Desafinado“ - Latin) . . . . .	65
5.1	Ablauf der Evaluation des Einzel-SVM-LDA-Klassifikators . . . . .	70
5.2	Ablauf der Evaluation des Ensemble-SVM-LDA-Klassifikators . . . . .	74
5.3	Ablauf der Evaluation des RSP-Klassifikators . . . . .	78
5.4	Ergebnisse der LDA-SVM-Klassifikatoren (Vergleich : MIDI - Audio) . . . . .	80
5.5	Ergebnisse der LDA-SVM-Klassifikatoren (Vergleich : MIDI - AUDIO) . . . . .	81
5.6	Ergebnisse der RSP-Klassifikatoren . . . . .	82
5.7	Ergebnisse des Hörtests . . . . .	84
5.8	Korrespondierende Ergebnisse: Hörtests und LDA-SVM-Klassifikator . . . . .	85
5.9	Ergebnisse der Musikerklassifikation . . . . .	87



# Tabellenverzeichnis

2.1	Intervalle innerhalb einer Oktave (Auswahl) . . . . .	10
2.2	Drei- und Vierklänge (Auswahl) . . . . .	11
2.3	Kirchentonleitern in C-Dur . . . . .	12
2.4	Typische Instrumentierungen in verschiedenen Genres . . . . .	18
2.5	Durch die Transkription gewonnene musikalische Parameter . . . . .	29
4.1	Strukturanalyse - Untersuchte musikalische Parameter . . . . .	64
5.1	Testsets für die Musikerklassifikation . . . . .	87
B.1	Internetquellen und Umfang des MIDI-Testsets zur Genreklassifikation	97
B.2	Testset - Audio - Genreklassifikation . . . . .	102
B.3	Testset - Audio - Musikerklassifikation (E-Gitarre) . . . . .	104
B.4	Testset - Audio - Musikerklassifikation (Saxophon) . . . . .	106
C.1	Verteilung der High-Level-Merkmale auf die vier Instrumentalspuren . .	108
D.1	Akronyme . . . . .	113
D.2	Formelzeichen . . . . .	115



# 1 Einleitung

## 1.1 Motivation

In keiner Musikkultur werden Musikstücke exakt in der Form gespielt, in der sie schriftlich notiert oder überliefert vorliegen. Jeder Musiker besitzt eine individuelle Auffassung von Timing, Dynamik und Artikulation, die sich unmittelbar auf sein Spiel auswirkt. All diese Nuancen lassen die Musik lebendig, aufregend und vor allem abwechslungsreich wirken. Bisherige Forschungsansätze im *Music Information Retrieval*<sup>1</sup> zielen vor allem auf die kontextuelle Beschreibung aufgenommener Musikstücke in Form von Audiodaten mit all ihren Charakteristika wie z.B. Tempo, Taktart, Klang, Struktur sowie Instrumentierung ab. Die Semantik der Aufführung dieser Musikstücke sowie die individuellen Spielweisen der beteiligten Musiker wurden dabei bisher jedoch wenig beachtet und analysiert.

Den Untersuchungsgegenstand dieser Arbeit stellen Instrumentalsoli aus verschiedenen kontemporären Musikgenres wie z.B. Blues, Swing, Funk und Rock dar. Musikstücke dieser Genres sind typischerweise strukturiert in in sich abgeschlossene Teile wie Strophe, Refrain, Interlude bzw. Intro und Outro. Diese sind in ihrer Länge und in ihrem Ablauf durch das Arrangement eines Stückes vorgegeben und lassen den an der Aufführung beteiligten Musikern nur geringen Spielraum für eine individuelle Art der Interpretation und Ausführung. In Soloparts, in denen meist ein Musiker am Melodieinstrument improvisiert und von den anderen beteiligten Musikern begleitet wird, hat der Solist trotz des oftmals vorgegebenen harmonischen Ablaufs einen großen Freiraum und kann größtenteils spontan musizieren. Soloparts ermöglichen – so eine der wesentlichen Hypothesen dieser Arbeit – von allen Teilen, aus denen ein Musikstück besteht, die meisten Rückschlüsse auf die individuelle Spielweise des Solisten. Die meisten zeitgenössischen Musiker waren bzw. sind im Laufe ihrer Schaffenszeit musikalisch in einem oder wenigen verwandten Genres aktiv. Dies lässt im Gegenschluss die Vermutung zu, dass für jedes Genre neben generellen melodischen, harmonischen und rhythmischen Eigenschaften typische individuelle Spielweisen der entsprechenden Musiker existieren. Wenn es gelingt, den Vorgang des Musizierens im jeweiligen musikalischen Kontext detaillierter zu beschreiben, so erschließen sich eine Vielzahl von Anwendungsszenarien wie z.B. eine verbesserte Erkennung einzelner Musiker in musikalischen Aufnahmen.

---

<sup>1</sup>(Automatisierte und inhaltsbasierte) Informationsgewinnung aus Musik.

## 1.2 Aufgabenstellung

### Themabeschreibung

In der Abteilung Metadaten am Fraunhofer IDMT wird intensiv an Verfahren zur automatischen Annotation von Audiosignalen und dabei insbesondere Musiksignalen geforscht. Besonderes Augenmerk liegt dabei auf der Gewinnung von semantisch hochwertigen Beschreibungen des musikalischen Inhalts. Als Beispiele hierfür sind für den menschlichen Zuhörer verständliche Begriffe wie Genre, Tempo, Tonart, Perkussivität und Gesangsanteil zu nennen. In der Vergangenheit wurden auch verschiedene Ansätze zur Transkription von Melodie- und Schlagzeuginstrumenten im polyphonen Musiksignal entwickelt.

Im Rahmen dieser Arbeit sollen eine geeignete Darstellungsform zur Beschreibung der rhythmischen, melodischen und harmonischen Spielweise innerhalb eines Instrumentalsolos gefunden werden. Dabei sollen Merkmale zur Beschreibung eines Solos im vorgegebenen harmonischen und rhythmischen Kontext abgeleitet werden und verschiedene Ansätze zur Beschreibung der Interaktion zwischen den beteiligten Instrumenten (Melodie-, Harmonie-, Bass- und Rhythmusinstrumente) untersucht werden. Die entwickelten Merkmale sollen auf ihre Anwendbarkeit für Genreerkennung, Segmentierung und Klassifikation von Musikern hin evaluiert werden. Dabei kann auf bereits bestehende Softwaremodule zur Merkmalsextraktion, Modellbildung und Klassifikation zurückgegriffen werden. Die erzielten Ergebnisse sind vorzustellen und anhand eigens aufbereiteter Testdaten (MIDI und Audio) quantitativ zu bewerten.

### Betreuer

Dipl.-Ing. Holger Großmann

Dipl.-Ing. (FH) Christian Dittmar

## 1.3 Aufbau

Im Kap. 2 werden die musiktheoretischen sowie die mathematischen und signaltheoretischen Begriffe erläutert, die grundlegend für das Verständnis dieser Arbeit sind.

Im darauf folgenden Kap. 3 wird ein Überblick über den Stand der Technik gegeben. Dabei liegt der Schwerpunkt neben den in der Literatur beschriebenen melodischen und rhythmischen High-Level-Merkmalen auf bereits existierenden Ansätzen zur Genre- und Musikerklassifikation sowie der Beschreibung von Improvisation und Interaktion in Musikstücken.

Die Umsetzung aller Verarbeitungsschritte vom Einlesen der Audiodaten, ihrer Umwandlung und Transkription, der rhythmischen und harmonischen Analyse bis hin zur Merkmalsextraktion wird im Kap. 4 im Detail dargestellt. Das Vorgehen bei der Evaluation der entwickelten Merkmale im Rahmen der Genre- und Musikerklassifikation, verwendete Klassifikatoren sowie die dabei erreichten Ergebnisse werden in Kap. 5 erläutert und diskutiert.

Abschließend erfolgt eine Zusammenfassung sowie einen Ausblick auf mögliche Verbesserungsansätze und Einsatzszenarien.





## 2 Grundlagen

### 2.1 Musiktheorie

#### 2.1.1 Melodik und Harmonik

##### Ton und Tonhöhe

Ein Ton bezeichnet ein Schallereignis, welches aus einer periodischen Schallschwingung resultiert. Natürliche Töne, die von Musikinstrumenten, aber auch von der menschlichen Stimme erzeugt werden, sind in der Regel komplexe zusammengesetzte Schwingungen, die aus einem Grundton und einem Gemisch von Obertönen bestehen. Dies sind Töne, deren Frequenz  $f_o$  ( $[f] = 1 \text{ Hz}$ ) einem ganzzahligen Vielfachen der Grundfrequenz  $f_0$  des Grundtones entsprechen. Die empfundene Tonhöhe steigt mit der Frequenz des Grundtones – nach jeder Verdopplung der Frequenz klingt der Ton eine Oktave höher (vgl. auch Kap. 2.1.1). Als Bezugsfrequenz zum Stimmen verschiedener Instrumente wird in der Regel der Kammerton a mit einer Frequenz von  $f = 440 \text{ Hz}$  verwendet.

Die Notenbezeichnungen in der modernen Notenschrift basieren den sieben Stammtönen C, D, E, F, G, A und B. Die Bezeichnung B entspricht dabei der internationalen Notierung<sup>1</sup>, welche in dieser Arbeit verwendet wird. Alle aufeinander folgenden Paare von Stammtönen haben einen Tonhöhenabstand von zwei Halbtonschritten, die beiden Tonpaare B und C bzw. E und F von jeweils einem Halbtonschritt. Das auf diese Art definierte Tonsystem wird als diatonisch bezeichnet. Eine Oktave setzt sich gemäß der temperierten Stimmung<sup>2</sup> aus 12 Halbtönen zusammen. Durch die beiden Vorzeichen  $\sharp$  und  $\flat$  können die Stammtöne alteriert – also um einen Halbtonschritt erhöht oder erniedrigt werden. Liegen zwei benachbarte Stammtöne zwei Halbtonschritte auseinander, so bezeichnet man die beiden durch die Verwendung von Vorzeichen abgeleiteten Töne – z.B.  $D\sharp$  und  $E\flat$  – als enharmonisch verwechselt.

Ausgehend vom Ton C sind in Abb. 2.1 alle Stammtöne einer Oktave inklusive der chromatisch verwechselten Zwischentöne dargestellt – sie bilden zusammen die chromatische Tonleiter. Der Begriff Partitur bezeichnet eine Darstellungsform, in der die Stimmen aller in einem Musikstück gespielten Instrumente in Notenform notiert sind.

---

<sup>1</sup>Dieser Ton wird als B notiert, jedoch im Deutschen als „H“ bezeichnet.

<sup>2</sup>Gleichstufige Stimmung nach Werkmeister, A. (1691) – vgl. [Hai00], S. 29.



Abb. 2.1: Die chromatische Tonleiter

Intervall Stufe	Name	Größe	Halbtonschritte	Beispiel
1	Prime	rein	0	C - C
2	Sekunde	klein	1	C - D $\flat$
		rein	2	C - D
3	Terz	kleine	3	C - E $\flat$
		groß	4	C - E
4	Quarte	rein	5	C - F
5	Quinte	vermindert	6	C - G $\flat$
		rein	7	C - G
6	Sexte	klein	8	C - A $\flat$
		groß	9	C - A
7	Septime	klein	10	C - B $\flat$
		groß	11	C - B
8 (1)	Oktave	rein	12	C - c

Tab. 2.1: Intervalle innerhalb einer Oktave (Auswahl)

## Intervall und Akkord

Werden zwei Töne gleichzeitig oder nacheinander gespielt, so bezeichnet das Intervall das Verhältnis bzw. den Abstand ihrer Tonhöhen. In Tab. 2.1 sind alle Intervalle im Bereich einer Oktave sowohl anhand ihrer musikalischen Bezeichnung als auch ihrer Größe in Halbtonschritten aufgelistet. Weitere Intervalle, die sich aus enharmonischen Verwechslungen ergeben (wie z.B. die übermäßige Prime: C - C $\sharp$ ), wurden der Übersichtlichkeit halber hier nicht betrachtet. Intervallpaare wie z.B. die reine Quarte und die reine Quinte, die sich zu einer reinen Oktave ergänzen, werden Komplementärintervalle genannt.

Ein Akkord bezeichnet das Zusammenklingen mehrerer Töne, deren Intervallabstand untereinander charakteristisch für den Typ des Akkordes ist. Um einen Akkord zu bilden, werden auf den Grundton des Akkordes weitere Akkordtöne (zumeist in Terzabständen) übereinander geschichtet. Bezogen auf den Grundton sind für Dreiklänge also die Intervalle Terz und Quinte bzw. für Vierklänge die Intervalle Terz, Quinte und

Name	Beispiel Akkord	Töne	Intervalle		
			Terz	Quinte	Septime
Dur	C	C, E, G	groß	rein	
Moll	Cm	C, Eb, G	klein	rein	
Vermindert	C <sup>-</sup> , Cdim	C, Eb, Gb	klein	vermindert	
Übermäßig	C <sup>+</sup> , Caug	C, E, G <sup>♯</sup>	klein	übermäßig	
Dur-Major-7	Cmaj <sup>7</sup> , C <sup>Δ</sup>	C, E, G, B	groß	rein	groß
Dur-7	C <sup>7</sup>	C, E, G, Bb	groß	rein	klein
Moll-7	Cm <sup>7</sup>	C, Eb, G, Bb	klein	rein	klein
Moll-7-b5	Cm <sup>7</sup> b <sup>5</sup>	C, Eb, Gb, Bb	klein	vermindert	klein
Vermindert-7	C <sup>o</sup> , C	C, Eb, Gb, Bbb	klein	vermindert	vermindert

Tab. 2.2: Drei- und Vierklänge (Auswahl)

Septime charakteristisch. Vor allem im jazzverwandten Genres wie Swing und Latin werden oft durch Akkorderweiterungen weitere Akkordtöne hinzugefügt. Hierbei werden bezogen auf den Grundton des Akkordes größere Intervalle wie beispielsweise die None (9) oder die Undezime (11) eingesetzt. In der Tab. 2.2 sind die häufigsten Drei- und Vierklänge anhand ihrer Bezeichnung und ihrer charakteristischen Intervallstruktur aufgelistet. Akkorde werden oftmals anstatt in der Grundstellung in einer Umkehrung gespielt. Diese unterscheidet sich in der Reihenfolge, mit der die Akkordtöne in ihrer Tonhöhe angeordnet sind. So liegt beispielsweise in der ersten Umkehrung eines Akkordes die Terz unter der Quinte und der Grundton wird eine Oktave höher gespielt. Die Ursache dafür liegt zumeist in einer besseren Stimmführung<sup>3</sup>.

### Tonart, Skale und funktionale Tonhöhe

In den meisten Musikstücken werden nicht alle 12 Töne der chromatischen Tonleiter gleichhäufig verwendet. Durch die Tonart wird eine Auswahl von sieben Tönen in Form einer Tonleiter vorgegeben, die das Material für Melodie- und Akkordbildung und somit die Tonalität eines Musikstückes oder eines Abschnittes eines Musikstückes bestimmt. Werden die Töne einer Tonleiter in Terzen übereinander geschichtet, so können auf jeder Stufe der Tonart Drei- bzw. Vierklänge gebildet werden. In Abb. 2.3 ist dies anhand der Vierklänge dargestellt, die sich durch eine Harmonisierung der C-Dur-Tonleiter (siehe Abb. 2.2) ableiten lassen.

Ausgehend von jedem der sieben Töne der Tonleiter lassen sich verschiedene Skalen ableiten. In der sogenannten „Akkord-Skalen-Theorie“<sup>4</sup>, die in der melodisch-harmonischen Ausbildung von Jazzmusikern eine zentrale Rolle spielt, bilden sie die Grundlage für

<sup>3</sup>Vgl. [Hai00].

<sup>4</sup>Vgl. [Jun01].



Abb. 2.2: Töne der C-Dur-Tonleiter

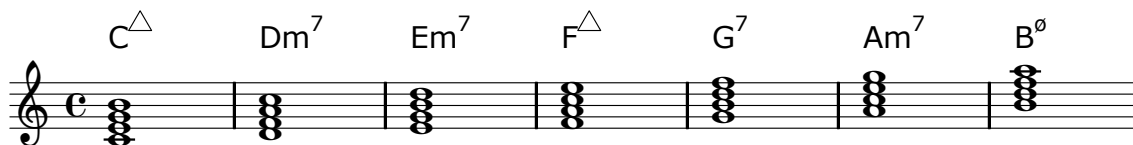


Abb. 2.3: Harmonisierung der C-Dur-Tonleiter in Vierklänge

Stufe	Name	Töne	Intervallstruktur	Akkord
1	Ionisch	C, D, E, F, G, A, B	G, G, H, G, G, G, H	Cmaj <sup>7</sup> , C <sup>Δ</sup>
2	Dorisch	D, E, F, G, A, B, C	G, H, G, G, G, H, G	Dm <sup>7</sup>
3	Phrygisch	E, F, G, A, B, C, D	H, G, G, G, H, G, G	Em <sup>7</sup>
4	Lydisch	F, G, A, B, C, D, E	G, G, G, H, G, G, H	Fmaj <sup>7</sup> , F <sup>Δ</sup>
5	Mixolydisch	G, A, B, C, D, E, F	G, G, H, G, G, H, G	G <sup>7</sup>
6	Äolisch	A, B, C, D, E, F, G	G, H, G, G, H, G, G	Am <sup>7</sup>
7	Lokrisch	B, C, D, E, F, G, A	H, G, G, H, G, G, G	Bm <sup>7</sup> b <sup>5</sup> , B

Tab. 2.3: Kirchentonleitern in C-Dur

die Improvisation über vorgegebene Akkordfolgen. Die sieben Skalen, die sich aus der Dur-Tonleiter ableiten, werden auch als Kirchentonleitern<sup>5</sup> bezeichnet. Jede Skale hat eine typische Intervallstruktur zwischen aufeinander folgenden Tönen. In Tab. 2.3 sind die sieben Skalen für C-Dur mit ihrer Bezeichnung, ihrer Intervallstruktur (G bedeutet Ganz- und H Halbtonschritt) und dem jeweils korrespondierenden Septakkord (Vierklang) der Tonleiter dargestellt. Vor allem in Jazz-verwandten Genres werden neben der Durtonleiter bzw. der damit verwandten Tonleiter von natürlich Moll<sup>6</sup> vor allem die Tonleitern harmonisch Moll und melodisch Moll zum Ableiten weiterer Akkorde und Skalen verwendet. Außerdem kommen einige symmetrische Skalen wie z.B. die Ganzton-Halbton-Skala sowie Skalen mit weniger als sieben Tönen wie z.B. die pentatonische Moll-Skala bestehend aus 5 Tönen zum Einsatz.

Die Information über den Harmonieverlauf kann in der Regel aus den gespielten Noten des Harmonie- bzw. des Bassinstrumentes gewonnen werden. Die Tonart- und Akkorderkennung stellt ein eigenständiges Forschungsgebiet innerhalb des Music Information Retrieval dar. Da sie nicht Teil der Aufgabenstellung dieser Arbeit, für die Analyse des

<sup>5</sup>Vgl. [Hem97], S. 134.

<sup>6</sup>Diese entspricht den Tönen der äolischen Skale auf der 6. Stufe der Dur-Tonleiter.

gespielten Solos jedoch von großem Nutzen ist, wurde eine stark vereinfachte harmonische Analyse implementiert, die in Kap. 4.1.4 näher beschrieben wird.

Es wird eine funktionale Tonhöhe definiert, die den Intervalltyp des Tones bezogen auf den Grundton des Akkordes bezeichnet. Die Angabe erfolgt hierbei als musikalisches Intervall ohne Berücksichtigung der Intervallgröße – so erfolgt beispielsweise nur die Angabe „Terz“ anstatt „großer“ oder „kleiner Terz“. Dazu wird die absolute Tonhöhe  $p_A$  (in Halbtönen) per Modulo-12-Operation auf eine Oktave ( $\tilde{p}_A \in [0, 11]$ ) projiziert und anschließend entsprechend der vierten und ersten Spalte aus Tab. 2.1 in eine funktionale Tonhöhe  $p_F$  umgerechnet ( $p_F \in [1, 7]$ ). Dies ermöglicht eine tonartgeschlechtunabhängige Betrachtung der Tonhöhe eines Tones in einem gegebenen harmonischen Kontext.

### Improvisation

In der Musik beschreibt der Begriff Improvisation die spontane Darbietung von musikalischem Material, welches im Vorfeld nicht in notierter Form vorliegt. Die Improvisation stellt weltweit die am häufigsten praktizierte Form des Musizierens dar und ist ein zentrales Element in nahezu allen musikalischen Kulturkreisen – von der klassischen indischen Musik über afrikanische und nahöstliche Musikformen bis hin zu spanischem Flamenco und keltischer Musik in Europa. Der mit der Improvisation eng verknüpfte kreative Prozess verlangt vom Musiker ein hohes Maß an musikalischem Können sowie eine sehr gute Beherrschung des eigenen Instrumentes. Im Kontext der Improvisation findet auch immer eine Kommunikation und Interaktion des Solisten mit den begleitenden Musikern statt.

Sowohl in der klassischen europäischen Musik seit der Renaissance als auch in einer Vielzahl kontemporärer Musikgenres wie z.B. Swing, Bebop, Cool Jazz, Fusion Jazz, Soul, Funk und Rock<sup>7</sup>, die sich im Laufe des 20. Jahrhunderts entwickelten, basiert die Improvisation in der Regel auf einem sich wiederholenden harmonischen und rhythmischen Ablauf. Vor allem in den jazzverwandten Genres gibt es Themen, also Hauptmelodien eines Stückes, die der improvisierende Musiker umspielt, mit Verzierungen versieht (ornamentiert) oder sie in ihrem melodischen und rhythmischen Verlauf variiert. Besonders in kontemporären Genres wie Pop, Rock und Funk spielt das Arrangement eines Liedes die zentrale Rolle, die Improvisation findet hier meist nur noch in sogenannten Soloparts statt, deren zeitliche Dauer sowie harmonische und rhythmische Grundstruktur festgelegt ist.

Neben der freien Improvisation, die vor allem im Free Jazz und der Avantgarde-Musik

---

<sup>7</sup>Vgl. [Ber07], S. 19.

seit Beginn der 1960er Jahre<sup>8</sup> eine große Verbreitung gefunden hat, spielen die akkordische und die modale Improvisation in den in dieser Arbeit betrachteten Genres (siehe Kap. 2.1.3) die Hauptrolle. Hierbei kann der solierende Musiker das zu verwendende Tonmaterial aus den jeweiligen Akkorden bzw. den dazu korrespondierenden Skalen im jeweiligen harmonischen Kontext ableiten (siehe Akkord-Skalen-Theorie, Kap. 2.1.1). Wie einleitend bereits beschrieben, ist die Improvisation untrennbar mit dem Musiker verbunden, der sie ausführt. Berendt schreibt in [Ber07], S. 206 passend, wozu der Versuch der Imitation eines bereits gespielten Solos durch einen anderen Musiker führt: „Dann verliert es seinen Charakter, und es bleibt nichts als die nackte Formel der Noten.“

## 2.1.2 Rhythmik

### Metrum, Tempo und Taktart

Der Begriff Metrum bezeichnet die Struktur und Gewichtsverteilung der Notenwerte in der Musik. „Das Metrum ist der verborgene Pulsschlag der Musik, der den ständigen Wechsel zwischen schwer und leicht, betont und unbetont regelt.“<sup>9</sup> Das Tempo eines Musikstückes beschreibt den zeitlichen Abstand aufeinander folgender Hauptzählzeiten. Es ist in der Regel nicht konstant sondern variiert im Laufe des Stückes. Die Angabe erfolgt üblicherweise in Schlägen pro Minute (*beats per minute* - BPM).

In der musikalischen Notation wird durch die Taktart eine rhythmische Struktur festgelegt, auf die die Startzeitpunkte und Längen aller Noten bezogen werden. Sie wird durch die Taktbezeichnung (Taktmensur) am Anfang eines Taktes als Bruch angegeben. Der Wert des Nenners gibt die Länge eines Grundschlages des Taktes als Notenwert an, der des Zählers entspricht der Anzahl der entsprechenden Grundschläge innerhalb des Taktes. Entsprechend des Zählers der Taktmensur wird zwischen geraden und ungeraden Taktarten unterschieden. In Abb. 2.4 sind die beiden einfachen Taktarten  $\frac{2}{4}$  und  $\frac{3}{4}$  sowie die in der musikalischen Praxis am häufigsten eingesetzten (zusammengesetzten) Taktarten  $\frac{4}{4}$  und  $\frac{6}{8}$  mit den jeweils betonten Zählzeiten dargestellt<sup>10</sup>.

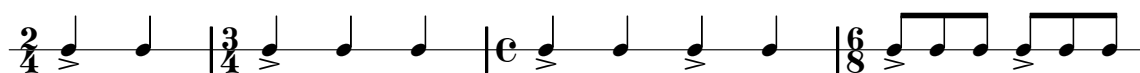


Abb. 2.4: Verschiedene Taktarten - betonte Zählzeiten

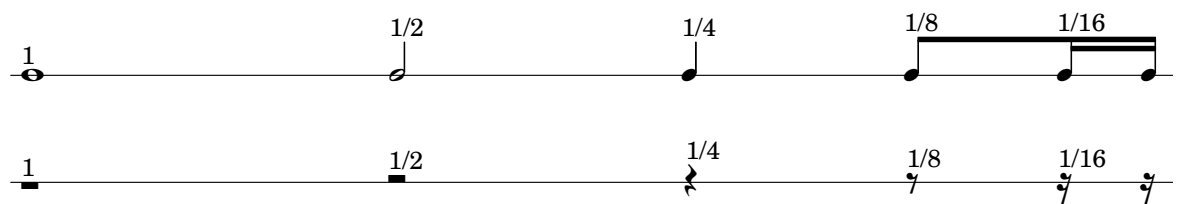
<sup>8</sup>Vgl. [For98].

<sup>9</sup>Vgl. [Hai00], S. 125.

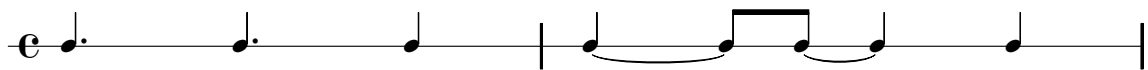
<sup>10</sup>Der  $\frac{4}{4}$ -Takt wird wie in Abb. 2.4 durch das Alla-breve-Zeichen repräsentiert.

## Notenlänge

Neben der Tonhöhe und dem Startzeitpunkt wird eine Note durch ihre zeitliche Dauer charakterisiert. Diese wird in der Notation durch ihren Notenwert dargestellt. In Abb. 2.5(a) sind die wichtigsten Notenwerte und die jeweils korrespondierenden Pausenwerte anhand ihrer Notationsform und Länge dargestellt. Durch eine Punktierung wird eine Note um ihren halben Notenwert verlängert, durch einen Haltebogen wird das Weiterklingen einer Note über ihre Länge hinaus gefordert (siehe Abb. 2.5(b)).



(a) Notenwerte und korrespondierende Pausen



(b) Punktierung und Haltebogen

Abb. 2.5: Notenlängen

## Polyrhythmik

Durch das Zusammenspiel mehrerer Instrumente, die sowohl binäre als auch ternäre Metren<sup>11</sup> einsetzen, kommt es zur Wahrnehmung der sogenannten Polyrhythmik. Hierbei wird eine bestimmte zeitliche Dauer – meistens ein oder mehrere Takte – unterschiedlich unterteilt. Häufig treten eine „2 gegen 3“- bzw. eine „4 gegen 3“-Polyrhythmik auf (siehe Abb. 2.6). Diese unterschiedlichen rhythmischen Ebenen überlagern sich und „treffen“ sich nach sechs bzw. 12 Zeiteinheiten – diese entspricht im dargestellten Beispiel der Länge einer Achtelnote – auf einer gemeinsamen Betonung wieder. Der Autor nimmt an, dass durch die Verwendung von Polyrhythmik die wahrgenommene rhythmische Komplexität der Musik durch den Hörer ansteigt.

## Synkopen

Eine Synkope bezeichnet die Verlagerung einer rhythmischen Betonung von einer Hauptzählzeit des Taktes auf eine Position zwischen zwei Hauptzählzeiten. Die metrische Struktur des Taktes bleibt dabei unverändert. Eine rhythmische Betonung, die auf

<sup>11</sup>Vgl. [Uhl05], S.15.

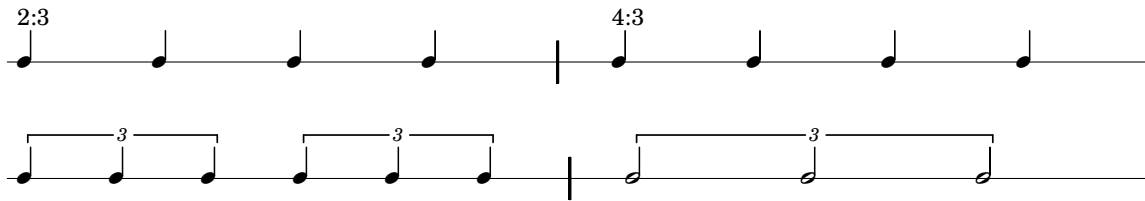


Abb. 2.6: Polyrhythmik : „2 gegen 3“ sowie „4 gegen 3“

den Hauptzählzeiten liegt, wird in der englischsprachigen Literatur als *downbeat*, solche zwischen den Hauptzählzeiten als *offbeat* bezeichnet. Diese Bezeichnungen werden in dieser Arbeit übernommen. Die Verwendung von Synkopen ist typisch für einzelne Genres, wie z.B. für lateinamerikanische Musikgenres.

In Abb. 2.7 ist dies anhand eines Beispiels dargestellt. In beiden Takten wird aufgrund der Haltebögen die dritte Zählzeit nicht gespielt. Im ersten Takt verlagert sie sich um eine Achtelnote auf die Zählzeit „2 und“. Im zweiten Takt bewirkt die synkopierte Spielweise zwei verschobene (betonte) Zählzeiten auf „2 und“ sowie „3 und“.

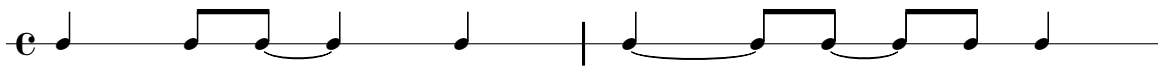


Abb. 2.7: Synkopen

## Swing

Der Begriff Swing bezeichnet neben einem musikalischen Genre, welches in den USA in den 1930er Jahren entstand<sup>12</sup>, ein rhythmisches Phänomen, welches typisch für jazzverwandte Musik ist. Charakteristisch ist zum einen die Verlagerung der Betonung von den Zählzeiten „1“ und „3“ auf die Zählzeiten „2“ und „4“. Weiterhin erfolgt eine zeitliche Verschiebung jeder zweite Achtelnote nach hinten, wie in Abb. 2.8-(b) dargestellt ist. Dieser Rhythmus wird häufig von den Becken oder der Hi-Hat des Schlagzeuges vorgegeben und von den anderen Instrumenten übernommen. In der musikalischen Praxis ist die Stärke dieser Verschiebung tempoabhängig (vgl. [GFB03]). Sie bewegt sich im Bereich zwischen den beiden in Abb. 2.8 notierten Ausführungen und wird umso intensiver wahrgenommen, je stärker sich die Rhythmik der gespielten Noten derer des 2. Notenbeispiels annähert.

<sup>12</sup>Vgl. [Ber07], S. 19.



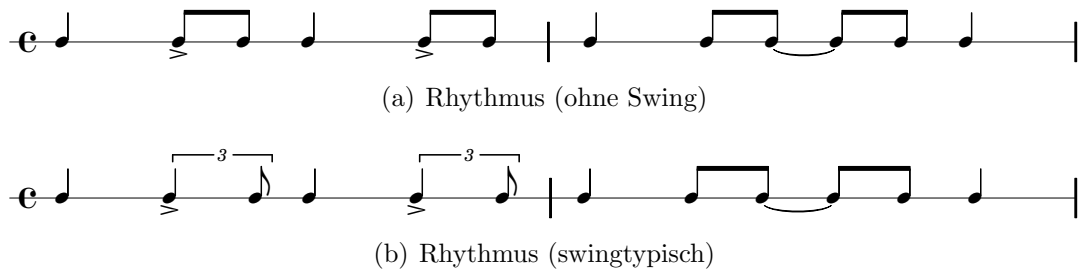


Abb. 2.8: Swing

### 2.1.3 Musikalische Genres

Durch die Einteilung von Musikstücken in verschiedene Genres erfolgt eine Klassifizierung entsprechend musikalischer – also sowohl rhythmischer, melodischer, harmonischer als auch instrumentaler – Eigenschaften und Charakteristika. Jedes Genre hat sich musikalisch aus vielen Einflüssen sowie anderen Genres entwickelt, somit besteht immer auch ein historisch-kultureller Zusammenhang zur jeweiligen Entstehungszeit und zum Entstehungsort – dies soll hier jedoch nicht näher erläutert werden. In dieser Arbeit wurden die sechs Genres Swing (SWI), Latin (LAT), Funk (FUN), Blues (BLU), Pop-Rock (POP) sowie Metal-Hardrock (MHR) ausgewählt, da repräsentative Musikstücke trotz großer musikalischer Unterschiede oftmals Instrumentalsoli in ihrem Arrangement enthalten und damit als Untersuchungsgegenstand im Kontext dieser Arbeit geeignet sind.

### 2.1.4 Instrumentierung

In dieser Arbeit wird ein Instrumentierungsmodell verwendet, welches relativ einfach ist, dennoch in der musikalischen Praxis genreübergreifend häufig anzutreffen ist. Der Solist am Melodieinstrument wird von drei Begleitmusikern am Harmonie-, Bass- und Schlaginstrument begleitet. Alle Musikstücke, die in dieser Arbeit untersucht wurden, erfüllen dieses Kriterium.

Die Begleitmusiker definieren durch ihr Spiel den rhythmischen und harmonischen Kontext, an dem sich der Solist während des Improvisierens orientiert. Vor allem Schlag- und Bassinstrument geben Tempo, Taktart sowie wichtige – oft genretypische – rhythmische Akzente vor. Der harmonische Kontext wird durch das Spielen von bestimmten Akkorden durch das Harmonieinstrument sowie der Basslinie des Bassinstrumentes definiert. Der harmonische Ablauf innerhalb eines Stückes ist meist durch die Komposition des Musikstückes festgelegt und besteht üblicherweise aus mehrfach wiederholten Strukturen. Diese treten auch während des Solos auf – im Genre Swing wird eine solche Struktur als Chorus bezeichnet. In Tab. 2.4 sind einige typische Instrumente der

vier Gruppen aufgelistet, die in den sechs Genres häufig vorkommen. Auf die Frage, welchen Einfluss ein Fehlen einer oder mehrere Begleitinstrumente für die Analyse hat, wird im Kap. 5.1 näher eingegangen.

Genre	Melodieinstrument	Harmonieinstrument
Blues	Saxophon, E-Gitarre	E-Gitarre, Piano
Swing	Trompete, Saxophon	Gitarre, Piano
Funk	Trompete, Saxophon, Gitarre	E-Gitarre, Keyboard, E-Piano
Latin	Trompete, Saxophon, Gitarre	Akkustik-Gitarre, Piano
PopRock	E-Gitarre, Keyboard	E- und Akkustik-Gitarre, Keyboard
Metal-Hardrock	E-Gitarre	E-Gitarre, Keyboard
Genre	Bassinstrument	Schlaginstrument
Blues	E-Bass, Kontrabass	Schlagzeug
Swing	Kontrabass	Schlagzeug
Funk	E-Bass	Schlagzeug
Latin	Kontrabass, E-Bass	Schlagzeug, Perkussion
PopRock	E-Bass	Schlagzeug
Metal-Hardrock	E-Bass	Schlagzeug

Tab. 2.4: Typische Instrumentierungen in verschiedenen Genres

## 2.2 Mathematik

In den folgenden Kapiteln werden die wichtigsten Grundbegriffe aus der Stochastik, Statistik und Informationstheorie erläutert, die in den implementierten High-Level-Merkmalen bei der Beschreibung der Soloparts Verwendung finden.

### 2.2.1 Stochastik

#### Zufallsgröße

Eine Zufallsgröße  $X$  beschreibt eine Funktion, die jedem Elementarereignis  $\omega$  aus einem Definitionsbereich  $\Omega$  eine reelle Zahl  $X(\omega) \in \mathbb{R}$  zuweist. Für den Wertebereich der Wahrscheinlichkeitsfunktion  $P(X)$  aller Ereignisse gilt

$$P(X) \in \mathbb{R}, \quad P(X) \in [0, 1] .$$

Eine Zufallsgröße heißt diskret, wenn ihr Wertebereich eine endliche oder höchstens abzählbare Menge ist. Die Summe aller Einzelwahrscheinlichkeiten  $p_i$  berechnet sich

aus

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} P(X = x_i) = 1 . \quad (2.1)$$

### Verteilungskennwerte

Der Erwartungswert einer diskreten Zufallsgröße ist definiert als

$$E(X) = \sum_{i=1}^{\infty} p_i x_i . \quad (2.2)$$

Die Varianz  $\sigma^2(X)$  beschreibt die mittlere quadratische Abweichung einer Zufallsgröße  $X$  von ihrem Erwartungswert  $E(X)$  und stellt somit ein Maß für ihre Streuung dar.

$$\sigma^2(X) = E[(X - E(X))^2] = \sum_{i=1}^{\infty} [x_i - E(X)]^2 p_i \quad (2.3)$$

Die Quadratwurzel der Varianz wird als Standardabweichung  $\sigma$  bezeichnet. Die Kovarianz zweier Vektoren ergibt sich aus

$$\text{cov}(\mathbf{x}_i, \mathbf{x}_j) = E[(\mathbf{x}_i - \bar{\mathbf{x}}_i)(\mathbf{x}_j - \bar{\mathbf{x}}_j)] . \quad (2.4)$$

## 2.2.2 Statistik

### Empirische Kennwerte

Liegen für eine unbekannte Größe  $X$  eine Menge von  $N$  Observationen  $x_i$  vor, so ermöglichen diese die Berechnung verschiedener statistischer Kenngrößen zur ihrer Beschreibung.

Der Mittelwert  $\bar{x}$  oder auch das arithmetische Mittel der vorliegenden Observationen berechnet sich aus

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i , \quad (2.5)$$

ihre Varianz aus

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 . \quad (2.6)$$

Der Median  $\tilde{x}$  ergibt sich in Abhängigkeit der Observationsanzahl  $N$  aus der Gleichung

$$\tilde{x} = \begin{cases} x_{\frac{N+1}{2}} & , \text{ falls } N \text{ ungerade, bzw. als} \\ \frac{1}{2}(x_{\frac{N}{2}} + x_{\frac{N}{2}+1}) & , \text{ falls } N \text{ gerade.} \end{cases} \quad (2.7)$$

Hierbei wird angenommen, dass die Observationen  $x_i$  der Größe nach sortiert sind. Bei unsymmetrischen Werteverteilungen weicht der Median in der Regel stark vom Mittelwert der Verteilung ab.

In [LI07] wird die D’Agostino-Maßzahl  $D$  als Maß für die Stärke der Normalverteilung der Observationen  $x_i$  eingesetzt. Dieser Ansatz wird in der vorliegenden Arbeit übernommen.

$$D = \frac{\sum_{i=1}^N (i - \frac{N+1}{2}) x_i}{\sqrt{N^3 (\sum_{i=1}^N x_i^2 - \frac{1}{N} (\sum_{i=1}^N x_i)^2)}} \quad (2.8)$$

### 2.2.3 Informationstheorie

#### Nachrichtenquelle

In der Informationstheorie wird eine Nachrichtenquelle durch eine diskrete Zufallsvariable  $X$  modelliert, die alle Werte aus einem endlichen Alphabet  $Z = \{z_1, z_2 \dots z_N\}$  mit den Wahrscheinlichkeiten  $p_i$  annimmt, wobei die Stochastizitätsbedingung  $\sum_i p_i = 1$  gilt. Eine Quelle ohne Gedächtnis wird dabei als Bernoulli-Quelle bezeichnet, zwischen dem Auftreten zweier aufeinander folgender Zeichen besteht keine statistische Abhängigkeit.

Markovquellen bezeichnen Quellen mit Gedächtnis. Für Markovquelle erster Ordnung ist die Auftretenswahrscheinlichkeit eines Zeichens  $X = x_j$  nach dem Auftreten des vorherigen Zeichens  $X = x_i$  von der Verbundwahrscheinlichkeit

$$p(x_i, x_j) = p(x_i) p(x_j | x_i) \quad (2.9)$$

abhängig.

#### Informationsgehalt

Der Informationsgehalt  $I(p_i)$  ( $[I] = 1$  bit), der mit dem Auftreten des Wertes  $X = x_i$  mit der Auftretenswahrscheinlichkeit  $p_i$  verknüpft ist, entspricht

$$I(p_i) = \frac{1}{\log_2 p_i} = -\log_2 p_i . \quad (2.10)$$

#### Entropie

Die Entropie bezeichnet den mittleren Informationsgehalt eines Zeichens. Für eine Bernoulli-Quelle ist die Entropie  $H_0(X)$  ( $[H_0] = 1$  bit/Zeichen) definiert als

$$H_0(X) = \sum_{i=1}^N p_i I(p_i) = - \sum_{i=1}^N p_i \log_2 p_i . \quad (2.11)$$

Je geringer der Wert der Entropie, desto höher ist die Redundanz und Vorhersagbarkeit der Quelle. Die Entropie nimmt ihren Maximalwert für den Fall an, dass jedes der  $N$  möglichen Quellenzeichen die gleiche Auftretenswahrscheinlichkeit  $p_i = 1/N$  besitzt. In diesem Fall ist die Vorhersagbarkeit der Quelle bezogen auf das Auftreten des folgenden Zeichens minimal und der damit verbundene Informationsgewinn maximal.

Aus der bedingten Entropie

$$H_i(x_i) = - \sum_{j=1}^N p_{j|i} \log_2 p_{j|i} \quad (2.12)$$

lässt sich die Entropie  $H_1(X)$  der Markovquelle erster Ordnung berechnen.

$$H_1(X) = \sum_{i=1}^N p_i H_i(x_i) = - \sum_{i=1}^N p_i \sum_{j=1}^N p_{j|i} \log_2 p_{j|i} \quad (2.13)$$

## 2.3 Statistische Mustererkennung



Abb. 2.9: Typischer Ablauf der Mustererkennung

Die statistische Mustererkennung bezeichnet die Suche nach Strukturen und Gesetzmäßigkeiten, die in Signalen in Form von Mustern vorliegen. Typische Anwendungsbeispiele sind die Sprach- und Sprecherdetektion aus Audiodaten, die Suche nach bestimmten Objekten in Bildern und die Detektion von Bewegungslinien (Trajektorien) in Videoaufnahmen. Im folgenden Kapitel wird auf typische Aspekte eingegangen, die dabei in der Audiosignalverarbeitung eine wichtige Rolle spielen.

### 2.3.1 Merkmal und Merkmalsextraktion

Merkmale repräsentieren semantische Teilaspekte, die der Beschreibung eines Signal dienen. Dies führt zu einer Reduktion der bei der anschließenden Analyse zu verarbeitenden Datenmenge. Merkmale, die den Verlauf sowie die Form und Zusammensetzung eines physikalischen Signals beschreiben, werden als *Low-level*-Merkmale bezeichnet. Sie werden mit Hilfe verschiedener Verfahren der digitalen Signalverarbeitung direkt aus dem zu beschreibenden Signal oder einem in einen Bildbereich (z.B. Frequenz- oder Cepstralbereich) transformierten Signal gewonnen. Ihre Berechnung erfolgt in der Regel

gefenstert über sehr kurze zeitliche Abschnitte des Signales (*frames*). Häufig eingesetzte *Low-level*-Merkmale sind die Nulldurchgangsrate<sup>13</sup> (*zero crossing rate*) und Kurzzeitenergie<sup>13</sup> (*short time energy*) im Zeitbereich, spektrale Flachheit<sup>14</sup> (*spectral flatness measure*) und spektrale Schwerpunkt<sup>15</sup> (*spectral centroid*) im Frequenzbereich sowie die Mel-Cepstrum-Koeffizienten<sup>15</sup> (MFCC) im Cepstralbereich.

Aus einer Kombination bestehender *Low-level*-Merkmale sowie unter Anwendung von grundlegendem semantischen Wissen (z.B. aus der Musiktheorie) können *Mid-level*-Merkmale berechnet werden. Sie werden häufig im Kontext der musikalischen Ähnlichkeitsberechnung eingesetzt. Beispielhaft sind chromagrammbasierte Merkmale ([BP04]), die sich aus der Häufigkeit bestimmter Tonhöhen ableiten, sowie sogenannte *rhythmical patterns* ([DGW04]) zu nennen, die wiederkehrende Fragmente der Rhythmusinstrumente abbilden.

Merkmale, die sich wie beispielsweise Tonart, Taktart oder Tempo direkt aus der Semantik der Musiktheorie ableiten, werden als *High-level*-Merkmale bezeichnet. Sie können auch ohne signaltheoretisches Wissen über das Audiosignal von einem Hörer in ihrer Bedeutung erfasst werden. Ihre Extraktion bedarf meist mehrerer Verarbeitungsschritte und ist in der Regel aufwendiger und fehleranfälliger als die Berechnung von *Low*- und *Mid-level*-Merkmalen. Die Merkmalsextraktion bezeichnet die Abbildung eines Audiosignales in einen Merkmalsvektor  $\mathbf{x}_F \in \mathbb{R}^{N_F}$  bestehend aus  $N_F$  Merkmalen.

### 2.3.2 Dimensionsreduktion des Merkmalsraumes

#### Hauptachsentransformation (PCA)

Die Hauptachsentransformation<sup>16</sup> – beschreibt ein Verfahren zur Redundanzreduktion durch Verringerung der Korrelation zwischen mehrdimensionalen Merkmalsvektoren. Es findet eine Überführung in einem niederdimensionalen Vektorraum mit neuer Basis statt. Hierbei handelt es sich um ein Verfahren, welches in unüberwachten Lernsituationen angewandt wird, da für die einzelnen Merkmalsvektoren keine Klassenzugehörigkeit bekannt ist. Ziel ist es, den Informationsverlust bezüglich der Verteilung im Merkmalsraum so gering wie möglich zu halten. Ausgangspunkt der Betrachtung ist die Matrix  $\mathbf{X}_F \in \mathbb{R}^{N_S \times N_F}$ , die zeilenweise alle  $N_S$  Merkmalsvektoren enthält, die aus den Objekten im vorliegenden Datensatz extrahiert wurden.

Die Kovarianzmatrix  $\Sigma \in \mathbb{R}^{N_F \times N_F}$  von  $\mathbf{X}_F$  setzt sich entsprechend Gleichung (2.4) aus

<sup>13</sup>Vgl. [ST95].

<sup>14</sup>Vgl. [Dit02].

<sup>15</sup>Vgl. [Pee04].

<sup>16</sup>Karhunen-Loève-Transformation (KLT) – *Principle Component Analysis* (PCA).

den Kovarianzen zwischen den einzelnen Spaltenvektoren  $\mathbf{x}_i$  der Matrix  $\mathbf{X}_F$  zusammen.

$$\mathbf{\Sigma} = \begin{bmatrix} \text{cov}(\mathbf{x}_1, \mathbf{x}_1) & \cdots & \text{cov}(\mathbf{x}_1, \mathbf{x}_{N_F}) \\ \vdots & \ddots & \vdots \\ \text{cov}(\mathbf{x}_{N_F}, \mathbf{x}_1) & \cdots & \text{cov}(\mathbf{x}_{N_F}, \mathbf{x}_{N_F}) \end{bmatrix} \quad (2.14)$$

Die Werte auf der Hauptdiagonale von  $\mathbf{\Sigma}$  entsprechen den Varianzen  $\sigma_{\mathbf{x}_i}^2$ .

$$\text{cov}(\mathbf{x}_i, \mathbf{x}_i) \equiv \sigma_{\mathbf{x}_i}^2. \quad (2.15)$$

Die Kovarianzmatrix  $\mathbf{\Sigma}$  ist symmetrisch, da  $\text{cov}(\mathbf{a}, \mathbf{b}) \equiv \text{cov}(\mathbf{b}, \mathbf{a})$  gilt. Um eine Dekorrelation aller Merkmale zu erreichen, müssen alle Kovarianzen gleich Null werden und  $\mathbf{\Sigma}$  in eine Diagonalmatrix  $\mathbf{\Lambda}$  überführt werden. Dies wird durch eine Rotation des Merkmalsraumes mit Hilfe der Transformationsmatrix  $\mathbf{\Gamma}_{KLT}$  erreicht.

$$\mathbf{\Lambda} = \mathbf{\Gamma}_{KLT} \cdot \mathbf{\Sigma} \cdot \mathbf{\Gamma}_{KLT}^T \quad (2.16)$$

Eine symmetrische Matrix  $\mathbf{A}$  kann über eine Eigenwertzerlegung gemäß

$$\mathbf{A} = \mathbf{V}_A \cdot \mathbf{\Gamma}_A \cdot \mathbf{V}_A^{-1} = \mathbf{V}_A \cdot \mathbf{\Gamma}_A \cdot \mathbf{V}_A^T \quad (2.17)$$

zerlegt werden. Die Matrix  $\mathbf{V}_A$  enthält spaltenweise alle Eigenvektoren von  $\mathbf{A}$ .  $\mathbf{\Gamma}_A$  ist eine Diagonalmatrix mit allen Eigenwerten  $\lambda_i$  auf ihrer Hauptdiagonale. Da die Eigenvektoren einer symmetrischen Matrix zueinander orthogonal sind, kann  $\mathbf{V}_A^{-1}$  durch  $\mathbf{V}_A^T$  ersetzt werden.

Durch Umstellen von Gleichung (2.17) nach  $\mathbf{\Gamma}_A$  ergibt sich für die Transformationsmatrix  $\mathbf{\Gamma}_{KLT}$  in Gleichung (2.16) in analoger Betrachtung die Forderung, dass ihre Zeilen gleich den Eigenvektoren von  $\mathbf{\Sigma}$  sein müssen. Nach der Transformation der Kovarianzmatrix  $\mathbf{\Sigma}$  enthält  $\mathbf{\Lambda}$  alle Eigenwerte von  $\mathbf{\Sigma}$  auf ihrer Hauptdiagonale.

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_i \end{bmatrix} \quad (2.18)$$

Gemäß Gleichung (2.15) entsprechen die Eigenwerte  $\lambda_i$  den Anteilen an der Gesamtvarianz, die der Prozess in Richtung der korrespondierenden Eigenvektoren von  $\mathbf{\Lambda}$  besitzt. Werden sie der Größe nach sortiert, so können die Eigenwerte mit sehr geringem Anteil an der Gesamtvarianz sowie ihre korrespondierenden Eigenvektoren aus der Transfor-

mationsmatrix verworfen und  $N_{PC}$  Hauptkomponenten (*principle components*) beibehalten werden. Je nach Anforderung an die Transformation ist dabei eine maximale Reduktion der Varianz des Prozesses bis auf einen Wert von höchstens 70% des Ausgangswertes zulässig<sup>17</sup>, damit der Informationsverlust bezüglich der Verteilungseigenschaften von  $\mathbf{X}_F$  nicht zu groß ist. Die Hauptachsentransformation bewirkt somit eine Projektion der Merkmalsvektoren  $\mathbf{x}_F$  in einen niederdimensionalen Merkmalsraum  $\mathbb{R}^{N_{PC}}$ . Die Matrix  $\widetilde{\mathbf{X}}_F$ , die die dekorrelierten Merkmalsvektoren enthält, berechnet sich aus der Multiplikation

$$\widetilde{\mathbf{X}}_F = \mathbf{X}_F \cdot \mathbf{\Gamma}_{KLT}^T. \quad (2.19)$$

### Lineare Diskriminanzanalyse (LDA)

Liegen für die extrahierten Merkmalsvektoren Informationen bezüglich der Klassenzugehörigkeit der repräsentierten Objekte vor, so handelt es sich um eine überwachte Lernsituation. Ein Verfahren zur Dimensionsreduktion des Merkmalsraumes, welches dabei die Trennbarkeit der einzelnen Klassen im transformierten Merkmalsraum maximiert, ist die lineare Diskriminanzanalyse (LDA). Für jeden Merkmalsvektor  $\mathbf{x}_i$  liegt ein Klassenlabel  $c_i$  vor, das die Zugehörigkeit zu einer bestimmten Klasse abbildet. Um die Trennbarkeit der Klassen im Merkmalsraum für die anschließende Klassifikation zu maximieren, werden zwei Ziele verfolgt:

1. Maximierung der Innerklassenvarianz (Varianz zwischen den klassenbezogenen Zentroiden  $\boldsymbol{\mu}_c$  im Merkmalsraum)
2. Minimierung der Intraklassenvarianz (Varianz zwischen den Merkmalsvektoren aller Objekte einer Klasse im Merkmalsraum)

Beide Varianzen werden durch die *Between-class*-Streumatrix  $\mathbf{S}_B$  sowie die *Within-class*-Streumatrix  $\mathbf{S}_W$  beschrieben<sup>18</sup>.

$$\mathbf{S}_B = \sum_c N_c (\boldsymbol{\mu}_c - \bar{\mathbf{x}})(\boldsymbol{\mu}_c - \bar{\mathbf{x}})^T \quad (2.20)$$

$$\mathbf{S}_W = \sum_c \sum_{i \in c} (\mathbf{x}_i - \boldsymbol{\mu}_c)(\mathbf{x}_i - \boldsymbol{\mu}_c)^T \quad (2.21)$$

<sup>17</sup>Vgl. Webb, A. R. (2002), *Statistical Pattern Recognition*. John Wiley and Sons Ltd., S. 324.

<sup>18</sup>Vgl. [Wel07].



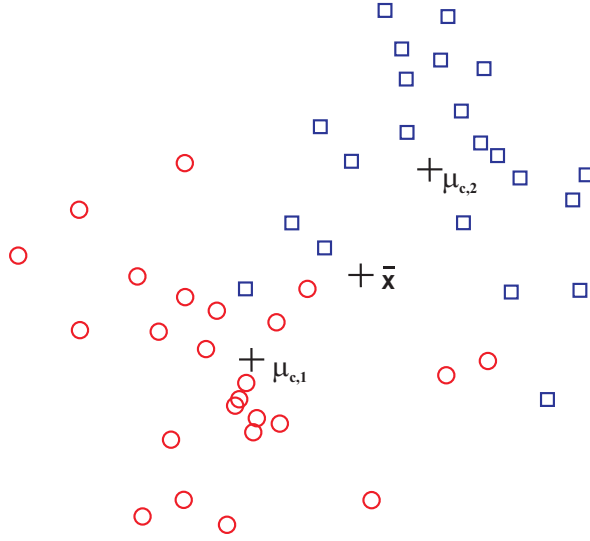


Abb. 2.10: Zentroiden für ein 2-Klassen-Problem

$\mu_c$  entspricht dem Schwerpunkt aller  $N_c$  Merkmalsvektoren einer Klasse  $c$  und  $\bar{x}$  dem Schwerpunkt aller Merkmalsvektoren des Datensatzes.

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} \mathbf{x}_i \quad (2.22)$$

$$\bar{x} = \sum_i \mathbf{x}_i \quad (2.23)$$

Dies ist in Abb. 2.10 anhand eines 2-Klassen-Problems beispielhaft dargestellt. Der Rayleigh-Koeffizient  $J(\Gamma)$  gibt das Verhältnis beider Streumatrizen nach einer Transformation mit der Matrix  $\Gamma$  an. Ziel der LDA ist es, diesen Koeffizienten zu maximieren, um eine optimale Trennbarkeit aller Klassen nach der Transformation zu gewährleisten.

$$\Gamma_{opt} = \max_{\Gamma} J(\Gamma) = \max_{\Gamma} \frac{\Gamma^T \mathbf{S}_B \Gamma}{\Gamma^T \mathbf{S}_W \Gamma} \quad (2.24)$$

Für die LDA wird im Folgenden eine Implementierung vorgestellt, in der die optimale Transformationsmatrix über einen zweistufigen PCA-Prozess berechnet wird. Ausgangspunkt ist eine spaltenweise varianznormierte Merkmalsmatrix, die für jede Spalte den Mittelwert 0 und die Varianz 1 besitzt.

Im ersten Schritt erfolgt eine Eigenwertzerlegung der *Within-scatter*-Matrix  $\mathbf{S}_W$  entsprechend Gleichung (2.17), um eine Dekorrelation der Merkmalsvektoren innerhalb der einzelnen Klassen zu erreichen. Die Eigenvektoren von  $\mathbf{S}_W$  werden gemäß Gleichung (2.18) spaltenweise in der Matrix  $\mathbf{V}_W$ , ihre Eigenwerte in der Diagonalmatrix  $\Lambda_W$  gespeichert. Anschließend erfolgt eine Transformation der *Between-scatter*-Matrix

$\mathbf{S}_B$  mit Hilfe der aus den Eigenvektoren der *Within-scatter*-Matrix abgeleiteten Transformationsmatrix.

Im zweiten Schritt erfolgt eine Varianznormierung im transformierten Merkmalsraum. Da alle Eigenvektoren  $\lambda_i$  von  $\mathbf{S}_B$  nach der Transformation auf den Hauptdiagonalen der transformierten Streumatrix  $\mathbf{S}_B$  liegen, wird über eine zweifache Multiplikation mit der Matrix

$$\mathbf{\Lambda}_W^{-1/2} = \begin{bmatrix} \lambda_1^{-1/2} & 0 & \dots & 0 \\ 0 & \lambda_2^{-1/2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_i^{-1/2} \end{bmatrix} \quad (2.25)$$

eine Varianz aller Merkmale im transformierten Merkmalsraum von 1 erreicht. Zusammengefasst erfolgen die ersten beiden Schritte der Transformation also gemäß der Gleichung

$$\hat{\mathbf{S}}_B = \mathbf{\Lambda}_W^{-1/2} \cdot \mathbf{V}_W^\top \cdot \mathbf{S}_B \cdot \mathbf{V}_W \cdot \mathbf{\Lambda}_W^{-1/2}. \quad (2.26)$$

Der dritte Schritt entspricht einer Dekorrelation der Klassenmittelpunkte  $\mu_c$ . Hierzu erfolgt eine Eigenwertzerlegung der Matrix  $\hat{\mathbf{S}}_B$ . Die Eigenvektoren von  $\hat{\mathbf{S}}_B$  werden analog zur ersten Eigenwertzerlegung spaltenweise in der Matrix  $\mathbf{V}_{\hat{B}}$  gespeichert.

Die Transformationsmatrix  $\mathbf{\Gamma}_{LDA}$  (vgl. Gleichung (2.19)) ergibt sich also aus einer Multiplikation der Transformationsmatrizen der drei Teilschritte.

$$\mathbf{\Gamma}_{LDA} = \mathbf{V}_{\hat{B}}^\top \cdot \mathbf{\Lambda}_W^{-1/2} \cdot \mathbf{V}_W^\top \quad (2.27)$$

Die Anzahl der Spalten der Transformationsmatrix  $\mathbf{\Gamma}_{LDA}$  kann verringert werden, um eine Dimensionsreduktion des Merkmalsraumes zu erreichen. Ausschlaggebend für die Wahl der beizubehaltenden Spalten ist die Größe der Eigenvektoren in  $\mathbf{\Lambda}_{\hat{B}}$ . Um den Informationsverlust bezüglich der Diskriminanz aller Klassen möglichst gering zu halten, werden – wie bereits in Kap. 2.3.2 beschrieben – die Spalten beibehalten, deren korrespondierende Eigenwerte zusammen einen Großteil der Gesamtvarianz enthalten. Im Allgemeinen wird die Gleichung

$$N_{PC} = N_C - 1 \quad (2.28)$$

für die Berechnung der Anzahl der verwendeten Hauptkomponenten eingesetzt. Wie in Kap. 5.1.2 beschrieben, bestätigt die Observation der Transformationsmatrix  $\mathbf{\Gamma}_{LDA}$  die Gültigkeit dieser Gleichung.

### 2.3.3 Klassifikation

Der Begriff Klassifikation beschreibt die Zuordnung eines unbekannten Objektes zu einer Klasse innerhalb einer vorgegebenen Klassenhierarchie. Das Objekt wird durch den Merkmalsvektor  $\mathbf{x}$  repräsentiert. Die einzelnen Klassen können je nach Anwendung und Aufgabenstellung musikalische Genres, einzelne Instrumente aber auch Musiker oder Bands sein. Die Zuordnung erfolgt mit Hilfe des Klassifikators, der in Form einer mathematischen Zuordnungsfunktion vorliegt. Er wird anhand der statistischen Eigenschaften einer vorgegebenen Lernstichprobe dimensioniert. Ist die Klassenzugehörigkeit jedes Elementes der Lernstichprobe bekannt, spricht man von überwachtem Lernen. Ziel des Klassifikatorentwurfes ist es, die Anzahl von fehlerhaften Klassenzuordnungen unbekannter Merkmalsvektoren zu minimieren. Daher ist es notwendig, dass in den einzelnen Merkmalen, aus denen sich der Merkmalsvektor zusammensetzt, ein möglichst hoher Informationsinhalt bezüglich der Klassenzugehörigkeit eines Objektes enthalten ist.

#### Support Vector Machines

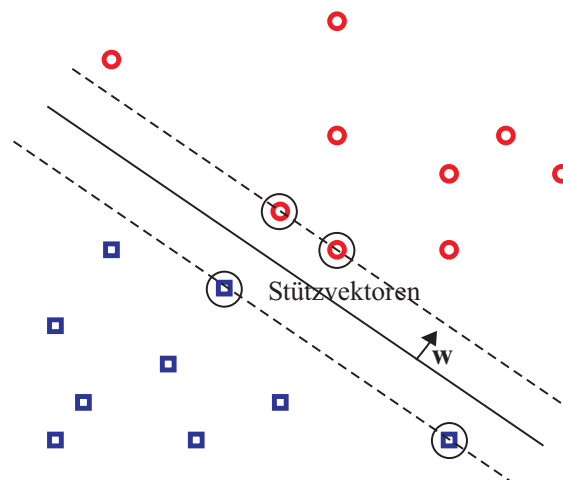


Abb. 2.11: 2-Klassen-Beispiel mit Trennebene

Die *support vector machine* beschreibt einen Klassifikator, der versucht die Vektoren verschiedener Klassen im mehrdimensionalen Merkmalsraum mit Hilfe einer Hyperebene zu trennen. Die Berechnung der Trennebene erfolgt dabei so, dass an den Klassengrenzen ein möglichst breiter Randbereich (*margin*) frei von Klassenvektoren bleibt – daher resultiert der Begriff Breiter-Rand-Klassifikator (*large margin classifier*). Die

Stützvektoren (*support vectors*) bezeichnen diejenigen Klassenvektoren, die der Trennebene jeweils am nächsten liegen und somit ihre Ausrichtung im Merkmalsraum bestimmen. Dies ist in Abb. 2.11 für ein 2-Klassen-Problem in einem zweidimensionalen Merkmalsraum beispielhaft dargestellt.

Für eine Menge von  $N_S$  Merkmalsvektoren  $\mathbf{x}_F \in \mathbb{R}^{N_F}$  und dem dazu korrespondierenden Vektor  $\mathbf{c} \in \{-1, 1\}$  mit den Klassenlabels löst das „C-SVM“-Verfahren<sup>19</sup>, welches in dieser Arbeit eingesetzt wird, das Optimierungsproblem

$$\min_{\mathbf{w}, b} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{N_s} \xi_i \quad (2.29)$$

unter der Bedingung

$$c_i(\langle \mathbf{w}, \phi(\mathbf{x}_F) \rangle + b) \geq 1 - \xi_i .$$

Der zweite Term beinhaltet die Summe über alle Entfernungen  $\xi_i$  der Klassenvektoren zur Trennebene multipliziert mit einem Bestrafungsfaktor (*penalty parameter*). Dieser Term ermöglicht es prinzipiell auch, dass einzelne Klassenobjekte in den Halbraum fallen, der der anderen Klasse zugeordnet ist. Dies vereinfacht häufig die Suche nach einer optimalen Trennebene.

Die zu bestimmende Trennebene wird durch den Normalenvektor  $\mathbf{w}$  und den Parameter  $b$  wie folgt beschrieben:

$$\langle \mathbf{w}, \phi(\mathbf{x}_F) \rangle + b = 0 . \quad (2.30)$$

Eine Hyperebene kann die beiden Klassen nur dann trennen, wenn die Bedingung der linearen Trennbarkeit erfüllt ist. Da diese Eigenschaft in der Regel nicht gegeben ist, wird der Merkmalsraum in einen höheren Merkmalsraum überführt, bis die lineare Trennbarkeit gegeben ist und eine entsprechende Trennfunktion gefunden werden kann. Um den hohen Rechenaufwand bei der Transformation zu umgehen, werden für die Trennfunktionen sogenannte Kernel-Funktion (*kernel function*)  $\phi(\mathbf{x})$  verwendet, die mit vergleichsweise geringem Aufwand zu transformieren sind. In dieser Arbeit wird die sogenannte *radial basis function* (RBF-Kernel) eingesetzt:

$$K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2); \quad \gamma > 0 . \quad (2.31)$$

Für weiterführende Beschreibungen sei auf [CL01] verwiesen.

---

<sup>19</sup>Vgl. [CL01].

## 2.4 Audiotranskription

Der Begriff Transkription bezeichnet die Extraktion einer musikalische Notation bzw. Partitur aus einem aufgenommenen Musikstück. In dieser symbolischen Repräsentation werden alle gespielten Noten sowohl anhand ihrer zeitlichen Parameter als auch anhand ihrer Tonhöhe und Lautstärke abgebildet. Die Aufgabenstellung, Musikstücke mit monophonen Melodien zu transkribieren, ist in der Vergangenheit bereits durch verschiedene Ansätze algorithmisch zufriedenstellend gelöst worden (vgl. [Kla98], S. 1). Polyphone Stücke, in denen mehrere Instrumente zeitgleich gespielt werden, stellen für die Transkription hingegen eine große Herausforderung dar, da hier eine getrennte Transkription aller beteiligter Instrumente erfolgen soll. Im Allgemeinen besteht das Audiosignal aus einer Vielzahl von Einzelkomponenten, die sich in vielen Frequenzbereichen überlagern, was ihre anschließende Separierung zusätzlich erschwert.

In dieser Arbeit wird das am Fraunhofer IDMT entwickelte *Transcription Toolbox* System zur Extraktion von Partiturinformatoren aus Audiodaten eingesetzt. Es ermöglicht die getrennte Transkription der vier in Kap. 2.1.4 genannten Instrumententypen Melodie-, Harmonie-, Bass- sowie Schlaginstrument. Nach der Transkription liegen die *Partiturinformatoren* aller detektierten Noten als Vektoren vor, sie sind in Tabelle 2.5 dargestellt. Da alle in dieser Arbeit beschriebenen Ansätze zur Merkmalsextraktion (siehe Kap. 4.2) auf der Analyse von Partiturinformatoren beruhen, besteht für die Beschreibung von realen Audiodaten eine starke Abhängigkeit von der Qualität der vorangegangenen Transkription. Werden einzelne gespielte Noten in Bezug auf ihre Tonhöhe oder ihre zeitlichen Parameter fehlerhaft oder unvollständig detektiert, so hat dies unmittelbare Auswirkung auf die Aussagekraft der gewonnenen High-Level-Merkmale. Die vier im *Transcription Toolbox* System eingebundenen Transkriptionsalgorithmen erreichten im internationalen Vergleich gute bis sehr gute Ergebnisse. Eine Übersicht über die Funktionsweise der einzelnen Algorithmen sowie über die Evaluationsergebnisse findet sich in [DDR07] auf den Seiten 5 bis 7.

$p_A$	Absolute Tonhöhe	MIDI <i>pitch</i>
$v$	Lautstärke	MIDI <i>velocity</i>
$t$	Notenanfangszeitpunkt	MIDI <i>onset</i>
$\Delta t$	Notendauer	MIDI <i>duration</i>

Tab. 2.5: Durch die Transkription gewonnene musikalische Parameter



## 3 Stand der Technik

### 3.1 High-level-Merkmale

#### Melodische High-level-Merkmale

McKay stellt in [McK04] ein System zur Genreklassifikation basierend auf einer Menge von 109 *High-level*-Merkmalen vor, die aus symbolischen Audiodaten extrahiert werden. Die Merkmale können 7 Gruppen zugeordnet werden – Merkmale bezogen auf die Instrumentierung, die musikalische Textur, den Rhythmus, die Dynamik, der Häufigkeit von Tonhöhen sowie auf Melodien und Akkorde, die innerhalb der untersuchten MIDI-Dateien auftreten. Die instrumentierungsbezogenen Merkmale beinhalten u.a. Aussagen über das Auftreten einzelner Instrumentengruppen wie Gitarren, Streicher oder Bläser sowie über die Anzahl der jeweils innerhalb der Gruppen gespielten Noten. Die musikalische Textur wird durch die Untersuchung der verschiedenen Stimmen innerhalb des Musikstückes, der Anzahl und Dauer der jeweils gespielten Noten sowie durch den Dynamikunterschied zwischen den Stimmen abgebildet. Neben den bereits genannten Merkmalen werden in [McK04] auch Ansätze zur Charakterisierung von Melodien und Akkorden innerhalb einer MIDI-Datei beschrieben. Über drei verschiedene Tonhöhenhistogramme erfolgt die Untersuchung der Tonalität. Ein Intervallhistogramm ermöglicht es weiterhin, die häufigsten Intervalle in Halbtönen und Intervalltypen (z.B. Terzen oder Quinten) zu erkennen. Außerdem kann das von ihm implementierte System Arpeggios<sup>1</sup> und Akkorde auf die Anzahl der Akkordtöne, die Häufigkeit bestimmter Akkordtypen und die mittlere Dauer von Akkordtönen untersuchen.

In [Li06] vergleicht Li zwei Ansätze zur Darstellung von Melodieverläufen. Das Intervallmodell (*interval*) beschreibt den Abstand zweier Töne in Halbtonschritten, das Tonleiterstufenmodell (*scale degree*) hingegen als musikalisches Intervall, wobei die Töne durch ihre Tonleiterstufe in der momentanen Tonart repräsentiert werden (vgl. Kap. 2.1.1). Hierbei wird lediglich zwischen Dur- und Molltonarten unterschieden. Um die Eignung beider Ansätze zur „Modellierung von Melodie“ zu vergleichen, versucht Li, basierend auf einem großen Archiv von Musikstücken<sup>2</sup> über die Auftretenshäufigkeit einzelner Notenabstände bei gegebener Tonart ihre Auftretenswahrscheinlichkeiten zu

---

<sup>1</sup>Akkorde, bei denen die Akkordtöne nacheinander angeschlagen werden.

<sup>2</sup>Vgl. Schaffrath, H. (1995): *The Essen Folk Song Collection*, hauptsächlich deutsche Volkslieder, 6252 Musikstücke.

berechnen. Nach der Trainingsphase erfolgt eine Evaluation beider Ansätze mittels Melodien aus einem weiteren Musikarchiv<sup>3</sup>. Damit wird versucht, die Töne innerhalb der Melodie „vorherzusagen“, um im Anschluss über eine statistische Auswertung mittels eines aggregierten Wahrscheinlichkeitsmaßes beide Ansätze vergleichen zu können.

In nahezu allen Melodien – unabhängig vom Tonartgeschlecht – machte das Intervallmodell bessere Vorhersagen. Vergleichend dazu wurde ein auf der Tonleiterstufen-Beschreibung basierendes Markovmodell erster Ordnung, welches zusätzlich zu den Auftretenswahrscheinlichkeiten die Übergangswahrscheinlichkeiten zwischen zwei Zuständen berücksichtigt. Dieses Modell war den anderen beiden Modellen bei fast allen Melodien überlegen und zeigt nach Aussage des Autors, dass die Berücksichtigung der Tonhöhenübergänge innerhalb einer Melodie für ihre Beschreibung sehr wichtig ist.

Eerola und North untersuchten in [EN00] die Wahrnehmung melodischer Komplexität in Musikstücken und führen mit dem *Expectancy-Based Model* (EBM) ein neues Modell zur Beschreibung dieses Phänomenes ein. Es beinhaltet mehrere Faktoren, die in die drei Gruppen tonal, intervallbezogen und rhythmisch aufgeteilt sind. Beispielhaft für die erste Gruppe sind die tonale Stabilität, die metrische Position und die Tondauer von Tönen zu nennen – drei Faktoren, mit denen die Betonung einzelner Noten und die damit verbundene verstärkte Wahrnehmung beschrieben werden kann. Die intervallbezogenen Faktoren Nachbarschaft (*proximity*), registrale Umkehr und Richtung (*registral return*, *registral direction*), Abschluss (*closure*), Intervallabstand und Gleichklang bzw. Konsonanz (*consonance*) erlauben eine Beschreibung der Fortsetzung einzelner Töne in Melodien. Die drei rhythmischen Faktoren Schwankung (*rhythmic variability*), Synkopierung und Aktivität (*rhythmic activity*) erhöhen die Komplexität von Melodien, da sie zu einer erschwerten Wahrnehmung führen. Anhand dieser Faktoren werden Melodien aus symbolischen Audiodateien im MIDI-Format untersucht und über eine gewichtete Summe eine Komplexitätsauswertung (*complexity score*) berechnet.

Um die Eignung des Modells zur Vorhersage der Komplexitätswahrnehmung von Melodien zu überprüfen, wurde zum Vergleich ein informationstheoretisches Beschreibungsmodell und ein weiteres Modell, welches auf den Übergangswahrscheinlichkeiten aufeinanderfolgender Töne basiert, zur Beschreibung des gleichen Sachverhaltes herangezogen. In einem Hörtest sollten 59 Probanden jeweils 41 Melodien hinsichtlich ihrer melodischen Komplexität beurteilen, wobei sowohl das Tempo als auch das Timbre des Melodieinstruments konstant gehalten wurde (Panflöte). Über das statistische Verfahren der Mehrfachregressionsanalyse wurde die Korrelation zwischen der Komplexitätswahrnehmung der Probanden und der entsprechenden Vorhersage der einzelnen

---

<sup>3</sup>Vgl. Eerola, T., Toivianen, P. (2004): *Suomen Kansan Sävelmiä - finnish folk song database*, finnische Volkslieder, 8249 Musikstücke.



auf den Faktoren des Modells basierenden Prädiktoren untersucht.

Die fünf Faktoren Tonalität, registrale Richtung, Intervalldifferenz, Synkopierung und rhythmische Variabilität erlaubten die besten Vorhersagen und erreichten die höchsten Korrelationswerte, das informationstheoretische Modell und das Modell basierend auf den Übergangswahrscheinlichkeiten geringere Werte. Abschließend resümieren die Autoren, dass zur generellen Komplexitätsbeschreibung eines Musikstückes neben einzelner Melodien auch andere Faktoren wie Tempo und Fülle des Arrangements zu berücksichtigen sind. Als eine weitere Einsatzmöglichkeit wird der direkte Vergleich zwischen Musikstücken genannt, wobei eine modifizierte Gewichtung der einzelnen Modellfaktoren voraussichtlich zu besseren Ergebnissen führen würde.

In [MW07] versuchen Madsen und Widmer, die Spur innerhalb einer mehrstimmigen MIDI-Datei zu identifizieren, die die Hauptmelodie des Stückes enthält. Ihre Annahme besteht darin, dass ein hohes Maß an Komplexität innerhalb einer Melodielinie eine hohe Wahrscheinlichkeit bedingt, dass diese als Hauptmelodie wahrgenommen wird. Die Komplexitätsberechnung erfolgt hierbei nach einem informationstheoretischen Ansatz über die Bestimmung der Entropie, die als umso größer angenommen wird, je komplexer und weniger vorhersagbar der Verlauf einer Melodie ist. Diese wird für die drei Parameter Tonhöhe, Interval und Inter-Onset-Intervall<sup>4</sup> berechnet. Aus diesen drei Entropien wurden durch gewichtete Addition zweier kombinierter Entropien und über die Untersuchung der Verbundwahrscheinlichkeiten zweier Parameter zwei verschiedene Verbundentropien abgeleitet. Polyphone Melodiefragmente (z.B. Akkorde) wurden vereinfacht durch die jeweils höchste gespielte Note repräsentiert, außerdem erfolgte eine zeitliche Quantisierung. Zur Evaluierung wurden drei Vorhersagemodelle benutzt, zwei davon leiten sich von der entropiebasierten Komplexitätsberechnung ab. Das erste Modell ermittelt über gefensterte Ausschnitte von 6 s bis 15 s (Schrittweite : 0,2 s pro Fensterposition) eine Gewinnerspur mit der höchsten Entropie. Die globale Gewinnerspur wird über Akkumulation der Einzelergebnisse ermittelt. Im zweiten Modell erfolgt eine globale Entropieberechnung für jede Spur über die gesamte Liedlänge. Das dritte Modell basiert auf einem anderen Ansatz. Werden Melodien als Zeichenketten dargestellt, so kann über ihre Komprimierbarkeit auf ihr Maß an Komplexität geschlossen werden – einfache, sich wiederholende und damit weniger komplexe Melodien lassen sich aufgrund der höheren Redundanz stärker komprimieren. Motivation für diesen Ansatz ist, dass das Modell der Entropie sich ausschließlich auf die Auftretenswahrscheinlichkeiten einzelner Zeichen, nicht jedoch auf ihre Reihenfolge bezieht. Diese hat für Noten natürlich einen großen Einfluss auf die Melodie, die sie bilden. Hierbei wird

---

<sup>4</sup>Zeitlicher Abstand zwischen den Anfangszeitpunkten zweier aufeinanderfolgender Noten.

mittels des LZW-Kompressionsalgorithmus<sup>5</sup> eine Kompressionsrate bestimmt, die die Länge der Zeichenkette vor und nach der Kompression ins Verhältnis setzt.

Die beiden untersuchten Testsets bestanden aus 79 bzw. 29 Musikstücken aus dem Bereich Pop & Rock, für die jeweils die Melodiespur bestimmt wurde. Die beiden entropiebasierten Ansätze erreichten eine maximale Erkennungsquote von 50,6 % bzw. 62,1 % für die beiden Testsets und schlugen damit das kompressionsbasierte Modell mit 43,0 % bzw. 51,7 %. Die Rhythmus- bzw. Timing-basierten Merkmale, die von der Entropie der Inter-Onset-Intervalle abgeleitet waren, erreichten die höchsten Erkennungsquoten, woraus die Autoren eine hohe Korrelation zwischen rhythmischer Komplexität und melodischer Wahrnehmung schließen. Problematisch für das Modell war das Auftreten von Soli – die ein hohes Maß an Komplexität erzeugten, auch wenn das Soloinstrument nicht die Hauptmelodie des Stückes spielt.

Léon und Iñesta stellen in [LI07] eine Reihe von *High-level*-Merkmale dar, die typische Eigenschaften von Melodien in Jazz- und Klassikstücken abbilden und eine spätere Genreklassifikation ermöglichen sollen. Die Merkmale basieren auf beschreibenden statistischen Eigenschaften von Tonhöhen, Intervallen, Notenstartzeitpunkten, Notenlängen und Pausen. Die allgemeinen Deskriptoren beinhalten elementare Merkmale, wie Anzahl der gespielten Noten, Länge bzw. Anteil von Pausen an der Gesamtlänge des untersuchten Bereich. Weiterhin werden der Tonhöhenumfang, mittlere Tonhöhe sowie ihre Standardabweichung in Halbtönen berechnet. Analog gehen die Autoren für die beiden musikalischen Parameter Notenanzfangszeitpunkt und Inter-Onset-Intervall vor, die beide jeweils nach einer vorherigen zeitlichen Quantisierung berechnet werden. Über eine stark vereinfachte Annahme (die gerade für Jazz-Stücke nur selten zutrifft), dass jedem untersuchten Stück eine konstante Tonart zugrunde liegt, werden alle gespielten Noten in diatonische (in der Tonart enthaltene) und nichtdiatonische Noten aufgeteilt. Entsprechend des Auftretens der nichtdiatonischen Noten werden ebenfalls statistische Deskriptoren abgeleitet. Die Autoren führen die Nutzung der D’Agostino-Maßzahl (vgl. Kap. 2.2) zur Beschreibung der Werteverteilung der untersuchten musikalischen Parameter ein, dieses Maß wird in der vorliegenden Arbeit übernommen.

Die Grundidee, gleiche statistische Kenngrößen aus unterschiedlichen musikalischen Parametern zu berechnen, um gespielte Melodien semantisch möglichst umfassend beschreiben zu können, wird wie in Kap. 4.2 dargestellt in dieser Arbeit übernommen und weiterentwickelt.

---

<sup>5</sup>Ziv, J., Lempel, A.(1977): *A universal algorithm for sequential data compression*. IEEE Transactions on Information Theory 23(3) S. 337343.

## Rhythmische High-level-Merkmale

Über ein sogenanntes *beat-histogram*<sup>6</sup>, das auf der Autokorrelationsfunktion eines aus den Notenanfangszeitpunkten gewonnenen Impulssignales beruht, untersucht McKay in [McK04] auftretende rhythmische Periodizitäten. Anhand der gefundenen Maxima kann das Tempo, sowie ein Maß für rhythmische Schwankung bzw. Ungenauigkeit ermittelt werden. Weiterhin nutzt McKay die Notendichte pro Takt, die Längen und Häufigkeiten von Pausen, das Anfangstempo, sowie die Taktart zu Beginn des Stückes als weitere rhythmische Merkmale.

Brosbol und Schubert stellen in [BS06] mit dem *articulation index* ein Maß für die rhythmische Artikulation, also für die Dauer einzelner gespielter Noten in Bezug auf ihren zeitlichen Abstand dar. Diese wird in der Musikwissenschaft oftmals durch die beiden Begriffe *staccato* und *legato* beschrieben. Die Untersuchung erfolgt anhand ein-kanaliger monophoner Audiosignale. Über eine Spektrogrammberechnung mit einer Abtastfrequenz von knapp 11 kHz sollen möglichst alle Fundamentalfrequenzen<sup>7</sup>, sowie das harmonische Spektrum der gespielten Töne erfasst werden.

Mittels des statistischen Verfahrens der *Probabilistic Latent Semantic Analysis* (PLSA)<sup>8</sup> wird in einem iterativen Prozess das Betragsspektrogramm  $S(f, t)$  in die Einzelspektren  $W(f)$  bzw. in die korrespondierenden Intensitätshüllkurven  $H(t)$  im Zeitbereich aufgespaltet. Nach einer On- und Offsetdetektion der gespielten Noten wird eine *note map*  $N(t)$  erstellt, anhand derer über das Verhältnis zwischen Notenlänge und Abstand zweier Noten ein über das gesamte Stück gemittelter Wert für den articulation index berechnet wird.

In [GFB03] untersuchen Gouyon et. al. die bei der Aufnahme von Musikstücken auftretenden rhythmischen Abweichungen gespielter Noten von der eigentlichen Partitur am Beispiel des Swings. Wie in Kap. 2.1.2 bereits erläutert, handelt es sich um genretypische, meist regelmäßig an der gleichen metrischen Position auftretende Abweichungen (innerhalb des Achtelnotenrasters).

Der Begriff *swing ratio* beschreibt das Verhältnis der Länge einer Achtelnote bezogen auf die Länge ihres Vorgängers. Dieser Wert ist stark tempoabhängig, er reicht von einem Verhältnis von 2 : 1 für langsame Tempi, das einem Triolenfeeling (vgl. Abb. 2.8(b)) entspricht, bis hin zu einem Wert von annähernd 1,2 : 1 für hohe Tempi, bei denen die charakteristische Verschiebung kaum mehr wahrgenommen werden kann.

Um den *swing ratio* für ein gegebenes Musikstück zu berechnen, erfolgt zuerst eine

<sup>6</sup>Vgl. Brown, J. C. (1993) : *Determination of meter of musical scores by autocorrelation*, Journal of the Acoustical Society of America 94 (4): 1953-7.

<sup>7</sup>Grundfrequenzen, vgl. Kap. 2.1.1.

<sup>8</sup>Vgl. Hoffman, T. (2001), *Unsupervised Learning by Probabilistic Latent Semantic Analysis*, Machine Learning, vol. 42, pp. 177-196.

Onset-Erkennung über die Untersuchung der Energie innerhalb 11 ms langer Zeitfenster. Anschließend wird aus den erkannten Onset-Zeitpunkten alle Inter-Onset-Intervalle – also die zeitlichen Abständen zwischen jeweils zwei verschiedenen Noten – berechnet. Über die Berechnung eines Histogramms, welches zusätzlich mit einem Gaußfilter geglättet wird, erfolgt in weiteren Verarbeitungsschritten die Bestimmung des *swing ratio*.

Herrara et. al. stellen in [HSG04] drei semantische Deskriptoren vor, die die Instrumentierung eines Musikstückes hinsichtlich Perkussionsinstrumenten beschreiben. Es werden reale Audiodaten analysiert. Nach einer instrumentenunabhängigen Onset-Detektion und anschließender Hochpass-Filterung wird für jeden erkannten Onset das direkt anschließende 100 ms lange Zeitsegment untersucht. Über die Berechnung verschiedener *Low-level*-Merkmale, wie z.B. *mel-frequency spectral coefficients* und *spectral centroid* und der anschließenden Auswertung wird für jedes Segment entschieden, ob die gespielte Note einem Perkussionsinstrument zuzuordnen ist oder nicht. Über den Anteil aller perkussionsbezogenen Elemente zur Gesamtzahl aller gefundenen Onsets wird der *percussion index* berechnet.

Eine feinere Unterteilung bezüglich der Art des auftretenden Schlaginstruments wird durch das sogenannte *percussion profile* ermöglicht. Anschließend erfolgt eine hierarchische Klassifizierung – zuerst in harmonische und rhythmische Anschläge, anschließend die rhythmischen Anschläge in die von Membranophonen und Schlagidiophone (z.B. Becken) bis letztlich zur Unterteilung von Bass-Drum und Snare. Somit kann das *percussion profile* die Häufigkeit einzelner Schlaginstrumente abbilden und ermöglicht eine detailliertere klangliche Beschreibung eines Musikstückes.

Die Häufigkeit von Wechseln zwischen Bass-Drum- und Snare-Drum-Schlägen, die häufig charakteristisch für einzelne Genres sind, wird im *Kick-snare*-Deskriptor abgebildet.

Die insgesamt wahrgenommene Stärke der Perkussivität wird durch den *Percussivity*-Deskriptor beschrieben, der sich an einem Algorithmus zur Berechnung der wahrgenommenen Stärke von Umgebungsgeräuschen orientiert.

Ein Ansatz, der verwandt mit dem in Kap. 4.2.3 vorgestellten rhythmischen Strukturprofil ist, wird in von Lidy et. al. in [LPR05] dargestellt. Ihr Ziel ist es, die rhythmische Repräsentation eines Musikstückes nicht ausschließlich anhand des gespielten Rhythmus, sondern anhand von Fluktuationen innerhalb verschiedener Frequenzbänder im auditiv wahrnehmbaren Frequenzbereich. Die von ihnen eingeführten *rhythm patterns* beschreiben die entsprechenden Lautheitsschwankungen.

Nach einer Kurzzeit-Fourriertransformation erfolgt eine anschließende Aufsummierung der Energien innerhalb der 24 Bänder der Bark-Skala. Um die spezifische, durch den Menschen wahrgenommene Lautstärke abbilden zu können, erfolgen weitere Trans-

formationen in dB, phon und sone. Über eine erneute Fourier-Transformation wird eine zeitunabhängige Darstellung berechnet. Dies ermöglicht für verschiedene Modulationsfrequenzen von  $f = 0,2$  Hz bis  $f = 10$  Hz und den dazu korrespondierenden Tempobereich von 12 BPM bis 600 BPM eine Angabe über die Stärke der entsprechenden periodischen Modulation in diesem Frequenzband.

Im Gegensatz zu den 60 statisch festgelegten Modulationsfrequenzen ermöglicht das in der vorliegenden Arbeit vorgestellte rhythmische Strukturprofil eine tempoadaptive Darstellung. Das Verfahren, sowie seine Vorteile werden in Kap. 4.2.3 beschrieben.

Neben den bereits genannten *rhythm patterns* setzen Lidy et. al. wie in [LRPI07] beschrieben weitere rhythmische *High-level*-Merkmale im Kontext der Genreklassifikation ein. Über eine Aggregation des *rhythm pattern* innerhalb jedes Frequenzbandes der Bark-Skala wird eine rhythmische Beschreibung mit deutlich weniger Dimensionen abgeleitet, die Berechnung erfolgt dabei über Zeitabschnitte von 6 s Länge. Über eine Histogrammbildung wird der *Rhythm-histogram*-Deskriptor berechnet. Dieser ermöglicht eine gemittelte rhythmische Beschreibung des gesamten Musikstückes.

Neben dieser Histogrammbildung werden sieben grundlegende statistische Kenngrößen wie Mittelwert, Median und Varianz des Amplitudenwertes für jedes kritische Band und über das gesamte Stück berechnet. Diese Größen werden als *statistical spectrum descriptors* bezeichnet.

Außerdem erfolgt die Auswertung einer *onset detection function*, die einem Maß für die Länge der Anstiegszeiten (*attack time*) einzelner Noten entspricht und für kurze und perkussive Töne einen höheren Wert annimmt als beispielsweise für gestrichene Geigentöne. Sowohl für die *onset detection function* als auch für die *Inter-onset*-Intervalle werden jeweils wieder Minimum, Maximum, Mittelwert, Median sowie Standardabweichung berechnet und als Merkmale verwendet.

Einen weiteren Ansatz, verschiedene Genres der Tanzmusik anhand elementarer wiederkehrender rhythmischer *patterns* zu unterscheiden, stellen Dixon et. al. in [DGW04] vor. Grundlage der Untersuchung sind knapp 700 Audiodateien aus acht Genres wie u.a. Cha Cha, Jive, Quickstep und Walzer. Nach einer automatisierten Detektion von Taktlänge und Taktdauer wird mit Hilfe der jeweiligen Amplitudenhüllkurven pro Takt ein rhythmisches *pattern* berechnet. Die Autoren verfolgen hierbei den vereinfachten Ansatz, dass Pegelspitzen der Hüllkurve *Onset*-Zeitpunkten der Rhythmusinstrumenten entsprechen, da die untersuchte Musik generell sehr rhythmusdominant ist. Um die am häufigsten auftretenden (und damit repräsentativen) drum patterns zu detektieren, werden mit Hilfe eines *K-Means-Cluster* Algorithmus „Ausreißer“, die nur selten vorkommen, ausgefiltert. Als Distanzmaß wird hierbei die euklidische Distanz verwendet. Neben den drum patterns werden weitere daraus abgeleitete Merkmale eingesetzt, hier sind u.a. Amplitudenmaximum und -mittelwert, das geschätzte Tempo, Synkopie-

rungsmaß und *swing factor* zu nennen. Gouyon versucht das in [DGW04] beschriebene Verfahren zu nutzen, um den typischen Rhythmus von „Samba de Roda“<sup>9</sup> zu extrahieren – die Datenbasis setzt sich aus 49 Stücken mit je 10 bis 30 s Länge zusammen. Nach einer *Onset*-Detektion werden alle Daten mittels *resampling* auf eine einheitliche Länge transformiert um eine Vergleichsbasis zu schaffen. Außerdem erfolgt eine Amplitudennormierung. Mit Hilfe des *K-Means-Clustering* konnten drei häufig vorkommende *drum patterns* extrahiert werden, die sich vor allem in der Betonungsstärke der einzelnen Zählzeiten unterscheiden. Die gewonnenen *drum patterns* ermöglichen zusätzlich die Untersuchung feiner zeitlicher Abweichungen (*micro-time*) um jede Zählzeit. Diese sind für den typischen *groove* der entsprechenden Stilistik verantwortlich.

## 3.2 Improvisation und Interaktion

Hidaka et. al. stellen in [HGM95] ein automatisches Begleitsystem für Solisten vor, welches die Intention des improvisierenden Musikers erkennen und durch eine angepasste musikalische Begleitung darauf reagieren soll. Das Programm ist für Jazz-Standards (deren harmonischer Ablauf dem System bekannt ist) im 4/4-Takt ausgelegt und verarbeitet symbolische MIDI-Daten, d.h. das Signal des Solisten muss per MIDI-Ausgang (Keyboard) oder MIDI-Tonabnehmer<sup>10</sup> (Gitarre) vor der Verarbeitung umgewandelt und dem Programm zugeführt werden. Anschließend wird die musikalische Begleitung in Form der Bass- und Schlagzeugspur generiert.

Aus dem Eingangssignal werden acht *musical primitives* bestimmt – elementare musikalische Parameter wie z.B. *louder note* (hohe Lautstärke der Noten), *many notes* (hohe Notendichte), *chord note* (Note im aktuellen Akkord enthalten) oder *tension note* (akkordfremde Note). Diese werden anschließend über unterschiedliche Gewichtungen zu fünf *intention parameters* verknüpft, welche eine Aussage über die momentane Intention des Solisten ermöglichen sollen. Durch *excitement* (Aufregung) werden starke Betonungen und überraschende Wendungen im Spiel repräsentiert, *tension* (Spannung) zeigt eine Abweichung vom harmonischen Kontext durch die Verwendung vieler akkordfremder Noten, *emphasis on chord* (Akkordbetonung) stellt im Gegensatz dazu die Unterstützung der Akkordentwicklung durch den Solisten dar, *chord substitution* (Akkordersetzung) beschreibt das häufige Verwenden von Erweiterungstönen und durch *theme reprise* wird das Zitieren oder Wiederholen des Themas des aktuellen Stückes beschrieben.

Ziel ist es, möglichst in Echtzeit auf die Intention des Solisten zu reagieren anstatt nur

---

<sup>9</sup>Samba-Stil, Rundtanz.

<sup>10</sup>Tonabnehmer für Saiteninstrumente zur direkten Umwandlung des akustischen Signales in ein MIDI-Steuersignal.

zwischen programmierten *patterns* – also typischen Phrasen der jeweiligen Begleitinstrumente – hin- und herzuschalten. Jeder *intension parameter* ist im System mit einigen typische Modifikationen der Begleitspuren verknüpft (*alteration of accompaniment*). So bewirkt ein hoher Wert für *excitement* beispielsweise eine Lautstärkerhöhung des Basses, sowie mehr Noten und das verstärkte Einsetzen von Becken in der Schlagzeugspur. Das System wurde sowohl von Pianisten als auch Gitarristen getestet und konnte nahezu in Echtzeit auf ihr Spiel reagieren.

Baggi beschreibt in [Bag90] und [Bag91] mit *Neurswing* ebenfalls ein System zur „Untersuchung von Swing in afroamerikanischen Jazz durch Simulation einer Rhythmusgruppe“ ([Bag91], S. 1). Der Autor weist einleitend auf die hohe Bedeutung des Swings in der improvisierten Jazzmusik hin, deren Natur – ganz im Gegensatz zu klassischer Musik – nicht in der Aufführung fertig komponierter Stücke besteht, sondern vielmehr in der Improvisation der Musiker über vorgegebene harmonische Abläufe, die dazu führt, dass keine Aufführung der anderen gleicht. Diese rhythmischen und melodischen Variationen lassen sich nicht vollständig in einer Partitur erfassen, sondern müssen anhand von Aufnahmen verschiedener Auftritte analysiert werden.<sup>11</sup>

Jazzstücke liegen meist in Form eines harmonischen Ablaufes (*harmonic grid*) vor, der die Orientierung für das Harmonie- und Bassinstrument bildet und der im Laufe des Stückes durch Substitutionen und Alterierungen<sup>12</sup> variiert wird. Dieser Ablauf wird dem System zur Verfügung gestellt, woraus es das sogenannte *harmonic net* ableitet. Dies entspricht einer netzartigen Ablaufstruktur, die neben dem eigentlich vorgegeben Ablauf verschiedene alternative „harmonische Wege“ beinhaltet, die durch Variation der bestehenden Akkordfolgen entstehen. Diese Substitutions- und Alterierungsoptionen basieren auf musiktheoretischem Wissen über harmonische Zusammenhänge des Jazz, das dem System in algorithmischer Form zur Verfügung gestellt wird.

Die Struktur ist kein selbstlernendes Netz, bildet jedoch die Parallelität mehrerer rhythmischer, melodischer und harmonischer Vorgänge innerhalb des Prozesses der Jazzimprovisation ab. Die zweite Komponente des Systems ist das *stylistic net*. Es schafft eine Verbindung zwischen drei Eingangsreglern, mit denen der Nutzer Einfluss auf die Stilistik der harmonischen Begleitung (entsprechend verschiedener Epochen der Jazzmusik im frühen 20. Jahrhundert) und den Auftretenswahrscheinlichkeiten, die die Auswahl konkreter Pfade im *harmonic net* bestimmen. Diese Eingriffspunkte (*knobs*) beeinflussen u.a. die Art der angewandten harmonischen Substitution, die Stilistik der Bassbegleitung und typische *rhythm patterns* der Schlagzeugbegleitung. Das *stylistic net* verändert die Akzentuierung durch die Rhythmusgruppe und das Harmonieinstrument über die Dynamik der gespielten Noten. Die Ausgabe des System liegt in Form

<sup>11</sup> „That’s why the notation for Jazz is not the score, but the record“ [Bag90].

<sup>12</sup>Meist genretypische Akkordersetzungen und -variationen.

von MIDI-Steuerdaten vor, die Verarbeitung erfolgt in Echtzeit.

Ein anderer Ansatz wird von Erkkilä et. al. in [ELL<sup>+</sup>04] beschrieben. Im Rahmen des finnischen Forschungsprojektes „*Intelligent music systems in music therapy*“ entwickelten sie Methoden zur Untersuchung von musikalischer Improvisation im Rahmen der klinischen Musiktherapie. Ziel war es, unter Anwendung von Wissen über die Wahrnehmung von Rhythmus, Melodie, Harmonie und Tonalität eine Qualitätsbeurteilung von aufgenommener Musik zu entwickeln, um damit die Qualität der Therapie zu verbessern. Die Improvisation erfolgt dabei zwischen Therapeut und Patienten.

Die von ihnen entwickelte *Music Therapy Toolbox* (MTTB) für MATLAB ermöglicht die Auswertung im MIDI-Format aufgenommener Musikstücke anhand verschiedener musikalischer Parameter wie beispielsweise Notendichte, Dynamik oder Definiertheit von Rhythmus (Puls) und Tonalität. Für die Darstellung des musikalischen Dialoges bzw. der Interaktion zwischen den beteiligten Musikern wurde das *Immitation Diagram* (ID) entwickelt. In einem Versuch wurde die gemeinsame Improvisation zweier Musikstudenten mit Hilfe von MIDI-Daten aufgezeichnet und anschließend ausgewertet. Hierbei sollten die Musiker versuchen, die von ihnen gespielte Musik innerhalb der beiden Wahrnehmungsebenen Aktivität (*activity*) und Intensität (*valence*) verbal zu beschreiben. Über das statistische Verfahren der linearen Regressionsanalyse wurde untersucht, welche aus den Musikdaten berechneten *High-level*-Merkmale (wie z.B. Notendichte, statistische Merkmale bezüglich Tonhöhe und Tondauer und rhythmische Definiertheit) am stärksten mit den beiden Wahrnehmungsebenen korrelieren. Hierbei stellte sich heraus, dass die momentane Notendichte, die mittlere Lautstärke sowie das Minimum der Tonhöhe die besten Vorhersagen bezüglich Aktivität und Intensität ermöglichen.

Mántaras und Arcos geben in [MA02] einen sehr umfangreichen Überblick über weitere wissenschaftliche Veröffentlichungen zum Thema computerbasierte Systemen zur Komposition, Improvisation und Performanceanalyse von Musik.

### 3.3 Genreklassifikation

Ein Großteil der wissenschaftlichen Veröffentlichungen, in der die Genreklassifikation von Musikstücken behandelt wird, basiert auf der Extraktion und Verarbeitung von *Low*- und *Mid-level*-Merkmalen. Im Folgenden werden jedoch ausschließlich Ansätze beschrieben, die sich dabei auf die Verwendung von *High-level*-Merkmalen stützen.

Für die Evaluation des von McKay implementierten Systems ([McK04]) kam eine zweistufige Genretaxonomie mit den Hauptgenres Jazz, Populärmusik und Klassik und den Subgenres Bebop, Soul, Swing bzw. Rap, Punk, Country sowie Barock, Moderne Klas-



sik und Romantik zum Einsatz. Der Datensatz für Training und Evaluation der 111 Merkmale bestand aus insgesamt 950 MIDI-Dateien aus allen untersuchten Genres. Für alle eindimensionalen Merkmale wurde ein KNN-Klassifikator ( $k$  nächste Nachbarn, *k-Nearest-Neighbor*), für alle mehrdimensionalen Merkmale ein neuronales *Feed-forward*-Netzwerk eingesetzt. Mit Hilfe von genetischen Algorithmen wurde für jeden Klassifikator eine Gewichtung der Merkmale durchgeführt. Anschließend erfolgte nach dem gleichen Prinzip eine Gewichtung der Entscheidungen aller Genreklassifikatoren, um eine Ensemble-Entscheidung abgeben zu können. Mit einer flachen Klassifizierung wurde eine Genauigkeit von 96 % für die Hauptgenres, sowie 86 % für die Subgenres erreicht. Mit Hilfe der hierarchischen (zweistufigen) Klassifizierung verbesserten sich die beiden Ergebnisse auf 98 % bzw. 90 %.

Dixon et. al. erreichten in [DGW04] mit den in Kap. 3.1 beschriebenen Merkmalen, bei Verwendung von acht Genres, eine genrebezogene Klassifizierungsgenauigkeit bis zu 96 %. Die Autoren weisen darauf hin, dass ausschließlich nichtspektrale Merkmale zum Einsatz kommen. Im Rahmen dieses Papers wurden verschiedene überwachte Klassifikationsalgorithmen wie z.B. KNN, *decision-tree* (Entscheidungsbaum) oder *AdaBoost* (Meta-Lern-Verfahren) untersucht, wobei die besten Ergebnisse mit dem letztgenannten Klassifikator erzielt wurden.

Die Eignung der statistischen Deskriptoren, die Léon und Iñesta in [LI07] beschrieben, wurde mit Hilfe eines Bayes-Klassifikators, eines KNN-Klassifikators und eines SOM-Klassifikators (selbstorganisierende Karte - *self organizing map*) für den Einsatz in der Genreklassifikation zwischen Jazz- und Klassikstücken evaluiert. Die Merkmalsextraktion erfolgte nicht global über das gesamte Stück, sondern über zeitliche Frames mit der Länge  $\omega$  und der Schrittweite  $\sigma$ . Die durchgeführten Experimente zeigten, dass die erreichten Klassifikationsergebnisse für verschiedene Wertepaare  $(\omega, \sigma)$  größtenteils konstant blieben. Für jedes Merkmal wurde ein Maß berechnet, welches die Separierbarkeit der beiden Genres anhand dieses Merkmals abbildet. Anhand dieser Maßzahl konnte eine Auswahl der höchstdiskriminierenden Merkmale erfolgen, mit der die Genreklassifikation durchgeführt werden konnte. Die besten Ergebnisse wurden mit dem Bayes- (bester durchschnittlicher Wert 91 % für 12 Merkmale) und dem KNN-Klassifikator (bester durchschnittlicher Wert 93 % für 28 Merkmale) erreicht.

Neben den in [LI07] beschriebenen Merkmalen entwickelten Pérez-Sancho in [PSLI06] einen weiteren Ansatz zur Genreerkennung. Die gespielten Melodien innerhalb des Stückes wurden mit Hilfe der Intervalle und relativen Zeitdauern zweier aufeinanderfolgender Noten in Zeichenketten (sogenannte *n-words*) umgewandelt. Jedes Melodiefragment, dass von zwei Pausen umgeben, jedoch höchstens vier Takte lang war, wurde in eine Zeichenkette konvertiert. Entsprechend ihrer Häufigkeit erfolgte eine Rangbildung der gewonnenen Zeichenketten und die Merkmalsextraktion wurde nur anhand der

Zeichenketten mit den höchsten Rangwerten durchgeführt. Neben Bayes- und KNN-Klassifikator kommen zusätzlich die in der vorliegenden Arbeit ebenfalls verwendeten SVM-Klassifikatoren sowie die *Multilayer-Perceptron*-Klassifikatoren zum Einsatz. Die Genreklassifikation erfolgt anhand einer Ensemble-Entscheidung (vgl. auch Kap. 5.1.2). Das beste Ergebniss von knapp 97 % Klassifikationsgenauigkeit wurde für das *quadratic best-worst wighted majority* Abstimmungsprinzip zwischen den Ergebnissen der Einzelklassifikatoren erzielt. Hierbei werden den Klassifikatoren entsprechend den von ihnen erzielten Klassifikationsgenauigkeiten Autoritäten zwischen 0 und 1 zugeordnet, wobei 1 dem besten, 0 dem schlechtesten und dazwischen liegende Werte den übrigen Klassifikatoren zugewiesen wird. Über eine zusätzliche Quadrierung erfolgt eine stärkere Wichtung besserer Klassifikationsergebnisse.

### 3.4 Musikerklassifikation

In [Ell07] beschreibt Ellis ein Verfahren zur Musikerklassifikation basierend auf der komplementären Nutzung von MFCC-Merkmalen und chromagrammbasierten Merkmalen, welche den melodischen und harmonischen Inhalt eines Musikstückes unabhängig von der Instrumentierung repräsentieren. Im sogenannten *Beat-chroma*-Merkmal werden je ein Chromavektor (vgl. Kap. 4.2.2) über den zeitlichen Abstand zweier benachbarter Hauptzählzeiten berechnet. Ziel dabei ist es, Folgen von Chromavektoren zu modellieren, um häufig eingesetzt Akkordfolgen in den Musikstücken eines Künstlers abzubilden. Hierbei wird ein sogenanntes *key-normalization frond end* eingesetzt, welches versucht, über eine Rotation der Chromavektoren den Einfluss von Transposition auszugleichen.

Mit Hilfe von GMM-Klassifikatoren sowie eines Testsets bestehend aus 1412 Musikstücken von 20 verschiedenen Künstler wurde eine Musikerklassifikation durchgeführt. Die Kombination des timbralen Modells (MFCC) und des chromagrammbasierten Modells erreichte eine mittlere Klassifikationsgenauigkeit von 59 %.

Saunders et. al. entwickelten, wie in [SHSTW04] beschrieben, ein Verfahren zur Erkennung verschiedener Interpreten klassischer Klavierstücke<sup>13</sup> anhand ihrer Spielweise. Mit Hilfe des von ihnen entwickelten Analysesystems *performance worm* werden Tempo- und Lautstärkeverlauf in realen Audiodaten analysiert und der Verlauf als Trajektorie<sup>14</sup> über viele kurze Zeitfenster hinweg dargestellt. Über eine Clusterung gleichlanger Abschnitte werden die am häufigsten gespielten Muster als sogenannte Prototypen erkannt und in ein *Performance*-Alphabet übernommen. Dies beinhaltet

---

<sup>13</sup>Sechs Sonaten von W. A. Mozart mit unterschiedlicher Takt- und Tonart.

<sup>14</sup>Lat.: Bahnkurve.

Informationen über wiederkehrende Artikulationsverläufe, die ein bestimmter Interpret häufig einsetzt.

Anschließend kann die gesamte Entwicklung eines Stückes bezogen auf Tempo und Lautstärke in eine Folge von Zeichen dieses Alphabetes umgewandelt werden. Für die Erkennung der Interpreten anhand der extrahierten Zeichenkette wurden verschiedene Kernel-Verfahren wie z.B. *string kernel* und *n-gram kernel* eingesetzt und verglichen. Bei der Evaluation wurden jeweils zwei Pianisten gegeneinander getestet und anschließend die Ergebnisse über alle Kombinationen der sechs untersuchten Pianisten gemittelt. Für eine Kombination der *string kernel* mit anschließender SVM-Klassifikation (vgl. Kap. 2.3.3) wurde die höchste Erkennungsquote von knapp 82 % erreicht.



## 4 Konzeption und Implementierung

### 4.1 Vorverarbeitung

#### 4.1.1 Konvertierung und Transkription

In Abb. 4.1 sind alle Verarbeitungsschritte dargestellt, die vor der Merkmalsextraktion durchlaufen werden. Ausgangspunkt sind die zu untersuchenden Audiodaten. Liegen diese in symbolischer Form (im MIDI-Format) vor, so können über eine Datenkonvertierung mit Hilfe der *MIDI-Toolbox* ([ET04]) die Partiturnformationen (vgl. Kap. 2.4) direkt aus der Datei gewonnen werden. Die Bestimmung der genauen zeitlichen Position des Solos innerhalb des Musikstückes, sowie die Zuordnung der MIDI-Kanäle zu den einzelnen Instrumenten des Instrumentierungsmodelles (siehe 2.1.4) erfolgt manuell und wird für jedes Solo zum späteren Auslesen in einer XML<sup>1</sup>-Datei gespeichert. Anschließend werden die zur Beschreibung des Soloparts relevanten Anteile der Partiturnformation extrahiert.

Liegen die Audiodaten in subsymbolischer Form als WAV- oder MP3-Dateien vor, so erfolgt nach einer manuellen zeitlichen Detektion des Soloparts und Ausschneiden des entsprechenden Ausschnittes die automatisierte Transkription mit Hilfe des am Fraunhofer IDMT entwickelten *Transcription Toolbox* Systems ([DDR07], [DA08]). Es ermöglicht eine getrennte Transkription des Melodie-, des Harmonie-, des Bass- und des Schlaginstruments sowie falls nötig eine anschließende manuelle Korrektur der Transkriptionsergebnisse.

#### 4.1.2 Zeitliche Projektion

Neben der Transkription erfolgt im *Transcription Toolbox* System eine rhythmische Analyse der Audiodaten, in der die zeitlichen Positionen  $\mathbf{t}_B$  und Phasen  $\alpha_B$  der Hauptzählzeiten<sup>2</sup> detektiert werden. Die Phase  $\alpha_B$  nimmt dabei den Wert 1 für die erste Zählzeit eines beliebigen Taktes und den Wert 0 für alle anderen Zählzeiten an. Diese Darstellung ermöglicht eine sowohl tempo- als auch taktartadaptive Repräsentation der rhythmischen Struktur eines Musikstückes. Das untersuchte Solo besteht aus  $N_N$  Noten, die das Melodieinstrument innerhalb von  $N_B$  Takten spielt. Die Startzeitpunkte der detektierten Noten werden nach der Transkription im Vektor  $\mathbf{t}_N$  gespeichert.

---

<sup>1</sup>Extensible Markup Language.

<sup>2</sup>*beat-grid*.

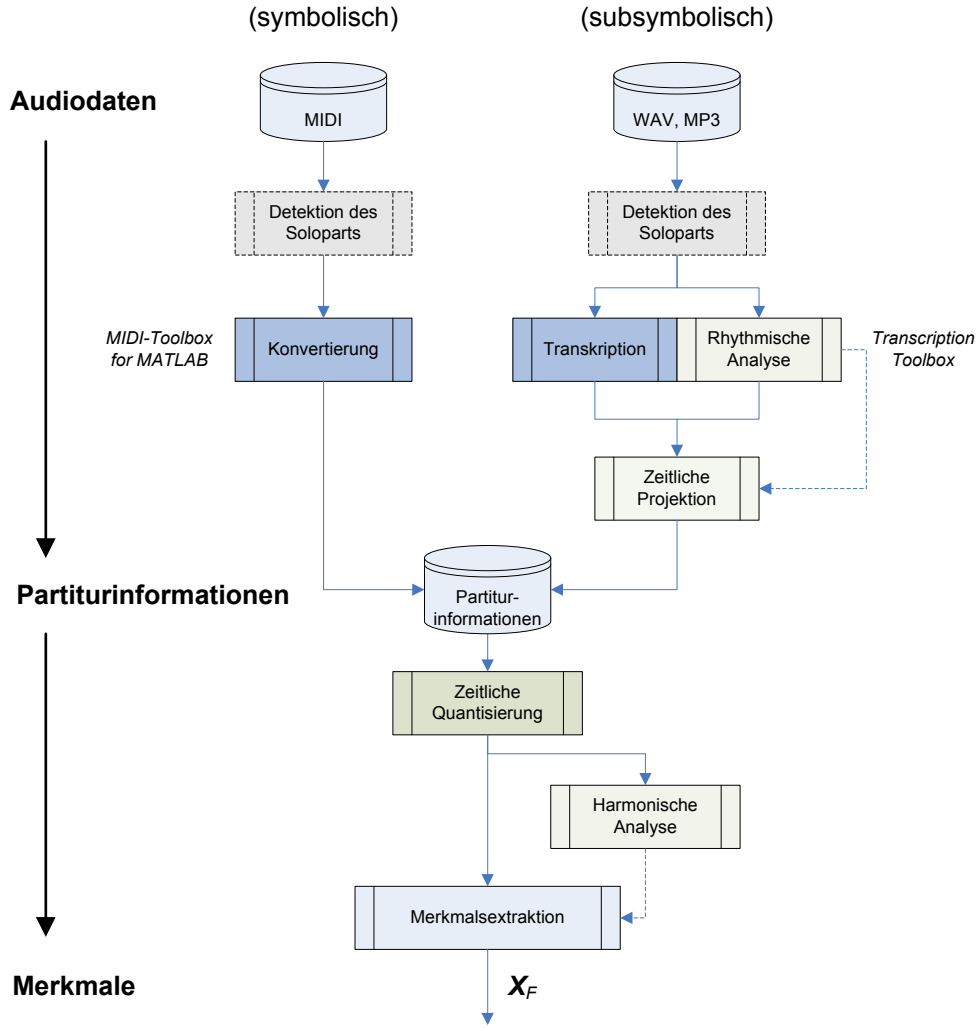


Abb. 4.1: Allgemeiner Ablauf der Analyse

Der Zeitpunkt einer jeden Zählzeit kann alternativ in einer taktlängenbezogene Repräsentation  $\tau_B(i)$  angegeben werden. Für die  $i$ -te Zählzeit gilt dabei

$$\tau_B(i) = \frac{i - i_1 + B_N}{B_N}; \quad 1 \leq i \leq N_B \cdot B_N. \quad (4.1)$$

Der Index  $i_1$  entspricht dem Index der ersten Zählzeit mit  $\alpha_B(i_1) = 1$ .  $B_N$  entspricht dem Zähler der Taktmensur. Die zeitliche Projektion ist in Abb. 4.2 anhand eines Beispiels dargestellt. Anschließend werden die Anfangszeiten aller transkribierten Noten  $t_N$  (in Sekunden) in eine taktlängenbezogene Repräsentation  $\tau_N$  umgerechnet. Hierzu wird als Näherung vorausgesetzt, dass das Tempo innerhalb des zeitlichen Abstands zweier benachbarter Zählzeiten konstant ist. Weiterhin wird gefordert, dass für jede detektierte Note die Position und Phase mindestens jeweils einer Zählzeit bekannt ist, die zeitlich

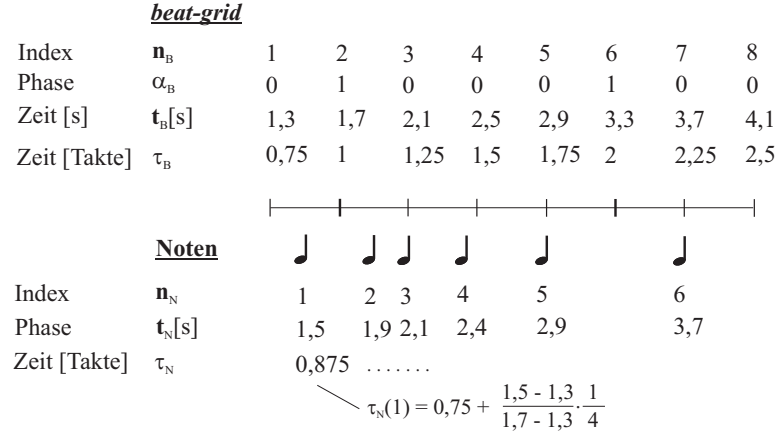


Abb. 4.2: Zeitliche Projektion (Beispiel)

vor bzw. hinter der Note liegt. Sind die aus der rhythmischen Analyse gewonnenen Informationen dafür nicht ausreichend, so werden über eine Interpolation die zeitlichen Positionen und Phasen der fehlenden Hauptzählzeiten berechnet, bis diese Bedingung erfüllt ist.

Für die  $i$ -te Note können die beiden direkt benachbarten Zählzeiten mit den Indize  $n_{B,\downarrow}$  und  $n_{B,\uparrow}$  bestimmt werden.

$$n_{B,\downarrow} = \min_{n_B} [\mathbf{t}_N(i) - \mathbf{t}_B(n_B)]; \quad \mathbf{t}_N(i) - \mathbf{t}_B(n_B) \geq 0 \quad (4.2)$$

$$n_{B,\uparrow} = \min_{n_B} [\mathbf{t}_B(n_B) - \mathbf{t}_N(i)]; \quad \mathbf{t}_B(n_B) - \mathbf{t}_N(i) \geq 0 \quad (4.3)$$

Die taktbezogene Anfangszeit  $\tau_N(i)$  einer Note wird aufgrund der angesetzten Näherung über eine Verhältnisgleichung bestimmt.

$$\tau_N(i) = \tau_B(n_{B,\downarrow}) + \frac{\mathbf{t}_N(i) - \mathbf{t}_B(n_{B,\downarrow})}{\mathbf{t}_B(n_{B,\uparrow}) - \mathbf{t}_B(n_{B,\downarrow})} \cdot \frac{1}{B_N} \quad (4.4)$$

Um die Notendauern  $\Delta \mathbf{t}_N$  in einer taktlängenbezogene Repräsentation umzurechnen wird ein Vektor  $\mathbf{t}_{N,Off}$  mit allen Notenendzeitpunkten (*offset times*) abgeleitet.

$$\mathbf{t}_{N,Off} = \mathbf{t}_N + \Delta \mathbf{t}_N \quad (4.5)$$

Analog zu  $\mathbf{t}_N$  erfolgt, wie in Gleichungen (4.2) bis (4.4) beschrieben, die zeitliche Projektion von  $\mathbf{t}_{N,Off}$  zu  $\tau_{N,Off}$ . Anschließend kann  $\Delta \mathbf{t}_N$  über

$$\Delta \mathbf{t}_N = \tau_{N,Off} - \tau_N \quad (4.6)$$

berechnet werden.

### 4.1.3 Zeitliche Quantisierung

Die zeitliche Quantisierung erfolgt entsprechend eines Quantisierungsrasters, welches auf einer Unterteilung jeder Taktlänge in  $Q$  äquidistante Rasterpunkte basiert. In dieser Arbeit wurde ein Wert von  $Q = 64$  gewählt. Die taktlängenbezogenen Zeitpunkte  $\tau_Q$  können analog zu Gleichung (4.1) berechnet werden.

$$\tau_Q(i) = \frac{i - 1 + Q}{Q}; \quad 1 \leq i \leq N_B \cdot Q \quad (4.7)$$

Anschließend kann für jede Note der nächste benachbarte Quantisierungsrasterpunkt mit dem Indiz  $n_{Q,i} = \min_{n_Q} |\tau_Q(n_Q) - \tau_N(i)|$  bestimmt und die zeitliche Quantisierung durchgeführt werden.

$$\tau_{N,q}(i) = \tau_Q(n_{Q,i}) \quad (4.8)$$

Die Quantisierung der Notendauer  $\Delta\tau_N$  erfolgt gemäß

$$\Delta\tau_{N,q}(i) = \frac{k}{Q}; \quad k = \min_k \left| \Delta\tau_N(i) - \frac{k}{Q} \right| \quad (4.9)$$

$$k \in \mathbb{Z}; \quad k \geq 1; \quad 1 \leq i \leq N_N.$$

Weiterhin gilt

$$\tau_{N,Off} = \tau_N + \Delta\tau_N. \quad (4.10)$$

### 4.1.4 Harmonische Analyse

Wie im Kapitel 2.1.1 beschrieben, erfolgt vor der Merkmalsextraktion eine stark vereinfachte harmonische Analyse. In Abb. 4.3 ist beispielhaft ein Ausschnitt aus der Harmoniespur des Stückes „Mack the Knife“ dargestellt. Die roten Balken stellen die gespielten Noten dar, blaue und grüne Balken die zeitlich verlängerten Noten. Hierbei werden Unterbrechungen, die durch Pausen zwischen aufeinanderfolgenden Akkorden entstehen, durch zeitliche Verlängerung des jeweiligen Vorgängerakkordes ausgefüllt. Ziel ist es dabei, auch für Noten des Melodieinstrumentes, die in diesen Pausen gespielt werden, einen harmonischen Bezug herstellen zu können.

Für jeden erkannten Akkord wird außerdem versucht, den entsprechenden Grundton (im Beispiel grün markiert) zu bestimmen. Dieser ist der Bezugspunkt bei der Bestimmung der funktionalen Tonhöhe des gespielten Solos (vgl. Kap. 2.1.1). Hierbei können bis zu vierstimmige Akkorde in allen Akkordumkehrungen untersucht werden. Die implementierte Funktion kann alle in Tabelle 2.2 dargestellten Drei- und Vierklänge sowie die Intervalle Terzen und Quinten und ihre entsprechenden Komplementärintervalle Sexten und Quarten richtig zuordnen und den Akkordgrundton entsprechend bestimmen. Für



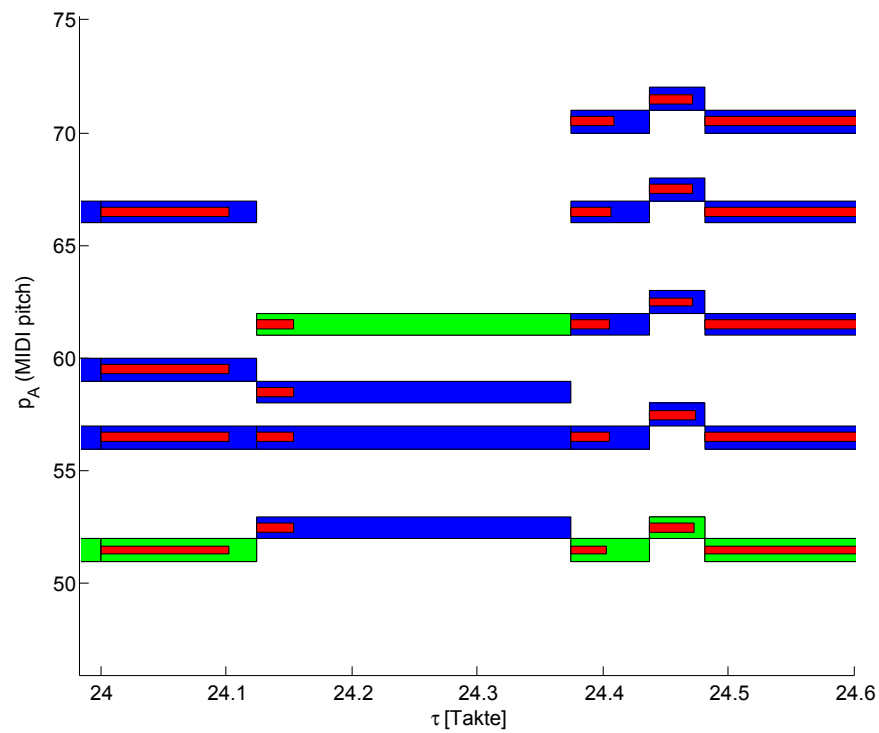


Abb. 4.3: Harmonische Analyse (Beispiel)

alle nicht erkannten Akkorde wird vereinfachend der tiefste gespielte Ton als Grundton angenommen.

In Kap. 4.2.2 werden die Merkmale, die sich aus dem Verlauf der funktionalen Tonhöhe ableiten, im Detail beschrieben.

## 4.2 Merkmalsextraktion

Um die Spielweise eines Musikers innerhalb eines Instrumentalsolos zu charakterisieren, werden in dieser Arbeit die folgenden Fragestellungen untersucht:

1. Was wird wie in welchem rhythmischen und harmonischen Kontext gespielt?
2. Wie ist das Solo strukturiert?
3. Wie und inwiefern interagiert der Solist mit den begleitenden Musikern?

Grundlage der Merkmalsextraktion sind die im Vorfeld gewonnenen Partiturinformationen des Melodie-, Harmonie-, Bass- und Schlaginstrumentes, sowie die Ergebnisse der harmonischen Analyse. Zur Ableitung entsprechender *High-level*-Merkmale, mit deren Hilfe die o. g. Fragestellungen innerhalb eines Instrumentalsolos untersucht werden sollen, werden fünf verschiedene Abstraktionsebenen verwendet:

1. Allgemeine Merkmale ( $\mathcal{G}$ ) (siehe 4.2.1)
2. Melodik ( $\mathcal{M}$ ) und Harmonik ( $\mathcal{H}$ ) (siehe 4.2.2)
3. Rhythmik ( $\mathcal{R}$ ) (siehe 4.2.3)
4. Struktur ( $\mathcal{S}$ ) (siehe 4.2.4)
5. Interaktion ( $\mathcal{I}$ ) (siehe 4.2.5)

Für jedes der vier Instrumente wird eine Auswahl der in diesem Kapitel vorgestellten Merkmale berechnet, von denen der Großteil auf der Charakterisierung von Melodie und Rhythmus nach musiktheoretischen Aspekten basiert. Motivierend dafür ist die Grundannahme, dass potenzielle Nutzer des implementierten Analysesystems über eine musikalische Grundbildung verfügen. Merkmale, die sich aus der Instrumentierung eines Stückes ableiten (vgl. z.B. [MF04]), spielen in dieser Arbeit keine Rolle, es ist lediglich die Zuordnung zu einer der vier in Kap. 2.1.4 genannten Instrumentenklassen von Bedeutung. Durch die Beschränkung auf *High-level*-Merkmale werden im Weiteren auch keine Merkmale verwendet, die das Timbre<sup>3</sup> einzelner Instrumente oder die Verwendung typischer Spielweisen wie Vibrato oder Tremolo abbilden. In den folgenden Kapiteln wird der Großteil aller implementierten *High-level*-Merkmale erläutert, eine Übersicht über die Zuordnung der Merkmale zu den vier untersuchten Instrumentalspuren wird in Anhang C gegeben.

---

<sup>3</sup>Klangbezogene Eigenschaften.

### 4.2.1 Allgemeine Merkmale

Zu den allgemeinen Merkmalen eines Musikstückes gehören das Tempo  $\mathcal{G}_{Tempo}$  (in BPM) sowie die mittlere Anzahl von Noten  $\mathcal{G}_{NumNotes}$  des untersuchten Instruments pro Takt.

$$\mathcal{G}_{NumNotes} = \frac{N_N}{N_B} \quad (4.11)$$

### 4.2.2 Melodische und harmonische Ebene

#### Absolute Tonhöhe (absolute pitch - PA)

**Polyphoniemaß (measure of polyphony)** Oftmals spielt der Solist kurze mehrstimmige (polyphone) Passagen innerhalb des Solos. Über ein Verhältnis aus deren zeitlichem Anteil bezogen auf die gesamte Dauer des Solos wird ein Polyphoniemaß berechnet.

$$\mathcal{M}_{PA,Poly} = \frac{t_{Polyphon}}{t_{Solo}} \quad (4.12)$$

Im Anschluss an die Berechnung von  $\mathcal{M}_{PA,Poly}$  wird der Einfachheit halber die mehrstimmigen Anteile des Solos in einstimmige (monophone) umgewandelt – hierbei wird jeweils die (auf die Tonhöhe bezogene) höchste Melodiestimme beibehalten.

**Chromavektorbezogene Merkmale** Durch eine Projektion der absoluten Tonhöhe  $p_A$  auf eine Grundoktave per Modulo-12-Operation kann der Chromavektor  $\mathbf{x}_{Chroma} \in \mathbb{R}^{12}$  aus den statistischen Häufigkeiten der einzelnen Werte von  $\tilde{p}_A$  berechnet werden:

$$\mathbf{x}_{Chroma}(i) = \frac{1}{N_N} \sum_{k=1}^{N_N} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \tilde{p}_A(i) = i \\ 0, & \text{sonst} \end{cases} \quad (4.13)$$

$$1 \leq i \leq 12$$

Wie in Kap. 2.1.1 beschrieben, werden gemäß der Akkord-Skalen-Theorie die zum Improvisieren hauptsächlich verwendeten Töne der Skale entnommen, die sich aus der aktuellen Tonart und dem aktuellen Akkord ableiten. Je nach Genre wird die Tonart unterschiedlich oft gewechselt. In Genres wie z.B. Pop-Rock bleibt sie oftmals über den gesamten Verlauf des Solos unverändert, während sie beispielsweise im Swing oft wechselt. Durch die Berechnung des Chromavektors erfolgt eine Mittelung der Häufigkeiten aller Töne unabhängig ihrer Oktavlage über das gesamte Solo. Die Werte von  $\mathbf{x}_{Chroma}$  sind für häufige Tonartwechsel näherungsweise gleichverteilt – für eine konstante Tonart jedoch besitzen vor allem die sieben Töne der korrespondierenden Skale hohe Häufigkeitswerte. Um die Werteverteilung innerhalb des Chromavektors zu beschrei-

ben, wird vereinfacht angenommen, dass die Werte der absoluten Tonhöhe durch die Zeichen eine Markovquelle erster Ordnung (vgl. Kap. 2.2.3) modelliert werden können. Es werden die Entropie nullter Ordnung im Merkmal  $\mathcal{M}_{PA,Chroma,0}$  (siehe Gleichung (2.11)) sowie die D’Agostino-Maßzahl der Verteilung entsprechend Gleichung (2.8) im Merkmal  $\mathcal{M}_{PA,Chroma,Dago}$  berechnet.

**Maß für die Häufigkeit von Akkordtönen** Aus der harmonischen Analyse des Solos, die vor der Merkmalsextraktion durchgeführt wird, können jedem  $i$ -ten Ton eines Solos eine Menge von Akkordtönen mit den Tonhöhen  $\tilde{\mathbf{p}}_{A,Harm,i}$  zugeordnet werden, die gleichzeitig vom Harmonieinstrument gespielt werden bzw. solche, die noch nachklingen. Entsprechend der Häufigkeit, mit der der Solist Akkordtöne des jeweils zuletzt gespielten Akkord in seinem Solo spielt, wird ein Merkmal  $\mathcal{M}_{PA,ChordTone}$  abgeleitet.

$$\mathcal{M}_{PA,ChordTone} = \sum_{k=1}^{N_N} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \mathbf{p}_A(k) \in \{\tilde{\mathbf{p}}_{A,Harm,k}\} \\ 0, & \text{sonst} \end{cases} \quad (4.14)$$

Ein hoher Wert dieses Merkmals impliziert einen hohen Harmoniebezug des gespielten Solos.

### Relative Tonhöhe (relative pitch - PR)

Die relative Tonhöhe (in Halbtönen) berechnet sich aus

$$\mathbf{p}_R(i) = \mathbf{p}_A(i+1) - \mathbf{p}_A(i) \quad (4.15)$$

$$\mathbf{p}_R \in \mathbb{Z}^{N_N-1} ; \quad i \in \mathbb{Z} ; \quad 1 \leq i \leq N_N - 1 .$$

Über eine Projektion auf eine oktavbezogene relative Tonhöhe  $\tilde{\mathbf{p}}_R$  ergibt sich:

$$\tilde{\mathbf{p}}_R(k) = \begin{cases} \mathbf{p}_R(k) \bmod 12 & \text{falls } \mathbf{p}_R(k) \geq 0 \\ -\{[-\mathbf{p}_R(k)] \bmod 12\} & \text{falls } \mathbf{p}_R(k) < 0 \end{cases} \quad (4.16)$$

$$\tilde{\mathbf{p}}_R \in [-11, 11] ; \quad 1 \leq k \leq N_N - 1 .$$

**Maß für die Häufigkeit konstanter Tonhöhen (measure of constant pitch)** Dieses Merkmal berechnet sich aus der Häufigkeit identischer aufeinanderfolgender Töne.

$$\mathcal{M}_{PA,ConstPitch} = \sum_{k=1}^{N_N-1} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \mathbf{p}_R(k) = 0 \\ 0, & \text{sonst} \end{cases} \quad (4.17)$$

**Verteilung** Die Werteverteilung von  $\mathbf{p}_R$  wird anhand seines Mittelwertes und seiner Standardabweichung in den beiden Merkmalen  $\mathcal{M}_{PR,MN}$  und  $\mathcal{M}_{PR,SD}$  abgebildet.

**Maß chromatischer Tonfolgen (measure of chromatics)** Eine chromatische Tonfolge besteht aus mindestens zwei identischen aufeinanderfolgenden Intervallen, die die Größe eines Halbtonschrittes besitzen. Da in keiner diatonischen Skale (vgl. Kap. 2.1.1) zwei Halbtonschritte unmittelbar hintereinander auftreten, lässt die Verwendung von chromatischen Tonfolgen auf den Einsatz tonartfremder Töne schließen. Dies kommt vor allem in jazz-verwandten Genres häufig vor und wirkt dem Höreindruck einer hohen Harmoniebezogenheit des Solos entgegen. Das Maß chromatischer Tonfolgen  $\mathcal{M}_{Chromatics}$  berechnet sich aus der Häufigkeit zweier aufeinanderfolgender Intervallwerte von 1 bzw.  $-1$ .

$$\mathcal{M}_{Chromatics} = \frac{1}{N_N} \sum_{k=1}^{N_N-2} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \tilde{\mathbf{p}}_R(k) = 1 \text{ und } \tilde{\mathbf{p}}_R(k+1) = 1 \\ 1, & \text{falls } \tilde{\mathbf{p}}_R(k) = -1 \text{ und } \tilde{\mathbf{p}}_R(k+1) = -1 \\ 0, & \text{sonst} \end{cases} \quad (4.18)$$

**Musikalisches Intervall** Die oktavbezogene relative Tonhöhe  $\tilde{\mathbf{p}}_R$  kann entsprechend der vierten und der ersten Spalte der Tab. 2.1 als musikalisches Intervall  $\tilde{\mathbf{p}}_{R,Mus}$  angegeben werden. Für den Wertebereich gilt hierbei:

$$\tilde{\mathbf{p}}_{R,Mus} \in \mathbf{p}_{R,Mus}$$

$$\mathbf{p}_{R,Mus} = [-7, \dots, -2, 1, 2, \dots, 7] .$$

Die Projektion erfolgt analog zur funktionalen Tonhöhe (vgl. Kap. 2.1.1). Dies ermöglicht eine verallgemeinerte Untersuchung der eingesetzten Intervalle innerhalb des Solos unabhängig der momentanen Tonalität. Die Einzelwahrscheinlichkeiten  $\mathbf{p}(\mathbf{p}_{R,Mus})$  werden über das gesamte Solo berechnet und als Merkmale  $\mathcal{M}_{PR,Mus,p,i}$  gespeichert. Weiterhin werden die aufeinanderfolgenden Intervalle als Markovprozess modelliert, aus den Einzelwahrscheinlichkeiten bzw. aus den Übergangswahrscheinlichkeiten können die beiden Entropien  $H_0$  und  $H_1$  entsprechend Gleichung (2.11) und (2.13) als Merkmale  $\mathcal{M}_{PR,Mus,H0}$  und  $\mathcal{M}_{PR,Mus,H1}$  berechnet werden.

**Maß der Häufigkeit kleiner Intervalle** Die häufige Verwendung von kleinen Intervallen innerhalb des Solos erzeugt den Höreindruck eines hohen melodischen Flusses innerhalb der gespielten Melodie. Um dies in einem Merkmal abzubilden, werden die Intervalle Prime, Sekunde und Terz mit einem Wert von 1 bewertet, alle größeren mit

einem Wert von 0.

$$\mathcal{M}_{PR,SmallInt} = \frac{1}{N_N - 2} \sum_{k=1}^{N_N-2} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } |\tilde{\mathbf{p}}_{R,Mus}(k)| \leq 3 \\ 0, & \text{sonst} \end{cases} \quad (4.19)$$

**Tendenz zu konstanter melodischer Richtung** Die Richtung der relativen Tonhöhe  $\mathbf{p}_R$  kann über Anwendung der Signumfunktion im Vektor  $\mathbf{p}_{R,Dir}$  abgebildet werden.

$$\mathbf{p}_{R,Dir} = \text{sgn} \{ \mathbf{p}_R \} \quad (4.20)$$

Im Merkmal  $\mathcal{M}_{PR,BalDir}$  wird der zeitliche Anteil von solchen Melodiefragmenten innerhalb des Solos abgebildet, in denen die Melodie in ihrer Tonhöhe kontinuierlich steigt oder fällt.

$$\mathcal{M}_{PR,BalDir} = \frac{1}{N_N - 2} \sum_{k=1}^{N_N-2} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \mathbf{p}_{R,Dir}(k) \equiv \mathbf{p}_{R,Dir}(k+1) \\ 0, & \text{sonst} \end{cases} \quad (4.21)$$

**Vorherrschende melodische Richtung** Um die dominante melodische Richtung über den gesamten Verlauf des Solos in einem Merkmal abzubilden, werden alle steigenden und fallenden Intervalle in Relation zueinander gesetzt:

$$\mathcal{M}_{PR,DomDir} = \frac{N_{PR,\uparrow}}{N_{PR,\downarrow}} \quad (4.22)$$

Es gilt:

$$N_{PR,\uparrow} = \sum_{k=1}^{N_N-1} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \mathbf{p}_{R,Dir}(k) \equiv 1 \\ 0, & \text{sonst} \end{cases} \quad (4.23)$$

bzw.

$$N_{PR,\downarrow} = \sum_{k=1}^{N_N-1} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \mathbf{p}_{R,Dir}(k) \equiv -1 \\ 0, & \text{sonst} \end{cases} \quad (4.24)$$

### Funktionale Tonhöhe

Nach der harmonischen Analyse (vgl. Kap. 4.1.4) wird eine mittlere Anzahl von Akkordtönen in allen detektierten Akkorden im Merkmal  $\mathcal{M}_{PF,ChordDegree}$  sowie die mittlere Anzahl von Akkorden pro Takt im Merkmal  $\mathcal{M}_{PF,ChordsPerBar}$  abgebildet.

Nach Bestimmung des Verlaufes der funktionalen Tonhöhe

$$p_F \in \mathbf{p}_F ; \quad \mathbf{p}_F = [-7, \dots, -2, 1, 2, \dots, 7]$$

können die Einzelwahrscheinlichkeiten  $\mathbf{p}(\mathbf{p}_F)$  berechnet und im Merkmal  $\mathcal{M}_{PF,p,i}$  gespeichert werden.

### 4.2.3 Rhythmische Ebene

#### Rhythmische Präzision

Um ein Maß für die wahrgenommene rhythmische Präzision der gespielten Melodie bezogen auf ein beliebiges rhythmisches Raster der Unterteilung  $k_G$  zu berechnen, wird eine mittlere prozentuale Abweichung aller Noten zu ihrem jeweils nächstliegenden Rasterpunkt berechnet und anschließend quadriert. Dies verringert die Wichtung von sehr kleinen und damit kaum wahrzunehmenden Abweichungen in diesem Merkmal.

$$\mathcal{R}_{ON, Prec, k_G} = \left( \frac{1}{N_N} \sum_{i=1}^{N_N} \left| \frac{\tau_N(i) - \tau_{N,q}(i)}{1/(2k_G)} \right| \right)^2; \quad R_{Pr} \in [0, 1] \quad (4.25)$$

In dem implementierten System werden die Präzisionswerte  $\mathcal{R}_{ON, Prec, k_G}$  für die Rasterunterteilungen  $k_G \in \{4, 8, 16, 32\}$  als Merkmale berechnet.

#### Maß für Swing-Faktor

Aus der Abweichung der Notenanfangszeitpunkte zu einem vorgegeben Quantisierungsraster  $Q$  der Swing-Faktor<sup>4</sup>  $\mathcal{R}_{ON, Swing}$  der gespielten Melodie berechnet. Alle Rasterpunkte mit den Indizes  $i = 2l - 1$  werden im Folgenden als *Downbeat*-Rasterpunkte und mit den Indizes  $i = 2l$  als *Offbeat*-Rasterpunkte bezeichnet. Es gilt jeweils

$$l \in \mathbb{Z}; \quad 1 \leq i \leq Q.$$

Bei der Quantisierung (vgl. Kap. 4.1.3) wird für jede Note  $i$  neben dem quantisierten Startzeitpunkt  $\tau_{N,q}(i)$  auch der Index  $l$  des Quantisierungsrasterpunktes gespeichert, auf den die Note projiziert wird. Entsprechend der Quantisierungsvorschrift gilt dabei  $\tau_{N,q}(i) \equiv \tau_Q(l)$ .

Die zeitliche Verschiebung  $\Delta\tau_{N,q}$  jeder Note durch die Quantisierung kann anschließend aus

$$\Delta\tau_{N,q} = \tau_N - \tau_{N,q}$$

berechnet werden. Für ein beliebiges rhythmisches Raster  $k$  kann nun der Swingfaktor, wie in Abb. 4.4 für ein Beispiel im  $\frac{4}{4}$ -Takt dargestellt, aus einem Verhältnis der beiden

---

<sup>4</sup>Vgl. [GFB03].

zeitlichen Verschiebungen  $\Delta\tau_1$  und  $\Delta\tau_2$  berechnet werden. Dazu werden entsprechend

$$\Delta\tau_{OB} \equiv \overline{\Delta\tau_{N,q,OB}} \text{ bzw. } \Delta\tau_{DB} \equiv \overline{\Delta\tau_{N,q,DB}}$$

die zeitlichen Verschiebungen aller auf *Downbeat*- bzw. aller auf *Offbeat*-Rasterpunkte projizierten Noten gemittelt. Der Swing-Faktor berechnet sich anschließend aus

$$\mathcal{R}_{ON,Swing,k} = \frac{\Delta\tau_1}{\Delta\tau_2} = \frac{1/Q + \Delta\tau_{OB} - \Delta\tau_{DB}}{1/Q + \Delta\tau_{DB} - \Delta\tau_{OB}}. \quad (4.26)$$

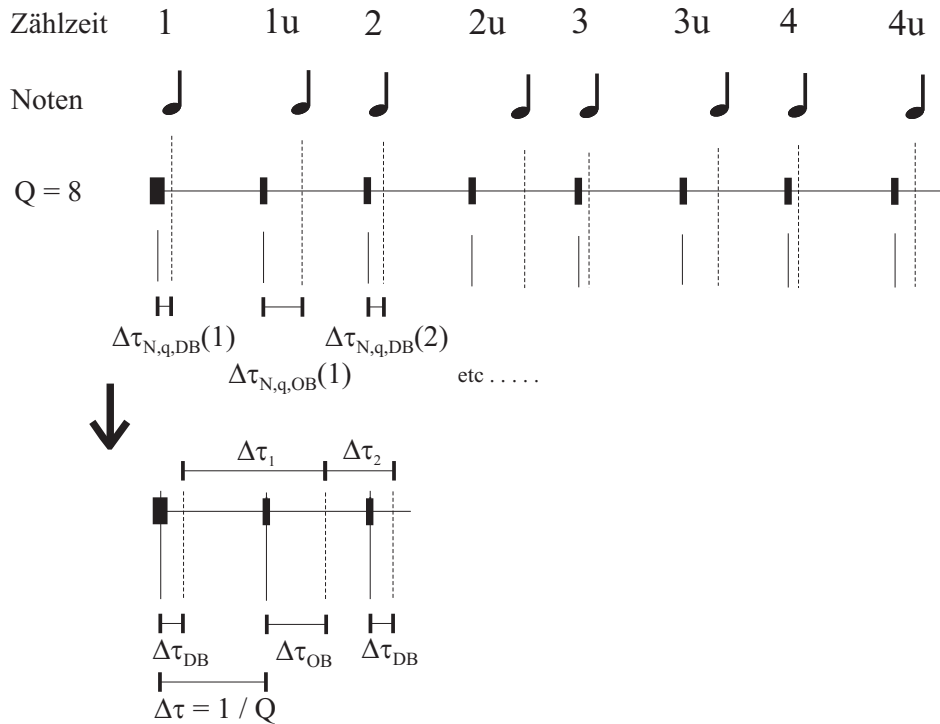


Abb. 4.4: Berechnung des Swing-Faktors im Achtelnotenraster (Q = 8)

### Rhythmisches Strukturprofil

Mit Hilfe des rhythmischen Strukturprofils wird die zeitliche Verteilung der Noten-anfangszeitpunkte untersucht, um daraus semantische rhythmische Merkmale wie z.B. ein Maß für Synkopierung oder das dominierende rhythmische Raster zu berechnen. Es ermöglicht eine tempo- und taktartunabhängige Beschreibung der zeitlichen Verteilung aller Noten innerhalb des Taktes, in der sie gespielt werden.

Das Verfahren basiert auf verschiedenen Unterteilungen der Taktlängen in jeweils  $k_G$



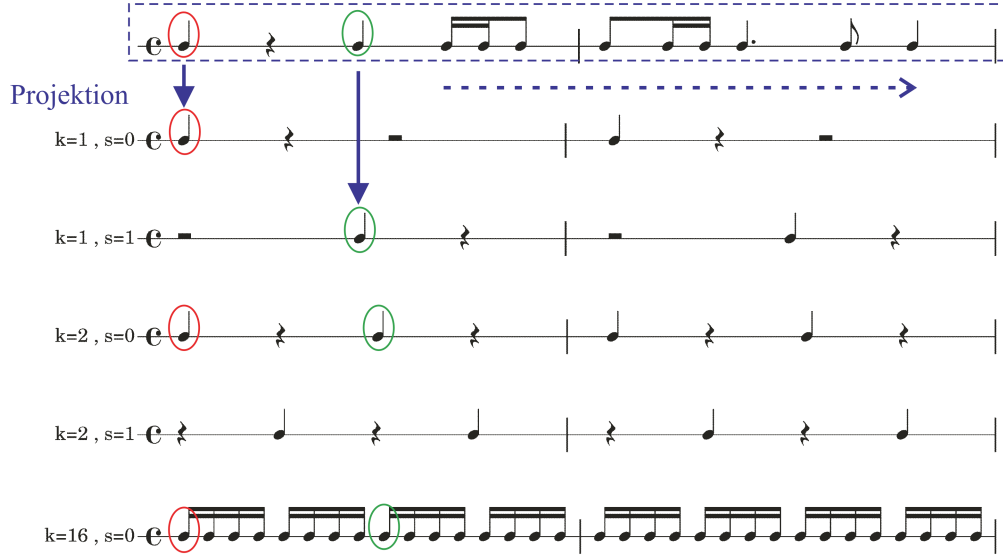


Abb. 4.5: Rhythmisches Strukturprofil - Projektion (Beispiel)

äquidistante Rasterzeitpunkte, wobei gilt:

$$1 \leq k_G \leq Q.$$

Musikwissenschaftlich relevant sind dabei besonders alle binären

$$k_G \in K_{G,bin}, \quad K_{G,bin} = \{2 \cdot 2^n\}, \quad n \in \mathbb{Z}, \quad 0 \leq n \leq n_{max}$$

und ternären

$$k_G \in K_{G,ter}, \quad K_{G,ter} = \{3 \cdot 2^n\}, \quad n \in \mathbb{Z}; \quad 0 \leq n \leq n_{max}$$

Unterteilungen, die im Folgenden betrachtet werden. In dieser Arbeit wird der Wert  $n_{max} = 5$  verwendet. In Abb. 4.5 ist ein zweitaktiges Notenbeispiel im  $\frac{4}{4}$ -Takt und darunter die Positionen der Rasterzeitpunkte für einige Werte von  $k = k_g$  und  $s = s_g$  dargestellt. Liegt der Anfangszeitpunkt  $\tau_N(i)$  einer Note in einem zeitlichen Bereich

$$\tau_G - \Delta\tau \leq \tau_N(i) \leq \tau_G + \Delta\tau \quad (4.27)$$

um einen Rasterpunkt, so wird sie jedem Raster zugeordnet, welches diesen Rasterpunkt enthält. Für  $\Delta\tau$  gilt:

$$\Delta\tau = \frac{1}{2k_{G,max}}; \quad k_{G,max} = \begin{cases} 2 \cdot 2^{n_{max}} & \text{falls } k_G \text{ binär} \\ 3 \cdot 2^{n_{max}} & \text{falls } k_G \text{ ternär.} \end{cases} \quad (4.28)$$

Im Folgenden bezeichnet die *Energie* einer Note ihre normierte Lautstärke:

$$\epsilon = \frac{1}{\max\{v\}} v \quad (4.29)$$

Im Energieprofil  $\Lambda_{R1}$  werden die Energien aller Noten aufsummiert, die auf das jeweili-

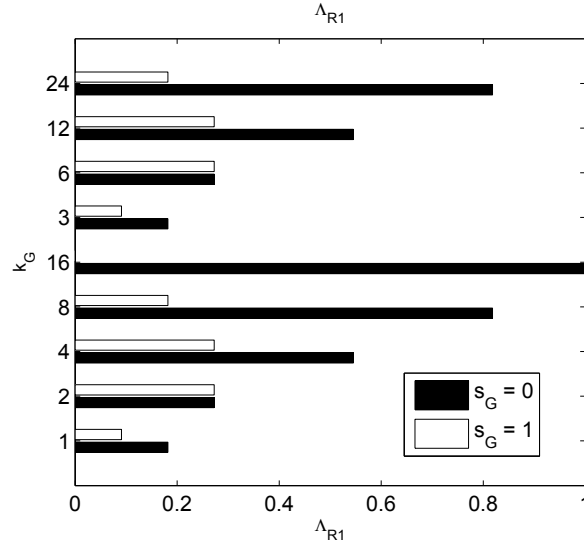


Abb. 4.6: Energieprofil  $\Lambda_{R1}$  (Beispiel)

ge Raster mit der Unterteilung  $k_G$  projiziert wurden, und anschließend auf die Gesamtenergie  $\epsilon_\Sigma$  aller Noten normiert. Es werden zwei Positionen jedes Rasters untersucht – unverschoben und um einen halben Rasterpunktastand nach hinten verschoben. Diese Verschiebung wird durch  $s_G = 0$  bzw.  $s_G = 1$  symbolisiert.

$$\Lambda_{R1}(k_G, s_G) = \frac{1}{\epsilon_\Sigma} \sum_i \epsilon_i \quad (4.30)$$

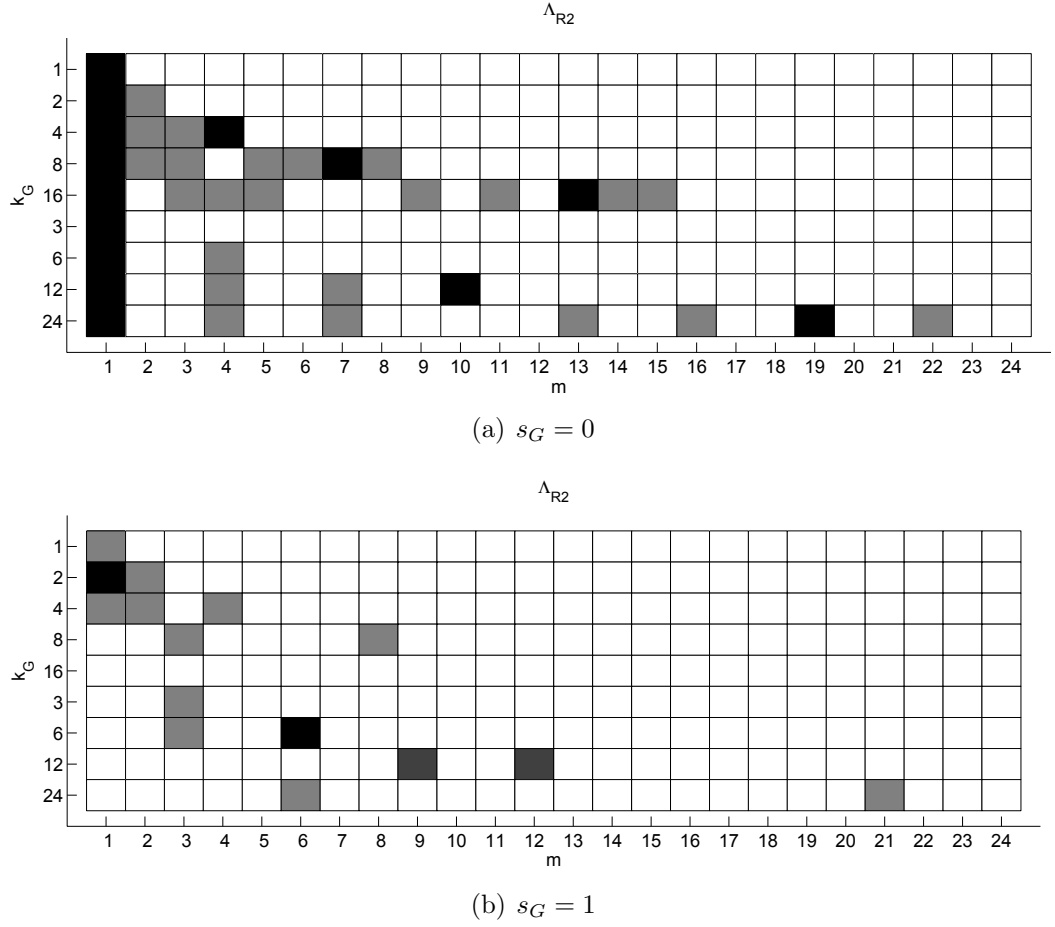
$$\forall i \begin{cases} [\tau_{N,q}(i) \bmod \Delta\tau_G] \leq \frac{\Delta\tau_G}{2}, & s_G = 0 \\ [(\tau_{N,q}(i) + \frac{\Delta\tau_G}{2}) \bmod \Delta\tau_G] \leq \frac{\Delta\tau_G}{2}, & s_G = 1 \end{cases}$$

Für den Zeitbereich  $\Delta\tau_G$  gilt dabei:

$$\Delta\tau_G = \frac{1}{k_G} \quad (4.31)$$

Das Energieprofil  $\Lambda_{R2}$  bildet die Verteilung der Noten innerhalb eines Rasters ab. Es entspricht der normierten Energie aller Noten, die dem Rasterpunkt  $m$  mit

$$1 \leq m \leq k_G$$

Abb. 4.7: Energieprofil  $\Lambda_{R2}$  (Beispiel)

innerhalb eines gegebenen Raster mit der Unterteilung  $k_G$  und der Verschiebung  $s_G$  zugeordnet sind.

$$\Lambda_{R2}(k_G, m, s_G) = \frac{1}{\epsilon_\Sigma} \sum_i \epsilon_i \quad (4.32)$$

$$(m - \frac{3}{2})\Delta\tau_G \leq [\tau_{N,q}(i) \bmod 1] \leq (m - \frac{1}{2})\Delta\tau_G \quad (4.33)$$

Aufgrund der Energienormierung gilt

$$\sum_{m=1}^{k_G} \frac{\Lambda_{R2}(k_G, m, s_G)}{\Lambda_{R1}(k_G, s_G)} = 1 .$$

In Abb. 4.6 und 4.7 sind die berechneten Energieprofile für das in Abb. 4.5 dargestellte Notenbeispiel abgebildet. Der Wert für  $\Lambda_{R2}(k_G, m, s_G)$  ist in dem Schwarzwert des entsprechenden Feldes in Abb. 4.7 kodiert. Die beiden Energieprofile  $\Lambda_{R1}$  und  $\Lambda_{R2}$  bilden zusammen das rhythmische Strukturprofil (RSP). Nach der Berechnung des

RSP anhand aller Noten innerhalb des betrachteten zeitlichen Bereiches, können daraus eine Vielzahl semantisch rhythmische Merkmale berechnet werden. Diese werden in den anschließenden Kapiteln erläutert.

### Rhythmisches Feeling

Ordnet man das Energieprofil  $\Lambda_{R1}(k_G, s_G)$  getrennt für binäre und ternäre Rasterunterteilungen  $k$  an, so kann über eine jeweilige Aufsummierung und Maximumsuche entschieden werden, welches das dominante rhythmische Feeling  $\mathcal{R}_{ON,RSP,Feel}$  des untersuchten Instrumentes ist.

$$\Lambda_{R1,\Sigma,Bin} = \sum_{s_G \in \{0,1\}} \sum_{k_G \in K_{G,bin}} \Lambda_{R1}(k_G, s_G) \quad (4.34)$$

$$\Lambda_{R1,\Sigma,Tern} = \sum_{s_G \in \{0,1\}} \sum_{k_G \in K_{G,ter}} \Lambda_{R1}(k_G, s_G) \quad (4.35)$$

Es gilt:

$$\mathcal{R}_{ON,RSP,Feel} = \begin{cases} 2, & \text{wenn } \Lambda_{R1,\Sigma,Bin} \geq \Lambda_{R1,\Sigma,Tern} \\ 3, & \text{wenn } \Lambda_{R1,\Sigma,Tern} > \Lambda_{R1,\Sigma,Bin} \end{cases} \quad (4.36)$$

Generell ist es denkbar, das Modell des rhythmischen Strukturprofils auch auf weitere (in der musikalischen Praxis aber seltener vorkommende) Taktarten wie z.B. den  $\frac{5}{4}$ - oder den  $\frac{7}{4}$ -Takt zu erweitern. In dieser Arbeit wird jedoch vereinfachend von binären oder ternären Taktarten ausgegangen.

### Rhythmische Charakteristik

Anschließend kann die dominante rhythmische Charakteristik  $\mathcal{R}_{ON,RSP,Char}$  bestimmt werden. Diese gibt an, ob die Noten überwiegend auf den durch die Unterteilung  $k_G$  vorgegebenen Rasterpunkten liegen (*downbeat*,  $\mathcal{R}_{ON,RSP,Char} = 0$ ) oder auf einem um einen halben Rasterpunkt abstand verschobenen Raster (*offbeat*,  $\mathcal{R}_{ON,RSP,Char} = 1$ ).

$$\Lambda_{R1,\Sigma,DB} = \sum_{k_G} \Lambda_{R1}(k_G, 0) \quad \begin{cases} k_G \in K_{G,bin}, & \text{wenn } R_F = 2 \\ k_G \in K_{G,ter}, & \text{wenn } R_F = 3 \end{cases} \quad (4.37)$$

$$\Lambda_{R1,\Sigma,OB} = \sum_{k_G} \Lambda_{R1}(k_G, 1) \quad \begin{cases} k_G \in K_{G,bin}, & \text{wenn } R_F = 2 \\ k_G \in K_{G,ter}, & \text{wenn } R_F = 3 \end{cases} \quad (4.38)$$

Über eine Maximumsuche ergibt sich:

$$\mathcal{R}_{ON,RSP,Char} = \begin{cases} 0, & \text{wenn } \Lambda_{R1,\Sigma,DB} \geq \Lambda_{R1,\Sigma,OB} \\ 1, & \text{wenn } \Lambda_{R1,\Sigma,OB} > \Lambda_{R1,\Sigma,DB} \end{cases} \quad (4.39)$$

### Rhythmisches Raster

Nach der Bestimmung des rhythmischen Feelings  $\mathcal{R}_{ON,RSP,Feel}$  und der rhythmischen Charakteristik  $\mathcal{R}_{ON,RSP,Char}$  kann ein drittes Merkmal zur Beschreibung der Verteilung der Notenanfangszeitpunkte abgeleitet werden – das dominante rhythmische Raster  $\mathcal{R}_{ON,RSP,Grid}$ . Es beschreibt, in welcher rhythmischen Auflösung sich der Großteil aller Noten befindet. Es gilt

$$\mathcal{R}_{ON,RSP,Grid} = \min_{k_G} \iff \Lambda_{R1}(k_G, R_C) > x_{min} . \quad (4.40)$$

Das Verhältnis  $x_{min}$  repräsentiert den geforderten prozentualen Anteil aller Noten, die mindestens im Raster  $k_G$  liegen müssen und wird im Folgenden mit  $x_{min} = 0.9$  angesetzt. Für das Notenbeispiel aus Abb. 4.5 gilt wegen  $\Lambda_{R1}(16, 0) = 1$  (vgl. Abb. 4.6) also  $\mathcal{R}_{ON,RSP,Grid} = 16$  gelten, was musikalisch dem zugrunde liegenden Sechzehntelnotenraster entspricht.

### Maß für Sykkopierung

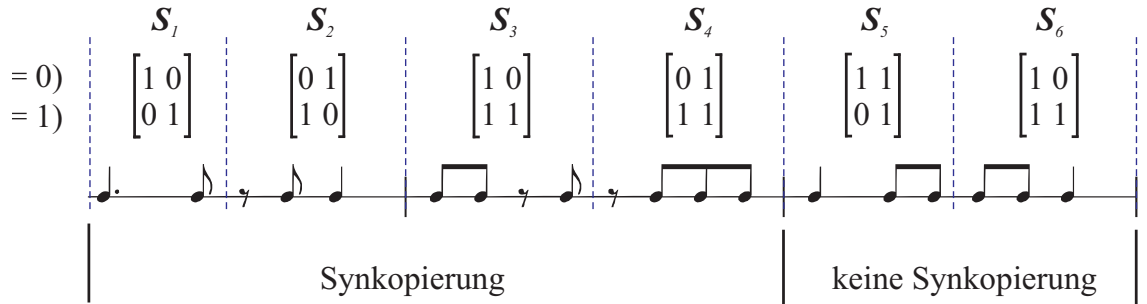


Abb. 4.8: Sykkopierungen und korrespondierende Matrizen  $S_i$  (Beispiele)

Wird das rhythmische Strukturprofil  $\Lambda_{R2}$  aus den quantisierten Anfangszeitpunkten  $\tau_{n,q}$  aller Noten eines Taktes gebildet, so kann anschließend gezielt nach den in Abb. 4.8 dargestellten Teilmatrizen  $S_i$  und damit nach dem Auftreten von Sykkopen innerhalb des jeweiligen Rasters  $k$  gesucht werden. Die Spalten der Matrizen  $S_i$  repräsentieren zwei beliebige benachbarte Positionen

$$m \in \{l, l+1\}; \quad l \in \mathbb{Z}; \quad 1 \leq l \leq k_g - 1$$

im Raster der Unterteilung  $k_G$  und die beiden Zeilen die beiden Rasterverschiebungen  $s \in \{0, 1\}$  innerhalb innerhalb der Matrix  $\mathbf{\Lambda}_{R2}$ . In der Abbildung ist ein Achtelnoten-raster dargestellt – es gilt  $k_G = 4$ . Da es sich um einen  $\frac{4}{4}$ -Takt handelt, liegen die Hauptzählzeiten bzw. die Betonungen auf den Viertelnoten. Für jede der folgenden Matrizen ist ein Notenbeispiel dargestellt, welches der Definition einer Synkope – also einer Verschiebung einer Betonung auf einen Zeitpunkt zwischen zwei Zählzeiten – entspricht. Es gilt

$$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{S}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \mathbf{S}_3 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \mathbf{S}_4 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}. \quad (4.41)$$

In den beiden letzten Notenbeispiele in der Abbildung mit den Matrizen  $\mathbf{S}_5$  und  $\mathbf{S}_6$  sind keine Synkopen enthalten, da die Viertelbetonungen jeweils gespielt werden.

Zur Detektion von Synkopen wird aus der Matrix  $\mathbf{\Lambda}_{R2}$  zuerst eine binäre Matrix  $\tilde{\mathbf{\Lambda}}$  der selben Größe abgeleitet, da die Lautstärke, mit der einzelne Noten gespielt werden, keinen Einfluss auf das Vorhandensein einer Synkope hat.

$$\tilde{\mathbf{\Lambda}}(k, m, s) = \begin{cases} 1 & \mathbf{\Lambda}_{R2}(k, m, s) > 0 \\ 0 & \mathbf{\Lambda}_{R2}(k, m, s) = 0 \end{cases} \quad (4.42)$$

Danach wird in allen Positionen

$$1 \leq m \leq k_G - 1$$

der Rasterunterteilung  $k_G$  nach dem Vorkommen einer der Matrizen  $\mathbf{S}_i$  gesucht. Eine Synkope gilt dabei als detektiert, wenn einer der vier Matrizen  $\mathbf{S}_i$  als Teilmatrix in  $\tilde{\mathbf{\Lambda}}$  gefunden wird, dazu muss gelten:

$$\begin{bmatrix} \tilde{\mathbf{\Lambda}}(k_G, m, 0) & \tilde{\mathbf{\Lambda}}(k_G, m + 1, 0) \\ \tilde{\mathbf{\Lambda}}(k_G, m, 1) & \tilde{\mathbf{\Lambda}}(k_G, m + 1, 1) \end{bmatrix} \equiv \mathbf{S}_i, \quad i \in \{1, 2, 3, 4\} \quad (4.43)$$

Um ein Maß für die Synkopierung im betrachteten rythmischen Raster  $k_G$  zu berechnen, wird die Anzahl  $N_S$  aller gefundenen Synkopen auf einen Maximalwert normiert, der der Anzahl aller Verschiebungspositionen  $k_G - 1$  im Raster entspricht.

$$\mathcal{R}_{ON, Syn, k_G} = \frac{N_S}{k_G - 1} \quad (4.44)$$

Der beschriebene Algorithmus ermöglicht es, in verschiedenen rhythmischen Ebenen nach Synkopen zu suchen, in der Arbeit erfolgt dies für  $k_G \in \{2, 4, 8, 16\}$ . Es werden entsprechend vier Merkmale  $\mathcal{R}_{ON, Syn, k_G}$  abgeleitet.

### Ausblick

Wie in Kap. 2.1.2 beschrieben, führt eine Überlagerung von binären und ternären Rhythmen zur Wahrnehmung von Polyrhythmik in Musikstücken. Dies könnte in einem weiteren rhythmischen Merkmal abgebildet werden, in dem untersucht wird, welche der vier untersuchten Instrumente ein rhythmisches Feeling von  $\mathcal{R}_{ON,RSP,Feel} = 2$  bzw. von  $\mathcal{R}_{ON,RSP,Feel} = 3$  besitzen.

Weitere Anwendungen der rhythmischen Strukturanalyse zur Untersuchung der Interaktion zwischen einzelnen Instrumenten werden im Kap. 4.2.5 erläutert. Die Anwendung des rhythmischen Strukturprofils für den Entwurf eines Klassifikators zur Genreklassifikation wird in Kap. 5.1.2 näher beschrieben.

### 4.2.4 Strukturelle Ebene

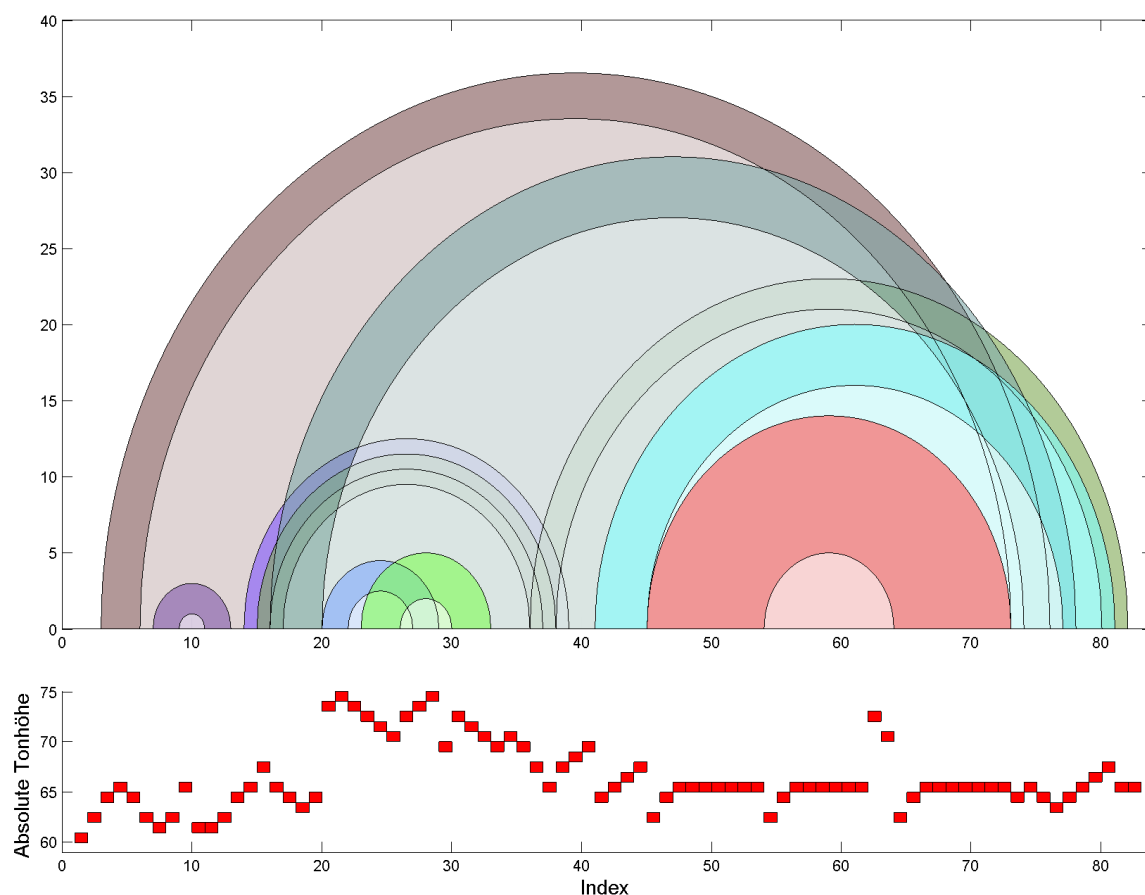


Abb. 4.9: Musterverteilung der absoluten Tonhöhe ("Desafinado" - Latin)

Um aus den Solo strukturbezogene Merkmale zu gewinnen, wurde ein Algorithmus zur Musterdetektion von sich wiederholenden Zeichenfolgen in einer gegebenen Zei-

$p_A$	Absolute Tonhöhe
$\widetilde{\tau}_{N,q}$	Notenanfangszeitpunkt (modifiziert)
$\widetilde{\Delta\tau}_{N,q}$	Notendauer (modifiziert)

Tab. 4.1: Strukturanalyse - Untersuchte musikalische Parameter

chenkette in MATLAB implementiert. Dieser wurde von Hsu et. al. in [HLC01] vorgeschlagen und ist im Anhang A.2 als Pseudocode aufgeführt. Ein Muster bezeichnet hierbei Folgen von mindestens drei Zeichen, die mindestens zweimal (in identischer Form) innerhalb der Zeichenkette enthalten sind.

Ziel der Verwendung dieses Algorithmus im Kontext der Merkmalsextraktion ist es, sowohl melodische als auch rhythmische Wiederholungen innerhalb des Solos erkennen und auswerten zu können. Als Zeichenketten werden die drei in Tab. 4.1 musikalische Parameter des Solos untersucht.

Die modifizierten Notenanfangszeitpunkte  $\widetilde{\tau}_{N,q}$  werden per Modulo-1-Operation aus den taktlängenbezogenen, quantisierten Notenanfangszeitpunkten  $\tau_{N,q}$  berechnet, so dass nur noch die Position der Noten innerhalb des jeweiligen Takt von Bedeutung ist. Zusätzlich werden sowohl Notenanfangszeitpunkte als auch Notendauern jeweils mit dem Quantisierungsraster  $Q$  multipliziert, damit die zu verarbeitende Zeichenkette aus ganzen Zahlen besteht. Es gilt

$$\widetilde{\tau}_{N,q} = [\tau_{N,q} \bmod 1] \cdot Q \quad (4.45)$$

$$\widetilde{\Delta\tau}_{N,q} = \Delta\tau_{N,q} \cdot Q. \quad (4.46)$$

Der Algorithmus gibt für jeden der musikalischen Parameter  $k$  eine Liste aller  $\mathbf{n}_{Pat}(k)$  detektierten Muster aus, von denen jedes durch die folgenden drei Parameter beschrieben wird:

1. Auftretenshäufigkeit  $\mathbf{F}_{Pat}(k, i)$
2. Musterlänge  $\mathbf{L}_{Pat}(k, i)$
3. Mittlere Distanz zwischen identischen Mustern  $\mathbf{D}_{Pat}(k, i)$

Zusätzlich werden redundante Muster – also solche, die in längeren Mustern enthalten sind und dabei die gleiche Auftretenshäufigkeit und -position besitzen – aus der Liste entfernt.

Aus den drei genannten Parametern wird ein Relevanzmaß für jedes detektierte Muster



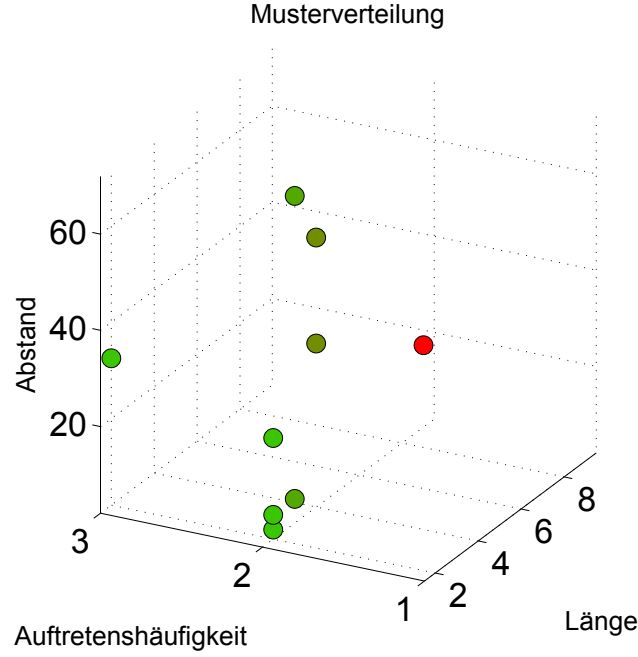


Abb. 4.10: Musterverteilung der absoluten Tonhöhe ("Desafinado" - Latin)

berechnet.

$$\mathbf{R}_{Pat}(k, i) = \left[ 1 - \frac{\mathbf{D}_{Pat}(k, i)}{\max \{ \mathbf{D}_{Pat} \}} \right]^2 + \frac{\mathbf{L}_{Pat}(k, i)}{\max \{ \mathbf{L}_{Pat} \}} + \frac{\mathbf{F}_{Pat}(k, i)}{\max \{ \mathbf{F}_{Pat} \}} \quad (4.47)$$

$$1 \leq i \leq \mathbf{n}_{Pat}(k)$$

Durch die Normierung auf den jeweiligen Maximalwert erfolgt für jeden der drei Parameter eine Projektion des Wertebereiches auf den Bereich  $[0, 1]$ . Der Einfluss der mittleren Distanz identischer Muster im Vergleich zu dem der anderen beiden Parameter wird durch eine zusätzliche Quadrierung des normierten Wertes verringert. Die Grundidee für die Verwendung eines Relevanzmaßes ist die Annahme, dass ein Muster einen umso größeren Wiedererkennungswert hat, desto länger es ist, desto häufiger es auftritt und desto geringer der mittlere Abstand  $\mathbf{D}_{Pat}(k, i)$  dabei ist.

Für jeden musikalischen Parameter werden die Anzahl der gefundenen Muster  $\mathbf{n}_{Pat}(k)$  auf die Gesamtzahl der Noten  $N_N$  normiert und als Merkmal verwendet:

$$\mathcal{S}_{k, Num} = \frac{\mathbf{n}_{Pat}(k)}{N_N} \quad (4.48)$$

$$\forall k \in [1, 4]$$

In Abb. 4.9 sind beispielhaft für ein Solo aus dem Latin-Stück „Desafinado“ die detektierten Muster für die Untersuchung der absoluten Tonhöhe  $\mathbf{p}_A$  dargestellt. Zur Veran-

schaulichung sind diese jeweils durch Halbkreissegmente verbunden<sup>5</sup>, wobei die Farbgebung hier willkürlich geschieht. Alle gefundenen Muster sind als Punkte entsprechend ihrer Parameter  $[f_{Pat}, d_{Pat}, l_{Pat}]$  im Koordinatensystem in Abb. 4.10 abgebildet. Der jeweilige Wert für die berechnete Relevanz ist im Farbwert kodiert – rot entspricht einer hohen Relevanz, grün einer geringen. Um die Verteilung zu charakterisieren, werden für jeden der vier Musterparameter die statistischen Größen Minimum, Maximum, statistischer Mittelwert, Median sowie Standardabweichung (siehe Kap. 2.2.2) berechnet und als weitere Merkmale gespeichert.

**Ausblick** In [HLC01] wird neben dem *Correlative-Matrix*-Algorithmus ein zweiter Algorithmus vorgeschlagen, der ein besseres Zeitverhalten besonders für längere Zeichenketten besitzt. Der sogenannte *String-Join*-Algorithmus basiert auf der wiederholten Verknüpfung kurzer Muster zur Detektion aller Muster in einer Zeichenkette, seine Implementierung ist jedoch wesentlich komplexer. Neben identischen Mustern ist auch die Suche nach ähnlichen Mustern vor allem für größere Musterlängen anstrebenswert, da Motive oder Themen, die der Solist im Verlauf seiner Improvisation aufgreift, oftmals variiert werden. Ein erweiterter Algorithmus sollte über eine Ähnlichkeitssuche in die Lage versetzt werden, diese leicht abgewandelten oder transponierten Melodien als Muster zu erkennen und in die Auswertung mit einzubeziehen. Vor allem in Soli über sogenannte *standards*<sup>6</sup> spielt das Zitieren von Hauptmotiven innerhalb des Solos eine große Rolle und kann deshalb gezielt untersucht werden.

### 4.2.5 Interaktionsebene

Zur Berechnung interaktionsbezogener Merkmale wurden zwei verschiedene Ansätze implementiert. Um zu untersuchen, ob zwei Instrumente rhythmisch unisono oder komplementär zueinander spielen, werden anhand der taktweise berechneten rhythmischen Strukturprofile der beiden Instrumente der Differenzwert

$$\Delta\Lambda_{I_1, I_2} = \sqrt{\sum_k \sum_j \sum_k \{\Lambda_{I_1}(i, j, k) - \Lambda_{I_2}(i, j, k)\}^2} \quad (4.49)$$

für jeden Takt gebildet und im Vektor  $\Delta\Lambda_{I_1, I_2}$  gespeichert. Der Mittelwert sowie die Standardabweichung dieses Vektors werden als Merkmale  $\mathcal{I}_{RSP, I_1, I_2, MW}$  sowie  $\mathcal{I}_{RSP, I_1, I_2, SD}$  gespeichert und charakterisieren die rhythmische Interaktion zwischen den beiden Instrumenten. Diese Merkmale werden für das Melodieinstrument bezogen auf

<sup>5</sup>Vgl. auch : *The Shape Of Song*, <http://www.turbulence.org/Works/song/> (zuletzt aufgerufen: 23.07.2008).

<sup>6</sup>Musikstücke mit großer Verbreitung und hohem Bekanntheitsgrad aus den Genres Swing, Blues und Latin.

das Harmonie-, Bass- und Schlaginstrument berechnet.

Um die melodisch-harmonische Interaktion zwischen Melodie- und Harmonieinstrument zu beschreiben, wird ähnlich dem in Kap. 4.2.2 beschriebenen Merkmal

$\mathcal{M}_{PA,ChordTone}$  für jeden Takt berechnet und zu einem Vektor zusammengefaßt. Analog zu  $\Delta\Lambda_{I_1,I_2}$  werden wieder Mittelwert und Standardabweichung berechnet und in den Merkmalen  $\mathcal{I}_{ChordTone,MW}$  sowie  $\mathcal{I}_{ChordTone,SD}$  gespeichert.

Über eine Erweiterung des in Kap. 4.2.4 beschriebenen Algorithmus zur Musterdetektion für die Anwendung auf die Spur des Harmonieinstrumentes – also auf polyphone Instrumentalspuren – könnte in Zukunft ebenso untersucht werden, inwieweit Solo- und Harmonieinstrument bestimmte Melodien unisono spielen.

**Ausblick** Bisher werden alle beschriebenen Merkmale global, d.h. über die gesamte Länge jedes Solos berechnet. Gerade für längere Instrumentalsolos macht eine lokale, *frame*-weise Verarbeitung durchaus Sinn. Werden jedoch zu wenige Noten pro *frame* betrachtet, so verlieren vor allem die statistischen Merkmale wie z.B. die berechneten Entropien ihre Aussagekraft, da zu wenig statistische Daten für ihre Berechnung vorliegen. Denkbar ist eine schrittweise Evaluation verschiedener Parameterwerte für die Schrittweite und Länge eines *frames* (vgl. auch [LI07], Kap. 3.3). Alternativ ist auch die Entwicklung eines Systems vorstellbar, welches je nach Merkmal eine globale bzw. eine lokale Merkmalsextraktion ermöglicht.

## 4.3 Implementierung

Im Rahmen der Diplomarbeit wurden alle in Abb. 4.1 dargestellten Verarbeitungsschritte in MATLAB implementiert. MATLAB ist eine plattformunabhängige Software für die Entwicklung von mathematischen Programmen vor allem für numerische Berechnungen und Modellbildungsaufgaben. Es wurden verschiedene Module implementiert, um die *MIDI-Toolbox* ([ET04]) zur Analyse von Audiodaten in symbolischem Format und die *Transcription Toolbox* ([DDR07]) zur Analyse von Audiodaten im sub-symbolischen Format zu verarbeiten. Analog dazu entstanden weitere Funktionen, um die rhythmische und harmonische Analyse, die zeitliche Projektion und alle Schritte der Merkmalsextraktion durchzuführen. Die Evaluation der Merkmale im Kontext der Genre- und Musikerklassifikation mit Hilfe der LibSVM-Toolbox([CL01]) wird in Kap. 5 beschrieben.

**Ausblick** Da das in dieser Arbeit entwickelte System auf den Transkriptionsergebnissen des im Fraunhofer IDMT entwickelten *Transcription Toolbox* System aufbaut, ist es naheliegend, die implementierten Funktionalitäten in einem Plugin zu bündeln,

welches es dem Nutzer erlaubt, im Anschluss an die Transkription eines Soloparts oder einer Melodie weitere semantisch hochwertige Beschreibungen zu erhalten. Dies könnte z.B. eine Einordnung des Solos in ein bestimmtes Genre sein.

# 5 Evaluation

## 5.1 Genreerkennung

### 5.1.1 Aufgabenstellung

Aus dem Untersuchungsgegenstand dieser Arbeit leitet sich die Fragestellung ab, inwieweit eine Genreklassifikation eines Musikstückes ausschließlich anhand des untersuchten Soloparts mit Hilfe der in dieser Arbeit dargestellten *High-level*-Merkmale möglich ist. Bei der Wahl des Klassifikators wurden zwei Ansätze gewählt:

1. Klassifikation mit Hilfe eines SVM-Klassifikators (vgl. Kap. 2.3.3) mit vorausgegangener LDA (vgl. Kap. 2.3.2)
2. Klassifikation anhand der rhythmischen Strukturprofile (vgl. Kap. 4.2.3) aller Begleitinstrumente

Zur Evaluation der Genreklassifikation wurden zwei verschiedene Testsets zusammengestellt – eines bestehend aus 300 MIDI-Dateien, das zweite aus 240 MP3-Dateien. Die Testsets enthalten jeweils 50 bzw. 40 repräsentative Soloparts aus den sechs in Kap. 2.1.3 genannten musikalischen Genres. Im Anhang B ist die genaue Zusammenstellung der Testsets sowie die entsprechenden Quellen aufgelistet. Ein zusätzliches Kriterium für die Auswahl der Soli stellte die in Kap. 2.1.4 geforderte Instrumentierung dar. Pro Testset und Klassifikator erfolgte eine 50-fache Kreuzvalidierung (*cross validation*), dabei werden die Ergebnisse von 50 voneinander unabhängigen Aufteilungen der Testsets in Trainings- und Testdatensatz gemittelt.

### 5.1.2 Klassifikatoren

#### LDA-SVM-Klassifikator

Der erste Ansatz beschreibt einen SVM-Klassifikator, der anhand einer Auswahl der extrahierten Merkmalsvektoren eines Datensatzes (Trainingsdatensatz) dimensioniert wird. Zur Dimensionsreduktion des Merkmalsraumes wird die in Kap. 2.3.2 beschriebene LDA angewandt. Die in Kap. 4 erläuterten *High-level*-Merkmale können nicht nur für das Melodieinstrument, sondern zum Teil auch für die drei Begleitinstrumente berechnet werden – die genaue Zuordnung ist in Anhang C dargestellt. Auf diese Weise lassen sich mit Hilfe aller Merkmale, die jeweils einem Instrument zugeordnet sind,

semantische Einzel-Klassifikatoren trainieren. Diese bilden jeweils einen Teil des musikalischen Wissens ab, welches auf der semantischen Ebene der *High-level*-Merkmale aus dem Solopart extrahiert werden kann. Über eine Mehrheitsentscheidung kann zusätzlich ein Ensemble-Klassifikator eine Gesamtentscheidung treffen.

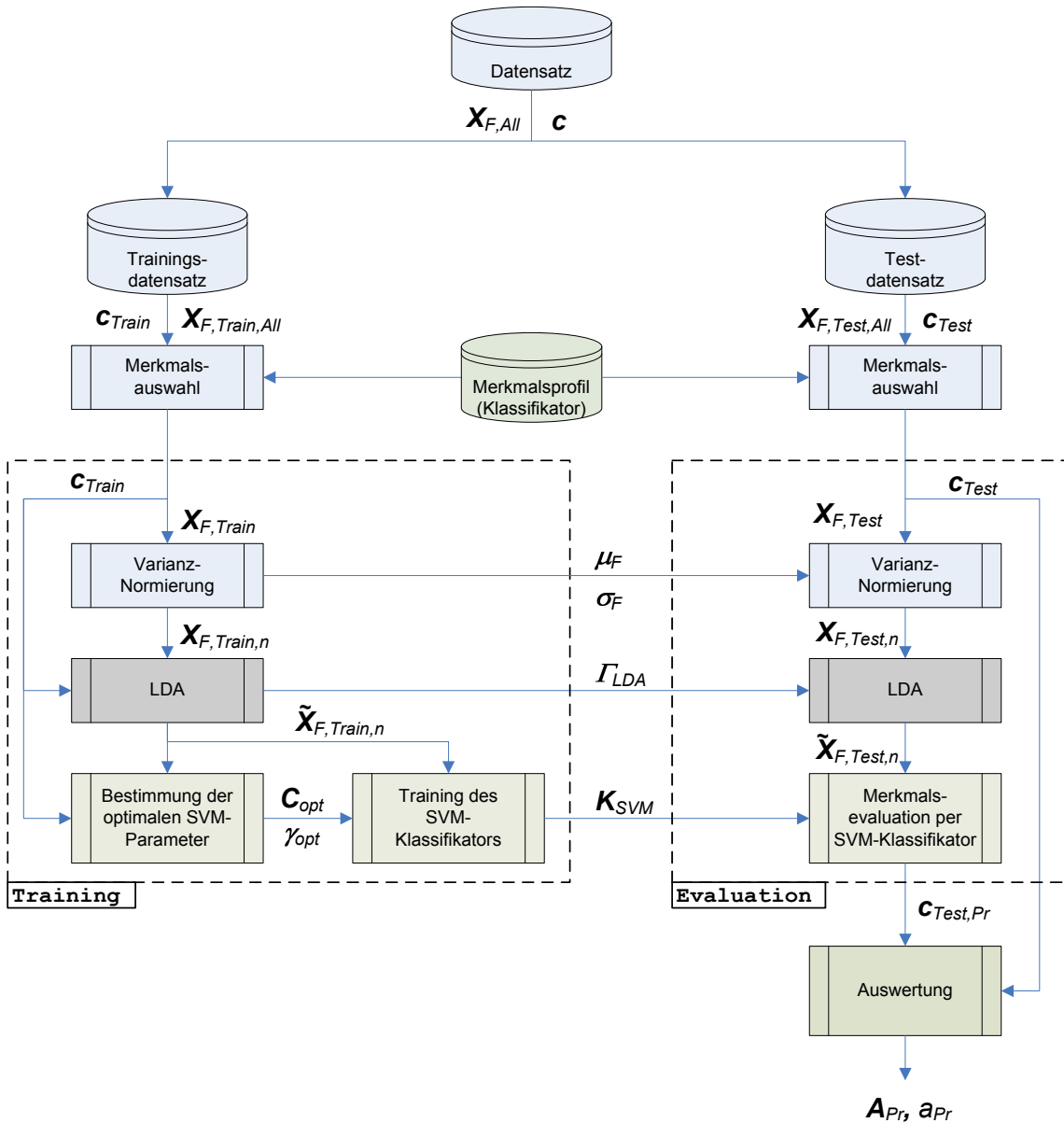


Abb. 5.1: Ablauf der Evaluation des Einzel-SVM-LDA-Klassifikators

**Evaluation der Einzel-Klassifikatoren** Der Ablauf der Evaluation ist in Abb. 5.1 dargestellt. Für die Implementierung unter MATLAB wurde neben einem bestehenden Skript zur Berechnung der LDA die LibSVM-Toolbox<sup>1</sup> zur Applikation des SVM-

<sup>1</sup>Vgl. [CL01].

Klassifikators verwendet.

Ausgangspunkt ist ein Datensatz, der sich aus einer Matrix  $\mathbf{X}_{F,All} \in \mathbb{R}^{N_S \times N_F}$  und einem Vektor  $\mathbf{c} \in \mathbb{Z}^{N_S}$  zusammensetzt. Die Matrix  $\mathbf{X}_{F,All}$  beinhaltet die extrahierten Merkmalsvektoren ( $N_F$  Merkmale) aller  $N_S$  Soli im jeweiligen Testset und der Vektor  $\mathbf{c}$  die Genrezuordnung aller Soli in Form eines Klassenlabels. Es gilt

$$\mathbf{c}(i) \in \mathbf{C}$$

$$\mathbf{C} = \{1, 2, 3, 4, 5, 6\} \Leftrightarrow \mathbf{C} = \{\text{SWI, LAT, FUN, BLU, POP, MHR}\} .$$

Zuerst erfolgt eine Aufteilung des Datensatzes in einen Trainings-

$$\mathbf{X}_{F,Train,All} \in \mathbb{R}^{N_S/2 \times N_F} ; \quad \mathbf{c}_{Train} \in \mathbb{Z}^{N_S/2}$$

sowie in einen Testdatensatz

$$\mathbf{X}_{F,Test,All} \in \mathbb{R}^{N_S/2 \times N_F} ; \quad \mathbf{c}_{Test} \in \mathbb{Z}^{N_S/2} .$$

Die Unterteilung erfolgt hierbei zufällig, jedoch so, dass sich alle Soli jedes Genre jeweils zu 50% auf beide Datensätze verteilen. Für jeden Einzel-Klassifikator wird nur eine Menge von  $N_{F,C}$  Merkmalen verwendet, die dem jeweils repräsentierten Instrument zugeordnet werden. Anschließend erfolgt eine Varianznormierung bezogen auf jedes Merkmal. Dafür werden die beiden Vektoren  $\boldsymbol{\mu}_F \in \mathbb{R}^{N_{F,C}}$  sowie  $\boldsymbol{\sigma}_F \in \mathbb{R}^{N_{F,C}}$  berechnet, die die spaltenweise Mittelwerte und Standardabweichungen der Matrix  $\mathbf{X}_{F,Train}$  enthalten. Es gilt dabei

$$\boldsymbol{\mu}_F(i) = \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{X}_{F,Train}(j, i) \quad (5.1)$$

$$\forall 1 \leq i \leq N_{F,C}$$

sowie

$$\boldsymbol{\sigma}_F(i) = \sqrt{\frac{1}{N_S} \sum_{j=1}^{N_S} [\mathbf{X}_{F,Train}(j, i) - \boldsymbol{\mu}_F(i)]^2} \quad (5.2)$$

$$\forall 1 \leq i \leq N_{F,C} .$$

Die Varianznormierung der Merkmalsmatrix auf einen spaltenweisen Mittelwert von 0, sowie auf eine spaltenweise Varianz von 1 erfolgt anschließend elementweise.

$$\mathbf{X}_{F,Train,n}(i, j) = \frac{\mathbf{X}_{F,Train}(i, j) - \boldsymbol{\mu}_F(j)}{\boldsymbol{\sigma}_F(j)} \quad (5.3)$$

$$\forall 1 \leq i \leq N_S \text{ sowie } \forall 1 \leq j \leq N_{F,C}$$

Im folgenden Schritt wird die Dimension des Merkmalsraumes mit Hilfe der LDA reduziert. Die Observation der Transformationsmatrix

$$\mathbf{\Gamma}_{LDA} \in \mathbb{R}^{N_F \times N_{PC}}$$

zeigt, dass praktisch die gesamte Varianz im Merkmalsraum auf die ersten fünf Dimensionen verteilt waren (vgl. Kap. 2.3.2). Daraus resultiert die Wahl von  $N_{PC} = 5$  – dies entspricht einer Anzahl von fünf Dimensionen im transformierten Merkmalsraum.

$$\widetilde{\mathbf{X}}_{F,Train,n} = \mathbf{X}_{F,Train,n} \cdot \mathbf{\Gamma}_{LDA} , \quad (5.4)$$

wobei gilt:

$$\widetilde{\mathbf{X}}_{F,Train,n} \in \mathbb{R}^{N_S \times N_{PC}} .$$

Um die für den transformierten Trainingsdatensatz optimale Parameterkonfiguration  $[C_{opt}, \gamma_{opt}]$  der RBF-Kernelfunktion<sup>2</sup> zu bestimmen, wird das in [HCL08] vorgeschlagene Verfahren angewandt. Zuerst erfolgt eine Parametersuche in einem groben Raster (Schrittweite : 2).

$$C = 2^{-5}, 2^{-3}, \dots, 2^{15}$$

$$\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$$

Für jede Parameterkonfiguration wird mit Hilfe der MATLAB-Funktion `svmtrain` ein SVM-Klassifikator trainiert und seine Klassifikationsgenauigkeit über einen Prädiktionstest mit Hilfe der Funktion `svmpredict` ermittelt. Die angrenzenden Wertebereiche der beiden Parameterwerte mit der höchsten Klassifikationsquote werden anschließend nach dem selben Verfahren in zwei Schritten untersucht. Dies erfolgt mittels immer feiner unterteilten Raster – zuerst mit einer Schrittweite von 0,5 um  $[C_{opt''}, \gamma_{opt''}]$

$$C = 2^{C_{opt''}-2}, 2^{C_{opt''}-1,5}, \dots, 2^{C_{opt''}+2}$$

$$\gamma = 2^{\gamma_{opt''}-2}, 2^{\gamma_{opt''}-1,5}, \dots, 2^{\gamma_{opt''}+2}$$

und anschließend mit 0,125 um  $[C_{opt'}, \gamma_{opt'}]$

$$C = 2^{C_{opt'}-0,5}, 2^{C_{opt'}-0,375}, \dots, 2^{C_{opt'}+0,5}$$

$$\gamma = 2^{\gamma_{opt'}-0,5}, 2^{\gamma_{opt'}-0,375}, \dots, 2^{\gamma_{opt'}+0,5} .$$

Mit Hilfe der beiden optimalen Parameter  $[C_{opt}, \gamma_{opt}]$ , sowie der Matrix  $\widetilde{\mathbf{X}}_{F,Train,n}$  und dem Vektor  $\mathbf{c}_{Train}$  kann nun abschließend der Klassifikator mit Hilfe der Funktion

---

<sup>2</sup>Vgl. Kap. 2.3.3.



`svmtrain` trainiert werden. Diese Funktion gibt eine Datenstruktur  $\mathbf{K}_{SVM}$  zurück, welche den trainierten Klassifikator repräsentiert.

Die Merkmalsmatrix  $\mathbf{X}_{F,Test}$  des Testdatensatzes wird analog zu Gleichung 5.3 mit Hilfe der beiden Vektoren  $\boldsymbol{\mu}_F$  und  $\boldsymbol{\sigma}_F^2$  spaltenweise varianznormiert. Anschließend erfolgt mit Hilfe der Transformationsmatrix  $\boldsymbol{\Gamma}_{LDA}$  eine erneute LDA.

$$\widetilde{\mathbf{X}}_{F,Test,n} = \mathbf{X}_{F,Test,n} \cdot \boldsymbol{\Gamma}_{LDA} \quad (5.5)$$

Über die Funktion `svmpredict` wird nun abschließend der mit den Trainingsdaten trainierte Klassifikator  $\mathbf{K}_{SVM}$  mit Hilfe der Testdaten evaluiert. Die Prädiktionen des Klassifikators (Zuordnungshypothesen) werden in Form von Klassenlabels in einem Vektor  $\mathbf{c}_{Pred} \in \mathbb{Z}^{N_S/2}$  gespeichert. Die über alle Genres gemittelte prozentuale Klassifikationsgenauigkeit  $a_{Pr}$  ergibt sich aus

$$a_{Pr} = \frac{1}{N_S/2} \sum_{k=1}^{N_S/2} c_k ; \quad c_k = \begin{cases} 1, & \text{falls } \mathbf{c}_{Pred}(k) \equiv \mathbf{c}_{Test}(k) \\ 0, & \text{sonst.} \end{cases} \quad (5.6)$$

Anschließend kann die Konfusionsmatrix  $\mathbf{A}_{Pr} \in \mathbb{R}^{6 \times 6}$  der Genreklassifikation berechnet werden.

$$\mathbf{A}_{Pr}(m, n) = \sum_{k=1}^{N_S/2} c_{i,k} ; \quad c_{i,k} = \begin{cases} 1, & \text{falls } \mathbf{c}_{Test}(k) \equiv m \text{ und } \mathbf{c}_{Pred}(k) \equiv n \\ 0, & \text{sonst} \end{cases} \quad (5.7)$$

$$\forall 1 \leq m \leq 6 ; \quad \forall 1 \leq n \leq 6 .$$

Nach einer abschließenden zeilenweisen Normierung der Matrix  $\mathbf{A}_{Pr}$  auf eine Summe von 1 entspricht der Wert  $\mathbf{A}_{Pr}(m, n)$  dem Prozentsatz aller Soli des Genres  $m$ , welche der Klassifikator dem Genre  $n$  zugeordnet hat. Je besser der Klassifikator funktioniert, desto geringer sind alle Werte von  $\mathbf{A}_{Pr}$  für  $m \neq n$  und desto stärker nähert sich  $\mathbf{A}_{Pr}$  der Einheitsmatrix  $\mathbf{I}^{6 \times 6}$  an. Für die Evaluation wurden für jeden Klassifikator 50 Iterationen durchgeführt, in denen der Datensatz (wie bereits beschrieben) jeweils zufällig in Trainings- und Testdatensatz aufgeteilt wurde. Abschließend erfolgt eine Mittelung über alle Ergebnisse.

$$\overline{\mathbf{A}}_{Pr} = \frac{1}{N} \sum_{k=1}^N \mathbf{A}_{Pr,k} \quad (5.8)$$

$$\overline{a}_{Pr} = \frac{1}{N} \sum_{k=1}^N a_{Pr,k} \quad (5.9)$$

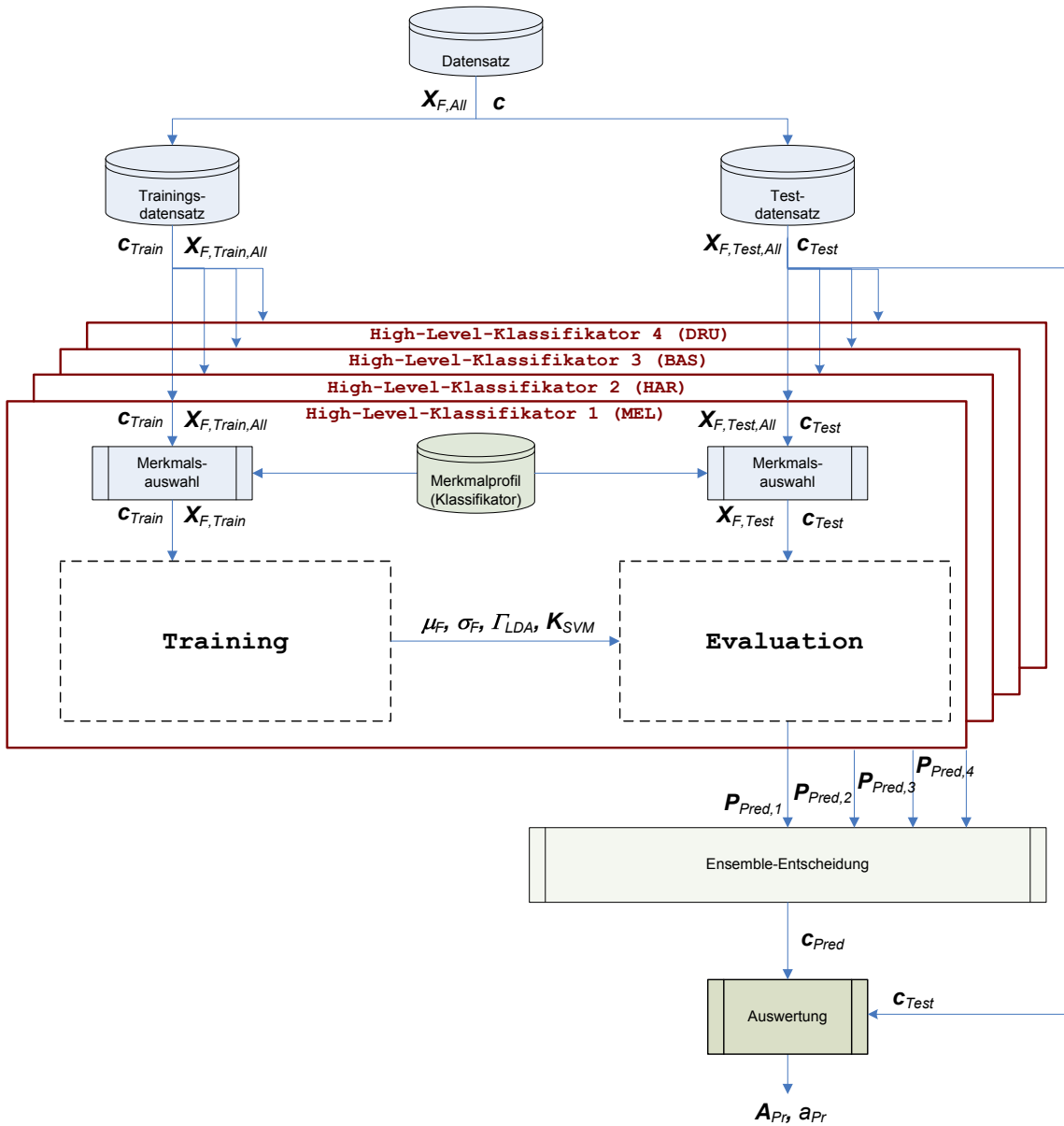


Abb. 5.2: Ablauf der Evaluation des Ensemble-SVM-LDA-Klassifikators

**Evaluation des Ensemble-Klassifikators** Neben den im vorherigen Abschnitt beschriebenen Einzel-Klassifikatoren wurde ein Ensemble-Klassifikator implementiert, der sich aus den vier semantischen Einzelklassifikatoren zusammensetzt und anschließend eine Ensemble-Entscheidung durchführt. Wie in Abb. 5.2 dargestellt, erfolgen Training und Evaluation analog zu den Einzelklassifikatoren<sup>3</sup>. Hierbei greifen alle Klassifikatoren auf den gleichen Trainings- und Testdatensatz zurück. Nach einer Mehrheitsentscheidung erfolgt eine Auswertung der Klassifikationsergebnisse. Die Funktion `svmpredict` der LibSVM-Toolbox ermöglicht die Ausgabe einer Matrix  $P_{Pred,i} \in \mathbb{R}^{N_S \times N_C}$ , welche

<sup>3</sup>Vgl. Abb. 5.1.

die Zuordnungswahrscheinlichkeiten aller  $N_S$  Soli im Testdatensatz zu den  $N_C$  Klassen (Genres) enthält.

Für die abschließende Ensemble-Entscheidung bezüglich der Klassenzuordnung werden die vier Matrizen  $\mathbf{P}_{Pred,i}$  in einer dreidimensionalen Matrix  $\mathbf{P}_{Pred} \in \mathbb{R}^{N_S \times N_C \times 4}$  gespeichert und anschließend der in Anhang A.1 beschriebene Algorithmus durchlaufen: Über einen Vergleich mit dem Vektor  $\mathbf{c}_{Test}$  können entsprechend der (5.7) die gemittelte prozentuale Klassifikationsgenauigkeit  $a_{Pr}$  und die Konfusionsmatrix  $\mathbf{A}_{Pr}$  berechnet werden. Sollten ein oder mehrere Instrumente aus dem Instrumentierungsmodell fehlen, können die verbleibenden Klassifikatoren trotzdem für eine Ensembleentscheidung genutzt werden, da so immer noch mehr semantische Information als bei der Verwendung von Einzel-Klassifikatoren für die Klassifikation berücksichtigt wird. Für den Ensemble-Klassifikator wurde ebenfalls wieder eine 50-fache Kreuzvalidierung zur Evaluation und Berechnung von  $\overline{\mathbf{A}}_{Pr}$  und  $\overline{a}_{Pr}$  gemäß den beiden Gleichungen in (5.9) durchgeführt.

### RSP-Klassifikator

Das in Kap. 4.2.3 beschriebene rhythmische Strukturprofil – konkreter das Energieprofil  $\mathbf{\Lambda}_{R2}$  (im Folgenden als  $\mathbf{\Lambda}$  bezeichnet) beschreibt die rhythmische Verteilung der Noten eines Instrumentes bezogen auf die Taktlänge. Es ist geeignet, um den rhythmischen Kontext<sup>4</sup> abzubilden, der von den Begleitinstrumenten während des Solos vorgegeben wird. Da es für jedes Genre typische rhythmische Muster der begleitenden Instrumente gibt, ist zu erwarten, dass eine Genreklassifikation des Soloparts anhand der korrespondierenden RSP der Begleitinstrumente möglich ist. Das im Folgenden beschriebene Verfahren ist in Abb. 5.3 dargestellt. Um den Klassifikator zu trainieren, werden für Harmonie-, Bass- und Schlaginstrument aus allen Soli innerhalb eines Trainingsdatensatzes global (über die gesamte Länge des Solos) die Energieprofile  $\mathbf{\Lambda}$  berechnet. Das Schlaginstrument wird hierbei getrennt für Bass-Drum und Snare-Drum betrachtet, damit die wichtigsten Rhythmen, die auf den verschiedenen Schlaginstrumenten des Schlagzeugsets gespielt werden, über das jeweilige RSP repräsentiert werden. Die rhythmischen Strukturprofile lassen sich jeweils in einer Matrix

$$\mathbf{\Lambda} \in \mathbb{R}^{13 \times 96 \times 2}$$

speichern (vgl. Kap. 4.2.3). Die Grundidee des RSP-Klassifikators besteht darin, alle aus einem Trainingsdatensatz gewonnenen Strukturprofile, die jeweils einem Instrument und einem Genre zugeordnet werden können, in einem Container

$$\mathbf{C}_{G,I} \in \mathbb{R}^{13 \times 96 \times 2 \times N_{G,I}}$$

---

<sup>4</sup>Vgl. Kap. 2.1.4.

zu speichern.  $N_{G,I}$  entspricht dabei der Anzahl an Strukturprofilen, die nach einer Trainingsphase des Klassifikators für das Genre  $G$  und das Instrument  $I$  vorliegen. Um ein unbekanntes Solo einem konkreten Genre zuzuordnen, werden die Strukturprofile  $\Lambda_n$  aller enthaltenen Begleitinstrumente extrahiert. Anschließend wird für jedes Profil  $\Lambda_I$  die minimale Distanz  $d_{min,I,G}$  zu jedem der entsprechenden sechs Container  $\mathbf{C}_{G,I}$  gesucht. Hierbei wird die euklidischer Distanz

$$d = d_{Eukl}(\Lambda_{I,1}, \Lambda_{I,2}) = \sqrt{\sum_k \sum_j \sum_k \{\Lambda_{I,1}(i, j, k) - \Lambda_{I,2}(i, j, k)\}^2} \quad (5.10)$$

verwendet. Alle ermittelten Distanzen  $d_{min,I,G}$  werden jeweils in einem Vektor  $\mathbf{d}_{min,I} \in \mathbb{R}^6$  gespeichert und können anschließend in Zuordnungswahrscheinlichkeiten bezüglich des jeweiligen Genre umgerechnet werden:

$$\tilde{\mathbf{d}}_{min,I} = \frac{1}{\max\{\mathbf{d}_{min,I}\}} [\max\{\mathbf{d}_{min,I}\} \cdot \mathbf{I}^{1 \times 6} - \mathbf{d}_{min,I}] \quad (5.11)$$

$$\mathbf{p}_{Pred,I} = \frac{1}{\sum_{k=1}^6 \tilde{\mathbf{d}}_{min}(k)} \tilde{\mathbf{d}}_{min} \quad (5.12)$$

Je geringer die Distanz, desto höher die Ähnlichkeit zu einem bestehenden RSP, das einem bestimmten Genre zugeordnet wurde. Der hier vorgestellte Klassifikator entspricht also einem *Nearest-Neighbour*-Klassifikator. Die Vektoren  $\mathbf{p}_{Pred,I}$  werden für alle Instrumente berechnet und zeilenweise zu einer Matrix  $\hat{\mathbf{P}}_{Pred,RSP} \in \mathbb{R}^{6 \times 5}$  zusammengefasst. Nachdem dies für alle  $N_S$  Soli des Testdatensatzes geschehen ist, wird mit Hilfe der Matrix  $\mathbf{P}_{Pred,RSP} \in \mathbb{R}^{N_S \times 6 \times 5}$  und dem im Anhang A.1 beschriebenen Algorithmus die Ensemble-Entscheidung  $\mathbf{c}_{Pred}$  gewonnen. Die Evaluation des RSP-Klassifikators erfolgte ebenfalls über eine 50-fache Kreuzvalidierung (vgl. Kap 5.1.2).

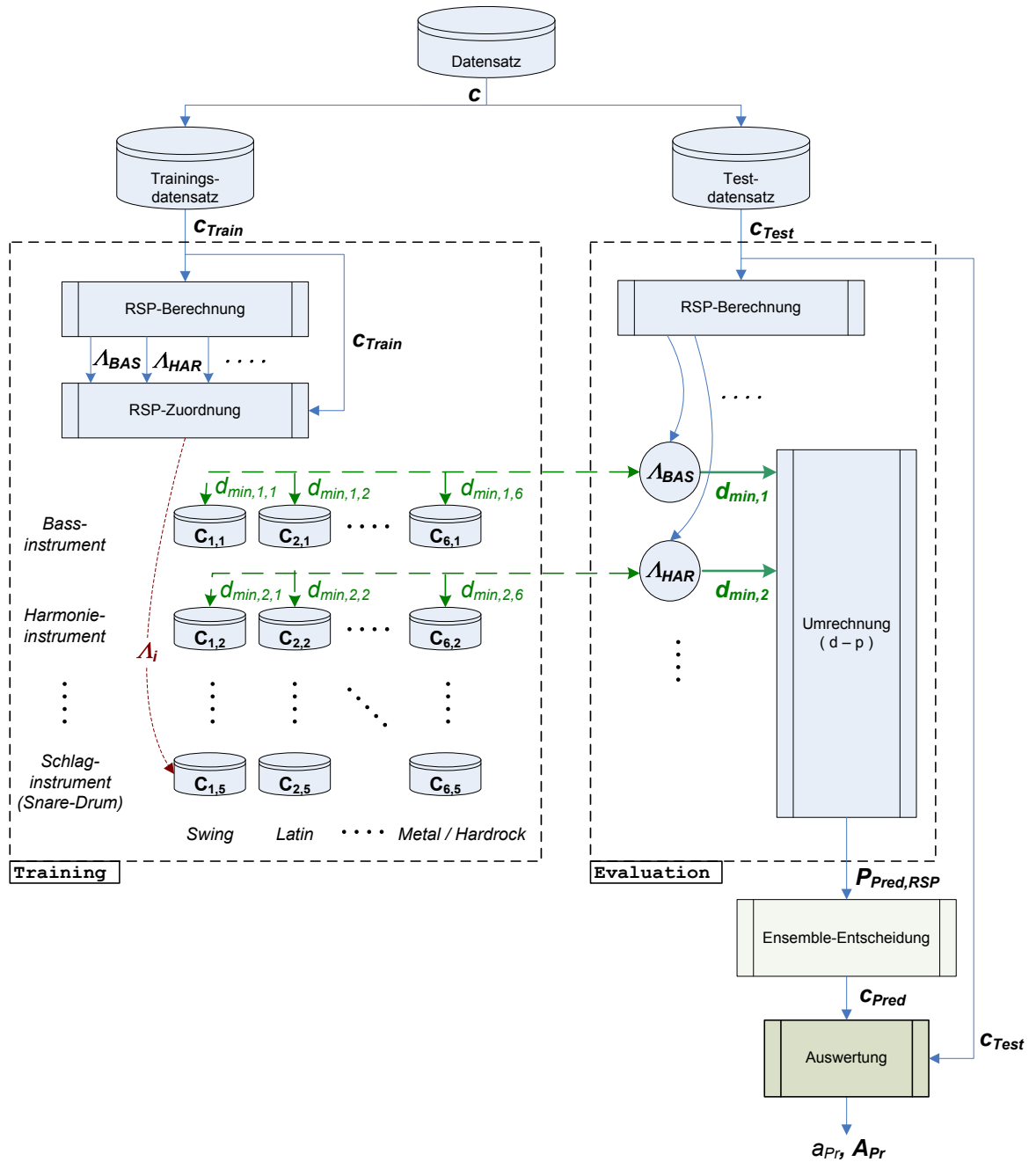


Abb. 5.3: Ablauf der Evaluation des RSP-Klassifikators

### 5.1.3 Hörtest

Um die Genauigkeit, die die beiden vorgestellten Klassifikatoren bei der Genreerkennung erreichten, qualitativ besser einordnen zu können, wurde im Rahmen dieser Arbeit zusätzlich ein Hörtest durchgeführt. Hierbei hatten 25 Probanden ebenfalls die Aufgabenstellung, ein unbekanntes Musikstück anhand seines Soloparts einem von sechs Genres zuzuordnen. Die Zusammenstellung der Probanden bezogen auf Geschlecht (männlich: rot, weiblich: grün), Alter und Jahre der musikalischen Vorbildung bzw. Praxis ist in Abb. 5.7(a) dargestellt.

Der Hörtest wurde über eine Internetseite durch Anmeldung und das anschließende Ausfüllen mehrerer Formulare absolviert. Das Web-Interface wurde in der Sprache HTML<sup>5</sup> sowie der Skriptsprache PHP<sup>6</sup> implementiert. Vor dem eigentlichen Hörtest wurden dem Probanden die Möglichkeit gegeben, sich jeweils zwei repräsentative Soloparts der sechs verschiedenen Genres anzuhören. Jedem Probanden wurden nacheinander 15 Soloparts vorgespielt, die zufällig aus dem erstellten MIDI-Testset<sup>7</sup> ausgewählt wurden. Jedes Solo musste genau einem Genre zugeordnet werden.

Unabhängig von der originalen Instrumentierung des jeweiligen Musikstückes, wurden zur akustischen Wiedergabe die Soloparts in der Art modifiziert, dass sowohl das Melodie-, Harmonie- als auch das Bassinstrument von einem Piano gespielt wurden. Ziel hierbei war es, eine Erkennung des Solos anhand einer genretypischer Instrumentierung – wie z.B. der E-Gitarre in den Genres Pop-Rock oder Metal-Hardrock – zu verhindern. Zusätzlich wurde ein Metronom zur rhythmische Orientierung hinzugefügt. In den ersten fünf Soli wurde ausschließlich die Melodiespur, in weiteren fünf die Melodie- und Harmoniespur und in den letzten fünf zusätzlich dazu noch die Bass- und Schlagzeugspur abgespielt. Ziel dieser Unterteilung war es herauszufinden, welchen Einfluss eingeschränkte musikalische Information in Form einer begrenzten Instrumentierung auf die Genreerkennung hat.

Neben den erreichten Klassifikationsergebnissen der Probanden werden in Kap. 5.1.4 zum Vergleich auch die Ergebnisse eines modifizierten LDA-SVM-Ensemble-Klassifikators dargestellt. Dieser konnte in jedem der betrachteten Szenarios im Rahmen Merkmalsextraktion auf die gleichen Instrumentalspuren zurückgreifen, die jeweils auch den Probanden zur Verfügung standen.

---

<sup>5</sup>Hypertext Markup Language.

<sup>6</sup>PHP Hypertext Preprocessor / Personal Home Page Tools.

<sup>7</sup>Vgl. Anhang B.

### 5.1.4 Ergebnisse

#### LDA-SVM-Klassifikator

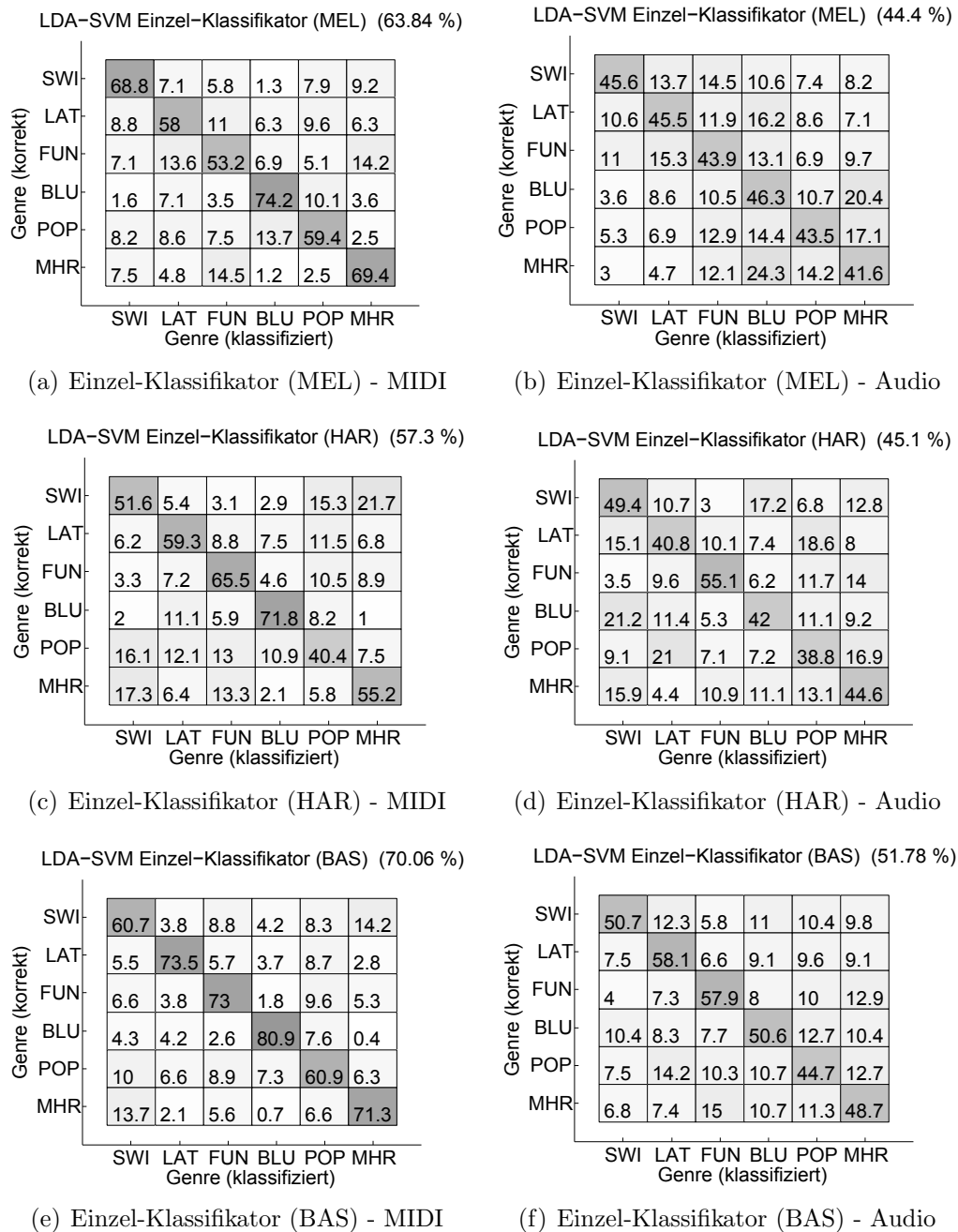


Abb. 5.4: Ergebnisse der LDA-SVM-Klassifikatoren (Vergleich : MIDI - Audio)

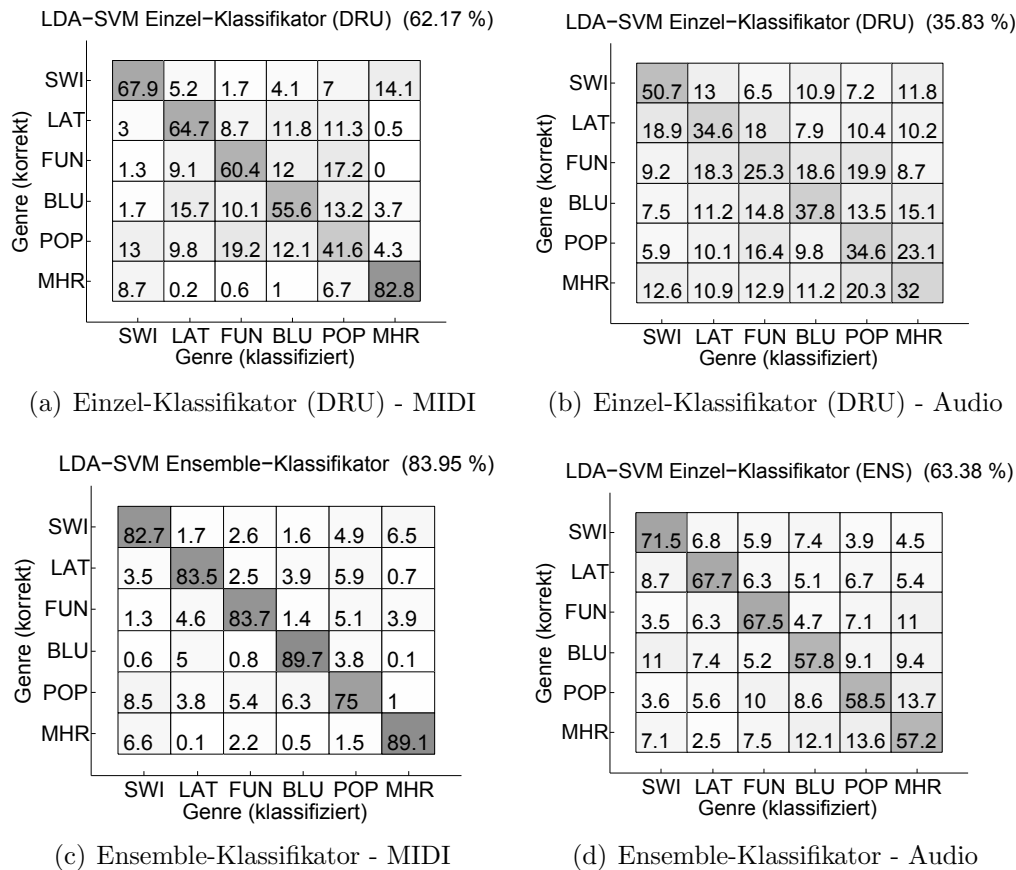


Abb. 5.5: Ergebnisse der LDA-SVM-Klassifikatoren (Vergleich : MIDI - AUDIO)

Die Evaluationsergebnisse zeigen, dass die in dieser Arbeit beschriebenen *High-level*-Merkmale für die Aufgabenstellung der Genreklassifikation anhand von Soloparts generell nützlich sind. Die Einzel-Klassifikatoren, welche jeweils nur die Einzelspuren der beteiligten Instrumente betrachten, erreichen durchschnittliche Erkennungsquoten von etwa 60% (vgl. Abb. 5.4(a), 5.4(c) und 5.5(a)), für die Bassspur von rund 70% (vgl. Abb. 5.4(e)). Durch eine Ensemble-Entscheidung wird die Information, die aus allen beteiligten Instrumenten gewonnen werden kann, berücksichtigt. Die Klassifizierungsergebnisse verbessern sich dadurch für jedes Genre im Vergleich zum jeweils besten Einzel-Klassifikator um 7 bis 15% auf durchschnittlich rund 84% (vgl. Abb. 5.5(c)). Die jeweils korrespondierenden Ergebnisse des Audio-Genre-Testsets sind im Durchschnitt um knapp 20 % schlechter (vgl. Abb. 5.4(b), 5.4(d) und 5.4(f)), vor allem für die Merkmale, die die Schlagzeugspur charakterisieren, sind die Ergebnisse unbefriedigend (vgl. Abb. 5.5(b)). Auch hier erreicht der Ensemble-Klassifikator die beste mittlere Erkennungsquote (63,4 %).



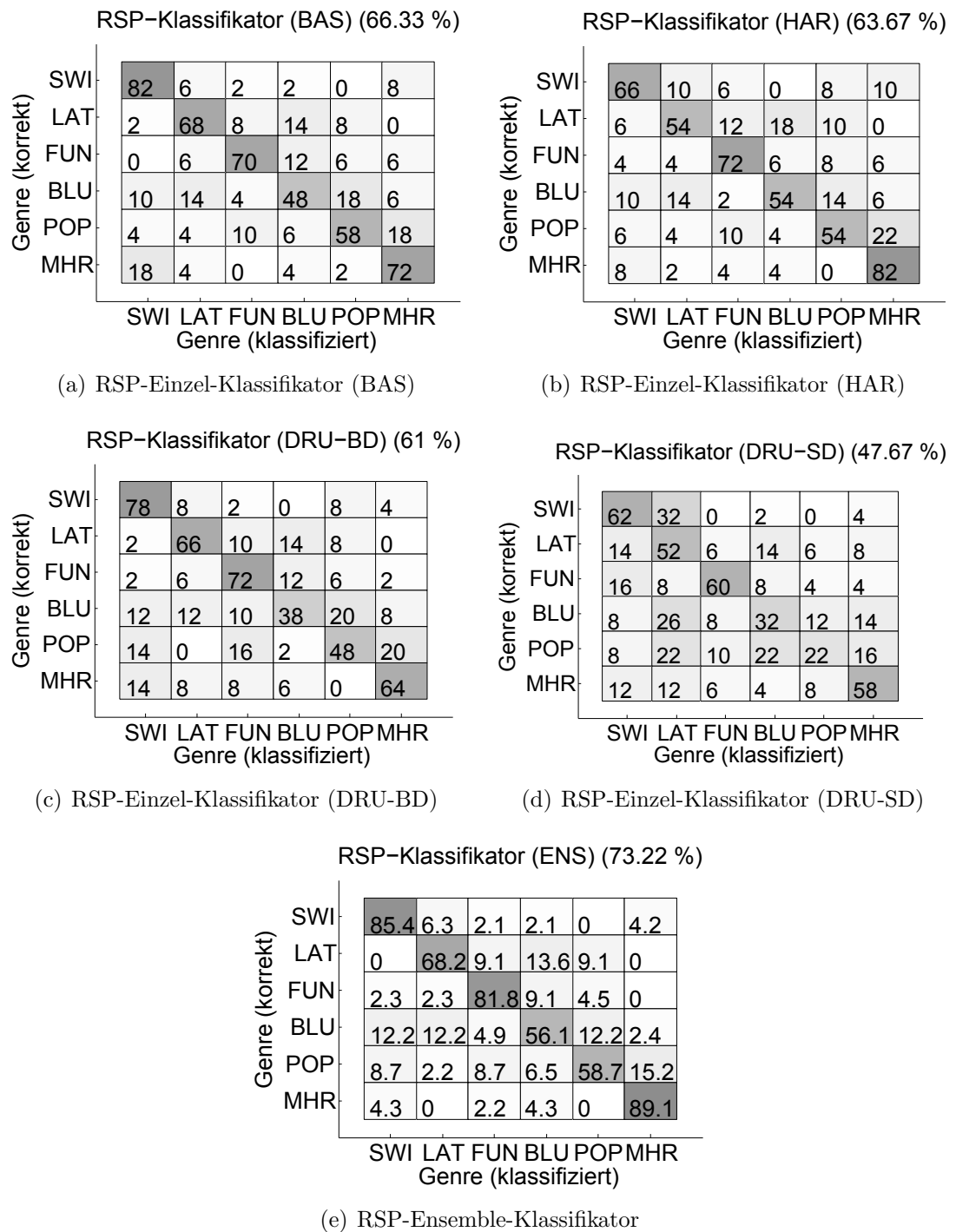


Abb. 5.6: Ergebnisse der RSP-Klassifikatoren

## RSP-Klassifikator

Die Ergebnisse der RSP-Klassifikatoren sind stark genreabhängig. Wie in Abb. 5.6(a) zu sehen ist, ermöglicht die Beschreibung des durch das Bassinstrument definierten rhythmischen Kontext für die Genre Swing, Latin, Funk sowie Metal-Hardrock eine qualitativ gute Genreklassifikation. Für das Harmonieinstrument (vgl. Abb. 5.6(b)) ist dies für Swing, Funk und Metal-Hardrock der Fall. Die bereits erläuterte Aufteilung des Schlagzeugs in die zwei Gruppen Bass-Drum und Snare-Drum erweist sich als günstig, um entsprechende Klassifikatoren anhand der einzelnen Spuren zu trainieren. Wie in Abb. 5.6(c) und 5.6(d) zu sehen ist, werden wieder für die drei Genres Swing, Funk sowie Metal-Hardrock gute Klassifikationsergebnisse gewonnen.

Für die beiden Genres Blues und Pop-Rock wird von den RSP-Klassifikatoren eine maximale Erkennungsquote von lediglich 56% bzw. 58% erreicht. Hier scheint der rhythmische Kontext der Begleitinstrumente nur einen kleinen Anteil der diskriminativen Eigenschaften zu enthalten, die charakteristisch für die beiden Genres sind. Der LDA-SVM-Ensemble-Klassifikator (vgl. Abb. 5.5(c)) erreicht im Vergleich mit 89% bzw. 75% deutlich bessere Ergebnisse. Der RSP-Klassifikator wurde ausschließlich mit dem MIDI-Genre-Testset evaluiert.

## Hörtest

Die Ergebnisse des Hörtest (Abb. 5.8) zeigen, dass der Mensch trotz verhältnismäßig hoher musikalischer Vorbildung (vgl. Abb. 5.7(a)) nur bedingt in der Lage ist, ein unbekanntes Musikstück ausschließlich anhand des Soloparts einem Genre zuzuordnen. Wie in Abb. 5.7(c) und 5.7(d) zu sehen ist, gibt es keine erkennbare Korrelation zwischen erreichter Erkennungsgenauigkeit und Alter bzw. musikalischer Vorbildung. Die vereinheitlichte Instrumentierung unabhängig des jeweiligen Genres sowie das Fehlen bestimmter Instrumentalspuren in den ersten beiden Szenarios erschwerte die Erkennung. Dies bestätigten viele Probanden im Anschluss an den Hörtest.

Die Auswertung des individuellen Musikgeschmackes der Probanden zeigt, dass nur ein geringer Teil Musikstücke aller untersuchten Genres regelmäßig hört – in der Regel trifft dies nur für zwei bis drei Genres zu. Somit ist das Wissen, mit dem ein maschineller LDA-SVM- bzw. RSP-Klassifikator trainiert wird, wesentlich ausgewogener als jenes, welches einem Probanden mit „durchschnittlichem“ Musikgeschmack für die Bewältigung der ihm gestellten Aufgabe zur Verfügung steht. Um den Musikgeschmack der Probanden quantitativ zu erfassen, sollte für jedes Genre eine Bewertung  $x_G \in [1, 5]$  abgegeben werden, wobei  $x_G = 1$  zum Ausdruck bringt, dass der Proband dieses Genre praktisch nie hört und  $x_G = 5$ , dass er es sehr häufig hört. In Abb. 5.7(b) ist ähnlich einem *box plot* der Mittelwert plus minus die Standardabweichung von  $\overline{x_G}$  abgetragen.

Wie zu erkennen ist, sind die Mittelwerte für die Genre Swing, Latin, Funk und Blues fast identisch, diese werden in Durchschnitt seltener gehört. Für die beiden anderen Genres – vor allem aber Pop-Rock – erreichen sie höhere Werte. Die Standardabweichung ist bis auf Metal-Hardrock nahezu konstant über die restlichen Genres.

In den ersten beiden Szenarios wurde das Genre Pop-Rock am besten erkannt, was darauf zurückgeführt werden kann, dass die Probanden die entsprechenden Musikstücke oftmals aufgrund deren hohen Bekanntheitsgrades wiedererkannten. Weiterhin ist zu sehen, dass die Stücke der beiden Genres Swing und Blues dem entsprechend anderen Genre zugeordnet wurden – beide Genres werden durchschnittlich eher selten gehört. Die hohe Erkennungsquote des Genres Latin im dritten Szenario (mit kompletter Instrumentierung) führt der Autor auf die genretypische Präsenz von Perkussions-Instrumenten zurück, aufgrund derer die Stücke sehr häufig richtig zugeordnet wurden.

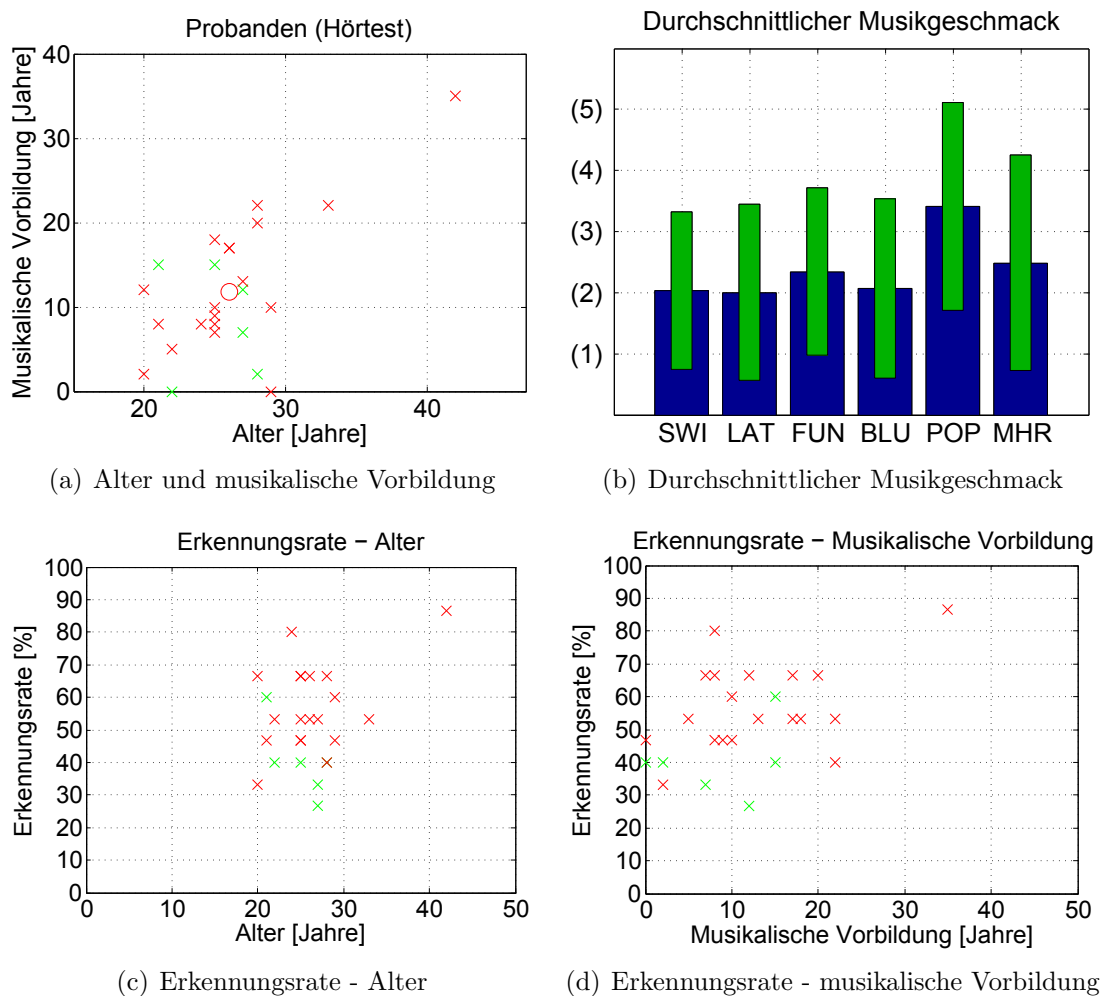


Abb. 5.7: Ergebnisse des Hörtests

Hörtest – Szenario 1 (37.55 %)

Genre (korrekt)	SWI	47.6	14.3	4.8	33.3	0	0
LAT	28	20	0	28	20	4	
FUN	30	30	20	5	10	5	
BLU	20	13.3	6.7	46.7	13.3	0	
POP	7.7	3.8	3.8	19.2	57.7	7.7	
MHR	5.6	16.7	11.1	11.1	22.2	33.3	
	Genre (klassifiziert)	SWI	LAT	FUN	BLU	POP	MHR

(a) Szenario 1 : Hörtest

LDA-SVM Szenario 1 (63.08 %)

Genre (korrekt)	SWI	65	6.4	6.9	3.2	11.3	7.2
LAT	5.5	62.1	11.4	5.2	8.3	7.5	
FUN	5.8	13.1	52.5	6.4	6.8	15.3	
BLU	4	6.7	2.5	72.1	12.1	2.6	
POP	13.5	8.2	7.6	10.6	57	3.1	
MHR	4.3	5.8	12.5	2.6	4.9	69.9	
	Genre (klassifiziert)	SWI	LAT	FUN	BLU	POP	MHR

(b) Szenario 1 : LDA-SVM-Klassifikator

Hörtest – Szenario 2 (58.79 %)

Genre (korrekt)	SWI	58.8	11.8	11.8	17.6	0	0
LAT	28	48	4	4	16	0	
FUN	23.5	5.9	47.1	5.9	17.6	0	
BLU	23.5	5.9	0	64.7	0	5.9	
POP	4	0	16	0	80	0	
MHR	8.3	4.2	8.3	4.2	20.8	54.2	
	Genre (klassifiziert)	SWI	LAT	FUN	BLU	POP	MHR

(c) Szenario 2 : Hörtest

LDA-SVM Szenario 2 (71.69 %)

Genre (korrekt)	SWI	69.9	1.9	6.2	2.5	9.5	10
LAT	4.5	70.7	8.3	4.8	8.4	3.3	
FUN	4.5	7.4	67.7	3.6	5.6	11.2	
BLU	1.2	4.9	1.6	85.1	6.3	1	
POP	7.4	8.6	8.3	10.9	62.4	2.6	
MHR	7.1	3.4	9.4	2.6	3	74.4	
	Genre (klassifiziert)	SWI	LAT	FUN	BLU	POP	MHR

(d) Szenario 2 : LDA-SVM-Klassifikator

Hörtest – Szenario 3 (63.06 %)

Genre (korrekt)	SWI	56.5	13	8.7	21.7	0	0
LAT	10	80	10	0	0	0	
FUN	4.3	26.1	43.5	4.3	8.7	13	
BLU	25	0	5	65	5	0	
POP	0	5.6	0	27.8	66.7	0	
MHR	0	0	4.8	0	28.6	66.7	
	Genre (klassifiziert)	SWI	LAT	FUN	BLU	POP	MHR

(e) Szenario 3 : Hörtest

LDA-SVM Ensemble-Klassifikator (83.95 %)

Genre (korrekt)	SWI	82.7	1.7	2.6	1.6	4.9	6.5
LAT	3.5	83.5	2.5	3.9	5.9	0.7	
FUN	1.3	4.6	83.7	1.4	5.1	3.9	
BLU	0.6	5	0.8	89.7	3.8	0.1	
POP	8.5	3.8	5.4	6.3	75	1	
MHR	6.6	0.1	2.2	0.5	1.5	89.1	
	Genre (klassifiziert)	SWI	LAT	FUN	BLU	POP	MHR

(f) Szenario 3 : LDA-SVM-Klassifikator

Abb. 5.8: Korrespondierende Ergebnisse: Hörtests und LDA-SVM-Klassifikator

## 5.2 Musikererkennung

### 5.2.1 Aufgabenstellung

Um die Eignung der entwickelten Merkmale zur Erkennung einzelner Musiker anhand von ihnen gespielter Soli zu evaluieren, wurden zwei weitere Testsets bestehend aus realen Audiodaten im MP3-Format zusammengestellt. Sie setzen sich jeweils aus 30 Ausschnitten aus Soli von je vier Künstlern zusammen, die das gleiche Instrument spielen und verwandten Genres zugeordnet werden können (Bebop und Modern Jazz bzw. Blues und Blues-Rock). Die Zusammenstellung der beiden Testsets ist in Tab. 5.1 sowie im Anhang B dargestellt.

Die Transkription, die anschließenden Vorverarbeitungsschritte sowie die Merkmalsextraktion erfolgt gemäß Abb. 4.1 genau wie für das Audio-Testset für die Genreklassifikation (vgl. Kap. 5.1). Da das Harmonie-, Bass- und Schlaginstrument für die Musikerklassifikation keine Rolle spielen, wird der bereits beschriebene LDA-SVM-Klassifikator eingesetzt, der auf den aus der Melodiespur extrahierten High-Level-Merkmalen basiert. Zusätzlich werden die in Kap. 4.2.5 beschriebenen Merkmale dem Klassifikator zur Verfügung gestellt, die die rhythmische und harmonische Interaktion zwischen Melodieinstrument und den Begleitinstrumenten abbilden.

### 5.2.2 Ergebnisse

Die Konfusionsmatrizen der zwei untersuchten Musikerkonstellationen für die beiden Instrumente E-Gitarre und Saxophon sind in Abb. 5.9 dargestellt, es wurden mittlere Erkennungsquoten von 58,7% und 56% erreicht.

Da die Musiker jeweils dem gleichen Instrument und dem gleichen Genre zuzuordnen sind, erlauben die verwendeten High-Level-Merkmale, die dem Klassifikator zur Verfügung stehen, nur eine unbefriedigende Klassifikation. Nach Meinung des Autors spielen hier vor allem klangliche Aspekte (Timbre des Instrumentes, Klang typischer Verstärkereinstellungen und Effekte bei den E-Gitarristen) sowie instrumentaltypische Spieltechniken bei der Unterscheidung eine große Rolle. Diese wurden in dieser Arbeit jedoch nicht berücksichtigt. Eine Kombination der implementierten High-Level-Merkmale mit weiteren Low- und Mid-Level-Merkmalen, welche die o.g. Aspekte abbilden können, verspricht eine starke Verbesserung der Klassifikationsgenauigkeit im Rahmen der Musikererkennung.

Testset	Interpreten		Instrument
Audio-Artist-Sax	John Coltrane	(COL)	Saxophon
	Dexter Gordon	(GOR)	
	Charlie Parker	(PAR)	
	Joshua Redman	(RED)	
Audio-Artist-Git	Eric Clapton	(CLA)	E-Gitarre
	Rory Gallagher	(GAL)	
	Jimi Hendrix	(HEN)	
	Steve Ray Vaughan	(VAU)	

Tab. 5.1: Testsets für die Musikerklassifikation

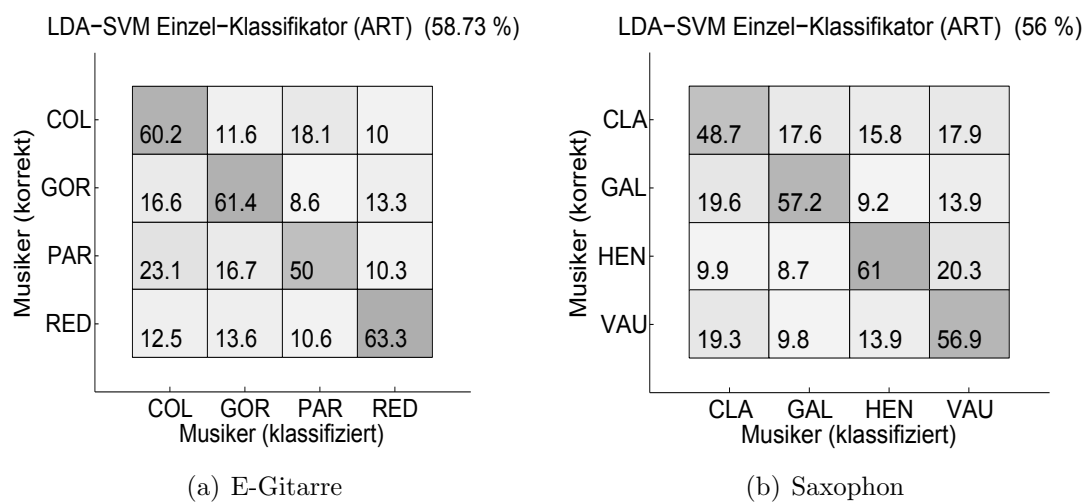


Abb. 5.9: Ergebnisse der Musikerklassifikation

## 6 Zusammenfassung und Ausblick

### 6.1 Fazit

Im Rahmen der vorliegenden Diplomarbeit wurden eine Vielzahl verschiedener *High-level*-Merkmale implementiert, welche Instrumentalspuren innerhalb verschiedener semantischer Ebenen beschreiben können (vgl. Kap. 4.2). Die Eignung dieser Merkmale zur Genre- und Musikerklassifikation wurde anhand von vier zusammengestellten Testsets (vgl. Anhang B) mit Hilfe verschiedener Klassifikatoren evaluiert. Die dabei erzielten Ergebnisse sind in Kap. 5.1 sowie 5.2 dargestellt und diskutiert.

Für die Genreklassifikation wird zusätzlich in Kap. 5.1.2 ein auf dem rhythmischen Strukturprofil basierender neuer Ansatz für einen *Nearest-Neighbour*-Klassifikator dargestellt, der gute Ergebnisse erzielte. Der Vergleich mit den Resultaten eines durchgeführten Hörtest zeigt, dass menschliche Hörer nur bedingt in der Lage sind, einen Solopart einem Genre zuzuordnen, wenn ihnen dabei keine Informationen bezüglich der Instrumentierung vorliegen.

Über die Ensemble-Entscheidung wurden sowohl für den LDA-SVM- als auch für den RSP-Klassifikator die besten Ergebnisse erreicht, sie ermöglicht die Auswertung der gesamten musikalischen Information, die durch die entsprechenden Einzelklassifikatoren abgebildet werden. Auch wenn im Allgemeinen der Höreindruck des Soloinstrumentes innerhalb eines Soloparts dominiert, erfolgt die Genrezuordnung tendenziell stärker über die (eigentlich hintergründige) musikalische Begleitung.

### Segmentierung

In der in Kap. 1.2 genannten Aufgabenstellung wurde neben der Genre- und Musikerklassifikation die Evaluation der *High-level*-Merkmale im Rahmen der Segmentierung von Musikstücken gefordert. Diese wurde in der vorliegenden Arbeit jedoch nicht näher untersucht. Der Grund ist vor allem die Tatsache, dass das *Transcription Toolbox* System, welches für die Verarbeitung von realen Audiodaten verwendet wurde, momentan nur für die Transkription von Audiostücken mit einer Länge von höchstens 30 Sekunden konzipiert ist. Musikstücke aus den in dieser Arbeit betrachteten Genres haben in der Regel aber mindestens 2 Minuten Spielzeit.

Nach Meinung des Autors sind besonders Informationen über den Dynamikverlauf und die Instrumentierung von großer Bedeutung bei der Dektektion von Soloparts. Beide Aspekte werden durch die hier vorgestellten *High-level*-Merkmale nicht abgebildet

(vgl. Kap. 4.2). In wiederkehrenden Teilen wie Strophe und Refrain identifiziert das *Transcription Toolbox* System häufig – falls vorhanden – die Gesangsmelodie als Melodieinstrument. Gerade in Musikstücken mit Soloparts, die wenige langsam gespielte Noten des Melodieinstrumentes enthalten, ist es denkbar, dass sich ohne genauere Information über das Melodieinstrument die extrahierten *High-level*-Merkmale mitunter für Solopart und Strophe oder Refrain nur geringfügig unterscheiden würden. Hierbei wird angenommen, dass sich die Spuren der Begleitinstrumente kaum ändern. Ähnliche Probleme sind bei der Detektion von Soloparts in MIDI-Dateien zu erwarten.

## 6.2 Ausblick

Neben der Genreentscheidung gemäß der Ensemble-Entscheidung der vorgestellten Klassifikatoren ist eine detailliertere Auswertung denkbar. Spielt beispielsweise die Rhythmusgruppe (Bass- und Schlaginstrument) einen eher latin-typischen Rhythmus während Harmonie- sowie Melodieinstrument swing-typisch spielen, so ist die Genreentscheidung des Ensemble-Klassifikators immer als unvollständig anzusehen, da sie auf einem Kompromiss beruht. Um ein Solo mit Hilfe der Klassifikatoren zu beschreiben, ist also eine zusätzliche Angabe der Ergebnisse aller Einzel-Klassifikatoren sinnvoll, da dies eine vollständigere und damit semantisch hochwertigere Beschreibung des analysierten Solos ermöglicht.

Alternativ zu dem in Kap. 5.1.2 vorgestellten Algorithmus der Ensemble-Entscheidung könnten anhand der Ergebnisse der Einzelklassifikatoren genrebezogene Wichtungsfaktoren ermittelt werden, mit denen die Aussagekraft einzelner Klassifikatoren bezüglich der Genreklassifikation unbekannter Soli abgebildet wird.

Um eine bessere Genreklassifikation zu ermöglichen, müssen weitere semantische Informationen über Merkmale abgebildet werden. Hierbei sind vor allem Merkmale bezüglich der Instrumentierung, der instrumententypischen Spielweise und des Timbres zu nennen. Der logische Schritt ist eine Kombination entsprechender *Low*- und *Mid-level*-Merkmale mit den in dieser Arbeit vorgestellten *High-level*-Merkmalen. Im folgenden Kapitel werden einige mögliche Anwendungsszenarien skizziert.



## 6.3 Einsatzszenarien

**Musikerprofil** Über die Extraktion von im Solo eingesetzten Spielweisen wie beispielsweise Vibrato oder Tremolo sowie dem typischen Klang der von einzelnen Musikern gespielten Instrumenten sowie häufig eingesetzter Effekte kann ein Musikprofil bestimmt werden, das den Solisten in seinem Spiel charakterisiert. Es ist zu erwarten, dass dies eine deutliche Verbesserung der Erkennungsraten im Kontext der Musikerklassifikation zur Folge hat.

Eine Charakterisierung von Gesangsspuren anhand ähnlicher timbraler und musikalischer Merkmale ist ebenfalls denkbar.

**Strukturanalyse** Wenn zukünftige Systeme eine zeiteffiziente und qualitativ zufriedenstellende Transkription von kompletten Musikstücken ermöglichen, können (lokal) extrahierten *High-level*-Merkmale für die Strukturanalyse von großem Nutzen sein. Nach Meinung des Autors eignen sich dafür besonders die rhythmischen Strukturprofile der transkribierten Spuren, da vor allem Musikstücke aus kontemporären Genres wie Pop oder Rock auf der Wiederholung typischer Rhythmen und einer klaren Unterteilung im Arrangement beruhen. Hier könnten verschiedene Clusterverfahren zur Extraktion dieser Rhythmen eingesetzt werden, um anschließend eine Segmentierung des untersuchten Musikstückes durchzuführen.

**Live-Visualisierung** Bei Live-Konzerten werden in der Regel alle Instrumente bzw. die entsprechenden Verstärker auf der Bühne getrennt mikrofoniert und gelangen auf verschiedenen Signalwegen zum Mischpult des Tontechnikers. Werden nun diese Spuren einzeln dem Transkriptionssystem zugeführt, so ist eine qualitativ bessere Transkription möglich, da die störenden Frequenzanteile von nicht gewünschten Instrumenten minimiert werden, die das jeweils zu transkribierende Signal überlagern.

Denkbar wäre eine Transkription einer Auswahl von Instrumenten – beispielsweise nach der Aufteilung Melodie-, Harmonie-, Bassinstrument und Schlagzeug. Bei Untersuchung kurzer zeitlicher Abschnitte des Signales sollte es möglich sein, eine Visualisierung (z.B. projiziert an eine Seitenwand neben der Bühne oder über den Musikern) anhand der in dieser Arbeit vorgestellten Mid- und Highlevelmerkmale umzusetzen. Am vielversprechendsten sind nach Meinung des Autors die interaktions- und rhythmusbezogenen Merkmale, da eine Visualisierung, welche sich an ihnen orientiert, die Zuhörer bzw. Zuschauer des Konzertes sehr intuitiv ansprechen wird.

**Visualisierung von musikalischer Entwicklung** Es ist vorstellbar, die zeitliche musikalische Entwicklung einzelner Genres anhand jeweils typischer Vertreter und den

von ihnen gespielten Soli zu untersuchen. Für das Genre Jazz könnte so beispielsweise versucht werden, die Entwicklung über alle Stile, die sich im Laufe des letzten Jahrhunderts entwickelt haben, abzubilden und so auch Verwandtschaften und Ähnlichkeiten untereinander darzustellen.

**Melodische Kontur** Um eine zeiteffiziente Analyse von Musikstücken auf High-Level-Ebene zu ermöglichen, wäre es denkbar eine Art „Melodische Kontur“ zu berechnen. Diese würde dem Verlauf der jeweils höchsten Notenlinie entsprechen – unabhängig davon, von welchem Instrument sie gespielt wird. Eine Analyse dieser Kontur mit Hilfe der in dieser Arbeit vorgestellten *High-level*-Merkmale könnte für die Strukturanalyse und Segmentierung von Musikstücken hilfreich sein, und hätte den Vorteil, dass die vorab nötige Transkription effizienter zu berechnen wäre.

# A Algorithmen

## A.1 Entscheidungsalgorithmus des Ensemble-Klassifikators

**Algorithmus** :  $[c_{Pred}] = \text{Ensemble-Entscheidung } (P_{Pred})$

**Eingabe** :  $P_{Pred}$  - Matrix mit Zuordnungswahrscheinlichkeiten der 4 Einzel-Klassifikatoren

**Ausgabe** :  $c_{Pred}$  - Vektor mit Klassenzuordnungen aller Stücke des Testdatensatzes

```

für  $k \leftarrow 1$  bis  $N_S$  tue
  wenn Eindeutige Mehrheitsentscheidung bezüglich der Klassifikation des
 $k$ -ten Teststückes möglich dann
     $c_{Pred}(k) = \max_{c_i} f(c_i)$  %  $f$  - Häufigkeit
  sonst
    % Mittelwertbildung der Zuordnungswahrscheinlichkeiten des  $k$ -ten
    % Teststückes über alle 4 Klassifikatoren
     $p_{Pred,k} \in \mathbb{R}^6$ 
    für  $c \leftarrow 1$  bis 6 tue
       $p_{Pred,k}(c) = 1/4 \sum_{n=1}^4 P_{Pred}(k, c, n)$ 
      % Klassenzuordnung nach Maximumsuche
     $c_{Pred}(k) = \max_c p_{Pred,k}(c)$ 

```

## A.2 Correlative-Matrix-Algorithmus

Der folgende Algorithmus wurde von Hsu et. al. in [HLC01] näher beschrieben und soll hier nur als Pseudocode aufgeführt werden.

**Algorithmus :**  $[RP] = \text{Pattern\_Search\_Correlative\_Matrix}(S)$

**Eingabe :**  $S$  - Zeichenkette der Länge  $n$

**Ausgabe :**  $RP = \{p_i, f_i\}$  - Liste aller wiederholten Muster  $p_i$  sowie ihrer Frequenzen  $f_i$

- 1  $[T] = \text{Construct\_Correlative\_Matrix}(S)$
- 2  $[RP] = \text{Generate\_Repeating\_Pattern}(T, S)$

**Algorithmus :**  $[T] = \text{Construct\_Correlative\_Matrix}(S)$

**Eingabe :**  $S$  - Zeichenkette der Länge  $n$

**Ausgabe :**  $T$  - *Correlative Matrix* der Dimension  $n \times n$

- 1 % Ausfüllen der ersten Zeile von  $T$
- 2 **für**  $j \leftarrow 2$  **bis**  $n$  **tue**
- 3     **wenn**  $S(1) \equiv S(j)$  **dann**
- 4          $T(1, j) = 1$
- 5     **sonst**
- 6          $T(1, j) = 0$
- 7 % Ausfüllen der restlichen Zeilen von  $T$
- 8 **für**  $i \leftarrow 2$  **bis**  $n - 1$  **tue**
- 9     **für**  $j \leftarrow i + 1$  **bis**  $n$  **tue**
- 10         **wenn**  $S(i) \equiv S(j)$  **dann**
- 11              $T(i, j) = T(i - 1, j - 1) + 1$
- 12         **sonst**
- 13              $T(i, j) = 0$

**Algorithmus :**  $[RP] = \text{Generate\_Repeating\_Pattern}(T, S)$

**Eingabe :**  $T$  - *Correlative Matrix* der Dimension  $n \times n$

**Eingabe :**  $S$  - Zeichenkette der Länge  $n$

**Ausgabe :**  $RP = \{p_i, f_i\}$  - Liste aller wiederholten Muster  $p_i$  sowie ihrer Frequenzen  $f_i$

```

1  % CS - Liste (candidate set) an Musterkandidaten
2  CS =
3  für  $i \leftarrow 1$  bis  $n$  tue
4      für  $j \leftarrow 1$  bis  $n$  tue
5          wenn  $T(i, j) \neq 0$  dann
6              für  $l \leftarrow 1$  bis  $T(i, j)$  tue
7                  % Aktuelles Muster
8                   $pat = S[(j - l + 1) : j]$ 
9                  wenn  $T(i + 1, j + 1) \equiv 0$  dann
10                     wenn  $(pat, f_i, s_i) \in CS$  dann
11                         ersetze  $(pat, f_i, s_i)$  in  $CS$  durch  $(pat, f_i + 1, s_i)$ 
12                     sonst
13                          $CS = CS \cup \{(pat, 1, 0)\}$ 
14                 sonst
15                     wenn  $(pat, f_i, s_i) \in CS$  dann
16                         ersetze  $(pat, f_i, s_i)$  in  $CS$  durch  $(pat, f_i + 1, s_i + 1)$ 
17                     sonst
18                          $CS = CS \cup \{(pat, 1, 1)\}$ 
19             sonst
20 % Berechnung der Auftretenshäufigkeiten der Muster
21 für  $\forall (p_i, f_i, s_i) \in CS$  tue
22     wenn  $f_i \equiv s_i$  dann
23         lösche  $(p_i, f_i, s_i)$  aus  $CS$ 
24     sonst
25          $f_i = 1 + \sqrt{1 + 8 f_i} / 2$ 
26         füge  $(p_i, f_i)$  in die Liste  $RP$  ein

```



## B Testsets für Genre- und Musikerklassifikation

### Testset - MIDI - Genreklassifikation

URL	zuletzt aufgerufen
<a href="http://www.dmoz.org/Arts/Music/Sound_Files/MIDI/">http://www.dmoz.org/Arts/Music/Sound_Files/MIDI/</a>	18.07.2008
<a href="http://www.funkville.net/midi.htm">http://www.funkville.net/midi.htm</a>	18.07.2008
<a href="http://www.geocities.com/BourbonStreet/1114/justjazz2.htm">http://www.geocities.com/BourbonStreet/1114/justjazz2.htm</a>	18.07.2008
<a href="http://www.gnu-bee.com/dablues.shtml">http://www.gnu-bee.com/dablues.shtml</a>	18.07.2008
<a href="http://www.gnu-bee.com/latin.shtml">http://www.gnu-bee.com/latin.shtml</a>	18.07.2008
<a href="http://www.midiworld.com/jazz.htm">http://www.midiworld.com/jazz.htm</a>	18.07.2008
<a href="http://diverse.freepage.de/funkenstein/midi/funk.html">http://diverse.freepage.de/funkenstein/midi/funk.html</a>	18.07.2008
<a href="http://home5.swipnet.se/~w-54954/midi.htm">http://home5.swipnet.se/~w-54954/midi.htm</a>	18.07.2008
Anzahl Soli :	300 (6 × 50)
Gesamtlänge :	194:58 min
Mittlere Länge pro Solo :	39,0 s

Tab. B.1: Internetquellen und Umfang des MIDI-Testsets zur Genreklassifikation

### Testset - Audio - Genreklassifikation

Titel	MEL HAR BAS DRU BG				
	Anzahl der transkribierten Noten des Melodieinstrumentes - <b>MEL</b> , Harmonieinstrumentes - <b>HAR</b> , Bassinstrumentes - <b>BAS</b> , Schlaginstrumentes - <b>DRU</b> , rhythmischen Rasters ( <i>beat grid</i> ) - <b>BG</b>				
<i>Blues</i>					
BLU-001-Bassing-Blues_Jut-1.mp3	76	178	50	162	46
BLU-002-Bassing-Blues_Jut-2.mp3	86	251	46	222	23
BLU-003-Bassing-Blues_Jut-3.mp3	72	281	45	189	43
BLU-004-Bassing-Blues_Jut-4.mp3	80	59	41	280	38
BLU-005-Deep_Purple-Lazy.mp3	161	278	7	353	112
BLU-006-Eric_Clapton-Before_You_Accuse_Me.mp3	137	66	64	149	47
BLU-007-Deep_Purple-Maybe_I_am_a_Leo.mp3	113	440	25	178	48
BLU-008-Eric_Clapton-001-Ain_t_that_loving_you-1.mp3	92	145	65	67	89
BLU-009-Eric_Clapton-002-Ain_t_that_loving_you-2.mp3	112	155	56	63	41
BLU-010-Eric_Clapton-003-Ain_t_that_loving_you-3.mp3	136	106	73	76	26
BLU-011-Eric_Clapton-025-Have_you_ever_loved_a_woman-1.mp3	52	59	28	139	62
BLU-012-Eric_Clapton-026-Have_you_ever_loved_a_woman-2.mp3	107	102	41	196	86
BLU-013-Eric_Clapton-027-Have_you_ever_loved_a_woman-3.mp3	74	60	41	225	65
BLU-014-Eric_Clapton-028-Have_you_ever_loved_a_woman-4.mp3	78	54	44	149	30

BLU-015-Eric.Clapton-029-Have_you_ever_loved_a_woman-5.mp3	87	159	48	171	74
BLU-016-Eric.Clapton-030-Have_you_ever_loved_a_woman-6.mp3	71	105	43	146	69
BLU-017-Stevie.Ray.Vaughan-001-Say_What-1.mp3	93	99	37	340	53
BLU-018-Stevie.Ray.Vaughan-002-Say_What-2.mp3	141	199	35	211	28
BLU-019-Stevie.Ray.Vaughan-003-Say_What-3.mp3	121	111	34	196	54
BLU-020-Stevie.Ray.Vaughan-004-Say_What-4.mp3	102	186	31	151	47
BLU-021-Stevie.Ray.Vaughan-005-Say_What-5.mp3	87	266	38	211	54
BLU-022-Stevie.Ray.Vaughan-006-Look_At_Little_Sister.mp3	108	42	31	235	41
BLU-023-Stevie.Ray.Vaughan-007-Aint_Gone_N_Give_Up_On_Love-1.mp3	47	106	21	100	51
BLU-024-Stevie.Ray.Vaughan-008-Aint_Gone_N_Give_Up_On_Love-2.mp3	72	91	33	49	67
BLU-025-Stevie.Ray.Vaughan-009-Aint_Gone_N_Give_Up_On_Love-3.mp3	77	144	29	132	67
BLU-026-Stevie.Ray.Vaughan-010-Aint_Gone_N_Give_Up_On_Love-4.mp3	81	69	37	147	39
BLU-027-Stevie.Ray.Vaughan-011-Aint_Gone_N_Give_Up_On_Love-5.mp3	92	115	32	120	70
BLU-028-Stevie.Ray.Vaughan-012-The_House_Is_Rockin.mp3	118	282	58	91	30
BLU-029-Stevie.Ray.Vaughan-023-Leave_My_Girl_Alone-1.mp3	162	158	47	185	122
BLU-030-Stevie.Ray.Vaughan-024-Leave_My_Girl_Alone-2.mp3	180	171	47	364	117
BLU-031-Stevie.Ray.Vaughan-025-Wall.Of.Denial-1.mp3	166	153	78	353	72
BLU-032-Stevie.Ray.Vaughan-026-Wall.Of.Denial-2.mp3	168	314	76	170	38
BLU-033-Stevie.Ray.Vaughan-027-Wall.Of.Denial-3.mp3	178	166	67	328	65
BLU-034-Stevie.Ray.Vaughan-028-Love.Me.Darlin-1.mp3	105	184	45	195	51
BLU-035-Stevie.Ray.Vaughan-029-Love.Me.Darlin-2.mp3	101	49	34	171	46
BLU-036-Stevie.Ray.Vaughan-030-Love.Me.Darlin-3.mp3	134	188	57	294	69
BLU-037-Rory.Gallagher-Sinnerboy.mp3	145	78	107	223	28
BLU-038-Rory.Gallagher-Calling_Card-1.mp3	82	96	66	118	82
BLU-039-Rory.Gallagher-Calling_Card-2.mp3	67	166	49	61	66
BLU-040-Rory.Gallagher-Calling_Card-3.mp3	40	77	52	50	33
<i>Funk</i>					
FUN-001-Funkadelic-Cosmic.Slop-1.mp3	127	123	66	155	66
FUN-002-Funkadelic-Cosmic.Slop-2.mp3	108	134	64	130	75
FUN-003-Funkadelic-Standing_On_The_Verge_Of_Getting_It_On.mp3	83	157	63	163	41
FUN-004-Funkadelic-Alice_In_My_Fantasies.mp3	83	136	71	142	25
FUN-005-Funkadelic-Red_Hot_Mama-1.mp3	82	140	72	139	34
FUN-006-Funkadelic-Red_Hot_Mama-2.mp3	82	266	77	149	31
FUN-007-Funkadelic-Red_Hot_Mama-3.mp3	81	189	79	228	35
FUN-008-Joshua_Redman-Molten_Soul-2.mp3	59	70	69	98	30
FUN-009-Joshua_Redman-Molten_Soul-3.mp3	80	67	64	146	40
FUN-010-Joshua_Redman-Molten_Soul-4.mp3	72	77	64	214	37
FUN-011-Marcus.Miller-Scoop-1.mp3	58	51	66	201	36
FUN-012-Marcus.Miller-Scoop-2.mp3	66	171	73	202	41
FUN-013-Tower_Of_Power-Cant_Stand_To_See_The_Slaughter.mp3	132	108	71	346	94
FUN-014-Marcus.Miller-Stevland-1.mp3	70	96	63	330	44
FUN-015-Marcus.Miller-Stevland-2.mp3	57	157	60	272	41
FUN-016-Marcus.Miller-Stevland-3.mp3	63	63	50	285	34
FUN-017-Marcus.Miller-Stevland-4.mp3	70	100	78	378	47
FUN-018-Marcus.Miller-Panther.mp3	88	64	60	245	34
FUN-019-Marcus.Miller-Juju.mp3	60	97	50	137	30
FUN-020-Marcus.Miller-Tales.mp3	51	81	78	169	39
FUN-021-Marcus.Miller-True_Geminis-1.mp3	46	33	49	165	48
FUN-022-Marcus.Miller-True_Geminis-2.mp3	68	86	63	221	66
FUN-023-Marcus.Miller-Erik-1.mp3	48	104	52	165	24
FUN-024-Marcus.Miller-Erik-2.mp3	59	43	56	186	25
FUN-025-Marcus.Miller-Come_Together.mp3	96	178	97	255	44
FUN-026-Marcus.Miller-Burning_Down_The_House-1.mp3	38	246	66	126	34
FUN-027-Marcus.Miller-Burning_Down_The_House-2.mp3	61	67	60	120	30
FUN-028-Marcus.Miller-Burning_Down_The_House-3.mp3	53	46	54	127	28
FUN-029-Marcus.Miller-Lonny_s_Lament-1.mp3	92	81	71	147	36
FUN-030-Marcus.Miller-Lonny_s_Lament-2.mp3	68	67	62	162	29



FUN-031-Marcus_Miller-Red_Baron-1.mp3	82	76	76	183	66
FUN-032-Marcus_Miller-Red_Baron-2.mp3	76	62	77	182	62
FUN-033-Marcus_Miller-Red_Baron-3.mp3	52	70	66	217	32
FUN-034-Matalex-B_Funkt_And_Bang_Your_Head.mp3	86	37	74	185	81
FUN-035-Matalex-X_Perience-1.mp3	63	49	52	164	34
FUN-036-Matalex-X_Perience-2.mp3	92	105	36	167	32
FUN-037-Matalex-X_Perience-3.mp3	96	65	50	166	33
FUN-038-Joshua_Redman-Swunk-1.mp3	55	82	64	176	36
FUN-039-Joshua_Redman-Swunk-2.mp3	108	58	85	143	39
FUN-040-Joshua_Redman-Swunk-3.mp3	104	102	94	170	47
<i>Latin</i>					
LAT-001-Buena_Vista_Social_Club-Chan_Chan-1.mp3	93	73	38	202	52
LAT-002-Buena_Vista_Social_Club-Chan_Chan-2.mp3	88	62	38	186	38
LAT-003-Compay_Segundo-Huellas_Del_Pasado-1.mp3	70	87	29	395	34
LAT-004-Compay_Segundo-Huellas_Del_Pasado-2.mp3	113	87	46	323	48
LAT-005-Compay_Segundo-Orgullecida.mp3	58	58	34	318	32
LAT-006-Compay_Segundo-Chicharrones-1.mp3	103	49	46	259	38
LAT-007-Compay_Segundo-Chicharrones-2.mp3	85	42	45	288	35
LAT-008-Compay_Segundo-Chicharrones-3.mp3	84	52	46	279	38
LAT-009-Compay_Segundo-Mujeres_De_Mayari.mp3	128	64	58	589	54
LAT-010-Compay_Segundo-Sambrosa.mp3	119	57	65	425	44
LAT-011-Compay_Segundo-Quien_Te_Bautizo.mp3	85	78	34	204	33
LAT-012-Compay_Segundo-Mi_Calderito.mp3	90	45	32	245	32
LAT-013-Compay_Segundo-Saludo_Compay.mp3	83	58	37	325	31
LAT-014-Compay_Segundo-Como_La_Avellaneda.mp3	63	64	26	235	26
LAT-015-Compay_Segundo-Idilio-1.mp3	111	58	39	488	52
LAT-016-Compay_Segundo-Idilio-2.mp3	116	71	29	450	79
LAT-017-Cubanismo-El_Platanal_De_Bartolo-1.mp3	97	47	63	253	40
LAT-018-Cubanismo-Mambo_UK.mp3	109	191	54	372	36
LAT-019-Marcos_Valle-The_Answer-1.mp3	151	63	13	175	57
LAT-020-Luis_Bonfa-Tristeza-1.mp3	61	145	19	71	37
LAT-021-Cubanismo-El_Paso_de_Encarnación.mp3	122	91	70	451	40
LAT-022-Stan_Getz-Desafinado-2.mp3	69	73	12	209	37
LAT-023-Stan_Getz-Desafinado-3.mp3	56	69	10	179	44
LAT-024-Stan_Getz-Desafinado-4.mp3	84	62	10	241	33
LAT-025-Cubanismo-El_Paso_de_Encarnación-2.mp3	98	209	80	275	44
LAT-026-Walter_Wanderley-Beach_Samba.mp3	62	85	14	170	50
LAT-027-Stan_Getz-Doralice-1.mp3	60	181	16	84	93
LAT-028-Stan_Getz-Doralice-2.mp3	49	88	24	135	44
LAT-029-Stan_Getz-Doralice-3.mp3	54	178	28	91	26
LAT-030-Cubanismo-Jamming_In_Nijmegen-1.mp3	169	121	77	450	51
LAT-031-Cubanismo-Jamming_In_Nijmegen-2.mp3	118	470	67	337	52
LAT-032-Stan_Getz-Samba_Triste-2.mp3	49	73	10	77	41
LAT-033-Stan_Getz-Samba_Triste-4.mp3	35	113	12	202	35
LAT-034-Stan_Getz-Samba_Triste-5.mp3	51	181	12	130	37
LAT-035-Cubanismo-Donde_Esta_Coto-1.mp3	80	258	39	236	32
LAT-036-Cubanismo-Mimi.mp3	115	103	73	129	107
LAT-037-Cubanismo-Mimi-2.mp3	107	101	65	191	58
LAT-038-Cubanismo-Con_Mana_Se_Rompe.mp3	138	80	74	234	41
LAT-039-Stan_Getz-Samba_De_Uma_Nota_So-1.mp3	37	151	31	123	31
LAT-040-Stan_Getz-Samba_De_Uma_Nota_So-2.mp3	70	87	30	137	35
<i>Metal_Hardrock</i>					
MHR-001-Deep_Purple-Highway_Star-1.mp3	150	191	19	140	31
MHR-002-Cherry_Moon_Drive-Aeroplane.mp3	81	81	48	121	46
MHR-003-Deep_Purple-Highway_Star-2.mp3	177	297	17	117	33
MHR-004-Alice_In_Chains-No_Excuses.mp3	66	44	44	111	33
MHR-005-Eric_Clapton-Sunshine_of_your_love-1.mp3	75	122	24	136	27

MHR-006-Eric_Clapton-Sunshine_of_your_love-2.mp3	98	191	13	108	48
MHR-007-Eric_Clapton-Tales_of_brave_Ulysses.mp3	175	105	37	138	49
MHR-008-ACDC-Hells_Bells.mp3	137	225	67	272	51
MHR-009-Eric_Clapton-The_core-2.mp3	130	79	12	191	64
MHR-010-Jimi_Hendrix-All_along_the_watchtower-1.mp3	71	15	45	119	21
MHR-011-Jimi_Hendrix-All_along_the_watchtower-2.mp3	70	49	52	99	60
MHR-012-Jimi_Hendrix-All_along_the_watchtower-3.mp3	68	67	50	116	42
MHR-013-Jimi_Hendrix-Foxy_lady.mp3	72	166	28	115	68
MHR-014-Jimi_Hendrix-Machine_gun-1.mp3	119	199	61	138	62
MHR-015-Jimi_Hendrix-Machine_gun-2.mp3	116	181	65	131	44
MHR-016-Jimi_Hendrix-Machine_gun-3.mp3	90	101	59	196	30
MHR-017-Jimi_Hendrix-Message_to_love-1.mp3	94	89	69	183	29
MHR-018-Jimi_Hendrix-Message_to_love-2.mp3	96	108	85	273	50
MHR-019-Jimi_Hendrix-Power_of_love-1.mp3	78	88	63	140	33
MHR-020-Jimi_Hendrix-Power_of_love-2.mp3	59	84	58	69	55
MHR-021-Jimi_Hendrix-Power_of_love-3.mp3	66	123	54	45	24
MHR-022-Jimi_Hendrix-Power_of_love-4.mp3	108	103	78	88	32
MHR-023-Jimi_Hendrix-Voodoo_chile.mp3	124	245	60	345	59
MHR-024-Jimi_Hendrix-Who_knows-1.mp3	96	99	65	133	34
MHR-025-Jimi_Hendrix-Who_knows-2.mp3	115	84	67	206	28
MHR-026-Joe_Satriani-Summer_Song-1.mp3	86	89	80	222	35
MHR-027-Joe_Satriani-Summer_Song-2.mp3	124	115	60	100	62
MHR-028-Joe_Satriani-Summer_Song-3.mp3	106	58	65	116	65
MHR-029-Joe_Satriani-Summer_Song-4.mp3	103	12	58	93	28
MHR-030-Joe_Satriani-The_Mystical_Potato_Head_Groove_Thing-1.mp3	101	129	73	141	60
MHR-031-Joe_Satriani-The_Mystical_Potato_Head_Groove_Thing-2.mp3	106	111	74	129	29
MHR-032-Joe_Satriani-The_Mystical_Potato_Head_Groove_Thing-3.mp3	84	101	71	135	29
MHR-033-Metallica-Enter_Sandman-1.mp3	112	272	25	138	19
MHR-034-Metallica-Enter_Sandman-2.mp3	68	34	27	121	34
MHR-035-Metallica-Fuel.mp3	100	191	68	76	31
MHR-036-Metallica-The_Call_Of_Ktulu-1.mp3	209	106	54	167	60
MHR-037-Metallica-The_Call_Of_Ktulu-2.mp3	137	259	52	182	31
MHR-038-Metallica-The_Unforgiven-1.mp3	118	40	25	95	22
MHR-039-Metallica-The_Unforgiven-2.mp3	124	63	37	79	30
MHR-040-Jimi_Hendrix-Power_of_love-5.mp3	93	156	83	60	30
<i>Pop-Rock</i>					
POP-001-Deep_Purple-Nerer_Before.mp3	120	54	36	165	31
POP-002-Cherry_Moon_Drive_Track_03-1.mp3	87	71	42	89	31
POP-003-Sting-Brand_New_Day-1.mp3	118	57	62	301	36
POP-004-Alanis_Morissette-Head_Over_Feet.mp3	133	39	51	146	31
POP-005-Eric_Clapton-Lay_down_Sally-1.mp3	57	278	35	136	33
POP-006-Eric_Clapton-Lay_down_Sally-2.mp3	47	23	43	99	31
POP-007-Eric_Clapton-Next_time_you_see_her-1.mp3	86	27	18	256	32
POP-008-Santana-Lightnin-1.mp3	111	75	61	119	69
POP-009-Eric_Clapton-Next_time_you_see_her-3.mp3	84	21	29	150	32
POP-010-Eric_Clapton-Bad_love.mp3	98	75	62	81	59
POP-011-Eric_Clapton-Badge-1.mp3	64	36	17	127	20
POP-012-Eric_Clapton-Badge-2.mp3	82	45	16	98	40
POP-013-Eric_Clapton-Peaches_and_diesel-1.mp3	59	49	15	153	35
POP-014-Eric_Clapton-Little_wing-1.mp3	106	104	28	135	30
POP-015-Eric_Clapton-Little_wing-2.mp3	176	68	55	183	36
POP-016-Eric_Clapton-Little_wing-3.mp3	184	65	45	151	31
POP-017-Eric_Clapton-Little_wing-4.mp3	190	82	47	150	68
POP-018-Eric_Clapton-Cocaine-1.mp3	89	137	17	137	33
POP-019-Eric_Clapton-Cocaine-2.mp3	98	52	24	192	31
POP-020-Eric_Clapton-Cocaine-3.mp3	86	89	15	169	33
POP-021-Jimi_Hendrix-Highway_chile.mp3	208	51	67	223	64

POP-022-Jimi_Hendrix-Bold_as_love-1.mp3	121	71	18	121	54
POP-023-Jimi_Hendrix-Bold_as_love-2.mp3	62	79	32	110	47
POP-024-Jimi_Hendrix-Drivin_South-1.mp3	54	63	47	303	59
POP-025-Jimi_Hendrix-Drivin_South-2.mp3	75	75	23	127	34
POP-026-Stevie_Ray_Vaughan-Crossfire-1.mp3	106	127	57	169	52
POP-027-Stevie_Ray_Vaughan-Crossfire-2.mp3	109	73	47	162	51
POP-028-Stevie_Ray_Vaughan-Crossfire-3.mp3	116	121	56	171	53
POP-029-Funkadelic-Ill_Stay-1.mp3	67	67	47	197	51
POP-030-Funkadelic-Ill_Stay-2.mp3	109	40	58	206	24
POP-031-Porcupine_Tree-Track_08-1.mp3	70	47	26	118	29
POP-032-Porcupine_Tree-Track_08-2.mp3	55	31	30	66	24
POP-033-Porcupine_Tree-Track_08-3.mp3	82	48	44	100	41
POP-034-Porcupine_Tree-Track_08-4.mp3	86	48	37	86	25
POP-035-Porcupine_Tree-Track_05-1.mp3	91	75	50	312	38
POP-036-Porcupine_Tree-Track_05-2.mp3	97	95	53	286	80
POP-037-Porcupine_Tree-Track_05-3.mp3	105	60	52	225	40
POP-038-Porcupine_Tree-Track_05-4.mp3	118	75	60	214	81
POP-039-Santana-Lightnin-2.mp3	114	79	73	142	35
POP-040-Al_Steward-It_Doesnt_Come_Naturally.mp3	83	176	66	244	59

*Swing*

SWI-001-Kenny_Baron-Voyage-1.mp3	105	99	30	224	103
SWI-002-Kenny_Baron-Voyage-2.mp3	93	269	21	184	36
SWI-003-Charlie_Parker-I_Cant_Get_Started-3.mp3	79	110	19	227	65
SWI-004-Charlie_Parker-I_Got_Rhythm-1.mp3	100	74	13	242	88
SWI-005-Charlie_Parker-I_Got_Rhythm-2.mp3	98	141	19	129	40
SWI-006-Kenny_Baron-Voyage-3.mp3	72	74	30	189	67
SWI-007-Kenny_Baron-Voyage-4.mp3	108	127	30	132	101
SWI-008-Miles_Davis-So_What-1.mp3	92	114	27	237	38
SWI-009-Dexter_Gordon-Blues_Up_And_Down-1.mp3	83	86	16	153	45
SWI-010-Dexter_Gordon-Blues_Up_And_Down-2.mp3	72	53	33	195	27
SWI-011-Dexter_Gordon-Blues_Up_And_Down-3.mp3	87	101	22	223	32
SWI-012-Dexter_Gordon-Cheesecake-1.mp3	62	80	39	179	25
SWI-013-Dexter_Gordon-Cheesecake-2.mp3	101	69	43	245	29
SWI-014-Dexter_Gordon-Cheesecake-3.mp3	69	109	49	218	37
SWI-015-Dexter_Gordon-Cheesecake-4.mp3	75	239	60	335	96
SWI-016-Dexter_Gordon-Cheesecake-5.mp3	60	90	38	198	53
SWI-017-Miles_Davis-So_What-2.mp3	66	111	27	270	64
SWI-018-Dexter_Gordon-Secret_Love-2.mp3	56	93	19	141	49
SWI-019-Dexter_Gordon-Secret_Love-3.mp3	49	65	17	150	40
SWI-020-Dexter_Gordon-The_End_Of_A_Love_Affair-1.mp3	87	209	43	216	28
SWI-021-Dexter_Gordon-The_End_Of_A_Love_Affair-2.mp3	85	107	58	248	38
SWI-022-John_Coltrane-Bakai-1.mp3	75	109	43	180	54
SWI-023-Miles_Davis-So_What-3.mp3	67	231	42	258	35
SWI-024-John_Coltrane-Bakai-3.mp3	74	105	31	127	25
SWI-025-John_Coltrane-Bakai-4.mp3	78	67	42	149	27
SWI-026-John_Coltrane-Bakai-5.mp3	79	52	37	159	44
SWI-027-John_Coltrane-Bakai-6.mp3	62	157	41	180	25
SWI-028-John_Coltrane-Chronic_Blues-1.mp3	67	230	26	149	58
SWI-029-John_Coltrane-Chronic_Blues-2.mp3	49	65	30	215	25
SWI-030-John_Coltrane-Locomotion-1.mp3	88	94	33	200	30
SWI-031-John_Coltrane-Locomotion-2.mp3	89	31	36	208	32
SWI-032-John_Coltrane-Locomotion-3.mp3	109	197	44	231	60
SWI-033-Joshua_Redman-Leap_Of_Faith-1.mp3	84	95	76	207	51
SWI-034-Joshua_Redman-Leap_Of_Faith-3.mp3	147	151	116	327	30
SWI-035-Joshua_Redman-Leap_Of_Faith-4.mp3	89	110	81	196	29
SWI-036-Joshua_Redman-Leap_Of_Faith-5.mp3	81	170	80	226	29
SWI-037-Joshua_Redman-Turnaround-1.mp3	42	183	33	120	57

SWI-038-Joshua_Redman-Turnaround-2.mp3	91	224	40	234	72
SWI-039-Joshua_Redman-Turnaround-3.mp3	57	44	46	149	29
SWI-040-Joshua_Redman-Turnaround-5.mp3	68	306	37	190	50
Anzahl Soli :	240 ( $6 \times 40$ )				
Gesamtlänge :	93:06 min				
Mittlere Länge pro Solo :	23,3 s				

Tab. B.2: Testset - Audio - Genreklassifikation

## Testset - Audio - Musikerklassifikation (E-Gitarre)

Titel	MEL	HAR	BAS	DRU	BG
<i>Eric Clapton</i>					
CLA-001-Ain.t.that_loving_you-1.mp3	92	145	65	67	89
CLA-002-Ain.t.that_loving_you-2.mp3	112	155	56	63	41
CLA-003-Ain.t.that_loving_you-3.mp3	136	106	73	76	26
CLA-004-Lay_down.Sally-1.mp3	57	278	35	136	33
CLA-005-Lay_down.Sally-2.mp3	47	23	43	99	31
CLA-006-Next_time.you.see.her-1.mp3	86	27	18	256	32
CLA-007-Next_time.you.see.her-2.mp3	97	33	22	164	35
CLA-008-Next_time.you.see.her-3.mp3	84	24	29	150	32
CLA-009-The_core-1.mp3	145	85	89	191	46
CLA-010-Bad_love.mp3	98	75	62	81	59
CLA-011-The_core-2.mp3	142	90	105	301	62
CLA-012-Sunshine_of.your_love-1.mp3	75	122	24	136	27
CLA-013-Sunshine_of.your_love-2.mp3	108	277	62	185	89
CLA-014-Tales_of.brave_Ulysses.mp3	175	105	37	138	49
CLA-015-Badge-1.mp3	58	42	43	37	53
CLA-016-Badge-2.mp3	76	33	49	82	30
CLA-017-Peaches_and.diesel-1.mp3	65	53	31	104	35
CLA-018-Little.wing-1.mp3	106	104	28	135	30
CLA-019-Little.wing-2.mp3	176	68	55	183	36
CLA-020-Little.wing-3.mp3	184	65	45	151	31
CLA-021-Little.wing-4.mp3	190	82	47	150	68
CLA-022-Cocaine-1.mp3	84	57	33	131	32
CLA-023-Cocaine-2.mp3	98	52	24	192	31
CLA-024-Cocaine-3.mp3	86	99	30	106	32
CLA-025-Have.you.ever_loved_a.woman-1.mp3	52	59	28	139	62
CLA-026-Have.you.ever_loved_a.woman-2.mp3	107	102	41	196	86
CLA-027-Have.you.ever_loved_a.woman-3.mp3	74	60	41	225	65
CLA-028-Have.you.ever_loved_a.woman-4.mp3	78	54	44	149	30
CLA-029-Have.you.ever_loved_a.woman-5.mp3	87	159	48	171	74
CLA-030-Have.you.ever_loved_a.woman-6.mp3	71	105	43	146	69
<i>Rory Gallagher</i>					
GAL-001-Shadow_Play-1.mp3	85	52	81	148	24
GAL-002-Shadow_Play-2.mp3	122	47	85	144	24
GAL-003-Shadow_Play-3.mp3	105	21	95	158	38
GAL-004-Shadow_Play-4.mp3	123	55	118	205	57
GAL-005-Shadow_Play-5.mp3	79	30	74	213	32
GAL-006-Bad.Penny-1.mp3	87	72	47	120	32
GAL-007-Bad.Penny-2.mp3	111	53	48	155	66
GAL-008-Tattooed.Lady-1.mp3	122	87	86	230	63

GAL-009-Tattood_Lady-2.mp3	131	21	85	193	45
GAL-010-Tattood_Lady-3.mp3	111	81	73	172	40
GAL-011-Whats_Going_On-1.mp3	141	144	123	116	42
GAL-012-Whats_Going_On-2.mp3	108	92	114	74	26
GAL-013-Big_Guns-1.mp3	122	27	116	181	59
GAL-014-Big_Guns-2.mp3	78	55	127	199	57
GAL-015-Big_Guns-3.mp3	150	59	165	268	78
GAL-016-Kickback_City-1.mp3	83	85	64	101	35
GAL-017-Kickback_City-2.mp3	100	27	80	127	24
GAL-018-Bullfrog_Blues-1.mp3	191	79	217	266	49
GAL-019-Used_To.Be.mp3	86	73	76	150	19
GAL-020-Sinnerboy.mp3	145	78	107	223	28
GAL-021-Bourbon-1.mp3	112	18	123	182	36
GAL-022-Bourbon-2.mp3	137	99	124	184	31
GAL-023-Calling_Card-1.mp3	82	96	66	118	82
GAL-024-Calling_Card-2.mp3	67	166	49	61	66
GAL-025-Calling_Card-3.mp3	40	77	52	50	33
GAL-026-A_Million_Miles_Away-1.mp3	83	43	65	101	25
GAL-027-A_Million_Miles_Away-2.mp3	68	71	61	189	22
GAL-028-Born_On_The_Wrong_Side_Of_Time-1.mp3	86	165	49	155	40
GAL-029-Born_On_The_Wrong_Side_Of_Time-2.mp3	73	45	58	136	24
GAL-030-Daughter_Of_The_Everglades-1.mp3	116	97	101	120	57
<i>Jimi Hendrix</i>					
HEN-001-Who_knows-1.mp3	96	99	65	133	34
HEN-002-Who_knows-2.mp3	115	84	67	206	28
HEN-003-Machine_gun-1.mp3	119	122	61	138	62
HEN-004-Machine_gun-2.mp3	116	233	65	131	44
HEN-005-Machine_gun-3.mp3	90	95	59	196	30
HEN-006-Power_of_love-1.mp3	78	88	63	140	33
HEN-007-Power_of_love-2.mp3	59	84	58	69	55
HEN-008-Power_of_love-3.mp3	66	123	54	45	24
HEN-009-Power_of_love-4.mp3	108	182	78	88	32
HEN-010-Power_of_love-5.mp3	93	93	83	60	30
HEN-011-Message_to_love-1.mp3	94	89	69	183	29
HEN-012-Message_to_love-2.mp3	96	216	85	273	50
HEN-013-All_along_the_watchtower-1.mp3	71	15	45	119	21
HEN-014-All_along_the_watchtower-2.mp3	70	49	52	99	60
HEN-015-All_along_the_watchtower-3.mp3	68	67	50	116	42
HEN-016-Foxy_lady.mp3	72	166	28	115	68
HEN-017-Free_spirit-1.mp3	115	67	74	114	52
HEN-018-Free_spirit-2.mp3	112	68	87	255	36
HEN-019-Free_spirit-3.mp3	80	204	59	228	30
HEN-020-Highway_chile.mp3	208	51	67	223	64
HEN-021-House_of_the_rising_sun-1.mp3	115	106	78	245	28
HEN-022-House_of_the_rising_sun-2.mp3	112	78	82	247	61
HEN-023-House_of_the_rising_sun-3.mp3	104	104	69	280	28
HEN-024-House_of_the_rising_sun-4.mp3	153	70	85	288	32
HEN-025-Voodoo_chile.mp3	124	245	60	345	59
HEN-026-Little_Miss_Lover.mp3	59	89	24	196	34
HEN-027-Bold_as_love-1.mp3	121	71	18	121	54
HEN-028-Bold_as_love-2.mp3	62	79	32	110	47
HEN-029-Drivin_South-1.mp3	54	63	47	303	59
HEN-030-Drivin_South-2.mp3	75	75	23	127	34
<i>Steve Ray Vaughan</i>					
VAU-001-Say_What-1.mp3	93	99	37	340	53
VAU-002-Say_What-2.mp3	141	199	35	211	28
VAU-003-Say_What-3.mp3	121	111	34	196	54

VAU-004-Say_What-4.mp3	102	186	31	151	47
VAU-005-Say_What-5.mp3	87	93	38	211	54
VAU-006-Look_At_Little_Sister.mp3	108	42	31	235	41
VAU-007-Aint_Gone_N_Give_Up_On_Love-1.mp3	47	106	21	100	51
VAU-008-Aint_Gone_N_Give_Up_On_Love-2.mp3	72	91	33	49	67
VAU-009-Aint_Gone_N_Give_Up_On_Love-3.mp3	77	144	29	132	67
VAU-010-Aint_Gone_N_Give_Up_On_Love-4.mp3	81	69	37	147	39
VAU-011-Aint_Gone_N_Give_Up_On_Love-5.mp3	92	115	32	120	70
VAU-012-The_House_Is_Rockin.mp3	118	289	58	91	30
VAU-013-Crossfire-1.mp3	106	127	57	169	52
VAU-014-Crossfire-2.mp3	109	73	47	162	51
VAU-015-Crossfire-3.mp3	116	121	56	171	53
VAU-016-Tightrope-1.mp3	92	61	48	82	38
VAU-017-Tightrope-2.mp3	136	315	46	262	40
VAU-018-Tightrope-3.mp3	153	241	48	170	48
VAU-019-Tightrope-4.mp3	129	150	41	270	83
VAU-020-Tightrope-5.mp3	74	13	27	137	31
VAU-021-Tightrope-6.mp3	72	70	39	60	23
VAU-022-Tightrope-7.mp3	80	35	37	130	52
VAU-023-Leave_My_Girl_Alone-1.mp3	162	158	47	185	122
VAU-024-Leave_My_Girl_Alone-2.mp3	180	171	47	364	117
VAU-025-Wall_Of_Denial-1.mp3	166	153	78	353	72
VAU-026-Wall_Of_Denial-2.mp3	168	314	76	170	38
VAU-027-Wall_Of_Denial-3.mp3	178	166	67	328	65
VAU-028-Love_Me_Darlin-1.mp3	105	184	45	195	51
VAU-029-Love_Me_Darlin-2.mp3	101	49	34	171	46
VAU-030-Love_Me_Darlin-3.mp3	134	188	57	294	69
Anzahl Soli :	120 (4 × 30)				
Gesamtlänge :	46:00 min				
Mittlere Länge pro Solo :	23,0 s				

Tab. B.3: Testset - Audio - Musikerklassifikation (E-Gitarre)

## Testset - Audio - Musikerklassifikation (Saxophon)

Titel	MEL	HAR	BAS	DRU	BG
<i>John Coltrane</i>					
COL-001-Bakai-1.mp3	75	109	43	180	54
COL-002-Bakai-2.mp3	76	55	68	186	55
COL-003-Bakai-3.mp3	74	105	31	127	25
COL-004-Bakai-4.mp3	78	67	42	149	27
COL-005-Bakai-5.mp3	79	52	37	159	44
COL-006-Bakai-6.mp3	62	157	41	180	25
COL-007-Blue_Train-1.mp3	83	80	36	201	29
COL-008-Blue_Train-2.mp3	65	85	27	194	47
COL-009-Blue_Train-3.mp3	82	66	24	89	42
COL-010-Blue_Train-4.mp3	91	162	33	166	43
COL-011-Blue_Train-5.mp3	77	182	27	183	22
COL-012-Blue_Train-6.mp3	104	89	22	212	26
COL-013-Blue_Train-7.mp3	104	153	36	196	50
COL-014-Blue_Train-8.mp3	75	153	25	156	69
COL-015_Chronic_Blues-1.mp3	67	230	26	149	58
COL-016_Chronic_Blues-2.mp3	49	65	30	215	25

COL-017_Chronic_Blues-3.mp3	62	89	37	162	51
COL-018_Im_Old_Fashioned-1.mp3	24	27	18	70	29
COL-019_Im_Old_Fashioned-2.mp3	41	59	30	85	28
COL-020_Im_Old_Fashioned-3.mp3	43	64	28	90	43
COL-021_Im_Old_Fashioned-4.mp3	49	109	29	69	24
COL-022_Im_Old_Fashioned-5.mp3	47	40	24	84	27
COL-023_Lazy_Bird-1.mp3	77	74	24	185	40
COL-024_Lazy_Bird-2.mp3	72	144	63	69	63
COL-025_Lazy_Bird-3.mp3	96	77	35	239	29
COL-026_Lazy_Bird-4.mp3	85	21	35	165	71
COL-027_Lazy_Bird-5.mp3	72	37	24	163	27
COL-028-Locomotion-1.mp3	88	94	33	200	30
COL-029-Locomotion-2.mp3	89	31	36	208	32
COL-030-Locomotion-3.mp3	109	197	44	231	60
<i>Dexter Gordon</i>					
GOR-001-Secret_Love-1.mp3	37	149	54	162	57
GOR-002-Secret_Love-2.mp3	66	71	60	51	37
GOR-003-Secret_Love-3.mp3	64	81	71	69	38
GOR-004-Secret_Love-4.mp3	59	77	66	76	30
GOR-005-Secret_Love-5.mp3	75	78	30	198	46
GOR-006-Secret_Love-6.mp3	74	111	37	248	60
GOR-007-Secret_Love-7.mp3	93	109	63	268	49
GOR-008-Secret_Love-8.mp3	72	95	29	189	35
GOR-009-Secret_Love-9.mp3	69	80	40	201	43
GOR-010-Secret_Love-10.mp3	97	101	60	259	41
GOR-011-Secret_Love-11.mp3	87	101	54	225	33
GOR-012-Secret_Love-12.mp3	76	81	50	228	43
GOR-013-Secret_Love-13.mp3	64	51	53	249	28
GOR-014-Secret_Love-14.mp3	69	75	61	237	33
GOR-015-The_End_Of_A_Love_Affair-1.mp3	87	209	43	216	28
GOR-016-The_End_Of_A_Love_Affair-2.mp3	85	107	58	248	38
GOR-017-The_End_Of_A_Love_Affair-3.mp3	92	78	55	256	31
GOR-018-The_End_Of_A_Love_Affair-4.mp3	92	83	51	236	27
GOR-019-The_End_Of_A_Love_Affair-5.mp3	100	58	66	233	64
GOR-020-The_End_Of_A_Love_Affair-6.mp3	99	281	52	241	41
GOR-021-Blues_Up_And_Down-1.mp3	85	85	60	191	31
GOR-022-Blues_Up_And_Down-2.mp3	72	53	33	195	27
GOR-023-Blues_Up_And_Down-3.mp3	87	101	22	223	32
GOR-024-Blues_Up_And_Down-4.mp3	99	110	37	217	27
GOR-025-Blues_Up_And_Down-5.mp3	41	93	28	166	20
GOR-026-Cheesecake-1.mp3	62	80	39	179	25
GOR-027-Cheesecake-2.mp3	101	69	43	245	29
GOR-028-Cheesecake-3.mp3	69	109	49	218	37
GOR-029-Cheesecake-4.mp3	75	239	60	335	96
GOR-030-Cheesecake-5.mp3	60	90	38	198	53
<i>Charlie Parker</i>					
PAR-001-After_Youve_Gone-1.mp3	90	79	83	215	51
PAR-002-After_Youve_Gone-2.mp3	92	163	99	291	24
PAR-003-After_Youve_Gone-3.mp3	81	63	90	299	34
PAR-004-After_Youve_Gone-4.mp3	70	34	67	149	22
PAR-005-I_Cant_Get_Started-1.mp3	56	207	48	193	100
PAR-006-Blue_Bird.mp3	81	200	43	23	27
PAR-007-I_Cant_Get_Started-3.mp3	84	90	44	203	38
PAR-008-I_Cant_Get_Started-4.mp3	63	217	54	130	46
PAR-009-I_Cant_Get_Started-5.mp3	52	156	39	133	34
PAR-010-I_Cant_Get_Started-6.mp3	67	85	24	125	29
PAR-011-Oh_Lady_Be_Good-1.mp3	93	292	59	366	37

PAR-012-Oh_Lady_Be_Good-2.mp3	98	125	71	288	67
PAR-013-Oh_Lady_Be_Good-3.mp3	100	111	65	208	32
PAR-014-I_Got_Rhythm-1.mp3	103	119	86	423	44
PAR-015-I_Got_Rhythm-2.mp3	117	68	78	239	81
PAR-016-I_Got_Rhythm-3.mp3	102	175	81	328	77
PAR-017-The_Bird-1.mp3	72	106	35	200	29
PAR-018-The_Bird-2.mp3	65	359	45	202	28
PAR-019-The_Bird-3.mp3	97	164	51	203	81
PAR-020-The_Bird-4.mp3	112	124	72	179	36
PAR-021-The_Bird-5.mp3	74	123	61	146	90
PAR-022-The_Bird-6.mp3	84	350	57	244	29
PAR-023-Lover-1.mp3	85	47	38	273	42
PAR-024-Lover-2.mp3	74	71	78	182	51
PAR-025-Lover-3.mp3	72	64	47	244	24
PAR-026-Summertime-1.mp3	45	70	41	93	49
PAR-027-Summertime-2.mp3	29	49	35	82	33
PAR-028-Summertime-3.mp3	58	104	47	108	29
PAR-029-Summertime-4.mp3	37	72	33	78	33
PAR-030-Mango_Mangue-1.mp3	92	110	53	218	38

*Joshua Redman*

RED-001-Turnaround-1.mp3	42	183	33	120	57
RED-002-Turnaround-2.mp3	91	224	40	234	72
RED-003-Turnaround-3.mp3	57	44	46	149	29
RED-004-Turnaround-4.mp3	68	91	38	144	61
RED-005-Turnaround-5.mp3	68	306	37	190	50
RED-006-Chill-1.mp3	20	91	43	135	21
RED-007-Chill-2.mp3	42	45	24	110	38
RED-008-Make_Sure_You_re_Sure-1.mp3	42	74	27	65	30
RED-009-Make_Sure_You_re_Sure-2.mp3	46	101	46	23	105
RED-010-Make_Sure_You_re_Sure-3.mp3	47	297	21	84	38
RED-011-Make_Sure_You_re_Sure-4.mp3	35	79	28	27	32
RED-012-Neverend-1.mp3	38	40	20	44	24
RED-013-Neverend-2.mp3	48	74	17	70	26
RED-014-Neverend-3.mp3	46	64	22	84	33
RED-015-Neverend-4.mp3	36	74	22	53	35
RED-016-Neverend-5.mp3	62	56	15	45	32
RED-017-Leap_Of_Faith-1.mp3	84	95	76	207	51
RED-018-Leap_Of_Faith-2.mp3	113	111	91	179	55
RED-019-Leap_Of_Faith-3.mp3	147	151	116	327	30
RED-020-Leap_Of_Faith-4.mp3	89	110	81	196	29
RED-021-Leap_Of_Faith-5.mp3	81	170	80	226	29
RED-022-Mischief-1.mp3	57	88	30	144	64
RED-023-Mischief-2.mp3	43	152	53	183	27
RED-024-Mischief-3.mp3	64	81	41	144	57
RED-025-Mischief-4.mp3	71	53	40	259	31
RED-026-Mischief-5.mp3	63	107	52	194	61
RED-027-Molten_Soul-1.mp3	59	53	60	93	48
RED-028-Molten_Soul-2.mp3	59	70	69	98	30
RED-029-Molten_Soul-3.mp3	80	67	64	146	40
RED-030-Molten_Soul-4.mp3	72	77	64	214	37
Anzahl Soli :	120 (4 × 30)				
Gesamtlänge :	43:52 min				
Mittlere Länge pro Solo :	21,9 s				

Tab. B.4: Testset - Audio - Musikerklassifikation (Saxophon)



# C High-Level-Merkmale

Merkmal	Formelzeichen	MEL	HAR	BAS	DRU
GE_AV_NUM_NOTES_PER_BAR_DRU	$\mathcal{G}_{NumNotes}$	×	×	×	×
GE_TEMPO_BPM	$\mathcal{G}_{Tempo}$	×			
PA_MS_POLYPHONY	$\mathcal{M}_{PA,Poly}$	×			
PA_RANGE_HT	$\mathcal{M}_{PA,Range}$	×	×	×	
PA_CHROMA_DAGO	$\mathcal{M}_{PA,Chroma,Dago}$	×		×	
PA_CHROMA_ENT_0	$\mathcal{M}_{PA,Chroma,0}$	×		×	
PA_MS_CONSTANT_PITCH	$\mathcal{M}_{PA,ConstPitch}$	×		×	
PA_MS_CHORD_TONE_RATIO	$\mathcal{M}_{PA,ChordTone}$	×		×	
PR_ABS_MN	$\mathcal{M}_{PR,MN}$	×		×	
PR_ABS_SD	$\mathcal{M}_{PR,SD}$	×		×	
PR_MS_CHROMATICS	$\mathcal{M}_{PR,Chromatics}$	×		×	
PR_MUS_P.I.-7	$\mathcal{M}_{PR,Mus,p,-7}$	×		×	
PR_MUS_P.I.-6	$\mathcal{M}_{PR,Mus,p,-6}$	×		×	
PR_MUS_P.I.-5	$\mathcal{M}_{PR,Mus,p,-5}$	×		×	
PR_MUS_P.I.-4	$\mathcal{M}_{PR,Mus,p,-4}$	×		×	
PR_MUS_P.I.-3	$\mathcal{M}_{PR,Mus,p,-3}$	×		×	
PR_MUS_P.I.-2	$\mathcal{M}_{PR,Mus,p,-2}$	×		×	
PR_MUS_P.I.1	$\mathcal{M}_{PR,Mus,p,1}$	×		×	
PR_MUS_P.I.2	$\mathcal{M}_{PR,Mus,p,2}$	×		×	
PR_MUS_P.I.3	$\mathcal{M}_{PR,Mus,p,3}$	×		×	
PR_MUS_P.I.4	$\mathcal{M}_{PR,Mus,p,4}$	×		×	
PR_MUS_P.I.5	$\mathcal{M}_{PR,Mus,p,5}$	×		×	
PR_MUS_P.I.6	$\mathcal{M}_{PR,Mus,p,6}$	×		×	
PR_MUS_P.I.7	$\mathcal{M}_{PR,Mus,p,7}$	×		×	
PR_MUS_DAGO	$\mathcal{M}_{PR,Mus,Dago}$	×		×	
PR_MUS_VARIETY	$\mathcal{M}_{PR,Mus,Var}$	×		×	
PR_MUS_ENT_0	$\mathcal{M}_{PR,Mus,H0}$	×		×	
PR_MUS_ENT_1	$\mathcal{M}_{PR,Mus,H1}$	×		×	
PR_DIR_ENT_0	$\mathcal{M}_{PR,Dir,H0}$	×		×	
PR_DIR_ENT_1	$\mathcal{M}_{PR,Dir,H1}$	×		×	
PR_MS_SMALL_INTERVALS	$\mathcal{M}_{PR,SmallInt}$	×		×	
PR_MS_BAL_DIR	$\mathcal{M}_{PR,BalDir}$	×		×	
PR_MS_DOM_DIR	$\mathcal{M}_{PR,DomDir}$	×		×	
PR_MS_MEL_FLUENCY	$\mathcal{M}_{PR,MelFlu}$	×		×	
PF_P.I.1	$\mathcal{M}_{PF,p,1}$	×		×	
PF_P.I.2	$\mathcal{M}_{PF,p,2}$	×		×	
PF_P.I.3	$\mathcal{M}_{PF,p,3}$	×		×	
PF_P.I.4	$\mathcal{M}_{PF,p,4}$	×		×	
PF_P.I.5	$\mathcal{M}_{PF,p,5}$	×		×	
PF_P.I.6	$\mathcal{M}_{PF,p,6}$	×		×	
PF_P.I.7	$\mathcal{M}_{PF,p,7}$	×		×	
PF_ENT_0	$\mathcal{M}_{PF,H0}$	×		×	
PF_ENT_1	$\mathcal{M}_{PF,H0}$	×		×	
PF_DAGO	$\mathcal{M}_{PF,Dago}$	×		×	
PF_VARIETY	$\mathcal{M}_{PF,Var}$	×		×	
PF_DOM_CHORD_DEGREE_MN	$\mathcal{M}_{PF,ChordDegree,MN}$	×		×	
PF_DOM_CHORD_DEGREE_SD	$\mathcal{M}_{PF,ChordDegree,SD}$	×		×	
PF_MS_AV_CHORDS_PER_BAR_MN	$\mathcal{M}_{PF,ChordPerBar,MN}$	×		×	
PF_MS_AV_CHORDS_PER_BAR_SD	$\mathcal{M}_{PF,ChordPerBar,SD}$	×		×	

ON_MS_EVENT_DENSITY_MN	$\mathcal{R}_{ON,EvDens,MN}$	×	×	×	×
ON_MS_EVENT_DENSITY_SD	$\mathcal{R}_{ON,EvDens,SD}$	×	×	×	×
ON_SWING_RATIO_4	$\mathcal{R}_{ON,Swing,4}$	×	×	×	×
ON_SWING_RATIO_8	$\mathcal{R}_{ON,Swing,8}$	×	×	×	×
ON_SWING_RATIO_16	$\mathcal{R}_{ON,Swing,16}$	×	×	×	×
ON_SWING_RATIO_32	$\mathcal{R}_{ON,Swing,32}$	×	×	×	×
ON_MS_BIN_PLAY	$\mathcal{R}_{ON,Bin}$	×	×	×	×
ON_MS_TERN_PLAY	$\mathcal{R}_{ON,Tern}$	×	×	×	×
ON_MS_PERC_PREC_4	$\mathcal{R}_{ON,Prec,4}$	×	×	×	×
ON_MS_PERC_PREC_8	$\mathcal{R}_{ON,Prec,8}$	×	×	×	×
ON_MS_PERC_PREC_16	$\mathcal{R}_{ON,Prec,16}$	×	×	×	×
ON_MS_PERC_PREC_32	$\mathcal{R}_{ON,Prec,32}$	×	×	×	×
ON_MS_SYNC_2	$\mathcal{R}_{ON,Syn,2}$	×	×	×	×
ON_MS_SYNC_4	$\mathcal{R}_{ON,Syn,4}$	×	×	×	×
ON_MS_SYNC_8	$\mathcal{R}_{ON,Syn,8}$	×	×	×	×
ON_MS_SYNC_16	$\mathcal{R}_{ON,Syn,16}$	×	×	×	×
ON_BAR_DAGO	$\mathcal{R}_{ON,Dago}$	×	×	×	×
ON_BAR_ENT_0	$\mathcal{R}_{ON,H0}$	×	×	×	×
ON_BAR_ENT_1	$\mathcal{R}_{ON,H1}$	×	×	×	×
ON_BAR_VARIETY	$\mathcal{R}_{ON,Var}$	×	×	×	×
ON_MS_DOWN_BEAT_FEEL	$\mathcal{R}_{ON,Down}$	×	×	×	×
ON_MS_RSP_RHYTH_FEEL	$\mathcal{R}_{ON,RSP,Feel}$	×	×	×	×
ON_MS_RSP_RHYTH_CHAR	$\mathcal{R}_{ON,RSP,Char}$	×	×	×	×
ON_MS_RSP_RHYTH_GRID	$\mathcal{R}_{ON,RSP,Grid}$	×	×	×	×
IO_BAR_DAGO	$\mathcal{R}_{IO,Dago}$	×	×	×	×
IO_BAR_ENT_0	$\mathcal{R}_{IO,H0}$	×	×	×	×
IO_BAR_ENT_1	$\mathcal{R}_{IO,H1}$	×	×	×	×
DU_DAGO	$\mathcal{R}_{DU,Dago}$	×	×	×	×
DU_ENT_0	$\mathcal{R}_{DU,H0}$	×	×	×	×
DU_ENT_1	$\mathcal{R}_{DU,H1}$	×	×	×	×
PAT_PA (21 Merkmale)	$\mathcal{S}_{PA}$	×		×	
PAT_ON (21 Merkmale)	$\mathcal{S}_{ON}$	×		×	
PAT_DU (21 Merkmale)	$\mathcal{S}_{DU}$	×		×	

Tab. C.1: Verteilung der High-Level-Merkmale auf die vier Instrumentalspuren

# Literaturverzeichnis

- [Bag90] BAGGI, Denis L.: NeurSwing: A Connectionist Workbench For The Investigation Of Swing In Afro-American Jazz. In: *Proc. of the 23rd Asilomar Conference on Signals, Systems, and Computers* Bd. 1, 1990, S. 336–341
- [Bag91] BAGGI, Denis L.: Neurswing: An Intelligent Workbench for the Investigation of Swing in Jazz. In: *IEEE Computer* Bd. 24, 1991, S. 60–64
- [Ber07] BERENDT, Joachim-Ernst: *Das Jazzbuch - Von New Orleans bis in die achtziger Jahre*. Fischer Taschenbuch Verlag, Frankfurt a.M., 2007
- [BP04] BELLO, Juan P. ; PICKENS, Jeremy: A Robust Mid-level Representation for Harmonic Content in Music Signals. In: *Proc. of the 5th International Conference on Music Information Retrieval (ISMIR)*, 2004, S. 304–311
- [BS06] BROSBOL, Jens ; SCHUBERT, Emery: Calculating Articulation In Solo Music Performances. In: *Proc. of the 9th International Conference on Music Perception and Cognition (ICMPC)*, 2006, S. 1362–1365
- [CL01] CHANG, Chih-Chung ; LIN, Chih-Jen: LIBSVM: a library for support vector machines. In: <http://www.csie.ntu.edu.tw/~cjlin/libsvm> (zuletzt aufgerufen : 22.06.2008), 2001
- [DA08] DITTMAR, Christian ; ABESSER, Jakob: Automatic Music Transcription with User Interaction. In: *Proc. of the DAGA*, 2008
- [DDR07] DITTMAR, Christian ; DRESSLER, Karin ; ROSENBAUER, Katja: A Toolbox for Automatic Transcription of Polyphonic Music. In: *Proc. of the Audio Mostly*, 2007
- [DGW04] DIXON, Simon ; GOUXON, Fabien ; WIDMER, Gerhard: Towards Characterisation of Music via Rhythmic Patterns. In: *Proc. of the 5th International Conference on Music Information Retrieval (ISMIR)*, 2004
- [Dit02] DITTMAR, Christian: *Implementierung zweier Independent Subspace Analysis Verfahren und Überprüfung der Eignung für Blind Source Separation*, Fachhochschule Jena, Diplomarbeit, 2002

- [ELL<sup>+</sup>04] ERKKILÄ, J. ; LARTILLOT, O. ; LUCK, G. ; RIIKKILÄ, K. ; TOIVIAINEN, P.: Intelligent Music Systems In Music Therapy. In: *Music Therapy Today* Bd. 5, 2004
- [Ell07] ELLIS, Daniel: Classifying Music Audio with Timbral and Chroma Features. In: *Proc. of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007
- [EN00] EEROLA, Tuomas ; NORTH, Adrian C.: Expectancy-Based Model of Melodic Complexity. In: *Proc. of the 6th International Conference of Music Perception and Cognition (ICMPC)*, 2000
- [ET04] EEROLA, Tuomas ; TOIVIAINEN, Petri: MIDI Toolbox: MATLAB Tools for Music Research. In: *www.jyu.fi/musica/miditoolbox/* (zuletzt aufgerufen : 20.06.2008). Jyväskylä, Finland : University of Jyväskylä, 2004
- [For98] FORDHAM, John: *Das große Buch vom Jazz*. Christian Verlag, München, 1998
- [GFB03] GOUYON, Fabien ; FABIG, Lars ; BONADA, Jordi: Rhythmic Expressiveness Transformations Of Audio Recordings - Swing Modifications. In: *Proc. of the 6th International Conference on Digital Audio Effects (DAFX)*, 2003
- [Hai00] HAIDER, Kurt: *Einführung in die Musiktheorie*. Peter Lang, Frankfurt a.M., 2000
- [HCL08] HSU, Chih-Wei ; CHANG, Chih-Chung ; ; LIN, Chih-Jen: A Practical Guide to Support Vector Classification. In: *http://www.csie.ntu.edu.tw/~cjlin* (zuletzt aufgerufen : 23.06.2008), 2008
- [Hem97] HEMPEL, Christoph: *Neue Allgemeine Musiklehre*. Schott Musik International, 1997
- [HGM95] HIDAKA, Isao ; GOTO, Masataka ; MURAOKA, Yoichi: An Automatic Jazz Accompaniment System Reacting to Solo. In: *Proc. of the 1995 International Computer Music Conference (ICMC)* (1995), S. 167–170
- [HLC01] HSU, Jia-Lien ; LIU, Chih-Chin ; CHEN, Arbee L. P.: Discovering Nontrivial Repeating Patterns in Music Data. In: *IEEE Transactions on Multimedia* Bd. 3, 2001 (IEEE Transactions on Multimedia 3), S. 311 – 324

- [HSG04] HERRERA, Perfecto ; SANDVOLD, Vegard ; GOUYON, Fabien: Percussion-Related Semantic Descriptors Of Music Audio Files. In: *Proc. of 25th International AES Conference*, 2004
- [Jun01] JUNGBLUTH, Axel: *Jazz-Harmonielehre*. Schott, 2001
- [Kla98] KLAPURI, Anssi: *Automatic Transcription Of Music*, Tampere University of Technology, Finland, Diplomarbeit, 1998
- [Li06] LI, Yipeng: Melodic Modeling : A Comparison Of Scale Degree And Interval. In: *Proc. of the International Computer Music Conference (ICMC)*, 2006, S. 495–498
- [LI07] LÉON, Pedro J. P. ; INESTA, José M.: Pattern Recognition Approach For Music Style Identification Using Shallow Statistical Descriptors. In: *IEEE Transactions on System, Man and Cybernetics - Part C : Applications and Reviews* Bd. 37, 2007, S. 248–257
- [LPR05] LIDY, Thomas ; PÖLZBAUER, Georg ; RAUBER, Andreas: Sound Re-Synthesis from Rhythm Pattern Features - Audible Insight into a Music Feature Extraction Process. In: *Proc. of the International Computer Music Conference (ICMC)*, 2005
- [LRPI07] LIDY, Thomas ; RAUBER, Andreas ; PERTUSA, Antonio ; IÑESTA, José M.: Improving Genre Classification By Combination Of Audio And Symbolic Descriptors Using A Transcription System. In: *Proc. of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007
- [MA02] MÁNTARAS, Ramon López d. ; ARCOS, Josep L.: AI and Music: From Composition to Expressive Performances. In: *AI Magazine* 23 (2002), S. 43–57
- [McK04] MCKAY, Cory: *Automatic Genre Classification of MIDI Recordings*, McGill University, Montreal, Kanada, Diplomarbeit, 2004
- [MF04] MCKAY, Cory ; FUJINAGA, Ichiro: Automatic Genre Classification Using Large High-Level Musical Feature Sets. In: *Proc. of the International Conference in Music Information Retrieval (ISMIR)*, 2004, S. 525–530
- [MW07] MADSEN, Søren T. ; WIDMER, Gerhard: A Complexity-Based Approach to Melody Track Identification in MIDI Files. In: *Proc. of the International Workshop on Artificial Intelligence and Music (MUSIC-AI)*, 2007

- [Pee04] PEETERS, Geoffroy: A large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. In: *IRCAM, Paris, 2004*, 2004
- [PSLI06] PÉREZ-SANCHO, Carlos ; LÉON, Pedro J. P. ; INESTA, José M.: A Comparison of Statistical Approaches to Symbolic Genre Recognition. In: *Proc. of the International Computer Music Conference (ICMC)*, 2006, S. 545–550
- [SHSTW04] SAUNDERS, Craig ; HARDOON, David R. ; SHAW-TAYLOR, John ; WIDMER, Gerhard: Using String Kernels to Identify Famous Performers from their Playing Style. In: *Proc. of the 15th European Conference on Machine Learning (ECML)*, 2004, S. 384–395
- [ST95] SCHUKAT-TALAMAZZINI, Ernst G.: *Automatische Spracherkennung - Statistische Verfahren zur Musteranalyse*. Vieweg Verlag, 1995
- [Uhl05] UHLE, Christian: *Automatisierte Extraktion rhythmischer Merkmale zur Anwendung in Music Information Retrieval-Systemen*, Technische Universität, Diss., 2005
- [Wel07] WELLING, Max: Fisher Linear Discriminant Analysis. In: [http://www.ics.uci.edu/~welling/classnotes/papers\\_class/Fisher-LDA.pdf](http://www.ics.uci.edu/~welling/classnotes/papers_class/Fisher-LDA.pdf) (zuletzt aufgerufen : 25.07.2008), 2007

# D Akronyme und Formelzeichen

## Akronyme

IDMT	Fraunhofer-Institut für Digitale Medientechnologie
KLT	Karhunen-Loève-Transformation
LDA	Lineare Diskriminanzanalyse ( <i>Linear Discriminant Analysis</i> )
MFCC	<i>Mel-Frequency Cepstral Coefficients</i>
MIDI	<i>Musical Instrument Digital Interface</i>
MIR	<i>Music Information Retrieval</i>
MP3	<i>MPEG-1 Audio Layer 3</i>
PCA	Hauptachsentransformation ( <i>Principle Component Analysis</i> )
PHP	<i>PHP Hypertext Preprocessor, Personal Home Page Tools</i>
SVM	<i>Support Vector Machine</i>
XML	<i>Extensible Markup Language</i>

Tab. D.1: Akronyme

## Formelzeichen

$a, A, \alpha$	Skalare
$\mathbf{a}, \boldsymbol{\mu}$	Vektoren
$\mathbf{A}, \boldsymbol{\Sigma}$	Matrizen
$B_N$	Zähler der Taktmensur
$B_D$	Nenner der Taktmensur
$c$	Klassen-Label
$D$	D'Agostino-Maßzahl
$E$	Erwartungswert
$f$	Frequenz
$f_O$	Frequenz eines Obertones
$f_0$	Grundfrequenz eines Tones
$\mathcal{G}$	Allgemeine Merkmale
$\mathcal{H}$	Harmonische Merkmale
$H_0$	Entropie nullter Ordnung
$H_1$	Entropie erster Ordnung

$I$	Informationsgehalt
$\mathcal{I}$	Interaktionsbezogene Merkmale
$J$	Rayleigh-Koeffizient
$i, j, k$	Zählvariablen und Indizes
$k_G$	Unterteilung des rhythmischen Rasters
$\mathbf{K}_{SVM}$	Trainiertes Modell (SVM-Klassifikator)
$m, n$	Zählvariablen und Indizes
$\mathcal{M}$	Melodische Merkmale
$N_B$	Länge des Solos [Takte]
$N_N$	Anzahl der Noten
$p$	Wahrscheinlichkeit
$p_A$	Absolute Tonhöhe
$p_R$	Relative Tonhöhe
$p_F$	Funktionale Tonhöhe
$Q$	Quantisierungsauflösung [Rasterpunkte / Takt]
$\mathcal{R}$	Rhythmische Merkmale
$R_F$	Rhythmisches Feeling
$R_C$	Rhythmische Charakteristik
$R_R$	Rhythmisches Raster
$s_G$	Verschiebung des rhythmischen Rasters
$\mathbf{S}_B$	Innerklassen-Streumatrix
$\mathbf{S}_W$	Intraklassen-Streumatrix
$\mathcal{S}$	Strukturbezogene Merkmale
$t_N$	Startzeitpunkt einer Note [s]
$t_{N,Off}$	Endzeitpunkt einer Note [s]
$\Delta t_N$	Notendauer [s]
$t_B$	Zeitpunkt einer Hauptzählzeit [s]
$v$	Lautstärke
$\mathbf{x}_F$	Merkmalsvektor
$\tilde{\mathbf{x}}_F$	Merkmalsvektor (dimensionsreduziert)
$\mathbf{V}$	Eigenvektormatrix
$X$	Zufallsvariable
$\mathbf{X}_F$	Merkmalsmatrix
$\widetilde{\mathbf{X}}_F$	Merkmalsmatrix (dimensionsreduziert)
$Z$	Nachrichtenquelle
$\alpha$	Phase
$\alpha_B$	Phase der Zählzeiten



---

$\Gamma$	Transformationsmatrix
$\epsilon$	Energie
$\lambda$	Eigenwert
$\Lambda$	Eigenwert-Diagonalmatrix
$\Lambda_{R1}, \Lambda_{R2}$	Energieprofile im rhythmischen Strukturprofil
$\mu$	Mittelwert
$\mu_c$	Klassen-Zentroiden
$\sigma$	Standartabweichung
$\sigma^2$	Varianz
$\tau$	Zeit [Taktlängen]
$\Delta\tau$	Dauer [Taktlängen]
$\tau_N$	Startzeitpunkt einer Note [Taktlängen]
$\Delta\tau_N$	Dauer einer Note
$\tau_{Noff}$	Endzeitpunkt einer Note [Taktlängen]
$\tau_B$	Position einer Zählzeit [Taktlängen]
$\tau_Q$	Position eines Quantisierungsrasterpunktes [Taktlängen]
$\Sigma$	Kovarianzmatrix

---

Tab. D.2: Formelzeichen



## E Thesen

1. Die Definition einer funktionalen Tonhöhe ermöglicht es, unabhängig vom Tonartgeschlecht den Tönen einer Melodie eine Funktion im aktuellen harmonischen Kontext zuzuordnen.
2. Das rhythmische Strukturprofil erlaubt eine detaillierte rhythmische Beschreibung einer Instrumentalspur sowie die Berechnung weiterer rhythmischer High-Level-Merkmale wie z.B. einem Maß für Synkopierungen.
3. Die Detektion von rhythmischen und melodischen Wiederholungen (Mustern) sowie die Ableitung entsprechender statistischer Merkmale verbessert die strukturelle Beschreibung einer Instrumentalspur.
4. Fehler bei der Transkription von Musik in Form von fehlenden oder falsch detektierten Noten haben unmittelbare Auswirkung auf die Aussagekraft der aus den Partiturinformatoren extrahierten High-Level-Merkmale.
5. Über eine Ensembleentscheidung von vier voneinander unabhängigen Klassifikatoren, die mit den High-Level-Merkmalen aus den Spuren einzelner Instrumente trainiert wurden, werden im Vergleich die besten Ergebnisse bei der Genreklassifikation erreicht.
6. Eine Genreklassifikation, die ausschließlich auf der Beschreibung des rhythmischen Kontextes basiert, der durch die Begleitinstrumente in einem Solo vorgegeben wird, ist möglich. Auch hier erzielt eine Ensemble-Entscheidung die beste Klassifikationsgenauigkeit.
7. Wie in einem Hörtest gezeigt wurde, sind Probanden trotz verhältnismäßig hoher musikalischer Vorbildung nur bedingt in der Lage, ein Instrumentalsolo einem Genre zuzuordnen, wenn ihnen Informationen bezüglich des Timbres (Klangfarbe) und der ursprünglichen Instrumentierung nicht zur Verfügung stehen.
8. Die Musikerklassifikation mit den implementierten High-Level-Merkmalen ist prinzipiell möglich. Durch die zusätzliche Verwendung von Merkmalen bezüglich Timbre und Spieltechnik erwartet der Autor jedoch eine starke Verbesserung der Klassifikation.